



UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI GIURISPRUDENZA

Dipartimento di Scienze Giuridiche “Cesare Beccaria”
Dottorato di ricerca (XXXV Ciclo)
Curriculum di Diritto penale

**INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ:
TRA DIRITTO CIVILE E DIRITTO PENALE**

Tutor: Prof. Gian Luigi Gatta

Coordinatrice: Prof.ssa Francesca Poggi

Tesi di dottorato di:
Beatrice Fragasso
Matr. R12537

Anno Accademico 2021-2022

La macchina ha gettato a terra il conducente, e corre cieca nello spazio.

*Al culmine del processo di razionalizzazione,
la ragione è diventata irrazionale e stupida.*

M. HORKHEIMER, *Eclisse della ragione*, 1969¹

¹ M. HORKHEIMER, *Eclisse della ragione*, Einaudi, 1969, p. 113.

INDICE

Introduzione.....	VII

CAPITOLO I INTELLIGENZA ARTIFICIALE: IL FENOMENO, LE DEFINIZIONI, LA CASISTICA

1. L'intelligenza artificiale, un concetto in cerca di definizione: tra innovazione tecnologica e suggestioni fantascientifiche.....	1
1.1. Algoritmi tradizionali e algoritmi di intelligenza artificiale.....	6
1.2. Dall'i.a. simbolica al connessionismo, in tre fasi: logica, conoscenza, apprendimento	8
1.2.1. Intelligenza come applicazione deduttiva di principi logici.	9
1.2.2. Intelligenza come applicazione deduttiva di conoscenze specialistiche.....	11
1.2.3. Intelligenza come apprendimento induttivo.	12
2. <i>Machine learning</i> , <i>deep learning</i> , reti neurali.....	15
2.1. Le modalità di apprendimento nel <i>machine learning</i>	17
2.2. Anche i sistemi di <i>machine learning</i> sbagliano: l' <i>adversarial machine learning</i>	20
3. Il rischio di discriminazione algoritmica.	23
3.1. Il caso dei sistemi di supporto all'attività giudiziaria e all'attività di <i>law enforcement</i>	25
4. I.a. come "autore" di un reato e i.a. come strumento.	33
4.1. Il caso dei <i>deepfakes</i>	35
5. La nozione "penalisticamente orientata" di intelligenza artificiale.	40
5.1. L'autonomia dei sistemi di i.a.	41
5.2. L'interattività dei sistemi di i.a.	44
5.3. Sistemi di i.a. e opacità tecnologica (c.d. <i>black box</i>).	46
5.3.1. Dal disincanto weberiano all'incanto algoritmico?	53
6. Le definizioni normative.	56
6.1. La definizione fornita dall' <i>AI Act</i> (Commissione Europea).....	58
6.2. La definizione fornita in Italia dal Consiglio di Stato.....	61
7. Una proposta di classificazione per i sistemi di i.a. (con relativa casistica).	63
7.1. Sistemi che possono recare un'offesa <i>diretta</i> ai beni giuridici.	65

7.1.1.	<i>Embodied systems</i>	65
7.1.1.1.	<i>Autonomous vehicles</i>	65
7.1.1.1.1.	Quale grado di autonomia? La classificazione SAE.....	68
7.1.1.1.2.	Il quadro normativo (cenni).	72
7.1.1.1.3.	Incidenti che hanno coinvolto gli <i>autonomous vehicles</i> e prima casistica giudiziaria	76
7.1.1.2.	Sistemi di guida autonoma degli aerei	82
7.1.1.3.	Sistemi d'arma autonomi letali..	84
7.1.1.4.	<i>Healthcare robots</i>	90
7.1.2.	<i>Non-embodied systems</i>	94
7.1.2.1.	<i>Bots</i>	94
7.1.2.2.	Sistemi di <i>trading</i> algoritmico.	98
7.2.	Sistemi che possono recare un'offesa <i>indiretta</i> ai beni giuridici.	102
7.2.1.	Sistemi di supporto alla diagnostica e al trattamento medico.	102

CAPITOLO II

INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ CIVILE

1.	La responsabilità civile come territorio privilegiato per l' <i>enforcement ex post</i> dei danni da intelligenza artificiale.	109
1.1	Dal dogma della colpa alla responsabilità oggettiva per creazione del rischio: alcuni cenni storici.	112
1.2.	Ritorno al futuro: un declino del modello oggettivistico? Regolazione dell'i.a. e <i>policy</i> europea.	117
1.3.	Adattabilità della normativa vigente o necessità di introdurre una disciplina <i>ad hoc</i> ?	121
2.	La responsabilità oggettiva in capo all'utilizzatore del sistema di i.a.: i regimi speciali di responsabilità previsti dal codice civile.	125
2.1.	Responsabilità da cosa in custodia (art. 2051 c.c.) e responsabilità del custode dell'animale o di chi se ne serve (art. 2052 c.c.).	127
2.2.	Responsabilità da circolazione di veicoli (art. 2054 c.c.).....	130
2.3.	Responsabilità di padroni e committenti (art. 2049 c.c.).	133
3.	Responsabilità per danno da prodotto difettoso: cenni introduttivi.	134
3.1.	Sull'adeguatezza della nozione di "prodotto".....	138

3.2.	Sull'adeguatezza della nozione di "produttore".....	142
3.3.	Sull'adeguatezza della nozione di "difetto".....	145
3.3.1.	La delimitazione della rilevanza dei difetti: le esimenti.	149
3.3.2.	La classificazione tradizionale: difetti di fabbricazione, progettazione, informazione nei sistemi di i.a.....	152
3.3.3.	I difetti di progettazione: tra responsabilità per colpa e responsabilità assoluta.....	158
3.3.4.	La nozione di difetto di progettazione in relazione ai sistemi di i.a.: alcuni profili di inadeguatezza.	160
3.3.5.	Nuovi criteri per la valutazione del difetto: le proposte della bozza di revisione della <i>Product liability directive</i>	164
3.3.6.	Verso una nuova nozione di <i>design defect</i> in relazione ai sistemi di i.a.? Le proposte della dottrina.	166
3.3.6.1.	Sostanziale eliminazione del requisito del difetto.....	166
3.3.6.2.	L'agente modello umano.	168
3.3.6.3.	L'agente modello algoritmico.	169
3.4.	I rapporti tra <i>product safety</i> e <i>product liability</i> : verso un'adozione della <i>preemption doctrine</i> ?	170
3.4.1.	<i>Product Safety</i> e tecniche di armonizzazione. Cenni.	172
3.4.2.	Sistemi di i.a. e norme di armonizzazione: l' <i>AI Act</i>	177
3.4.3.	(segue) I sistemi di intelligenza artificiale ad alto rischio.	180
3.4.4.	La tesi della non difettosità del prodotto conforme agli standard tecnici...185	
3.4.5.	(segue) Separazione concettuale tra disciplina della <i>product liability</i> e disciplina della <i>product safety</i>	188
3.5.	L'accertamento del nesso di causalità. Profili sostanziali.....	191
3.5.1.	(segue) Profili processuali.	197
4.	Responsabilità per esercizio di attività pericolose (art. 2050 c.c.).....	200
5.	La personalità elettronica: una proposta europea, presto abbandonata.....	201
6.	<i>Manufacturer Enterprise Liability (MER)</i> e assicurazione obbligatoria.....	206
7.	Intelligenza artificiale e responsabilità civile: indicazioni e spunti per il prosieguo dell'indagine.....	208

CAPITOLO III

INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ PENALE

1. Diritto penale e intelligenza artificiale. Alcune premesse.....	215
2. La responsabilità penale diretta dei sistemi di i.a.: la tesi di Gabriel Hallevy.	218
2.1. Le critiche: il principio di personalità della responsabilità penale.....	223
2.2. (segue) Le funzioni della pena e il paragone con la responsabilità (para) penale delle persone giuridiche.....	227
3. La responsabilità dell'utilizzatore di un sistema di i.a.	231
3.1. Utilizzatore di sistema di i.a. totalmente autonomo: responsabilità colposa.....	231
3.2. Utilizzatore di sistema di i.a. parzialmente autonomo: responsabilità colposa.....	233
3.2.1. Il fatto tipico.	234
3.2.2. La colpa.	239
3.2.2.1. Il principio di affidamento.	246
3.2.2.2. Un limite alla rimproverabilità dell'utilizzatore? Prospettive <i>de jure condendo</i>	250
3.3. La divergenza tra voluto e realizzato: <i>aberratio ictus</i> e <i>aberratio delicti</i>	253
3.4. Quale rilievo per il <i>dolus generalis</i> ? Il caso degli <i>high frequency trading</i>	257
4. Responsabilità del produttore per l'evento lesivo derivante dal sistema di i.a.	262
4.1. Il diritto penale di fronte all'incertezza: principio di precauzione e tecniche di incriminazione.	263
4.2. La struttura del fatto tipico, tra commissione e omissione.....	268
4.3. L'accertamento del nesso di causalità.....	274
4.3.1. L'individuazione delle leggi scientifiche di copertura: un'incompatibilità tra ragionamento causale e approccio stocastico del <i>decision making</i> algoritmico?....	281
4.4. L'individuazione del soggetto personalmente responsabile: il c.d. <i>many hands problem</i>	290
4.4.1. La tipizzazione delle "aree di competenza" nella gestione del rischio ad opera dell' <i>AI Act</i>	295
4.5. La colpa del produttore.....	297
4.5.1. Le regole cautelari scritte: la colpa specifica.....	299
4.5.2. Il rapporto tra regole cautelari scritte e regole cautelari non scritte: quale spazio per il rischio consentito?	306
4.5.2.1. Individuazione delle aree di rischio consentito e limiti all'operatività dell'istituto.....	313

4.5.2.2. Gli obblighi di monitoraggio e il “fallimento” delle regole cautelari positive.....	317
4.5.2.3. Intelligenza artificiale e rischio consentito: verso un approccio “quantitativo”?	321
4.5.3. Le regole cautelari non scritte: la colpa generica.....	328
4.5.3.1. La prevedibilità dell’evento: è rimproverabile la <i>prevedibilità</i> dell’ <i>imprevedibilità</i> algoritmica?	329
4.5.3.2. Principio di precauzione e giudizio di prevedibilità dell’evento...	334
4.5.4. Intelligenza artificiale, rischio consentito ed individuazione delle cautele doverose	339
5. Una tutela anticipata nei confronti dei sistemi di i.a. pericolosi? Prospettive <i>de jure condito</i>	343
5.1. (<i>segue</i>) Prospettive <i>de jure condendo</i>	350
6. Riepilogo: intelligenza artificiale e responsabilità penale.....	352
Conclusioni.....	361
Bibliografia.....	367
Ringraziamenti.....	433

INTRODUZIONE

Una macchina a guida autonoma attraversa un incrocio con il semaforo rosso, scontrandosi con una vettura e causando il decesso del conducente di quest'ultima.

Un *account* in grado di pubblicare autonomamente dei contenuti sui *social network* (c.d. *social bot*), attivato da Microsoft su Twitter con l'obiettivo di testarne la capacità di apprendimento linguistico, inizia a pubblicare *tweet* oltraggiosi e di istigazione all'odio razziale.

Alcuni algoritmi di *trading* autonomo, interagendo tra loro sui mercati finanziari, effettuano acquisti e rivendite di titoli a velocità siderale, generando un improvviso calo dell'8% dell'indice *Dow Jones*.

Un sistema informatico programmato per effettuare acquisti casuali sul *Dark web* compra una partita di *ecstasy* e la recapita all'utilizzatore del sistema¹.

Pur nella loro diversità, questi esempi presentano, come minimo comune denominatore, un'offesa a beni giuridici tutelati penalmente, realizzata da sistemi che – in via di prima approssimazione – possiamo considerare come *autonomi*, ossia in grado di svolgere determinati compiti in base a processi decisionali razionali e indipendentemente dall'intervento di un essere umano. È quella che viene definita – non senza perplessità, da parte di alcuni studiosi² – intelligenza artificiale (i.a.): a differenza dei sistemi tecnologici che l'hanno preceduta, l'i.a. persegue gli obiettivi che le sono attribuiti in base a processi di *decision-making* autonomi, che non sono prestabiliti in fase di programmazione e che, talvolta, non sono nemmeno comprensibili da parte dell'uomo (c.d. *opacità tecnologica* o *black-box*, vd. *infra*, § 5.3).

Il salto di qualità registratosi, negli ultimi dieci anni, nello sviluppo dei sistemi di i.a. è legato a doppio filo con il pieno dispiegarsi delle potenzialità del c.d. *machine*

¹ Ci limitiamo, qui, ad una mera esemplificazione del campo di indagine che sarà oggetto della nostra trattazione. Questi casi, tra gli altri, saranno affrontati con maggiore dettaglio nella seconda parte di questo capitolo, ai § 7 ss., ove saranno forniti altresì gli appositi riferimenti bibliografici.

² K. CRAWFORD, *Né intelligente né artificiale. Il lato oscuro dell'IA*, Il Mulino, 2021, *passim*; L. FLORIDI, *Digital's Cleaving Power and Its Consequences*, in *Philosophy & Technology*, 2017, vol. 30, p. 126.

learning, una tecnica di programmazione che consente agli algoritmi di *apprendere* dall'esperienza e dall'ambiente circostante, modificando le proprie prestazioni nel corso del tempo (vd. *infra*, cap. I, par. 1.2.3. ss.).

Il cambiamento che l'intelligenza artificiale sta portando nelle nostre vite è epocale. Già oggi, molti sistemi di i.a. hanno dimostrato di essere più accurati, nello svolgere le attività loro delegate, degli esseri umani: prendiamo, ad esempio, i sistemi di supporto alla diagnostica e al trattamento (*Clinical Decision Support Systems*, c.d. CDSS, vd. *infra*, § 7.2.1), che in alcuni casi si sono rivelati più precisi nell'individuazione della corretta diagnosi rispetto ai medici³; si pensi, ancora, all'attività di riconoscimento biometrico, che già oggi i sistemi di i.a. sono in grado di svolgere in modo più efficace (oltre che rapido) rispetto agli esseri umani⁴; si pensi, infine, alle *self driving cars*, che, secondo alcune stime, se adottate da gran parte degli utenti della strada, potrebbero ridurre del 90% gli incidenti stradali⁵ – oltre che liberare uomini e donne da un'attività spesso percepita come noiosa e stressante.

Se la diffusione dell'intelligenza artificiale in ogni aspetto della vita quotidiana promette dunque di arrecare grandi benefici alla società, essa apre anche una serie di interrogativi sul piano politico, etico, economico, e, per quanto qui interessa, giuridico. Ricercatori, giornalisti ed esponenti di ONG specializzate hanno da tempo puntato il riflettore sul “lato oscuro” dell'i.a.⁶: approccio monopolistico⁷, rischio di calo

³ Vd. I. Steadman, *IBM's Watson is better at diagnosing cancer than human doctors*, in *Wired UK*, 11 febbraio 2013, disponibile a questo link: <https://www.wired.co.uk/article/ibm-watson-medical-doctor>, ove si legge che alcuni esperimenti condotti presso lo Sloan-Kettering Hospital negli Stati Uniti suggeriscono che le diagnosi effettuate dal CDSS Watson sono più accurate di quelle effettuate dai medici. La media di diagnosi di cancro corrette è del 90% per Watson, rispetto al 50% di un medico; vd. anche E. Blakemore, *AI outdoes radiologists when it comes to identifying hip fractures, study shows*, in *Washington Post*, 20 febbraio 2022, disponibile a questo link: <https://www.washingtonpost.com/health/2022/02/20/hip-fractures-ai/>.

⁴ Vd. il report ID R&D, *Human or Machine: AI Proves Best at Spotting Biometric Attacks*, 2022, disponibile a questo link: <https://www.idrnd.ai/wp-content/uploads/2022/01/ResearchBrief-HumanVsMachine-JAN2422.pdf>.

⁵ Vd. il report McKinsey & Company, *Ten ways autonomous driving could redefine the automotive world*, disponibile a questo link: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/ten-ways-autonomous-driving-could-redefine-the-automotive-world>.

⁶ K. CRAWFORD, *Né intelligente né artificiale. Il lato oscuro dell'IA*, cit.

⁷ T.N. NARECHANIA, *Machine Learning as Natural Monopoly*, 107 *Iowa Law Review*, 2022, p. 1543.

occupazionale⁸, discriminazione algoritmica⁹, inquinamento ambientale¹⁰ e protezione dei dati personali¹¹ sono solo alcune delle questioni problematiche che lo sviluppo di sistemi di i.a. pone.

In questa sede ci occuperemo di un tema che, quantomeno nel dibattito pubblico generalista, non ha ricevuto grande attenzione e che solo recentemente ha attirato l'interesse degli studiosi del diritto civile¹² e, in misura minore, del diritto penale¹³: quello della responsabilità per fatti lesivi realizzati mediante (o da?) sistemi di i.a.

⁸ J.M. WERBROUCK, *The Hypothesis of Technological Unemployment Caused by AI-Driven Automation and its Impact on Social Security Law*, in J. De Bruyne – C. Vanleenhove (eds.), *Artificial Intelligence and the Law*, Intersentia, 2021, p. 317. In realtà, è dubbio se lo sviluppo su larga scala di sistemi di i.a. avrà effetti negativi sull'occupazione; per una panoramica sulla letteratura esistente vd. F. BORDOT, *Artificial Intelligence, Robots and Unemployment: Evidence from OECD Countries*, in *Journal of Innovation Economics & Management*, vol. 37, n. 1, 2022, p. 117 ss.

⁹ Sulla discriminazione algoritmica vd. *infra*, § 3. e bibliografia ivi citata.

¹⁰ È stato calcolato, ad esempio, che un solo ciclo di apprendimento di un sistema di intelligenza artificiale può arrivare a produrre più di 280mila kg di anidride carbonica, circa 5 volte le emissioni di un'automobile americana nel suo ciclo di vita, vd. K. Hao, *Training a single AI model can emit as much carbon as five cars in their lifetimes*, in *MIT Technology Review*, disponibile a questo link: <https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/>.

¹¹ Sul tema vd. diffusamente G. FINOCCHIARO, *Riflessioni su intelligenza artificiale e protezione dei dati personali*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, Giuffrè, 2020, p. 237 ss.; F. PIZZETTI, *La protezione dei dati personali e la sfida dell'Intelligenza Artificiale*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018, p. 5; T.E. FROSINI, *Il discreto fascino delle forme di governo*, in *DPCE Online*, Vol 49, N. 4, apr. 2021; nello specifico settore dell'automotive, vd. T. MULDER – N.E. VELLINGA, *Exploring data protection challenges of automated driving*, in *Computer Law & Security Review*, Vol. 40, April 2021.

¹² In Italia vd., *ex multis*, G. ALPA, *Quale modello normativo europeo per l'intelligenza artificiale?*, in *Contr. impr.*, 2021, n. 4, p. 1003 ss.; A. FUSARO, *Quale modello di responsabilità per la robotica avanzata? Riflessioni a margine del percorso*, in *Nuova giur. civ. comm.*, 2020, n. 6, p. 1346; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, in *Oss. dir. civ. comm.*, 2020, fasc. 2, p. 445 ss.; il volume curato da A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1 e 2, Il Mulino, 2022, in particolare i contributi di U. RUFFOLO – A. AMIDEI, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design, strumenti di certificazione preventive e «autodisciplina» di settore*; U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*; A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*; nello specifico settore delle macchine a guida autonoma vd. A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, in *Europa e dir. priv.*, 2019, fasc. 4; A. DAVOLA – R. PARDOLESI, *In viaggio col robot: verso nuovi orizzonti della r.c. auto ("driverless")?*, in *Danno e resp.*, 2017, n. 5, p. 626; U. SALANITRO, *Intelligenza artificiale e responsabilità: la strategia della Commissione europea*, in *Riv. dir. civ.*, 2020, fasc. 6, p. 1246.

¹³ Per limitarci, per ora, all'Italia, vd. C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, 2020, n. 4, p. 1745; F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2019, fasc. 10; ID., *Intelligenza artificiale e diritto*

In breve, per richiamare gli esempi citati in precedenza: chi – e a quali condizioni – potrà essere chiamato a rispondere dei reati di omicidio stradale, istigazione all’odio razziale, manipolazione di mercato, acquisto di sostanze stupefacenti? Il problema, nelle sue linee generali, può essere riassunto in poche battute: la (parziale) perdita di controllo dell’operatore umano (dell’utilizzatore, così come del produttore, del programmatore, dello sviluppatore, etc.) sul processo decisionale e sul comportamento dell’algoritmo potrebbe comportare, nei termini che vedremo, un’attenuazione, se non un totale annullamento, delle istanze punitive.

I sistemi di i.a. cessano di essere un mero strumento nelle mani dell’uomo, mettendo in crisi quel dualismo tra oggetto e soggetto che, fin da Cartesio, ha caratterizzato la storia del pensiero moderno. La progressiva perdita di controllo sugli artefatti tecnologici “intelligenti” rischia insomma di scardinare i classici meccanismi imputativi del diritto penale, che sono forgiati su comportamenti umani e sono già stati

penale: qualche aggiornamento e qualche nuova riflessione, in G. Balbi e al. (a cura di), *Diritto penale e intelligenza artificiale. "Nuovi Scenari"*, Giappichelli, 2023, p. 1 ss.; L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, in M. Catenacci e al. (a cura di) *Studi in onore di Antonio Fiorella*, vol 1, Roma Tre Press, 2021, p. 813 ss.; P. SEVERINO, *Intelligenza artificiale e diritto penale*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 531 ss.; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, 2021, n. 1, p. 100; R. BORSARI, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *Medialaws*, 2019, n. 3, p. 262 ss.; A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 651 ss.; D. PIVA, *Machina discere, (deinde) delinquere et puniri potest*, *ivi*, p. 681 ss.; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, *ivi*, p. 713 ss.; R. BORGOGNO, *La responsabilità penale nei processi ad elevata automazione*, *ivi*, p. 727 ss.; A. CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018; ID., *Reati colposi e tecnologie dell'intelligenza artificiale*, in G. Balbi e al. (a cura di), *Diritto penale e intelligenza artificiale. "Nuovi Scenari"*, cit., p. 19 ss.; B. PANATTONI, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, in *Dir. inf.*, 2021, fasc. 1, p. 317; M.B. MAGRO, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in *Leg. pen.*, 10 maggio 2020; il volume curato da D. Provolò e al., *Genetics, Robotics, Law, Punishment*, Padova University Press, 2014, in particolare i contributi di M.B. MAGRO, *Biorobotica, robotica e diritto penale*, p. 499, e S. RIONDATO, *Robotica e diritto penale (robots, ibridi, chimere e "animali tecnologici")*, p. 599; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale: prove d'impatto a livello europeo*, in *Discrimen*, 2022; C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, in *Dir. pen. cont. – Riv. trim.*, 2022, n. 2, p. 50 ss.; sui rapporti tra responsabilità penale e intelligenza artificiale vd. anche i lavori di Ugo Pagallo, filosofo del diritto da sempre attento all'impatto delle nuove tecnologie sull'imputazione della responsabilità, vd. per tutti U. PAGALLO, *Saggio sui robot e il diritto penale*, in F. Vinciguerra – S. Dassano (a cura di), *Scritti in memoria di Giuliano Marini*, Edizioni Scientifiche Italiane, 2010, p. 595 ss.

sottoposti a “flessibilizzazione” – specie per quanto riguarda l’accertamento del nesso causale e della colpevolezza – nell’ambito della c.d. società del rischio¹⁴. Il diritto penale – fondato sul *mancato dominio*, da parte dell’agente, di un *fatto offensivo effettivamente dominabile*¹⁵ – rischia così di risultare inadeguato laddove tale *dominio* si venga a perdere e “autore” immediato del reato risulti essere proprio una macchina.

Il problema del c.d. *responsibility gap*¹⁶ è avvertito anche dalle istituzioni europee, che in diversi documenti hanno evidenziato come i “comportamenti emergenti”¹⁷ delle macchine “intelligenti” possano comportare un intollerabile vuoto di tutela nei confronti delle persone danneggiate. Già nel 2017, con la Risoluzione sul diritto civile e la robotica (2015/2103(INL))¹⁸, il Parlamento europeo esprimeva dubbi sull’applicabilità della normativa in materia di responsabilità civile da prodotto (dir. n. 85/374/CEE) anche ai danni causati da sistemi di i.a. In particolare, tra i considerando si leggeva che:

«considerando che più i robot sono autonomi, meno possono essere considerati come meri strumenti nelle mani di altri attori (quali il fabbricante, l’operatore, il proprietario, l’utilizzatore, ecc.); che ciò, a sua volta, pone il quesito se le regole ordinarie in materia di responsabilità siano sufficienti o se ciò renda necessari nuovi principi e regole volte a chiarire la responsabilità legale dei vari attori per azioni e omissioni imputabili ai robot, qualora le cause non possano essere ricondotte a un soggetto umano specifico, e se le azioni o le omissioni legate ai robot che hanno causato danni avrebbero potuto essere evitate;

¹⁴ Vd. C. PIERGALLINI, *Il paradigma della colpa nell’età del rischio: prove di resistenza del tipo*, in Riv. it. dir. proc. pen., 2005, p.1684; A. GARGANI, *La “flessibilizzazione” giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in Leg. pen., 2011, n. 2, p. 397 ss.

¹⁵ A. FIORELLA, *Responsabilità penale* (voce), in Enc. Dir., vol XXXIX, 1988, p. 1289.

¹⁶ Di *responsibility gap* si è parlato, inizialmente, soprattutto in dottrina. Vd., in particolare, i lavori seminali di A. MATTHIAS, *The responsibility gap: Ascribing responsibility for the actions of learning automata*, in *Ethics and Information Technology*, 2004, vol. 6, p. 175; R. SPARROW, *Killer robots*, in *J. Appl. Philos.*, 2007, vol. 24, p. 62; W. WALLACH – C. ALLEN, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, 2009, p. 198 ss. Recentemente si vedano le acute riflessioni di A. CAPPELLINI, *Reati colposi e tecnologie dell’intelligenza artificiale*, cit., *passim*.

¹⁷ Sul concetto di “*emergent behaviors*” – utilizzato in letteratura per descrivere i comportamenti “autonomi” dei sistemi di i.a. – vd., per tutti, R. CALO, *Robotics and the Lessons of Cyberlaw*, in 103 *Calif. L. Rev.*, 2015, *passim*, spec. p. 532 ss.

¹⁸ Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

considerando che, in ultima analisi, l'autonomia dei robot solleva la questione della loro natura alla luce delle categorie giuridiche esistenti e dell'eventuale necessità di creare una nuova categoria con caratteristiche specifiche e implicazioni proprie [...]

considerando che, nonostante l'ambito di applicazione della direttiva 85/374/CEE, l'attuale quadro giuridico non sarebbe sufficiente a coprire i danni causati dalla nuova generazione di robot, in quanto questi possono essere dotati di capacità di adattamento e di apprendimento che implicano un certo grado di imprevedibilità nel loro comportamento, dato che imparerebbero in modo autonomo, in base alle esperienze diversificate di ciascuno, e interagirebbero con l'ambiente in modo unico e imprevedibile»¹⁹.

Ancora, nell'ottobre 2020, il Parlamento Europeo è intervenuto nuovamente sul tema, con una Risoluzione recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale²⁰. Anche in quell'occasione, il Parlamento ribadiva la problematica applicazione della normativa sulla responsabilità aquiliana ai danni causati dai sistemi di i.a.

Al considerando 7, ad esempio, si legge che l'autonomia e l'opacità dei sistemi di i.a. più sofisticati «potrebbe rendere molto difficile ricondurre determinate azioni a specifiche decisioni umane prese durante la loro progettazione o il loro funzionamento. L'operatore di un siffatto sistema di IA potrebbe sostenere, ad esempio, che l'attività, il dispositivo o il processo fisico o virtuale che ha causato il danno o il pregiudizio fosse al di fuori del proprio controllo in quanto attivato da un'operazione autonoma del proprio sistema di IA. Inoltre, il semplice funzionamento di un sistema di IA autonomo non dovrebbe essere motivo sufficiente per ammettere l'azione di responsabilità. Di conseguenza, vi potrebbero essere casi di responsabilità in cui l'attribuzione della

¹⁹ Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, cit., vd. considerando AB, AC, AI. In quella sede veniva proposta, come possibile soluzione, la creazione di «*un regime di assicurazione obbligatorio, come già avviene, per esempio, con le automobili*» (§ 57), da integrarsi con «*un fondo per garantire la possibilità di risarcire i danni in caso di assenza di copertura assicurativa*» (§ 58), nonché «*l'istituzione di uno status giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi*». Entrambe le ipotesi saranno vagliate nel prossimo capitolo, ai §§ 5 e 6.

²⁰ Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, 2020/2014(INL).

responsabilità potrebbe essere iniqua o inefficiente o in cui la persona che ha subito un danno o un pregiudizio cagionato da un sistema di IA non possa dimostrare la colpa del produttore, di una terza parte che abbia interferito o dell'operatore, e non ottenga un risarcimento».

Nonostante la cristallina ricostruzione delle problematiche che concernono l'attribuzione della responsabilità (civile) per fatti illeciti commessi da sistemi di i.a., ad oggi, nell'Unione Europea, non si è ancora addivenuti ad una compiuta elaborazione di un quadro normativo organico in materia; vi sono, tuttavia, diverse proposte presentate dalla Commissione Europea, in corso di discussione presso il Parlamento europeo, che mirano a fornire un quadro giuridico di riferimento in relazione alla messa in commercio dei sistemi di i.a. e ai meccanismi di imputazione degli eventi lesivi algoritmici²¹. Su tali proposte torneremo ampiamente nel corso di questo lavoro.

Nell'ambito della nostra trattazione, dedicheremo il primo capitolo ad una sintetica panoramica sul fenomeno dell'intelligenza artificiale, cercando di districarci tra le numerose definizioni proposte dalla letteratura scientifica, al fine di individuare le specifiche caratteristiche tecniche che rendono l'intelligenza artificiale un fenomeno *dirompente* per l'attribuzione della responsabilità penale. L'obiettivo è quello di identificare una nozione “penalisticamente orientata” di intelligenza artificiale (vd. §§ 5 ss.), limitando così il nostro campo d'indagine ai soli sistemi che, in forza della loro imprevedibilità ed irriducibilità al codice algoritmico, possono entrare in frizione con le tradizionali categorie del diritto penale. Nel delimitare l'oggetto del nostro lavoro, brevi cenni saranno comunque dedicati ad alcuni sistemi di i.a. che, pur presentando problemi diversi da quelli che ci occupano in questa sede, possono interessare in vario modo l'ordinamento penale (algoritmi di supporto all'attività giudiziaria e *deep fakes*, vd., rispettivamente, i §§ 3.1. e 4.1.). La parte finale del primo capitolo (§ 7 ss.) sarà infine dedicata al tentativo di delineare una classificazione dei sistemi di i.a. – così come definiti ai fini del nostro lavoro – fondata sulle modalità con le quali essi interagiscono con

²¹ Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione, COM/2021/206 final, 21 aprile 2021 (cd. *AI Act*); Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale), COM (2022) 496 final, 28 settembre 2022; Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi, COM (2022) 495 final, 28 settembre 2022.

l'esterno e, di conseguenza, sul tipo di bene giuridico sul quale questi possono incidere. Tale classificazione si baserà sullo studio della casistica relativa alle ipotesi di danno da intelligenza artificiale già verificatesi in diversi ambiti.

Nel secondo capitolo, ci porremo invece l'obiettivo di valutare l'adeguatezza degli attuali meccanismi di imputazione della responsabilità civile in relazione ai danni derivanti dall'attivazione di sistemi di i.a. La scelta di dedicare, in una tesi di dottorato in *diritto penale*, un intero capitolo alla responsabilità *aquiliana* potrebbe apparire insolita. Tale decisione, a nostro modo di vedere, si giustifica tuttavia con il tentativo di verificare se – in un settore in cui, già *prima facie*, l'accertamento di una responsabilità penale pare particolarmente complesso – sussistano, in altri settori, modelli di responsabilità adeguati ed effettivi per i soggetti danneggiati. In particolare, preso atto che la clausola generale della responsabilità per colpa *ex art. 2043 c.c.* troverà probabilmente un'applicazione limitata, la nostra attenzione si concentrerà sui regimi di ascrizione oggettivi contenuti nel codice civile (artt. 2049-2054; vd. §§ 2. ss.) e nella legislazione speciale (dir. 85/374/CE sulla responsabilità del produttore, trasposta agli artt. 114-127 cod. cons., vd. §§ 3. ss.), nonché su alcuni meccanismi imputativi proposti dalla dottrina per affrontare specificamente il problema del *responsibility gap* nel settore dell'intelligenza artificiale (personalità elettronica e *Manufacturer enterprise liability*, rispettivamente §§ 5. e 6.). Alcuni dei temi che tratteremo nel secondo capitolo (es. accertamento del nesso di causalità, nozione di difetto di progettazione, requisiti stabiliti per la messa in commercio di sistemi di i.a. dalla Proposta di Regolamento europeo presentata il 21 aprile 2021²²) anticipano, d'altra parte, questioni che verranno poi affrontate, in ottica penalistica, nel terzo capitolo.

Quest'ultimo costituirà il fulcro del nostro lavoro e sarà finalizzato all'enucleazione di alcuni dei principali problemi che la qualificazione penalistica dei danni da intelligenza artificiale pone. Un primo sguardo sarà dedicato alla proposta, emersa in dottrina²³, di introdurre una responsabilità penale diretta dei sistemi di i.a. –

²² Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale, cit.

²³ La tesi della possibile responsabilità diretta delle macchine è stata avanzata in una serie di scritti dal giurista israeliano Gabriel Hallevy; le argomentazioni più compiute sono svolte nella monografia, vd. ID., *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015; vd. anche ID., “*I, Robot – I, Criminal*”—*When Science Fiction Becomes Reality: Legal Liability of AI Robots committing Criminal Offenses*, in 22 *Syracuse Sci. & Tech. L. Rep.* 1, 2010, p. 1 ss.; ID., *The Criminal Liability of Artificial*

una proposta che, lo anticipiamo, non incontra il nostro favore, stante il suo radicale contrasto con il principio di personalità della responsabilità penale (vd. § 2. e 2.1.). Ci soffermeremo poi, in particolare, sulla configurabilità di *reati causalmente orientati* e sulla possibile imputazione di essi all'utilizzatore (par. 4 ss.) e al produttore del sistema di i.a. (par. 5 ss.); soltanto in maniera cursoria ci occuperemo infine della possibile configurabilità di reati a forma vincolata (par. 4.2.4) e sulle tecniche di anticipazione della tutela penale (par. 6).

Intelligence Entities - from Science Fiction to Legal Social Control, in 4 *Akron Intellectual Property Journal*, 2010, p. 171 ss.; ID., *Dangerous Robots – Artificial Intelligence vs. Human Intelligence*, 11 febbraio 2018, in SSRN, disponibile a questo link: <https://ssrn.com/abstract=3121905>. Le suggestioni di Hallevy sono poi state riprese e approfondite da Y. HU, *Robot Criminals*, in 52 *U. Mich. J. L. Reform*, 2019, p. 487 ss.

CAPITOLO I

INTELLIGENZA ARTIFICIALE: IL FENOMENO, LE DEFINIZIONI, LA CASISTICA

SOMMARIO: 1. L'intelligenza artificiale, un concetto in cerca di definizione: tra innovazione tecnologica e suggestioni fantascientifiche. – 1.1. Algoritmi tradizionali e algoritmi di intelligenza artificiale – 1.2. Dall'i.a. simbolica al connessionismo, in tre fasi: logica, conoscenza, apprendimento. – 1.2.1. Intelligenza come applicazione deduttiva di principi logici. – 1.2.2. Intelligenza come applicazione deduttiva di conoscenze specialistiche. – 1.2.3. Intelligenza come apprendimento induttivo. – 2. *Machine learning*, *deep learning*, reti neurali. – 2.1. Le modalità di apprendimento nel *machine learning*. – 2.2. Anche i sistemi di *machine learning* sbagliano: l'*adversarial machine learning*. – 3. Il rischio di discriminazione algoritmica. – 3.1. Il caso dei sistemi di supporto all'attività giudiziaria e all'attività di *law enforcement*. – 4. I.a. come "autore" di un reato e i.a. come strumento. – 4.1. Il caso dei *deepfakes*. – 5. La nozione "penalisticamente orientata" di intelligenza artificiale. – 5.1. L'autonomia dei sistemi di i.a. – 5.2. L'interattività dei sistemi di i.a. – 5.3. Sistemi di i.a. e opacità tecnologica (c.d. *black box*). – 5.3.1. Dal disincanto weberiano all'incanto algoritmico? – 6. Le definizioni normative. – 6.1. La definizione fornita dall'*AI Act* (Commissione Europea). – 6.2. La definizione fornita in Italia dal Consiglio di Stato. – 7. Una proposta di classificazione per i sistemi di i.a. (con relativa casistica). – 7.1. Sistemi che possono recare un'offesa *diretta* ai beni giuridici. – 7.1.1. *Embodied systems* – 7.1.1.1. *Autonomous vehicles*. – 7.1.1.1.2. Quale grado di autonomia? La classificazione SAE. – 7.1.1.1.3. Il quadro normativo (cenni). – 7.1.1.1.4. Incidenti che hanno coinvolto gli *autonomous vehicles* e prima casistica giudiziaria. – 7.1.1.2. Sistemi di guida autonoma degli aerei. – 7.1.1.3. Sistemi d'arma autonomi letali. – 7.1.1.4. *Healthcare robots*. – 7.1.2. *Non-embodied systems*. – 7.1.2.1. *Bots*. – 7.1.2.2. Sistemi di *trading* algoritmico. – 7.2. Sistemi che possono recare un'offesa *indiretta* ai beni giuridici. – 7.2.1. Sistemi di supporto alla diagnostica e al trattamento medico.

1. L'intelligenza artificiale, un concetto in cerca di definizione, tra innovazione tecnologica e suggestioni fantascientifiche.

Tra tutti gli interrogativi (giuridici ed extragiuridici) che lo sviluppo e la messa in commercio di prodotti "intelligenti" sollevano, il nostro lavoro si propone di valutare l'impatto che le nuove forme dell'agire algoritmico possono avere sull'imputazione della responsabilità penale (e civile) per gli eventi lesivi scaturenti dai sistemi di intelligenza artificiale.

In via preliminare, cercheremo di ricostruire, nei prossimi paragrafi, quali sono le caratteristiche tecniche che rendono l'intelligenza artificiale un fenomeno *dirompente* per il diritto e, in particolare, per il diritto penale. Lo scopo, ovviamente, non è quello di fornire un *excursus* completo e approfondito, dal punto di vista tecnico-scientifico, del

funzionamento dei sistemi di i.a. – compito di cui, di certo, non siamo all'altezza; cercheremo, piuttosto, di utilizzare un approccio *strumentale*, delineando quelle nozioni che sono essenziali per comprendere le implicazioni che lo sviluppo dei sistemi di i.a. può avere sull'imputazione della responsabilità penale.

Partiamo, innanzitutto, dal problema definitorio. Nel mondo scientifico non si è ancora raggiunto un consenso su cosa si intenda per “intelligenza artificiale”¹, tanto che sono diversi gli autori che ritengono che il richiamo alla nozione di i.a. debba essere concepito più come metafora – per cogliere il *novum* di questi sistemi rispetto a precedenti tecnologie – che come categoria dai contorni ben definiti². Secondo una nota classificazione proposta da Stuart J. Russell e Peter Norvig, autori di uno dei più diffusi manuali sull'i.a.³, la nozione di intelligenza artificiale varia a seconda della *concezione di intelligenza* cui si aderisce e, in particolare, dipende dalla risposta a due interrogativi:

- per essere considerato intelligente, l'agente deve essere in grado di pensare e agire come un essere umano o è sufficiente un comportamento “razionale” (laddove la razionalità è concepita come la capacità di raggiungere, sulla base delle risorse disponibili, il miglior risultato possibile nel perseguimento di un obiettivo dato)? (*human vs rational*);

- l'intelligenza è un attributo del *pensiero* – ossia del processo di ragionamento interiore – o può invece ricavarsi dal mero *comportamento esteriore*? (*thinking vs acting*).

Dalla combinazione delle risposte a questi interrogativi possono derivare quattro diverse concezioni di “intelligenza” e, di conseguenza, di intelligenza artificiale:

1. Macchina che agisce come un essere umano (*human + acting*);
2. Macchina che pensa come un essere umano (*human + thinking*);

¹ Lo ammettono sia gli articoli scientifici che i documenti di *policy*, vd., rispettivamente, P.M. KRAFFT E AL., *Defining AI in Policy versus Practice*, in *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, p. 72 ss.; US NATIONAL SCIENCE AND TECHNOLOGY COUNCIL COMMITTEE, *Preparing for the future of Artificial Intelligence*, 12 ottobre 2016, p. 6.

² L. PICOTTI, «Nuovi» crimini cibernetici e possibile rilevanza dell'intelligenza artificiale. Gli Atti Digitali degli Stati Generali del Diritto di Internet, in *Diritto di Internet*, 2022, suppl. fasc. n. 1, p. 5; ID., *Cybercrimes and Criminal Relevance of Artificial Intelligence*, in F.J. Castro Toledo (coord.), *La transformación algorítmica del sistema de justicia penal*, Thomson Reuters, 2022, p. 69; G. LO SAPIO, *Intelligenza artificiale: rischi, modelli regolatori, metafore*, in *Federalismi*, 2022, n. 27, p. 245 ss., sostanzialmente anche C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 2019, fasc. 4, § 2.

³ S. RUSSELL – P. NORVIG, *Artificial Intelligence: A Modern Approach*, Pearson College Div., 4th ed., 2020, p. 1 ss.

3. Macchina che pensa razionalmente (*rational + thinking*);
4. Macchina che agisce razionalmente (*rational + acting*).

Ad oggi, gli studiosi – con alcune sfumature – tendono a ricondurre il concetto di intelligenza a quest’ultimo modello (c.d. *standard model*): un agente viene considerato intelligente quando è in grado di conseguire determinati obiettivi in un ambiente complesso e incerto⁴. La caratteristica qualificante dell’intelligenza artificiale rispetto agli altri sistemi tecnologici, dunque, sembra consistere nella capacità di agire adattandosi ad una serie indefinita di circostanze, prendendo decisioni che massimizzano il risultato atteso. Se su tale generica e astratta definizione gli esperti tendono a convergere, rimane tuttavia complessa l’individuazione concreta dei sistemi che presentano le suddette caratteristiche.

Alcuni studiosi, ad esempio, hanno rilevato come la concezione di intelligenza artificiale si modifichi nel tempo, a seconda della percezione che la società ha di una tecnologia in un determinato contesto storico: in breve, non appena ci abituiamo al fatto che uno specifico compito possa essere svolto da una macchina, quest’ultima cessa di essere considerata “intelligenza artificiale” (C.d. *AI effect*)⁵. Nessuno, oggi, ad esempio, considererebbe Google Maps o lo *spam filtering* come sistemi di i.a.: «[a]s soon as it

⁴ Vd. S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 4 e 36 ss. I due elementi citati – quello del perseguimento degli obiettivi e quello dell’ambiente incerto – ricorrono nella maggior parte delle definizioni della nozione di “intelligenza”: si vd., a titolo di esempio, J.S. ALBUS, *Outline for a theory of intelligence*, in *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 21, Iss. 3, May-June 1991, p. 474: «the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioral subgoals that support the system’s ultimate goal»; D.B. FOGEL, *Review of “Computational intelligence: Imitating life”*, in *Proc. of the IEEE*, vol. 83, iss. 11, 1995, p. 1590: «Any system [...] that generates adaptive behaviour to meet goals in a range of environments can be said to be intelligent»; A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability* – Study requested by the JURI Committee, July 2020, p. 19: «a machine or system is deemed intelligent when it can act in, adapt to, and learn from specific environments so as to achieve predefined goals»; J. HORST, *A Native Intelligence Metric for Artificial Systems*, *Performance Metrics for Intelligent Systems, Workshop*, in *Proceedings of the Performance Metrics for Intelligent Systems (PerMIS) Workshop*, 2002, disponibile a questo link: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=824478: «the successful (i.e., goal-achieving) performance of the system in a complicated environment»; D. POOLE, *Computational Intelligence: A logical approach*, Oxford University Press: «[An intelligent agent does what] is appropriate for its circumstances and its goal, it is flexible to changing environments and changing goals, it learns from experience, and it makes appropriate choices given perceptual limitations and finite computation».

⁵ A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 20; J. KAPLAN, *Artificial Intelligence: What Everyone Needs to Know*, Oxford University Press, 2016, p. 37.

works, no one calls it AI anymore»⁶. Talvolta, d'altra parte, il riferimento al preteso carattere "intelligente" di un sistema pare più un'operazione di *marketing* che una ponderata descrizione del *software* utilizzato: così, algoritmi caratterizzati da ben consolidati meccanismi di *automazione* vengono talvolta spacciati per sistemi dotati di *autonomia*⁷.

Di certo, per prima cosa, può essere utile sgombrare il campo da equivoci. I sistemi di i.a. ad oggi esistenti hanno ben poco in comune con gli androidi che hanno sempre popolato l'immaginario fantascientifico letterario e cinematografico, aventi sembianze e capacità intellettive paragonabili a quelle degli esseri umani (tanto che uno dei *topos* di questo immaginario consiste proprio nella lotta degli automi per sfuggire al controllo dell'uomo)⁸. Al contrario, quando oggi parliamo di sistemi di i.a. facciamo riferimento ad algoritmi in grado di svolgere compiti molto specifici (come guidare un autoveicolo, effettuare negoziazioni sui mercati finanziari, pubblicare contenuti sui *social network*, etc.), ma del tutto privi di *common sense* o, comunque, di un'intelligenza ad ampio spettro assimilabile a quella di un essere umano. Di fatto, tutti i sistemi di i.a. ad oggi esistenti sono *artificial narrow intelligence* (ANI)⁹, chiamati anche sistemi di i.a. "deboli"; non mancano, d'altra parte, gli studi che si pongono – come obiettivo di lungo periodo – quello di sviluppare sistemi di *artificial general intelligence* (c.d. AGI), detti anche sistemi di i.a. "forti"¹⁰. Nonostante gli odierni algoritmi "intelligenti" siano lontani

⁶ La frase è generalmente attribuita a John McCarthy; vd. B. MEYER, *John McCarthy*, in *Communications of the ACM*, 28 ottobre 2011.

⁷ Lo rileva A. BERTOLINI, *Artificial Intelligence Does Not Exist! Defying the Technology-Neutrality Narrative in the Regulation of Civil Liability for Advanced Technologies*, in *Europa e dir. priv.*, 2022, n. 2, p. 373. Per un'analisi dell'evoluzione dei sistemi tecnologici, dall'*automazione* all'*autonomia* si vd. B. PANATTONI, *AI and Criminal Law: The Myth of 'Control' in a Data-Driven Society*, in *92 International Review of Penal Law*, 2021, p. 125 ss.

⁸ Si pensi, a mero titolo di esempio, al mostro di Frankenstein (M. SHELLEY, *Frankenstein; or, The Modern Prometheus*, 1818) o agli androidi protagonisti del romanzo di P.K. DICK, *Do Androids Dream of Electric Sheep?*, 1968, dal quale è stato tratto nel 1982 il celebre film *Blade Runner*, diretto da Ridley Scott. Ancora, pensiamo al supercomputer HAL 9000 che – nel film *2001: Odissea nello spazio*, diretto da Stanley Kubrick nel 1968 – tenta di eliminare l'intero equipaggio presente sulla nave spaziale, avendo percepito che gli astronauti avevano intenzione di disattivarlo.

⁹ In questo senso vd. R. DOWELL, *Fundamental Protections for Non-Biological Intelligences or: How We Learn to Stop Worrying and Love Our Robot Brethren*, in *19 Minn. J.L. Sci. & Tech.*, 2018, p. 308; S. VALLOR – G.A. BEKEY, *Artificial Intelligence and the Ethics of Self-Learning Robots*, in P. Lin – K. Abney – R. Jenkins (eds.) *Robot Ethics 2.0*, Oxford University Press, 2017, p. 339.

¹⁰ La distinzione tra *weak AI* e *strong AI* risale a J.R. SEARLE, *Minds, brains, and programs*, in *Behavioral and Brain Sciences*, 1980, n. 3 p. 417. Il dibattito sulla fattibilità tecnica della creazione di sistemi di i.a. con capacità intellettive paragonabili a quelle dell'essere umano è in realtà ancora acceso; si va dagli

dal possedere le capacità intellettive ed emotive dell'essere umano, non è raro, nelle cronache giornalistiche, sentir parlare di sistemi di i.a. “senzienti” e “autocoscienti”. Recentemente, ad esempio, ha creato scalpore la notizia di un ingegnere di Google – poi sospeso dal suo incarico – che aveva affermato che la tecnologia per la comprensione del linguaggio sviluppata da Google (la c.d. LaMDA, *Language Model for Dialogue Applications*) avrebbe dimostrato coscienza e consapevolezza della propria esistenza, poiché in grado di esporre verbalmente le proprie paure e i propri desideri¹¹: una conclusione – quella dell'ingegnere – che pare facilmente confutabile, dato che l'obiettivo dei sistemi di *Natural Language Processing* è proprio quello di *leggere, prevedere ed imitare* le modalità espressive degli esseri umani, senza che sia necessaria l'acquisizione di una comprensione semantica del linguaggio.

studiosi che preconizzano un imminente sviluppo di forme di AGI (es. R. KURZWEIL, *The Singularity Is Near: When Humans Transcend Biology*, Viking Penguin, 2005) a coloro che, al contrario, ritengono che tale obiettivo sia ancora molto lontano, o addirittura irraggiungibile (es. gli informatici Craig Silverstein e Mark Andreessen, le cui dichiarazioni sono contenute in K. AULETTA, *Googled: The End of the World as We Know It*, Penguin Press, 2009, p. 327). La molteplicità dei punti di vista in merito è ben testimoniata da un sondaggio condotto tra alcuni esperti del settore, vd. S.D. BAUME AL., *How long until human-level AI? Results from an expert assessment*, in *Technological Forecasting and Social Change*, vol. 78, iss. 1, p. 185. L'attuale dibattito pare essere andato persino oltre, concentrandosi sul possibile sviluppo di sistemi di i.a. che *superino* le capacità intellettive umane (c.d. *superintelligenza*); sul tema vd. N. BOSTROM, *Superintelligenza*, Bollati Boringhieri, 2014, che in tale rivoluzione tecnologica vede una potenziale catastrofe per la società, dalla quale è necessario tutelarsi; vd. anche il sondaggio condotto su 550 esperti nel campo dell'i.a., ai quali è stato chiesto quando, a loro parere, si avranno le capacità tecnologiche per sviluppare una “superintelligenza”, vd. V.C. MÜLLER – N. BOSTROM, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*, in V.C. Müller (ed.), *Fundamental Issues of Artificial Intelligence*, Springer, 2016, p. 555 ss. Sulla possibile futura regolazione dei sistemi AGI vd. T. MAHLER, *Regulating Artificial General Intelligence (AGI)*, in B. Custers – E. Fosch-Villaronga (eds.), *Law and Artificial Intelligence*, Asser Press – Springer, p. 521 ss.

¹¹ La notizia è stata riportata per la prima volta dal Washington Post (N. Tiku, *The Google engineer who thinks the company's AI has come to life*, 11 giugno 2022, disponibile a questo link: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>) ed è stata poi ripresa da numerose testate, anche in Italia (vd. ad es. F. Bini, *"L'intelligenza artificiale sta diventando senziente", ingegnere Google lancia l'allarme. E Big G lo sospende*, in *La Repubblica*, 13 giugno 2022, disponibile a questo link: https://www.repubblica.it/economia/2022/06/13/news/google_intelligenza_artificiale_ingegnere_bot-353695551/).

1.1. Algoritmi tradizionali e algoritmi di intelligenza artificiale.

In via di prima approssimazione, possiamo affermare che l'i.a. è un algoritmo. Con tale termine, tradizionalmente, si intende una sequenza di istruzioni che prescrive di trasformare un *input* in un determinato *output*, sulla base di un procedimento predefinito¹²: *se l'input è A, l'output deve essere Z; se B, Y; se C, X*, etc. (c.d. schema dell'*if-this-then-that*). L'algoritmo ha dunque una funzione che potremmo definire *normativa*¹³.

L'algoritmo, di per sé, non necessita di un supporto informatico per essere implementato: possono essere considerati algoritmi anche le istruzioni per la navigazione in alto mare o le stesse norme giuridiche (che prescrivono al giudice, ad esempio, di applicare una *determinata pena* qualora venga accertata la commissione di un *determinato fatto di reato*, oppure, sul versante processuale, di *dichiarare nullo* un atto processuale qualora ricorrano *determinate circostanze*)¹⁴. Ciononostante, la nozione di algoritmo è andata sempre di più assumendo l'accezione di *istruzione impartita ad un sistema informatico (software)*, con conseguente e radicale traslazione semantica.

La prima differenza tra un'istruzione impartita ad un essere umano ed un'istruzione impartita ad una macchina è che, in questo secondo caso, la formulazione degli *step* procedurali deve essere precisa nei minimi dettagli, non lasciando nulla di sottinteso né di ambiguo. Se, infatti, il *sensus communis* consente alla mente umana di

¹² In questo senso vd. G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1, il Mulino, 2022, p. 50: «Semplificando al massimo, potremmo dire che gli algoritmi sono sequenze di istruzioni, linee di codice, che possiamo scrivere e che possiamo leggere»; vd. anche T. GILLESPIE, *The relevance of algorithms*, in T. Gillespie – P. Boczowski – K. Foot (eds), *Media Technologies: Essays on Communication, Materiality, and Society*, MIT Press, p. 167: «*encoded procedures for transforming input data into a desired output, based on specified calculations*»; N. DIAKOPOULOS, *Algorithmic Accountability: Journalistic investigation of computational power structures*, in *Digit. Journal*, 2015, 3, p. 400: «*a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome*»; queste ultime due definizioni sono state riprese dal Consiglio d'Europa nel report *Algorithms and Human Rights* (Council of Europe DGI(2017)12), 2018.

¹³ Così S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion*, Springer, 2020, p. 14.

¹⁴ Vd. D.R. DESAI – J.A. KROLL, *Trust But Verify: A Guide to Algorithms and the Law*, in *Harvard journal of law & technology*, 2017, Vol. 31, n. 1, p. 23: «*we can call the steps for brushing teeth an algorithm*».

colmare eventuali lacune nell'algoritmo, le macchine, al contrario, applicano alla lettera le istruzioni impartite, non essendo in grado di interpretarle sulla base del contesto¹⁵.

D'altra parte, l'esecuzione di una regola da parte di un essere umano può essere imprecisa e difettosa – a causa di fallacie cognitive o di valutazioni di opportunità inerenti al caso concreto – mentre il sistema informatico promette di eseguire *perfettamente* l'istruzione fornita in sede di programmazione, eliminando qualsiasi discrezionalità e valutazioni di opportunità nel *decision making*¹⁶. Nell'esecuzione dell'operazione algoritmica, dunque, il *software* pare *infallibile* (purché, com'è ovvio, il sistema informatico di supporto sia dotato di sufficiente capacità computazionale); ciò non toglie, tuttavia, che le capacità di calcolo non eliminano il problema dell'accuratezza dell'*input*, nonché quello della stessa correttezza dell'istruzione fornita attraverso algoritmo.

Come vedremo nei prossimi paragrafi (vd. *infra*, §§ 1.2.3. ss.), a differenza dei tradizionali algoritmi, i più recenti algoritmi di i.a., che utilizzano tecniche di *machine learning*, non recano istruzioni *nozionistiche* (il citato schema dell'*if-this-then-that*), ma sono piuttosto forniti di indicazioni su *come apprendere*, *come classificare*, *come generalizzare*.

Gli algoritmi – che si tratti di intelligenza artificiale o meno – possono essere incorporati in un supporto fisico (c.d. *robot*) o avere natura puramente virtuale (c.d. *bot*). Nel primo caso, il sistema sarà in grado di interagire con il mondo reale secondo il noto paradigma del «*sense, think, act*»: il *robot* percepisce l'ambiente circostante (*sense*), processa le informazioni (*think*) e agisce (*act*)¹⁷. Proprio questa abilità di agire direttamente nell'ambiente fisico in cui si trova rende il *robot* capace di causare danni alle persone o alle cose (si pensi alle *self driving cars*, ai droni da combattimento, alla chirurgia robotica): si tratta di quelle macchine che, nel corso della nostra trattazione, chiameremo *Embodied systems* (vd. *infra*, § 7.1.1., in questo cap.). Nel secondo caso, invece, gli algoritmi agiscono su sistemi informatici: in alcune ipotesi, come per gli algoritmi di *trading* finanziario o i *bot* sui *social network*, potranno incidere direttamente

¹⁵ D.R. DESAI – J.A. KROLL, *Trust But Verify*, cit., p. 23-24; sull'assenza di senso comune dei sistemi di i.a. vd. C.E.A. KARNOW, *The application of traditional tort law to embodied machine intelligence*, in R. Calo e al. (eds.), *Robot Law*, Edward Elgar Publishing, 2016, p. 75; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1194 ss.

¹⁶ S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., p. 14.

¹⁷ R. CALO, *Robotics and the Lessons of Cyberlaw*, cit., 2015, p. 529-530, e bibliografia ivi citata.

sull'ambiente di riferimento, in assenza dell'intervento umano; in altre, il sistema di i.a. fornirà un supporto alla decisione umana, come avviene nel caso degli algoritmi di giustizia predittiva o a quelli di supporto alla diagnostica.

1.2. Dall'i.a. simbolica al connessionismo, in tre fasi: logica, conoscenza, apprendimento.

Fin dalle origini, la storia dell'intelligenza artificiale è stata caratterizzata da periodi di grande entusiasmo, alternati ad altrettanti momenti di scetticismo, legati all'insuccesso di alcune tecnologie, all'abbassamento delle aspettative e alla corrispondente riduzione degli investimenti pubblici e privati (c.d. *AI Winters*). Lungi dal voler fornire un'approfondita "storia intellettuale" dell'evoluzione dei sistemi di i.a. – compito che di certo esula dalle nostre competenze, nonché dagli obiettivi di questo lavoro – ci limiteremo qui a brevi cenni sulle tappe fondamentali che hanno condotto allo sviluppo di macchine *autonome* e *adattive*, che costituiscono il vero motivo di interesse dell'intelligenza artificiale nel campo della responsabilità penale¹⁸.

In particolare, semplificando molto, ci pare che lo sviluppo dei sistemi di i.a. abbia vissuto tre fasi, corrispondenti a tre diverse concezioni dell'intelligenza e dell'apprendimento. Nella fase embrionale, gli studiosi ritenevano di poter creare macchine intelligenti fornendo ai sistemi una serie di nozioni *logiche*, nella convinzione che avrebbero potuto essere utilizzate per risolvere una serie indefinita di problemi concreti (vd. *infra*, § 1.2.1.); in una seconda fase, le macchine venivano dotate di nozioni specialistiche, applicabili ad un novero molto ristretto di problemi (§ 1.2.2.); infine, nella fase attuale, i sistemi non necessitano di nozioni fornite *a priori*, ma *apprendono* dall'esperienza e dagli *input* ricevuti (§ 1.2.3.).

Anticipiamo fin da subito che, in questo percorso evolutivo, si assiste altresì ad una traslazione degli obiettivi dell'intelligenza artificiale, intesa come campo di studi: non più la creazione di macchine aventi le medesime capacità intellettive dell'essere umano (la già citata AGI, *artificial general intelligence*), quanto lo sviluppo di sistemi

¹⁸ Per una sintesi delle tappe più significative nell'evoluzione dell'i.a. vd. J. KAPLAN, *Artificial Intelligence*, cit., p. 13 ss. e S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 17 ss.

che sono in grado di svolgere efficacemente compiti molto specifici (ANI, *artificial narrow intelligence*).

1.2.1. Intelligenza come applicazione deduttiva di principi logici.

La nascita del termine “intelligenza artificiale” – e del relativo campo di studi – viene tradizionalmente fatta risalire alla conferenza chiamata “*Dartmouth Summer Research Project on Artificial Intelligence*”, che si tenne nell’estate del 1956 al Dartmouth College di Hanover (New Hampshire), su impulso di quattro scienziati (John McCarthy, Marvin Minsky, Claude Shannon e Nathaniel Rochester) e che vide la partecipazione dei maggiori studiosi e ricercatori in materia di reti neurali, scienze cognitive ed informatica¹⁹. L’obiettivo che, in quella sede, si proponevano i ricercatori era quello di porre le basi per lo sviluppo di macchine che si comportassero in modi che sarebbero stati considerati “intelligenti” se posti in essere da persone («*For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving*»²⁰).

Già nel 1950, lo scienziato inglese Alan Turing si era interrogato sulle potenzialità cognitive delle macchine (“*Can machine think?*”²¹), sviluppando a tal proposito quello che rimarrà come uno dei più celebri esperimenti volto a verificare l’eventuale comportamento intelligente della macchina (oggi conosciuto come *test di Turing*)²². L’esperimento prevedeva un *imitation game*: una donna e un computer, collocati in due stanze separate, comunicavano per iscritto con un intervistatore umano, il quale poteva rivolgere loro ogni genere di domande. Al termine dello scambio, l’intervistatore doveva scoprire l’identità dei due conversanti: se non fosse stato in grado di discernere correttamente i risultati dei suoi interlocutori, la macchina avrebbe superato il test,

¹⁹ J. MCCARTHY – M.L. MINSKY – N. ROCHESTER – C.E. SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 31 agosto 1955, disponibile a questo link: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>.

²⁰ J. MCCARTHY – M.L. MINSKY – N. ROCHESTER – C.E. SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, cit.

²¹ A. TURING, *Computing Machinery and Intelligence*, in *Mind*, 49, 1950, p. 433.

²² A. TURING, *Computing Machinery and Intelligence*, cit., p. 433.

dimostrando di essere interscambiabile con il conversante umano e, dunque, intelligente²³.

Accanto a studi finalizzati alla riproduzione del *comportamento* umano, vi erano invece progetti che miravano a creare artificialmente il *pensiero* umano. Tra questi vi fu sicuramente il *General Problem Solver* (GPS)²⁴ di Allen Newell e Herbert Simon, un *software* che, imitando i modelli di *problem solving* tipicamente utilizzati dalla mente umana, aveva l'ambizione di porsi come "risolutore" universale di quesiti; in questo caso, i due studiosi, più che all'estrinsecazione dell'intelligenza, erano interessati a verificare se i processi cognitivi umani potessero essere riprodotti su supporto meccanico, studiando i procedimenti decisionali e i tempi di risposta delle macchine.

In ogni caso – che si trattasse di riprodurre i meccanismi cognitivi dell'uomo, o che ci si accontentasse, invece, di una mera replicazione esteriore di tali meccanismi – le prime fasi di ricerca nel campo dell'i.a. furono caratterizzate dal tentativo di progettare algoritmi che mimassero le *capacità umane*. In particolare, tra gli anni Cinquanta e gli anni Ottanta del Novecento, il paradigma dominante nel campo dell'Intelligenza artificiale è stato *funzionalista* e *simbolico*: *funzionalista*, poiché propugnava un'idea di intelligenza come *computazione*, slegata dal supporto cerebrale e dunque riproducibile anche su *hardware* non biologici, purché dotati di adeguata potenza di calcolo; *simbolico*,

²³ Per una descrizione più approfondita dell'esperimento vd. S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 984 ss. Molti studiosi, nel corso degli anni, hanno messo in dubbio il fatto che l'esperimento potesse fungere da affidabile criterio per riconoscere una macchina intelligente. L'obiezione principale contesta la parzialità del fattore comunicativo quale indice di intelligenza, vd. ad. es. J. HODGES, *Alan Turing*, in *Stanford Encyclopedia of Philosophy*, 2002, rev. 2013, § 7. La più nota critica all'affidabilità del test fu sviluppata dal noto filosofo statunitense John Searle in ID., *Minds, Brains and Programs*, in *Behavioral and Brain Science*, 1980, vol. 3, pp. 417-424, in cui si contestava l'idea che il corretto utilizzo del linguaggio e della sintassi possa dimostrare anche la comprensione della semantica (e, dunque, secondo un approccio simbolico, l'intelligenza stessa). Searle, in particolare, per confutare il test di Turing, proponeva l'esperimento mentale della "stanza cinese": una persona che parla, come unica lingua, l'inglese si trova chiuso in una stanza e deve interagire con qualcuno all'esterno che parla solo il cinese. Dentro la stanza, l'individuo trova però un libro d'istruzioni in inglese, che gli indicano come rispondere a qualsiasi domanda ricevuta in cinese. Queste gli rendono possibile mettere in relazione una serie di simboli formali con un'altra serie di simboli formali, cioè gli permettono di dare una risposta (*output*) a ogni domanda (*input*): è quello che Searle chiama il "*programma*". Così, l'individuo potrà essere considerato, dagli osservatori esterni, un madrelingua cinese, nonostante egli si sia limitato – proprio come un *computer* – a seguire le istruzioni fornite, senza che abbia in alcun modo potuto *interpretare* e *comprendere* i simboli cinesi.

²⁴ A. NEWELL – H.A. SIMON, *GPS, A Program that Simulates Human Thought*, in E.A. Feigenbaum – J. Feldman (eds.), *Computers and Thought*, 1963.

poiché considerava l'intelligenza come capacità logica di “manipolare simboli”, ossia di applicare razionalmente, nel caso concreto, principi e relazioni logiche preesistenti²⁵.

Quello *simbolico* era un approccio deduttivo (c.d. *top-down*): nel codice algoritmico venivano iscritte una serie di informazioni iniziali (assiomi, frasi dichiarative, metodi di ragionamento logico), che – nelle intenzioni dei programmatori – avrebbero dovuto essere sufficienti a risolvere l'intero spettro dei quesiti prospettabili.

1.2.2. Intelligenza come applicazione deduttiva di conoscenze specialistiche.

L'approccio simbolico si rivelò tuttavia fallimentare se applicato in ambiti non circostanziati, poiché la genericità delle informazioni e dei modelli di calcolo forniti in sede di programmazione consentiva di risolvere soltanto problemi elementari. È invece con riferimento alla soluzione di quesiti molto specialistici e complessi che l'i.a., sul finire degli anni Settanta, segnò il suo primo grande successo, grazie allo sviluppo degli *expert systems* – algoritmi che, sulla base degli anzidetti metodi deduttivi (vd. *supra*, § 1.2.1.), applicavano in maniera sistematica una serie di regole specifiche e settoriali, inserite nella stringa di codice a seconda del campo di interesse²⁶. Gli *expert systems* – inizialmente sviluppati nel campo dell'identificazione delle molecole organiche e della diagnosi di malattie infettive – si diffusero rapidamente in moltissimi settori, in particolare nell'ambito della gestione delle risorse d'impresa, dando avvio ad un periodo di ingenti investimenti nel settore dell'i.a., sia pubblici che privati²⁷.

La diffusione su larga scala degli *expert systems* segnò un primo distacco dal tentativo di sviluppare una *artificial general intelligence*; da quel momento in poi, gli sforzi della comunità scientifica – e i relativi finanziamenti – si concentrarono piuttosto sulla prospettiva di progettare algoritmi in grado di svolgere compiti specifici (c.d. *domain specific systems*).

²⁵ A. NEWELL – H.A. SIMON, *Computer Science as Empirical Enquiry: Symbols and Search*, in *Communications of the ACM*, 19, 1976, p. 113 ss.

²⁶ Lo sviluppo degli *expert systems* si deve soprattutto al lavoro portato avanti, a partire dal 1965, dallo *Stanford Heuristic Programming Project*, guidato da Edward Feigenbaum; vd. S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 22.

²⁷ Vd. J. KAPLAN, *Artificial Intelligence*, cit., p. 24.

Tuttavia, l'*hype* suscitato dallo sviluppo di *expert systems* non resse le aspettative: gli obiettivi promessi dalle case produttrici erano troppo ottimistici da poter essere conseguiti con le tecnologie esistenti ed il ritorno economico non riuscì a pareggiare gli investimenti effettuati nel settore. Le ragioni di questo *inverno dell'i.a.* erano d'altra parte strutturali: gli *expert systems*, infatti, potevano applicare regole di tipo consequenziale (*if-then rules*), semplificando e velocizzando le attività umane, ma non erano in grado di gestire situazioni di incertezza. D'altra parte, per funzionare efficacemente, gli algoritmi richiedevano un continuo aggiornamento, che poteva essere effettuato soltanto dai programmatori, sulle informazioni di partenza (c.d. *hand coding*).

1.2.3. Intelligenza come apprendimento induttivo.

Sebbene risalgano a non più di trent'anni fa, i modelli descritti fino ad ora costituiscono quella che potremmo chiamare l'*archeologia* dell'Intelligenza Artificiale²⁸. Il presente dell'i.a. – e, presumibilmente, il suo futuro – è il *machine learning* (ML), una tecnica che consente agli algoritmi di apprendere dai dati, utilizzando un metodo induttivo (c.d. *bottom-up*): attraverso il ML, «*a computer observes some data, builds a model based on the data, and uses the model as both a hypotheses about the world and a piece of software that can solve problems*»²⁹.

La capacità dei sistemi di *machine learning* è quella di riuscire a trarre delle generalizzazioni statistiche (c.d. *statistical modelling*) a partire dall'esame di alcuni esempi³⁰. Durante la fase di “addestramento” (*training*), il programmatore sottopone all'algoritmo diversi casi, da cui quest'ultimo inferisce il modo di agire. Tale approccio parte dalla constatazione che molte delle attività che l'essere umano è in grado di svolgere

²⁸ L'approccio simbolico ai sistemi intelligenti è infatti oggi comunemente chiamato, con un misto di derisione e nostalgia, “*Good Old-Fashioned AI*” (GOFAI), vd. W. WALLACH – C. ALLEN, *Moral Machines*, cit., p. 183; M.A. LEMLEY – B. CASEY, *Remedies for Robots*, in 86 *University of Chicago Law Review*, 2019, p. 1322-1323.

²⁹ S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 651.

³⁰ P. DOMINGOS, *A Few Useful Things to Know about Machine Learning*, in *Communications of the ACM*, vol. 55, n. 10, nov. 2012, p. 79; nella letteratura giuridica vd. per tutti A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence based practices e tutela dei diritti fondamentali*, in *Arch. pen. (web)*, 2021, n. 1, p. 6.

non sono formalizzabili attraverso l'esplicitazione di regole definite, ma sono frutto di esperienza ed imitazione³¹.

Il paradigma simbolico-deduttivo, con riferimento a queste attività, mostrava effettivamente tutti i suoi limiti. Si pensi, ad esempio, al riconoscimento facciale, un'attività che svolgiamo senza difficoltà, in maniera intuitiva, ma per la quale sarebbe arduo stilare un elenco di regole precise che possano essere applicate, in via deduttiva, dall'algoritmo; lo stesso vale, ad esempio, per la scrittura. Pensiamo, ancora, alla guida di un autoveicolo, che – anche non volendo considerare le necessarie capacità di coordinamento motorio, che di certo possono essere riprodotte dalla macchina – è un'attività che richiede una complessa capacità di riconoscimento e previsione dei comportamenti degli altri utenti della strada. Così, a volte può bastare la mera postura di un pedone sul ciglio della strada per comprendere che questi sta per attraversare la carreggiata, nonostante il semaforo rosso; o ancora, l'incedere incerto di un'autovettura al lato destro della strada può indicare che il conducente è alla ricerca di un parcheggio e che potrà frenare all'improvviso una volta trovatone uno. È evidente che la gestione efficace di queste situazioni è difficilmente sintetizzabile in una formula, ancorché dettagliata, trattandosi di capacità che si acquisiscono attraverso l'esperienza.

A partire dagli anni Novanta, la ricerca in materia di i.a. cominciò a concentrarsi sulla creazione di algoritmi che, invece di applicare al caso concreto le regole astratte fornite in sede di programmazione, fossero in grado di riconoscere *pattern* statistici nei dati acquisiti e di effettuare generalizzazioni. Così, per tornare al primo degli esempi citati, se ad un algoritmo di riconoscimento facciale che utilizza tecniche di *machine learning* forniamo un *database* di immagini di Tizio, Caio e Sempronio, il sistema sarà in grado di discernere i tratti del viso, creando distinte categorie riconducibili a ciascuno dei tre individui (c.d. *unsupervised learning*); ovviamente, maggiore sarà il numero di immagini acquisite dall'algoritmo, più accurata sarà la classificazione. Se invece il nostro obiettivo è riconoscere una determinata persona in un vasto *database* di immagini, possiamo fornire all'algoritmo un set di immagini già etichettate, così che il sistema possa

³¹ T. Ray, *A new experiment: Does AI really know cats or dogs -- or anything?*, in *Zdnet*, 28 gennaio 2022, disponibile a questo link: <https://www.zdnet.com/article/what-does-ai-know-of-cats-and-dogs-maybe-very-little/>.

identificare i tratti del viso di tale persona e riconoscerli all'interno delle immagini fornite come *input* (c.d. *supervised learning*)³².

Un simile risultato sarebbe difficilmente raggiungibile attraverso l'approccio simbolico e "causale", che richiederebbe la definizione di un procedimento logico astratto, da applicare nel caso concreto. D'altra parte, lo stesso algoritmo non è in grado di formulare una *regola di classificazione* – che consenta *in tutti i casi*, in maniera infallibile, di etichettare le immagini – ma ragiona *per analogia*: il sistema non sa spiegare quali sono stati i criteri che, in un *database* contenente milioni di immagini, gli hanno consentito di individuare il viso di una determinata persona³³.

Allo stesso modo, tornando al secondo esempio da noi prospettato, anche per la guida di autoveicoli pare difficile formalizzare un procedimento logico-astratto idoneo a predire i comportamenti degli altri utenti della strada. L'immagazzinamento e la classificazione di enormi quantità di dati sul traffico, tuttavia, consente oggi agli algoritmi di guida automatica di *prevedere* con un altissimo livello di accuratezza come si muoveranno veicoli e pedoni (vd. *infra*, § § 7.1.1.1., in questo cap.).

L'affermarsi di tecniche di *machine learning* ha determinato un vero e proprio cambio di paradigma nel campo dell'i.a., consentendo lo sviluppo di sistemi autonomi più efficienti in moltissimi settori. I benefici dell'utilizzo di tale tecnica sono molteplici. In primo luogo, il *machine learning* solleva i programmatori dall'attività – insieme complessa e costosa – di produrre istruzioni esplicite. Gli sviluppatori non sempre sono in grado di conoscere in anticipo tutte le situazioni che il sistema potrà trovarsi ad affrontare: così, ad esempio, un algoritmo di *trading* finanziario dovrà adattarsi a condizioni di mercato di volta in volta differenti, al fine di massimizzare i profitti. L'idea che sta alla base del *machine learning* è che *insegnare ad effettuare generalizzazioni* è più efficace, sul lungo periodo, di *impartire insegnamenti nozionistici*. In secondo luogo,

³² G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, cit., p. 50-51.

³³ In argomento vd. A.D. SELBST, *Negligence and AI's Human Users*, in 100 *Boston University Law Review*, 2020, p. 1318 ss. Sull'incapacità di comprensione simbolica dei sistemi di *machine learning* vd. per tutti M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 12; EAD., *Robot, cyborg e intelligenze artificiali*, in A. Cadoppi – S. Canestrari – A. Manna – M. Papa (a cura di), *Cybercrime*, Utet, 2019, p. 1184.

il ML consente ai sistemi di individuare *pattern* molto sottili nei dati, che gli esseri umani potrebbero trascurare o non riconoscere³⁴.

Nel caso in cui, poi, i *software* siano connessi tra loro attraverso meccanismi di *cloud computing*, i sistemi di i.a. potranno scambiarsi reciprocamente informazioni ed apprendere non solo dalla propria esperienza, ma anche da quella dei propri “simili”³⁵.

Nonostante già negli anni Sessanta ci fossero studi sull'apprendimento automatico delle macchine, soltanto a partire dalla fine degli anni Ottanta, con il declino degli *expert systems*, il *machine learning* cominciò ad acquisire un certo seguito tra i ricercatori nel campo dell'i.a. Dagli anni Dieci del Duemila, poi, l'accelerazione nello sviluppo dei sistemi di *machine learning* fu favorita sostanzialmente da due fattori: innanzitutto, l'ampia disponibilità di dati (c.d. *big data*), raccolti sia nell'ambiente virtuale, sia grazie allo sviluppo di sensori ad alta definizione e a basso costo; in secondo luogo, l'incremento di capacità computazionale dei computer, sia in termini di potenza che in termini di velocità e memoria³⁶.

2. Machine learning, deep learning, reti neurali.

La tecnica di *Machine learning* ad oggi più utilizzata è quella fondata sulle reti neurali artificiali (*artificial neural network*, ANN)³⁷, ossia sistemi di elaborazione dei dati il cui funzionamento trae ispirazione dall'apparato nervoso biologico. Il tentativo di utilizzare reti neurali artificiali al fine di sviluppare algoritmi più intelligenti risale agli

³⁴ Vd. A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, in 87 *Fordham Law Review*, 2018, p. 1094; vd. anche S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 651. Anche gli esempi che seguono sono tratti da qui.

³⁵ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 6.

³⁶ Tra i tanti vd. J. KAPLAN, *Artificial Intelligence*, cit., p. 39; J. SINGH E AL., *Responsibility & Machine Learning: Part of a process*, in SSRN, 2016, p. 2, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860048; R. CALO, *Artificial Intelligence Policy: A Primer and Roadmap*, in *University of Bologna Law Review*, 3(2), p. 402; G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, cit., p. 49; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1182.

³⁷ Altre tecniche di *Machine Learning* sono le seguenti: Decision Trees, Naïve Bayes, Support Vector Machines, Rule learners; vd. J. SINGH E AL., *Responsibility & Machine Learning*, cit., p. 4-9; vd. anche R. DEVILLÉ E AL., *Basic Concepts f AI for Legal Scholars*, in J. De Bruyne – C. Vanleenhove (eds.), *Artificial Intelligence and the Law*, cit., p. 12 ss.

anni Quaranta del Novecento³⁸, ma negli ultimi anni ha subito una forte accelerazione, grazie alla massiccia quantità di dati di addestramento oggi disponibili.

Le reti neurali sono *software* di modellazione statistica che possono essere installati direttamente su un *personal computer*; tuttavia, sempre più spesso, la complessità degli algoritmi che ne costituiscono il fondamento richiede l'utilizzo di *hardware* dedicati (c.d. *acceleratori hardware*), che consentono un addestramento e un'esecuzione delle funzioni algoritmiche più rapidi³⁹.

Le reti neurali ricevono informazioni esterne (*input*) su uno strato di “nodi d'ingresso” ed emettono risposte nell'ambiente esterno (*output*), attraverso l'elaborazione di dati compiuta da un diverso strato di nodi; in alcuni casi, tra le unità di *input* e di *output* si aggiungono uno o più strati di nodi (unità nascoste o *hidden*). Solitamente, le reti neurali sono dunque formate da tre strati (che, tuttavia, possono coinvolgere migliaia di neuroni e decine di migliaia di connessioni):

- lo strato degli ingressi (*input*), che riceve ed elabora i segnali in ingresso, adattandoli alle richieste dei neuroni della rete;
- lo strato *hidden* (*strato nascosto*), che procede all'elaborazione vera e propria e può essere, a sua volta, strutturato su più livelli di neuroni;
- lo strato di uscita (*output*), che raccoglie i risultati dell'elaborazione dello strato H e li “comunica” all'esterno, sotto forma di dati di *output*.

Quando le reti neurali sono organizzate e stratificate su più livelli si parla di *deep learning* – un campo di studi, dunque, che può essere considerato come una branca del *machine learning*. La potenza computazionale dei modelli di *deep learning* deriva dalla capacità dei neuroni, strato dopo strato, di apprendere da quanto già processato dai nodi nei livelli precedenti⁴⁰. In particolare, ogni strato della rete si concentra su un aspetto specifico dell'informazione di *input* da elaborare: così, in un sistema di *deep learning* chiamato ad analizzare un'immagine, il livello più basso potrebbe focalizzarsi sull'analisi

³⁸ In particolare, la nascita del campo di studi delle reti neurali artificiali viene comunemente fatto risalire al lavoro di W.S. MCCULLOCH – W. PITTS, *A logical calculus of the ideas immanent in nervous activity*, in *Bulletin of Mathematical Biophysics*. 1943, 5 (4), p. 115-133; vd. K.P. MURPHY, *Machine Learning. A Probabilistic Perspective*, MIT Press 2021, p. 568-569.

³⁹ Vd. M. Negretti, *Intelligenza artificiale, quali sono le architetture hardware dedicate*, in *AI4Business*, 20 gennaio 2021, disponibile a questo link: <https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-quali-sono-le-architetture-hardware-dedicate/>, che fornisce una spiegazione semplificata dei vantaggi dell'*hardware* nell'addestramento e nel funzionamento delle reti neurali.

⁴⁰ J.D. KELLEHER – B. TIERNAY, *Data Science*, MIT Press, 9th ed., 2018, p. 238.

del colore, quello successivo sulle linee e quello che segue ancora sulle ombreggiature; in conclusione, il processo di riconoscimento dovrebbe concludersi con la capacità del sistema di i.a. di identificare la forma rappresentata e di riconoscere l'immagine⁴¹. Il *deep learning* consente di acquisire previsioni molto più accurate rispetto alle tecniche tradizionali di *machine learning*, poiché coinvolge una rete più fitta di strati e connessioni neurali. Le applicazioni di intelligenza artificiale che, nel corso degli ultimi anni, hanno avuto un maggior successo in termini di efficienza sono proprio quelle dotate di sistemi di *deep learning*: è il caso, ad esempio, dei *software* di guida autonoma, dei *chatbots*, degli algoritmi di supporto alla diagnostica e al trattamento clinico.

2.1. Le modalità di apprendimento nel *machine learning*.

Abbiamo visto come lo sviluppo delle tecniche di *machine learning* nel campo dell'intelligenza artificiale abbia costituito un vero e proprio cambio di paradigma nel funzionamento delle macchine intelligenti, determinando un'accentuazione della capacità predittiva degli algoritmi. In questo paragrafo cercheremo di descrivere quali sono le modalità di apprendimento maggiormente utilizzate nel campo del *machine learning*, al fine di comprendere in che modo gli algoritmi giungono a formulare ipotesi predittive e a prendere decisioni. In letteratura, in particolare, si è soliti distinguere tra tre diverse modalità di apprendimento⁴²:

(i) *supervised learning* (apprendimento supervisionato): il sistema analizza la coppia *input-output* assegnata dall'addestratore e impara a riconoscere l'*output* che deve essere correlato a ciascun *input* (c.d. *classification* o *pattern recognition*⁴³). Pensiamo, ad esempio, al *training* di un algoritmo di *self-driving car*, finalizzato all'identificazione di pedoni e bus nel traffico stradale; l'agente impara che a ciascun *input* (immagine) corrisponde un determinato *output* classificatorio (c.d. *label*, etichetta; in questo caso,

⁴¹ L. DORMEHL, *Thinking Machines: The Quest for Artificial Intelligence and Where It's Taking Us Next*, Tarcherperigree, 2017, p. 50-51.

⁴² S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 653; K.P. MURPHY, *Machine Learning*, cit., p. 2; vd. anche J. BUYERS, *Artificial Intelligence. The practical legal issues*, Law Brief Publishing, 2018, p. 10 ss.

⁴³ K.P. MURPHY, *Machine Learning*, cit., p. 2

“bus” o “pedone”). Un esempio di *supervised learning* “diffuso”, proprio nel contesto dell’addestramento di algoritmi per la guida autonoma, può essere considerata l’applicazione ReCAPTCHA⁴⁴ di Google: il *software* chiede agli utenti di classificare alcune immagini – quasi sempre tratte dalla circolazione stradale (semafori, ponti, strisce pedonali e, per l’appunto, bus e pedoni) – al fine di riconoscere se l’utente è una persona o un *bot* che sta tentando di violare un account protetto da password⁴⁵. Nonostante la funzione del *software* sia quella di garantire la sicurezza informatica degli utenti, sono diverse le testate specializzate che hanno ipotizzato che i dati acquisiti con ReCAPTCHA siano utilizzati da Google per alimentare i propri algoritmi di guida autonoma installati sulle *Google Cars* (vd. fig. 1)⁴⁶.

⁴⁴ Acronimo di *Completely Automated Public Turing test to tell Computers and Humans Apart*, vd. la pagina di assistenza Google (*Che cos’è un CAPTCHA?*), disponibile al link: <https://support.google.com/a/answer/1217728?hl=it>.

⁴⁵ *Ibidem*.

⁴⁶ Così M. Lotkowski, *You Are Building a Self Driving AI Without Even Knowing About It*, Hackernoon, 25 febbraio 2017, disponibile a questo link: <https://hackernoon.com/you-are-building-a-self-driving-ai-without-even-knowing-about-it-62fadbfa5fdf>; R. Maruzani, *Are You Unwittingly Helping to Train Google’s AI Models?*, in *Towards Data Science*, 26 gennaio 2021, disponibile a questo link: <https://towardsdatascience.com/are-you-unwittingly-helping-to-train-googles-ai-models-f318dea53aee>; i vertici di Google, d’altra parte, avevano già ammesso di utilizzare i dati ottenuti attraverso ReCAPTCHA per addestrare *Google Street View*, vd. S. Perez, *Google Now Using ReCAPTCHA To Decode Street View Addresses*, in *TechCrunch*, 29 marzo 2012, disponibile a questo link: <https://techcrunch.com/2012/03/29/google-now-using-recaptcha-to-decode-street-view-addresses/>.

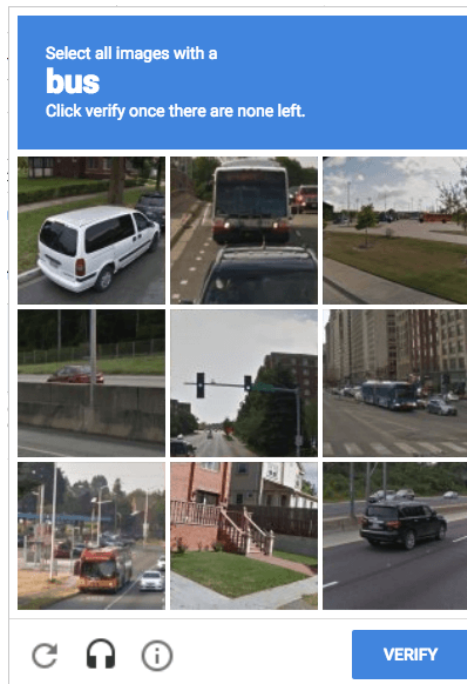


Fig. 1. Un esempio di riconoscimento immagini dell'applicazione ReCAPTCHA

(ii) *Unsupervised learning* (apprendimento non supervisionato): in questo caso, al sistema vengono forniti gli *input*, ma non gli *output* desiderati; l'obiettivo dell'algoritmo è trovare, attraverso metodi probabilistici o reti neurali, “*interesting patterns*”⁴⁷ nei dati (c.d. *knowledge discovery*⁴⁸). A differenza del *supervised learning*, questa tecnica consente di abbattere i costi derivanti dalla necessità di ottenere *labeled data* e, d'altra parte, consente la classificazione di enormi masse di dati. La più comune tecnica di *unsupervised learning* è il *clustering*, che individua categorie specifici all'interno di un *dataset (input)*: si pensi, ad esempio, agli algoritmi che classificano gli utenti di un sito *e-commerce* o di un motore di ricerca sulla base delle loro preferenze di acquisto o del loro comportamento online (pagine visitate, tempo di permanenza su un sito, etc.), al fine di inviare messaggi pubblicitari personalizzati; o, ancora, a piattaforme come Netflix o Spotify, che hanno interesse a segmentare i propri clienti al fine di fornire raccomandazioni personalizzate⁴⁹.

⁴⁷ K.P. MURPHY, *Machine Learning*, cit., p. 2.

⁴⁸ K.P. MURPHY, *Machine Learning*, cit., p. 2.

⁴⁹ G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, cit., p. 51.

(iii) *Reinforcement learning* (apprendimento con rinforzo): il sistema impara a risolvere problemi che richiedono di elaborare strategie molto sofisticate, attraverso il calcolo delle azioni più promettenti per raggiungere un particolare obiettivo. Il *reinforcement learning* viene utilizzato quando l'ambiente in cui opera l'agente di i.a. non è conosciuto a priori (si pensi al movimento di un *robot* in ambienti sconosciuti), o quando la modellazione computativa richiederebbe un impiego di risorse eccessivo (è il caso, ad esempio, della previsione di tutte le possibili mosse di una partita di scacchi). Uno degli elementi cardine del *reinforcement learning* è la predisposizione di incentivi e disincentivi da parte dei programmatori; grazie all'attribuzione di “premi” e “punizioni”, il sistema impara a riconoscere quali sono state le azioni più promettenti (quelle premiate) e quali quelle meno promettenti (quelle punite), così da modificare e migliorare le proprie *performance* in futuro⁵⁰.

2.2. Anche i sistemi di *machine learning* sbagliano: l'*adversarial machine learning*.

Come abbiamo già accennato, la ricerca nel campo dell'i.a. si è trasformata, nel corso del tempo, da *knowledge-based* a *data-based*. Se, da un lato, il *machine learning* consente agli algoritmi di individuare *pattern* ricorrenti che potrebbero non essere percepibili dall'uomo, dall'altro, il ragionamento analogico tipico dei sistemi di auto-apprendimento sconta una carenza assoluta di comprensione “semantica”, che talvolta può determinare la commissione di errori che, ad uno sguardo umano, appaiono come grossolani e macroscopici.

Così, può capitare che un sistema di riconoscimento di immagini basato sul *machine learning* scambi una tartaruga (o meglio, una stampa 3D che riproduce una tartaruga) per un fucile⁵¹. Ancora, alcuni ricercatori, ad esempio, hanno dimostrato che possono bastare degli adesivi o dei graffiti su un segnale stradale, una particolare inclinazione del cartello o una sua scarsa illuminazione per causare un errore di riconoscimento da parte di un sistema di guida autonoma; allo stesso modo, il sistema di

⁵⁰ G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, cit., p. 52.

⁵¹ A. ATHALYE E AL., *Synthesizing Robust Adversarial Examples*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

machine learning può scambiare per vero un poster che riproduce l'immagine del segnale stradale⁵².

Distance/Angle	Subtle Poster	Subtle Poster Right Turn	Camouflage Graffiti	Camouflage Art (LISA-CNN)	Camouflage Art (GTSRB-CNN)
5' 0°					
5' 15°					
10' 0°					
10' 30°					
40' 0°					
Targeted-Attack Success	100%	73.33%	66.67%	100%	80%

Fig. 2. Tratto da K. EYKHOLT E AL., *Robust Physical-World Attacks on Deep Learning Visual Classification*, cit.

Come emerge dalla tabella riassuntiva tratta dallo studio citato (fig. 2), piccolissime variazioni nell'immagine del segnale dello “stop” – insignificanti agli occhi di un conducente umano – possono determinare una percezione erranea da parte del sistema di i.a., proprio per il fatto che le tecniche di *machine learning* sono in grado di individuare associazioni statistiche nei dati, ma non sono invece capaci di tracciare modelli astratti di spiegazione causale di tali ricorrenze⁵³.

⁵² K. EYKHOLT E AL., *Robust Physical-World Attacks on Deep Learning Visual Classification*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

⁵³ D.F. LLORCA E AL., *Liability regimes in the age of AI: a use-case driven analysis of the burden of proof*, 3 novembre 2022, submitted to *Journal of Artificial Intelligence Research*, disponibile in Arxiv a questo link: <https://arxiv.org/abs/2211.01817>; A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1097; M. COECKELBERGH, *Artificial Intelligence, Responsibility Attribution, and a*

L'attenzione degli studiosi si è concentrata soprattutto sulla possibilità che gli algoritmi di ML siano intenzionalmente “ingannati” da parte di soggetti in grado di sfruttarne le vulnerabilità (c.d. *Adversarial machine learning*⁵⁴); a dire il vero, tuttavia, anche al di fuori di eventuali attacchi volontari, la varietà delle situazioni che possono occorrere nella vita reale è tale che anche il *dataset* più aggiornato e completo non sarà in grado di tenere conto dell'intera gamma di circostanze che potrebbero concretamente presentarsi, poiché ci saranno sempre casi specifici che sfuggono alla statistica.

Nonostante, dunque, i ricercatori stiano cercando di sviluppare modelli di *deep learning* resistenti rispetto a fenomeni di *adversarial machine learning*, vi sarà sempre una fisiologica percentuale di errori, benché sempre più bassa, via via che cresce l'accuratezza statistica degli algoritmi. Tornando all'esempio delle macchine a guida autonoma, il risultato paradossale sarà, dunque, che gli *autonomous vehicles* saranno generalmente in grado di reagire alle situazioni di emergenza in maniera molto più efficace rispetto agli esseri umani – contribuendo, così, ad una drastica riduzione degli incidenti su strada – ma potrebbero avere alcune difficoltà a riconoscere situazioni e simboli che per noi sono immediatamente carichi di significato. Come abbiamo già avuto modo di rilevare, infatti, i sistemi di *machine learning* si limitano a riconoscere *pattern* ricorrenti in un *dataset*, ma sono totalmente privi di capacità di discernimento e ragionamento.

Relational Justification of Explainability, in 26 *Science and Engineering Ethics*, 2020, p. 2060; Z.C. LIPTON, *The Mythos of Model Interpretability*, in 16 *Queue*, May-June 2018, p. 8-9; sul rapporto tra correlazioni e causalità, nell'ambito del ragionamento giuridico, vd. A. GARAPON – J. LASSÈGUE, *Giustizia digitale*, cit., p. 179 ss.

⁵⁴ *Ex multis* K. EYKHOLT E AL., *Robust Physical-World Attacks on Deep Learning Visual Classification*, cit.; A. NGUYEN E AL., *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*, in 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 427 ss.; A. ATHALYE E AL., *Synthesizing Robust Adversarial Examples*, cit.; H. XU, *Adversarial Attacks and Defenses in Images, Graphs and Text: A Review*, in *International Journal of Automation and Computing*, 17(2), April 2020, p. 151 ss.; nello specifico settore delle macchine a guida autonoma vedi il Report redatto dal TENCENT KEEN SECURITY LAB, *Experimental Security Research of Tesla Autopilot*, marzo 2019, disponibile a questo link: https://keenlab.tencent.com/en/whitepapers/Experimental_Security_Research_of_Tesla_Autopilot.pdf.

3. Il rischio di discriminazione algoritmica.

Nei sistemi che fanno utilizzo delle tecniche di *machine learning*, l'accuratezza dei risultati dipende sia dall'affidabilità del codice algoritmico, sia dalla precisione dei dati utilizzati per addestrare il modello (c.d. problema del *bias in, bias out* o *garbage in, garbage out*)⁵⁵. Dal momento che l'obiettivo dei sistemi di *machine learning* è quello di *generalizzare* – fornendo un *output* predittivo – quanto appreso dai dati di *input*, una prima, semplice, regola è che il modello non può imparare ciò che non ha mai visto. Così, se vogliamo sviluppare un algoritmo che riconosca le immagini di pomodori e, per addestrare il sistema, utilizziamo soltanto foto di pomodori rossi, il *software* non sarà in grado di riconoscere i pomodori *gialli*, poiché avrà appreso che il rosso è uno degli elementi che *definisce* i pomodori⁵⁶.

L'utilizzo di dati non completi o accurati in fase di addestramento ha già creato alcuni problemi di discriminazione nell'utilizzo dei sistemi di i.a.: in un caso, ad esempio, un algoritmo utilizzato per vagliare la rispondenza delle fototessere ai criteri stabiliti per il rilascio dei documenti di identità – che era stato addestrato principalmente con foto di persone di razza caucasica – aveva scartato quelle di alcune persone asiatiche, interpretando gli occhi a mandorla come occhi chiusi; ancora, l'algoritmo utilizzato dagli uffici HR di Amazon nella prima fase di selezione delle candidature discriminava le *cover letter* scritte dalle donne, poiché era stato addestrato, perlopiù, su lettere motivazionali scritte da uomini⁵⁷.

In altri casi, l'incompletezza dei dati forniti in sede di addestramento potrebbe causare problemi di *accuratezza* dei risultati predittivi proposti dal sistema algoritmico. È quanto è emerso, ad esempio, dalle sperimentazioni del sistema di supporto alla diagnostica clinica IBM Watson (vd. *infra*, § 7.2.1., in questo cap.) in alcuni stati europei

⁵⁵ Sul tema della discriminazione algoritmica la letteratura – specialmente statunitense – è molto ampia. Ci limitiamo qui a rinviare ai testi che hanno avuto un maggior impatto nel dibattito accademico (e non solo), vd. F. PASQUALE, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Harvard University Press, 2015; C. O'NEIL, *Armi di distruzione matematica*, Bompiani, 2016; S.U. NOBLE, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, 2018; J. CHENEY-LIPPOLD, *We Are Data: Algorithms and the Making of Our Digital Selves*, NYU Press, 2018; R. BENJAMIN, *Race After Technology: Abolitionist Tools for the New Jim Code*, Polity, 2019.

⁵⁶ R. DEVILLÉ E AL., *Basic Concepts f AI for Legal Scholars*, cit., p. 11.

⁵⁷ Entrambi gli esempi sono tratti da R. DEVILLÉ E AL., *Basic Concepts f AI for Legal Scholars*, cit., p. 11, a cui si rinvia per gli opportuni riferimenti bibliografici.

ed asiatici. Benché, infatti, tale strumento avesse conseguito un ottimo livello di accuratezza sul suolo statunitense, il medesimo risultato non è stato ottenuto quando il prodotto è stato utilizzato in contesti geografici diversi. Secondo alcuni studiosi, l'erroneità dei risultati sarebbe dovuta al fatto che la maggior parte degli articoli scientifici a cui il sistema attingeva erano stati scritti da autori americani, alla luce delle pratiche mediche accreditate negli Stati Uniti, e sulla base di un campione di pazienti limitato agli utenti delle strutture ospedaliere statunitensi⁵⁸.

D'altra parte, un rischio di *bias* è presente altresì in tutte quelle ipotesi in cui l'algoritmo, nel procedimento di *decision-making*, prende in considerazione elementi che dovrebbero essere ritenuti *irrilevanti*, giungendo a riservare un *diverso trattamento* in relazione a *situazioni uguali*. Il problema è che spesso gli utenti *non conoscono* i dati che sono presi in considerazione dall'algoritmo, né i meccanismi decisionali utilizzati per elaborarli: il rischio, dunque, è che gli algoritmi discriminino gli utenti e che questi, da parte loro, non siano nelle condizioni di contestare le decisioni automatizzate, perché ignari del preciso funzionamento dei sistemi computazionali⁵⁹. È evidente che tali discriminazioni – unite alla carenza di trasparenza – possono avere gravi conseguenze in termini di eguaglianza e mobilità sociale. Sempre più spesso, infatti, i sistemi di i.a. sono utilizzati per classificare attività socialmente rilevanti sulla base di dati personali: si pensi ai filtri *antispam*, al rilevamento delle frodi con le carte di credito, al *targeting* pubblicitario, al punteggio di credito assicurativo, al *credit scoring*.

Ben consapevole delle gravi disfunzionalità che può arrecare, in tutti i settori, la discriminazione algoritmica, la Commissione Europea ha tentato di rispondere al fenomeno prevedendo, all'art. 10, § 3, dell'*AI Act*, che «*I set di dati di addestramento, convalida e prova devono essere pertinenti, rappresentativi, esenti da errori e completi. Essi possiedono le proprietà statistiche appropriate, anche, ove applicabile, per quanto riguarda le persone o i gruppi di persone sui quali il sistema di IA ad alto rischio è destinato a essere usato*». Nonostante l'intento del legislatore europeo sia lodevole, la disposizione appare vaga e non del tutto rispondente al concreto funzionamento degli algoritmi più sofisticati. Non si capisce, innanzitutto, quali siano i criteri che debbano

⁵⁸ D. Gorski, *IBM's Watson versus cancer: Hype meets reality*, in *Science-based Medicine*, disponibile a questo link: <https://sciencebasedmedicine.org/ibm-watson-versus-cancer-hype-meets-reality/>.

⁵⁹ J. BURRELL, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, in *Big Data & Society*, January-June 2016, p. 1.

essere utilizzati per garantire *pertinenza* e *rappresentatività*. Come abbiamo già rilevato, infatti, i sistemi di *machine learning* si basano su tecniche statistiche che, per loro natura, non possono tenere conto delle singole *circostanze individuali*; la proposta di regolamento, tuttavia, non esplicita se ci siano dei metodi di analisi che siano da privilegiare rispetto ad altri.

D'altra parte, totalmente privo di effettività appare il richiamo all'assenza di errori nella raccolta di dati, dal momento che – stando alle testimonianze dei *data experts* – non esistono raccolte di dati (così come codici algoritmici) che possano considerarsi come assolutamente privi di errori⁶⁰. Insomma, per rifarci ad un brocardo caro ai penalisti (e non solo): *Ad impossibilia nemo tenetur!*

3.1. Il caso dei sistemi di supporto all'attività giudiziaria e all'attività di *law enforcement*.

Uno dei settori in cui sta emergendo in maniera plastica il rischio di *data bias* nel funzionamento degli algoritmi è quello della giustizia penale, ove già da tempo si stanno sperimentando sistemi di polizia e giustizia predittiva: i primi sono strumenti di applicazione statistica che hanno l'obiettivo di prevedere dove, quando e ad opera di chi potrà essere commesso un reato, al fine di prevenirne la commissione; i secondi sono, invece, algoritmi impiegati per valutare la pericolosità criminale di un soggetto che abbia commesso un reato, ossia la probabilità che questi commetta in futuro nuovi reati (c.d. *risk assessment tools*)⁶¹.

Se, da un lato, è vero che i sistemi computazionali possono contribuire a rivelare – e, di conseguenza, a correggere – l'incidenza di eventuali pregiudizi delle autorità di polizia o di quelle giudiziarie sulle deliberazioni assunte, dall'altro, gli algoritmi possono essere portatori di *implicit bias*, con il risultato di perpetuare, in questo modo, pregiudizi già esistenti.

L'*input* processato dall'algoritmo, infatti, può essere di per sé discriminatorio, rispetto all'obiettivo prefissato: si pensi, ad esempio, ad un *risk assessment tool* che

⁶⁰ G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, cit., p. 55.

⁶¹ F. BASILE, *Intelligenza artificiale e diritto penale*, cit, p. 10.

comprenda, tra i criteri di valutazione della pericolosità, dati riguardanti determinate caratteristiche etniche, religiose o sociali. La discriminazione può agire *direttamente* o *indirettamente*: direttamente, quando vengono presi in considerazione elementi di per sé discriminatori (es. il colore della pelle); indirettamente, quando un dato – di per sé *neutro* – viene valutato in senso negativo, perché lascerebbe presumere l'appartenenza ad un determinato gruppo etnico o sociale (es. l'indirizzo di residenza dell'imputato)⁶².

Sebbene si tratti di sistemi il cui utilizzo non sembra suscettibile di integrare alcuna fattispecie penale, ci preme, in questa sede, dare brevemente conto del loro funzionamento, stante il ruolo sempre maggiore che gli algoritmi predittivi stanno acquisendo nelle indagini e nei processi penali⁶³.

Quanto ai sistemi di polizia predittiva, si distinguono sostanzialmente due tipi di modelli. Da un lato, ci sono i sistemi che – elaborando dati ambientali, spaziali e temporali ricorrenti nei reati precedentemente commessi – individuano “zone calde” (c.d. *hotspots*),

⁶² L'esempio è tratto da S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta Etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Leg. pen.*, 18 dicembre 2018, p. 6, nota 14; vd. anche A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 12 ss.

⁶³ La letteratura penalistica e processual-penalistica sull'utilizzo di sistemi di i.a. nel settore della giustizia penale è ormai sterminata. Ci limitiamo qui a dar conto dei principali lavori pubblicati su questi temi: S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., D. KEHL E AL., *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Berkman Klein Center for Internet & Society, Harvard Law School, 2017; D.K. CITRON, *Technological Due Process*, 85 *Wash. U. L. Rev.*, 2008, p. 1249; in particolare, sulla polizia predittiva vd. A.G. FERGUSON, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, NYU Press, 2017; S. BRAYNE, *Predict and Surveil: Data, Discretion, and the Future of Policing*, Oxford University Press, 2020. In lingua italiana si vedano G. UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Dir. pen. cont. – Riv. Trim.*, 2020, n. 4; D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi*, Giappichelli, 2022; S. SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc.*, 2020, p. 604 ss.; G. CANZIO, *Intelligenza artificiale, algoritmi e giustizia penale*, in *Sist. pen.*, 8 gennaio 2021; O. DI GIOVINE, *Il judge-Bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Cass. pen.*, 2020, 951 ss.; V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit.; A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit.; P.P. PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen. (web)*, 2022, n. 2; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29 maggio 2019; L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, in *Dir. pen. cont.*, 2019, 2, 359 ss.; R. BICHI, *Intelligenza digitale, giurimetria, giustizia predittiva e algoritmo decisorio. Machina sapiens e il controllo sulla giurisdizione*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 423; M. AMISANO, *Prevedere -e non predire- attraverso gli algoritmi e le loro insidie*, in *Arch. pen.*, 2022, n. 2; in generale, sul rapporto tra processo e nuove tecnologie vd. A. GARAPON – J. LASSÈGUE, *Giustizia digitale*, Il Mulino, 2021 (ed. franc. 2018).

ossia luoghi ove è possibile prevedere che si realizzeranno attività criminose (c.d. *Place-Based Predictive Policing*⁶⁴). Ciò consente alle autorità di pubblica sicurezza di tracciare una mappa in cui vengono evidenziate le zone che, in determinati orari, sono a più alto rischio di criminalità, così da poter rendere più efficiente la predisposizione della forza pubblica e impedire la commissione di un maggior numero di reati, cogliendone in flagranza gli autori.

Dall'altro lato, sempre nell'ambito degli algoritmi utilizzati dalle forze dell'ordine, vi sono i sistemi di c.d. *crime linking*, che – attraverso l'elaborazione di dati di diversa provenienza, quali, ad esempio, frequenze di fotogrammi o informazioni relative a precedenti analoghi reati – consentirebbero di “profilare” il possibile autore di una serie criminale, prevedendone la successiva mossa (c.d. *Person-Based Predictive Targeting*⁶⁵). Gli algoritmi di *crime linking* si basano dunque sull'idea di fondo che alcune forme di criminalità sono solite manifestarsi in un arco temporale e in una zona geografica molto circoscritti (c.d. *near repeat crimes*, o reati a ripetizione ravvicinata)⁶⁶.

Il rischio discriminatorio, soprattutto per quanto riguarda i sistemi basati sugli *hotspots*, è evidente: gli algoritmi di *predictive policing*, infatti, in una certa misura si auto-alimentano con i risultati derivanti dal loro stesso utilizzo, innescando così veri e propri circoli viziosi: «se, ad esempio, un *software* predittivo individua una determinata “zona calda”, i controlli e i pattugliamenti della polizia in quella zona si intensificheranno, con inevitabile conseguente crescita del tasso dei reati rilevati dalla polizia in quella zona, che diventerà, quindi, ancora più “calda”, mentre altre zone, originariamente non ricondotte nelle “zone calde”, e quindi non presidiate dalla polizia, rischiano di rimanere, o di diventare, per anni zone franche per la commissione di reati»⁶⁷.

I *risk assessment tools*, invece, come già accennato, sono algoritmi impiegati per vagliare il pericolo di commissione di nuovi reati, da parte di chi ne abbia già commesso uno o più – una valutazione che spetta al giudice in diverse circostanze: quando si tratta di applicare una misura di sicurezza, una misura cautelare o una misura di prevenzione,

⁶⁴ A.G. FERGUSON, *The Rise of Big Data Policing*, cit., p. 62.

⁶⁵ A.G. FERGUSON, *The Rise of Big Data Policing*, cit., p. 34.

⁶⁶ Per l'analisi di alcuni sistemi di polizia predittiva utilizzati in Italia (in particolare, KeyCrime utilizzato dalla Questura di Milano e X-LAW utilizzato dalla Questura di Napoli) rinviamo a F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 12-13.

⁶⁷ F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 13.

o anche quando debba stabilire se concedere o meno la sospensione condizionale della pena o l'affidamento in prova al servizio sociale⁶⁸.

Fino ad oggi, a tale interrogativo i giudici hanno fornito risposte perlopiù intuitive, affidate alla propria sensibilità personale e a valutazioni di senso comune, nonché, quando consentito dalla legge, sulla base delle valutazioni cliniche fornite dai periti⁶⁹. Gli algoritmi predittivi promettono invece di effettuare una valutazione del rischio basata su elementi imparziali ed oggettivi⁷⁰, che dovrebbero essere indicativi di una più o meno elevata propensione alla criminalità dell'individuo: tali fattori possono riguardare l'età, il sesso, l'origine etnica, il livello di scolarizzazione, la situazione familiare e lavorativa, i precedenti penali, il consumo di sostanze stupefacenti o alcoliche etc.⁷¹. Una volta raccolti, l'algoritmo *combina* tali elementi – attribuendo un punteggio a ciascuno di essi – e predispone una *scala* di pericolosità, che può essere utilizzata dal giudice per pervenire alla decisione.

Negli Stati Uniti, i *risk assessment tools* sono già ampiamente utilizzati ai fini della concessione del *parole*, della *probation* e di altre misure alternative alla detenzione, così come ai fini di commisurazione della pena dell'imputato ritenuto colpevole (c.d. *sentencing*). L'algoritmo predittivo di gran lunga più utilizzato è il *Correctional Offender Management Profiling for Alternative Sanctions* (c.d. COMPAS), che prende in considerazione una serie di informazioni ottenute direttamente dall'imputato attraverso un'intervista, nonché i dati riguardanti il certificato del casellario e i carichi pendenti⁷².

Il sistema COMPAS è stato contestato in diverse occasioni a causa delle presunte discriminazioni insite nel meccanismo di valutazione adottato, nonché dell'irrelevanza – sotto il profilo criminologico – di alcuni fattori utilizzati dal modello. Da un'inchiesta pubblicata nel maggio 2016 sul sito *ProPublica* – condotta su un campione di circa settemila persone arrestate nella contea di Broward, in Florida, tra il 2013 e il 2014 –

⁶⁸ F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 16; A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 2 ss., a cui si rinvia per il riferimento ad alcuni dei sistemi di valutazione del rischio di recidiva già utilizzati in Italia (spec. p. 4).

⁶⁹ A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 7.

⁷⁰ Proprio tale pretesa oggettività ed imparzialità consentirebbe agli algoritmi predittivi di trovare un terreno fertile nel nostro Paese, ove pare ormai senza rimedio la crisi di legittimazione degli organi giurisdizionali, vd. V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 10-11.

⁷¹ F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 17.

⁷² S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cass. pen.*, 2019, p. 1748 ss.

emerge ad esempio che le previsioni di COMPAS favorivano sistematicamente gli imputati bianchi rispetto a quelli neri, sottostimando il rischio di recidiva dei primi e sovrastimando il rischio di recidiva dei secondi⁷³.

Critiche simili sono state sollevate nel c.d. caso Loomis, in cui l'imputato contestava l'entità della pena che il giudice di primo grado – avvalendosi del sistema COMPAS – gli aveva inflitto⁷⁴. Il ricorrente sosteneva che l'utilizzo processuale dell'algoritmo predittivo conduceva a risultati discriminatori, basati sul genere e sulla razza, e che l'opacità del suo meccanismo di funzionamento, dettato da ragioni di segreto industriale, contrastava con il principio del giusto processo (*right to due process*).

La Corte Suprema del Wisconsin, in quello che è diventato un vero e proprio *leading case* in materia, respingeva il ricorso di Loomis, allegando che – nonostante fosse effettivamente presente il rischio di una sopravvalutazione della pericolosità sociale di individui appartenenti a determinate minoranze etniche – l'utilizzo di algoritmi predittivi in sede processuale era legittimo, purché, come nel caso di specie, le valutazioni statistico-computazionali fossero sempre sottoposte al controllo e alla validazione di un giudice umano. Quanto all'eccezione fondata sull'assenza di trasparenza dell'algoritmo, la Corte affermava che Loomis aveva avuto comunque la possibilità di supervisionare gli *input* e gli *output* della decisione algoritmica: da un lato, infatti, l'imputato aveva accesso al manuale di COMPAS, che spiegava quali fossero le informazioni prese in

⁷³ J. Angwin e al., *Machine Bias*, in *ProPublica*, 23 maggio 2016, disponibile a questo link: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. In replica a tale studio, l'azienda produttrice di COMPAS ha commissionato un contro-studio ad alcuni ricercatori del proprio team, il quale avrebbe evidenziato una serie di errori nella metodologia utilizzata dai ricercatori di *ProPublica*, vd. A.W. FLORES E AL. *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks"*, in *Federal Probation Journal*, vol. 80, n. 2, 2016. In generale, sul rischio di discriminazione nell'utilizzo di algoritmi nell'ambito della giustizia penale vd. A.Z. HUQ, *Racial Equality in Algorithmic Criminal Justice*, in *Duke L.J.*, 2019, 1043 ss.

⁷⁴ *State of Wisconsin v. Loomis*, 881 NW 2d 749, 13 luglio 2016, disponibile a questo link: <https://case-law.vlex.com/vid/state-v-loomis-no-888404547>. Sul caso Loomis vd. K. FREEMAN, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, in *North Carolina Journal of Law & Technology*, vol. 18, 2016, p. 75 ss.; I.D.M. BERIAIN, *Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling*, in *Law, Probability & Risk*, 2018, p. 45; H. LIU E AL., *Beyond State v Loomis: artificial intelligence, government algorithmization and accountability*, in *International Journal of Law and Information Technology*, vol. 27, iss. 2, 2019, p. 122; C. DESKUS, *Fifth Amendment Limitations on Criminal Algorithmic Decision-Making*, in *NYU Journal of Legislation and Public Policy*, 2018, p. 237.

considerazione dall'algoritmo; dall'altro, egli aveva potuto visionare (e, dunque, contestare) i risultati finali di calcolo del rischio.

La “dottrina” Loomis presenta diverse criticità, che gli autori che si sono occupati del tema non hanno mancato di rilevare. Va innanzitutto sottolineato che l'utilizzo di algoritmi predittivi nell'ambito della giustizia penale non pone problemi di compatibilità soltanto con il principio di eguaglianza dei cittadini di fronte alla legge (art. 3 Cost.), ma anche con il modello moderno e liberale del diritto penale del fatto, fondato sui canoni di materialità del reato e di personalità della responsabilità penale⁷⁵. La valutazione del singolo caso alla luce di generalizzazioni statistiche sembra infatti prospettare un problematico scivolamento verso forme di criminalizzazione del “tipo d'autore”, in cui la pericolosità sociale viene desunta dal possesso di specifiche caratteristiche o dall'appartenenza a determinate comunità, in una sorta di ritorno agli schemi del determinismo biologico e sociale⁷⁶.

D'altra parte, il fatto che la misura della pena venga stabilita sulla base dei comportamenti adottati dalla comunità di appartenenza pare porsi in contrasto con il principio di individualizzazione del trattamento sanzionatorio, sancito dall'art. 27 Cost. ai commi 1 e 3, che impone che il giudice, nel determinare l'entità della pena, tenga conto delle specifiche esigenze del singolo caso⁷⁷.

Infine, sul piano più strettamente processuale, i *risk assessment tools* sono difficilmente compatibili con il canone del giusto processo e con il diritto alla difesa, che informano l'intero sistema processuale penale. Da un lato, infatti, la già citata opacità

⁷⁵ Sul diritto penale del fatto si vedano, nella dottrina italiana, G. DELITALA, *Il “fatto” nella teoria generale del reato*, Cedam, 1930; F. BRICOLA, *Teoria generale del reato*, in *Nov. dig. it.*, XIX, 1973; G. MARINUCCI, *Soggettivismo e oggettivismo nel diritto penale. Uno schizzo dogmatico e politico-criminale*, in *Riv. it. dir. proc. pen.*, 2011, p. 1 ss.

⁷⁶ V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 13-14; F. SGUBBI, *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Il Mulino, 2019, p. 41; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 21; A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 12-13; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, in C. Piergallini e al. (a cura di), *Studi in onore di Carlo Enrico Paliero*, vol. 1 – Politica criminale e teoria della pena, Giuffrè, 2023, p. 79; L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, cit., p. 364; per un esplicito riferimento al determinismo di stampo lombrosiano vd. A. GIANNINI, *Lombroso 2.0: On AI and Predictions of Dangerousness in Criminal Justice*, in *92 International Review of Penal Law*, 2021, p. 179.

⁷⁷ V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 13-14; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 21; A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 14.

dell'algoritmo impedisce una piena *controllabilità* della decisione: l'impossibilità di spiegare il risultato del procedimento algoritmico si traduce nell'impossibilità di contestarne l'attendibilità, in contrasto con il principio dell'*equality of arms*, che dovrebbe garantire la facoltà di contestare e criticare le prove contrarie⁷⁸; dall'altro, l'utilizzo di informazioni personali a fini predittivi parrebbe porsi in collisione con il principio del contraddittorio nella formazione della prova, nonché con il diritto al silenzio dell'imputato, risolvendosi in un potenziale contrasto con la garanzia del *nemo tenetur se detegere*⁷⁹.

Posto che, negli Stati membri dell'UE, il divieto di decisioni basate *esclusivamente* su trattamenti automatizzati, di cui all'art. 22 del Regolamento europeo in materia di protezione dei dati personali (c.d. GDPR, *General Data Protection Regulation*)⁸⁰, impedisce che il giudice possa venire integralmente sostituito dal calcolo algoritmico, la questione che si pone è quella di *normare* il possibile utilizzo di sistemi algoritmici come *supporto* all'attività giudiziaria.

Il tema ha già richiamato l'attenzione di alcuni organismi a livello sovranazionale. Innanzitutto, nel 2018 si è pronunciata la Commissione europea per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ), con l'emanazione della *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*⁸¹; successivamente, nel 2021, si è pronunciato anche il Parlamento Europeo, con la

⁷⁸ S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?*, cit., p. 1761; V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 15. Manes paragona la *black box decision* scaturente dal sistema algoritmico ad una decisione priva di motivazione o con motivazione meramente apparente, contrastante con l'art. 24 Cost. e con l'art. 111, co. 6 Cost.; vd. *ibidem*, p. 14-15.

⁷⁹ V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 18.

⁸⁰ Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), art. 22, co. 1: «L'interessato ha il diritto a non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona»; al comma 2 sono tuttavia previste delle deroghe. Sul tema vd., con alcuni accenni critici, M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 16.

⁸¹ CEPEJ, *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment*, 4 dicembre 2018; una traduzione a cura del Ministero della Giustizia è disponibile a questo link: <https://rm.coe.int/carta-etica-europea-sull-utilizzo-dell-intelligenza-artificiale-nei-si/1680993348>. Per un commento al testo normativo vd. S. QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit.

*Risoluzione sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*⁸².

La posizione assunta dalle due istituzioni pare sostanzialmente assimilabile: potendo rafforzare l'efficienza della giustizia penale, l'utilizzo di metodi computazionali nel settore processuale deve essere incoraggiato, purché tali sistemi siano trasparenti, accessibili, comprensibili, nonché rispettosi dei diritti fondamentali e dei valori europei, primo tra tutti il divieto di discriminazioni.

In particolare, la Carta Etica approvata dalla CEPEJ, enuncia una serie di principi che dovrebbero guidare l'eventuale introduzione di sistemi algoritmici nei sistemi giudiziari europei: i) principio del rispetto dei diritti fondamentali; ii) principio di non discriminazione; iii) principio di qualità e sicurezza; iv) principio di trasparenza, imparzialità e correttezza; v) principio di garanzia del controllo umano⁸³.

Va comunque ricordato che, in Italia, un altro ostacolo che si pone rispetto all'utilizzo degli algoritmi di giustizia predittiva è il divieto di perizia criminologica in fase cognitiva, stabilito dall'art. 220, co. 2, c.p.p. Qualora, dunque, i risultati dei calcoli algoritmici si ritenessero assimilabili alle perizie – e così parrebbe, trattandosi di valutazioni tecnico-scientifiche basate su principi e teorie psicologiche⁸⁴ – il giudice, in assenza di deroghe legislative, potrebbe farne uso soltanto «ai fini dell'esecuzione della pena o della misura di sicurezza» (art. 220, co. 2, c.p.p.).

⁸² Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale (2020/2016(INI)). Per un commento si rinvia a A. PINGEN, *EP Resolution on AI in Criminal Law and Policing*, in *Eucrim*, 17 novembre 2021; G. BARONE, *Intelligenza artificiale e processo penale: la linea dura del Parlamento Europeo. Considerazioni a margine della Risoluzione del Parlamento Europeo del 6 ottobre 2021*, in *Cass. pen.*, 2022, fasc. 3, p. 1180.

⁸³ Per l'analisi di questi principi si rimanda a S. QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., p. 4 ss.; D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi*, cit., p. 87 ss.; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 81 ss.

⁸⁴ S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?*, cit., p. 203; *contra* M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 20, che rileva come i *risk assessment tools* non sembrano finalizzati specificamente ad accertare «il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche» (art. 220, co. 2, c.p.p.).

4. I.a. come “autore” di un reato e i.a. come strumento.

Nel nostro lavoro ci occuperemo della qualificazione giuridica di quei fatti illeciti che coinvolgono sistemi di i.a. quali *decision-makers* – soli o con la partecipazione di esseri umani. È soltanto in questi casi, infatti, che emergono problemi di imputazione del fatto lesivo. Di minor rilievo, a nostro parere, sono invece le ipotesi in cui i sistemi di i.a. costituiscono dei *meri strumenti* – pur dotati, eventualmente, di un certo grado di autonomia – nelle mani degli utilizzatori: in questi casi, a rispondere del delitto commesso, a titolo di dolo, sarà colui che se ne è consapevolmente servito⁸⁵.

Emergono, in particolare, due possibili ipotesi di i.a. come *strumento* di attività illecite. Da un lato, vi è l'eventualità che un sistema di i.a. sia designato *appositamente* per commettere determinati delitti: si pensi ad esempio, ai *software* di i.a. utilizzati per effettuare truffe *online* e per trasmettere *virus* sui *device* o all'ipotesi – che, ad oggi, pare più fantascientifica che aderente allo stato della ricerca scientifica – del *robot-sicario*, progettato per uccidere determinate persone⁸⁶. In questi casi potrebbe essere opportuno

⁸⁵ S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame?*, cit., p. 425; A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 656-657; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, ivi, p. 713-714; L. PICOTTI, *Cybercrimes and Criminal Relevance of Artificial Intelligence*, cit., p. 72; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 69; S. MASSI, *Affidamento sull'intelligenza artificiale e “disimpegno morale” nella definizione dei presupposti della responsabilità penale*, ivi, p. 665 ss.; A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 7; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 100; S. RIONDATO, *Robotica e diritto penale (robots, ibridi, chimere e “animali tecnologici”*, in D. Provolo e al. (a cura di), *Genetics, Robotics, Law, Punishment*, Padova University Press, p. 601-602; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1206-1207; S. AIRES DE SOUSA, *“Não fui eu, foi a máquina”: teoria do crime, responsabilidade e inteligência artificial*, in A.M. Rodrigues (coord.), *Inteligência artificial no direito penal*, 2020, Almedina, p. 71-72; D. LIMA, *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law*, in 69 *South Carolina Law Review*, 2018, p. 690; F. BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 13.

⁸⁶ È l'ipotesi del c.d. “Picciotto Roboto”, termine coniato da C. REYNOLDS – M. ISHIKAWA, *Robotic Thugs*, in 2007 *ETHICOMP Proceedings*, 2007, p. 487 ss.; il termine è stato poi ripreso da U. PAGALLO, *The Adventures of Picciotto Roboto: AI & Ethics in Criminal Law*, in Aa. Vv., *The Social Impact of Social Computing. Proceedings of the Twelfth International Conference ETHICOMP 2011*, Sheffield Hallam University, 2011, p. 349 ss.; ID., *The Laws of Robots*, Springer, 2015, p. 66 ss. Per una rassegna degli attuali utilizzi illeciti dei sistemi di i.a., nonché dei prevedibili sviluppi futuri, vd. M. BRUNDAGE E AL., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Future of Humanity Institute, 2018; M. CALDWELL E AL., *AI-enabled Future Crime*, in *Crime Science*, 9:14, 2020, p. 1 ss.; T.A.

incriminare *la creazione in sé* del sistema, dal momento che tutti i possibili utilizzi della macchina sarebbero illeciti⁸⁷. Dall'altro lato, vi è l'ipotesi in cui sistemi di i.a. progettati per svolgere attività lecite siano utilizzati per perseguire finalità illecite: così, una *self driving car* potrà essere utilizzata per il trasporto di sostanze stupefacenti – come una sorta di “corriere robotico”⁸⁸ – o, ancora, potrà essere adoperata come “autobomba” per attacchi terroristici⁸⁹. In entrambi i casi, l'impiego di agenti artificiali costituirebbe soltanto una modalità di realizzazione di reati già tipizzati; ai sensi dell'art. 240, co. 1, c.p., i sistemi di i.a. potrebbero oltretutto essere confiscati, in quanto «cose che servirono o furono destinate a commettere il reato»⁹⁰.

In altre ipotesi, l'utilizzo con finalità illecite di sistemi di i.a. di per sé leciti potrebbe rendere opportuna l'introduzione di nuove norme incriminatrici: si pensi, ad esempio, alle immagini manipolate attraverso tecniche di i.a. (c.d. *deepfakes*, vd. *infra*, § 4.1., in questo cap.), che potrebbero essere diffuse a fini di estorsione (e, spesso, di “*sextorsion*”)⁹¹.

Anche qualora dovesse manifestarsi una divergenza tra il decorso causale prefigurato dall'agente e quello effettivamente verificatosi nel caso concreto (*aberratio causae*), si configurerà – come insegna la dottrina⁹² – un'ipotesi di responsabilità penale a titolo doloso⁹³. Ciò che è necessario e sufficiente, ai fini della configurazione del dolo è infatti che l'agente abbia attribuito alla sua azione l'attitudine a causare in concreto quell'evento, essendo irrilevante che abbia previsto un decorso causale diverso da quello

BLAUTH E AL., *Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI*, in *IEEE Access* 2022, Vol 10, p. 77110 ss.; T.C. KING E AL., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in *26 Science and Engineering Ethics*, 2020, p. 89 ss.

⁸⁷ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 7, nota 16.

⁸⁸ A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 343.

⁸⁹ I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 100.

⁹⁰ Così S. RIONDATO, *Robot: talune implicazioni di diritto penale*, in P. Moro – C. Sarra (a cura di), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Franco Angeli, 2017, p. 85.

⁹¹ B. PANATTONI, *Intelligenza artificiale*, cit., p. 329.

⁹² G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale. Parte generale*, Giuffrè, IX ed., 2022, p. 405; F. PALAZZO, *Corso di diritto penale. Parte generale*, VI ed., 2016, p. 295; M. GALLO, (voce) *Aberratio delicti, causae*, in *Enc. Dir.*, I, 1958, p. 66; G. FIANDACA – E. MUSCO, *Diritto penale. Parte generale*, Zanichelli, VIII ed., 2019, p. 377-378.

⁹³ I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 101; A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 7-8; *contra* S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 716-717.

che poi si è verificato⁹⁴. Se, ad esempio, a causa di un difetto di funzionamento, un drone dovesse colpire *non* la dimora della persona che l'utilizzatore intendeva uccidere, ma un edificio adiacente, l'utilizzatore del velivolo risponderà comunque di omicidio doloso, qualora l'incendio causato sul tetto di tale palazzo dovesse causare la morte della vittima designata⁹⁵.

Più complessi saranno invece i casi in cui l'agente intelligente realizzi un reato *diverso* da quello programmato dall'utilizzatore, oppure in cui quest'ultimo si limiti a fornire istruzioni *generali*, lasciando al *decision making* algoritmico la "scelta" di quando e come commettere il fatto illecito. In queste ipotesi, il sistema di i.a. cessa di essere un mero strumento nelle mani dell'utilizzatore, assumendo un vero e proprio ruolo decisionale nella realizzazione dell'illecito. Affronteremo questi aspetti più avanti (vd. *infra*, cap. III, § 3.3. ss.).

4.1. Il caso dei *deepfakes*.

Esulano dal nostro campo di indagine – in quanto *strumenti* in mano agli utilizzatori, più che *agenti autonomi* – anche i *deepfakes*, cui riteniamo comunque di dover dedicare qualche riflessione, stante la loro centralità nel dibattito pubblico, sia accademico⁹⁶ che generalista⁹⁷.

⁹⁴ G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 405.

⁹⁵ L'esempio è di I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 101.

⁹⁶ La letteratura è soprattutto di derivazione statunitense; tra i contributi più rilevanti si vd. M. BODI, *The First Amendment Implications of Regulating Political Deepfakes*, in 47 *Rutgers Computer & Tech. L.J.*, 2021, p. 143; N.I. BROWN, *Deepfakes and the weaponization of disinformation*, in 23 *Virginia Journal of Law & Technology*, no. 1, 2020, p. 1 ss.; A. YAMAOKA-ENKERLIN, *Disrupting Disinformation: Deepfakes and the Law*, in 22 *N.Y.U. J. Legis. & Pub. Pol'y*, 2020, p. 725; S. REID, *The Deepfake Dilemma: Reconciling Privacy and First Amendment Protections*, in 23 *U. Pa. J. Const. L.*, 2021, p. 209; per quanto concerne i problemi specifici derivanti dall'utilizzo di *deepfakes* per pratiche di *revenge porn* vd. M.A. FRANKS – A.E. WALDMAN, *Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions*, in 78 *Maryland Law Review*, 2019, p. 892; R.A. DELFINO, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, in 88 *Fordham L. Rev.*, 2019, p. 887; D. HARRIS, *Deepfakes: False Pornography is Here and the Law Cannot Protect You*, in 17 *Duke Law & Technology Review*, 2019, p. 99.

⁹⁷ Red. Il Post, *Sentirete parlare di video "deepfake"*, in *Il Post*, 24 novembre 2018, disponibile a questo link: <https://www.ilpost.it/2018/11/24/video-deepfake-informazione/>; C. Lavalle, *Le persone credono veri più i volti deepfake di quelli reali*, in *La Repubblica online*, 5 marzo 2022, disponibile a questo link:

I *deepfakes* sono contenuti (foto, video e audio) creati grazie a *software* di intelligenza artificiale che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce⁹⁸.

Ovviamente, in base alle tipiche caratteristiche del ML, maggiore è il numero di immagini o audio che alimentano l'algoritmo di *deep learning*, maggiore sarà l'accuratezza e la precisione del video o audio manipolato: ciò significa che i *deepfakes* saranno più facilmente scambiati per *contenuti veri* quando riguardano celebrità o personaggi pubblici, le cui immagini e voci registrate sono presenti in maniera massiccia sul web. Su YouTube, ad esempio, è possibile guardare Al Pacino che "recita" in *Taxi Driver* al posto di Robert De Niro⁹⁹ o Jim Carrey che interpreta il guardiano notturno di *The Shining* al posto di Jack Nicholson¹⁰⁰. Se in questi casi, tuttavia, il contesto palesemente ludico rende tali contenuti, nonostante il loro realismo, difficilmente credibili, lo stesso non può dirsi per i *deepfakes* realizzati a scopo di propaganda politica o per la produzione di pornografia non consensuale¹⁰¹.

Quanto al primo aspetto, ad esempio, ha avuto un certo impatto mediatico il video manipolato di Nancy Pelosi¹⁰² – una delle maggiori esponenti del *Democratic Party*

https://www.repubblica.it/tecnologia/2022/03/05/news/le_persone_credono_veri_piu_i_volti_umani_deepfake_di_quelli_reali-338859552/.

⁹⁸ La definizione è tratta dalla scheda informativa predisposta dal Garante della Privacy intitolata *Deepfake. Il falso che ti «ruba» la faccia (e la privacy)*, pubblicata nel dicembre 2020 e disponibile a questo link: <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9512226>. Questa la definizione adottata dall'*AI Act*: «sistema di IA che genera o manipola immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri per una persona» (art. 52, § 3).

⁹⁹ Vd. video *Taxi Driver starring Al Pacino [DeepFake]*, pubblicato dall'account Ctrl Shift Face e disponibile a questo link: <https://www.youtube.com/watch?v=9NkKj0aNB0s&t=205s>.

¹⁰⁰ Vd. video *The Shining starring Jim Carrey: Episode 2 - The Bat [DeepFake]*, pubblicato dall'account Ctrl Shift Face e disponibile a questo link: <https://www.youtube.com/watch?v=-ZRUZzZPGto&t=110s>.

¹⁰¹ In generale, sui principali utilizzi illeciti dei *deepfakes* vd. C.O. ONORATI, I.A., *politica e reati contro la personalità dello Stato*, in *Intelligenza artificiale e giurisdizione penale*, Atti del workshop organizzato dalla Fondazione Occorsio il 19 novembre 2021, in *Sist. pen.*, 13 giugno 2022, p. 108 ss. In ambito europeo, vd. il recente report di EUROPOL, *Facing reality? Law enforcement and the challenge of deepfakes*, 22 aprile 2022, disponibile a questo link: https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf.

¹⁰² Red. CBS, *Doctored Nancy Pelosi video highlights threat of "deepfake" tech*, in *CBS news*, 26 maggio 2019, disponibile a questo link: <https://www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25/> (a tale link: è possibile visualizzare anche il video manipolato).

statunitense – in cui la *speaker* della Camera appariva lenta e confusa nell’eloquio. Sebbene la falsità del video fosse evidente ad un occhio attento ed esperto, l’ampia circolazione del contenuto – specie a seguito della sua pubblicazione da parte dell’allora Presidente degli Stati Uniti, Donald Trump – contribuì a danneggiare la reputazione e l’immagine pubblica della nota esponente politica.

Nel contesto dell’invasione russa in Ucraina, poi, i *deepfakes* si sono moltiplicati, sul filo del sottile crinale che separa la satira dalla disinformazione: si va dal video manipolato che raffigura il Primo ministro ucraino Volodymyr Zelensky che invita i cittadini a deporre le armi e ad arrendersi di fronte all’invasione russa, a quello che riproduce il Presidente russo Vladimir Putin che ordina il ritiro delle truppe dal suolo ucraino¹⁰³.

Il mero sospetto che un video sia realizzato attraverso la tecnica del *deepfake* può, d’altra parte, avere ripercussioni sulla stabilità delle istituzioni politiche. Alla fine del 2018, ad esempio, fu postato in rete dal Governo del Gabon un video del Presidente Bongo Ondimba, che rilasciava i tradizionali auguri per il nuovo anno. Dal momento che il Presidente era stato assente per molti mesi dalla vita politica del Paese, e si sospettava che fosse affetto da una grave malattia o fosse addirittura deceduto, la sua insolita apparizione aveva portato molti utenti sui *social media* – inclusi alcuni politici – a diffidare della veridicità del video. Proprio il sospetto che il Governo stesse cercando di coprire un grave vuoto istituzionale era stato uno dei motivi alla base del tentativo di colpo di stato – poi represso – portato avanti da una frangia dell’esercito¹⁰⁴.

Oltre che per scopi di propaganda politica, i *deepfakes* sono già stati ampiamente sperimentati nella commissione di truffe e reati legati alla pornografia. È recente, ad esempio, la notizia di una serie di truffe perpetrate a danno di Binance, una società che gestisce una piattaforma di scambio di criptovalute, proprio attraverso l’utilizzo di *deepfake*: i truffatori organizzavano chiamate su Zoom nelle quali la loro immagine e la loro voce era manipolata per renderli indistinguibili dai dirigenti della società; in questo

¹⁰³ J. Wakefield, *Deepfake presidents used in Russia-Ukraine war*, in *BBC.com*, 18 marzo 2022, disponibile a questo link: <https://www.bbc.com/news/technology-60780142>.

¹⁰⁴ Vd. S. Cahlan, *How misinformation helped spark an attempted coup in Gabon*, in *The Washington Post* (online), 13 febbraio 2020, disponibile a questo link: <https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/> (a tale link: è possibile visualizzare anche il video manipolato). In ogni caso, da analisi forensi non sono però emerse tracce che provino che si trattasse di un *deepfake*.

modo, dietro corrispettivo, contrattavano con gli aspiranti clienti l’inserimento nel listino della piattaforma¹⁰⁵.

Quanto alla pornografia, negli ultimi anni si è registrata una preoccupante crescita dei c.d. *deepfake porn*, ossia di immagini manipolate in cui visi delle persone vengono “innestati” sui corpi di altri soggetti, nudi o impegnati in pose o atti di natura esplicitamente sessuale. Il fenomeno pare ancora più allarmante se si pensa che le fattispecie esistenti, nei vari ordinamenti, non sono normalmente idonee a coprire queste ipotesi. La fattispecie di *Diffusione illecita di immagini o video sessualmente espliciti* (c.d. *revenge porn*), ad esempio, introdotta all’art. 612-ter c.p. dal Codice Rosso (l. 19 luglio 2019 n. 69), parla di invio, consegna, cessione, pubblicazione o diffusione di «immagini o video a contenuto sessualmente esplicito, destinati a rimanere privati, senza il consenso delle persone rappresentate». Si tratta di una formula che difficilmente può comprendere – anche se interpretata estensivamente – i *deepfakes*, specie se si considera che il legislatore, quando ha voluto, ha dato esplicito e autonomo rilievo alle immagini manipolate: l’art. 600-quater.1 c.p., infatti, stabilisce che i reati di pornografia minorile (art. 600-ter e 600-quater c.p.) si applicano – con pena diminuita – anche quando il materiale pornografico rappresenta immagini virtuali realizzate utilizzando immagini di minori degli anni diciotto o parti di esse (co. 1), ove per “immagini virtuali” si intendono le «immagini realizzate con tecniche di elaborazione grafica non associate in tutto o in parte a situazioni reali, la cui qualità di rappresentazione fa apparire come vere situazioni non reali» (co. 2).

Una tutela penalistica contro i *deepfake porn*, dunque, parrebbe sussistere, *de jure condito*, soltanto nei confronti dei minori. Dell’inapplicabilità della disciplina del *revenge porn* alla diffusione di immagini pornografiche manipolate, d’altra parte, si è reso conto anche il legislatore: nel 2021 è stata infatti presentata una proposta di legge volta ad introdurre, all’art. 612-quater c.p., il reato di *Manipolazione artificiale di immagini di*

¹⁰⁵ K. Holt, *Hackers reportedly deepfaked a Binance exec to carry out listing scams*, in *Engadget*, 23 agosto 2022, disponibile a questo link: <https://www.engadget.com/binance-executive-deepfake-listing-scam-154729114.html>. In un’altra vicenda, risalente al 2019, la tecnica del *deepfake* era stata utilizzata per riprodurre, al telefono, la voce dell’amministratore delegato di una società e per convincere un suo sottoposto ad effettuare un versamento, vd. C. Stupp, *Fraudsters Used AI to Mimic CEO’s Voice in Unusual Cybercrime Case*, in *The Wall Street Journal*, 30 agosto 2019, disponibile a questo link: <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>.

persone reali allo scopo di ottenerne rappresentazioni nude che, se approvato, avrebbe punito «Salvo che il fatto costituisca più grave reato, chiunque invia, cede, pubblica o diffonde immagini di persone reali, comunque identificabili, manipolate artificialmente mediante l'uso di strumenti tecnologici o di sistemi di intelligenza artificiale allo scopo di ottenerne rappresentazioni nude idonee a trarre in inganno» (co. 1)¹⁰⁶.

In ogni caso, per tornare al nostro tema d'interesse, la casistica esaminata mostra che, in relazione ai *deepfakes* l'autonomia dell'i.a. svolge un ruolo servente. La capacità rappresentativa delle reti neurali consente di ottenere immagini estremamente somiglianti a quelle originali – immagini che potranno essere utilizzate per finalità legittime (es. arte, cinema) o illegittime (truffe, *revenge porn*, estorsione, sostituzione di persona, destabilizzazione politica). Il *deepfake*, tuttavia, è il prodotto di un *processo*: una volta concluso tale processo, il contenuto rimane uno *strumento* nelle mani del suo detentore.

Quello che si pone in relazione ai *deepfakes*, dunque, pare piuttosto un problema di *regolazione*: la questione che bisognerebbe porsi è perciò se sono idonee le fattispecie criminose esistenti a tutelare efficacemente i beni giuridici che potrebbero essere lesi dalla diffusione di contenuti *falsi* che sono spacciati per *veri* – stante, oltretutto, la facilità con cui oggi è possibile creare *deepfakes*, attraverso applicazioni scaricabili gratuitamente¹⁰⁷.

Le proposte normative, a tal proposito, si stanno muovendo in due direzioni. Da un lato, si richiedono obblighi minimi di trasparenza in relazione ai *deepfakes*, obbligando i produttori di foto o video manipolati ad “etichettare” il prodotto, avvertendo così gli utenti che si tratta di un falso¹⁰⁸. Dall'altro lato, alcuni ordinamenti stanno intervenendo

¹⁰⁶ Vd. Proposta di legge A.C. 2986, *Introduzione dell'articolo 612-quater del codice penale, in materia di manipolazione artificiale di immagini di persone reali allo scopo di ottenerne rappresentazioni nude*, presentata il 30 marzo 2021, il cui testo è disponibile sul sito della Camera dei deputati, a questo link: <https://www.camera.it/leg18/126?idDocumento=2986>.

¹⁰⁷ Si va dall'app *Zao*, che consente di sostituire il proprio volto a quello dell'attore protagonista di celebri sequenze cinematografiche, all'app *Deepnude* – ora ritirata dagli *app store* ma comunque facilmente scaricabile dai motori di ricerca – che permette di creare *deepfake* di nudo tramite il mero inserimento di foto di donne vestite, vd. G. Porro, *Sul web è caccia alle versioni pirata dell'app per creare finte foto di donne nude*, in *Wired Italia*, 4 luglio 2019, disponibile a questo link: <https://www.wired.it/internet/web/2019/07/04/deepnude-revenge-porn/>. Il Garante della Privacy, il 23 ottobre 2020, ha aperto un'istruttoria nei confronti di Telegram, la chat dove molte dei *deepfakes porn* erano circolati; il comunicato stampa è disponibile a questo link: <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9470722>.

¹⁰⁸ Così la Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale, 21 aprile 2021, cit., che, all'art. 52, co. 3, stabilisce che coloro che utilizzano *deepfake* «sono tenuti a rendere noto che il contenuto è stato generato o manipolato

in materie specifiche: negli Stati Uniti, ad esempio, alcuni interventi legislativi mirano ad estendere la disciplina prevista per il c.d. *revenge porn* anche alle immagini pornografiche create attraverso *deepfakes*¹⁰⁹; quanto alla propaganda politica, invece, vi sono stati, come la California e il Texas, che hanno vietato la pubblicazione di *deepfakes* “politici” nel periodo precedente alle elezioni¹¹⁰.

5. La nozione “penalisticamente orientata” di intelligenza artificiale.

Come abbiamo cercato di far emergere, la nozione di “intelligenza artificiale” soffre di una certa vaghezza, comprendendo – a seconda delle concezioni – una serie più o meno ampia di sistemi. In questa sede cercheremo dunque di delimitare l’oggetto del nostro lavoro, cercando di delineare le caratteristiche dell’i.a. che rendono tale fenomeno *dirompente* rispetto ai tradizionali modelli di imputazione della responsabilità penale. Escluderemo, dunque, dalle nostre prospettive di ricerca tutti quei sistemi che – essendo inquadrabili all’interno dei tradizionali schemi di *strumentalità* dell’oggetto rispetto all’utente umano – riproducono problematiche giuridiche note.

Il cuore del problema consiste nell’*imprevedibilità* del comportamento di alcuni dispositivi intelligenti. Il carattere di imprevedibilità, a sua volta, può essere meglio compreso se descritto attraverso il ricorso a tre aspetti, che si riscontrano nei più sofisticati sistemi di i.a.: l’*autonomia*, l’*interattività*, l’*opacità*.

artificialmente», sempre che l’uso non sia «autorizzato dalla legge per accertare, prevenire, indagare e perseguire reati» o sia «necessario per l’esercizio del diritto alla libera espressione e del diritto alla libertà delle arti e delle scienze garantito dalla Carta dei diritti fondamentali dell’UE, e fatte salve le tutele adeguate per i diritti e le libertà dei terzi».

¹⁰⁹ Assemb. 1280, 2019–2020, Cal. 2019, il cui testo – ancora in fase di discussione – è disponibile a questo link: <https://openstates.org/ca/bills/20192020/AB1280/>; il Virginia House Bill 2678, invece, è stato approvato il 8 marzo 2019; il testo è disponibile a questo link: <https://legiscan.com/VA/text/HB2678/2019>.

¹¹⁰ Vd. Assemb. 730 2019–2020, Cal. 2019, disponibile a questo link: <https://openstates.org/ca/bills/20192020/AB730/>, che vieta la pubblicazione di *deepfakes* nei 60 giorni precedenti le elezioni; Texas Sen. Bill 751, 2019–2020, 2019, disponibile a questo link: <https://legiscan.com/TX/text/SB751/id/2027638>, che li vieta nei 30 giorni precedenti.

5.1. L'autonomia dei sistemi di i.a.

L'autonomia è la caratteristica – sottolineata da tutta la dottrina che si è occupata del tema¹¹¹ – che rischia di scardinare i meccanismi classici di imputazione della responsabilità penale.

In una prima accezione, l'autonomia di un sistema robotico allude alla capacità della macchina di svolgere un determinato compito senza alcuna forma di controllo esterno: essa consisterebbe, dunque, nella capacità «*to operate in the real-world environment without any form of external control, once the machine is activated and at least in some areas of operation, for extended periods of time*»¹¹².

Aderendo a quest'impostazione, anche il termostato o i *robot* per la pulizia domestica dovrebbero considerarsi “autonomi”, poiché perfettamente in grado di svolgere il compito loro assegnato in assenza di intervento umano. In realtà, ciò che contraddistingue l'*autonomia* rispetto alla mera *automazione* – di cui il termostato e il *robot* domestico sono espressione – è la *capacità dei sistemi di assumere decisioni non ricavabili direttamente dagli input forniti dai programmatori*, ma che sono frutto dell'interazione con *dati ulteriori* provenienti dall'ambiente circostante.

Così, l'algoritmo che regola il funzionamento del termostato reca un'istruzione di questo tipo: quando la temperatura dell'aria (*input*) scende al di sotto di un valore *x*, il riscaldamento si accende (*output*); quando invece raggiunge il valore *y*, il riscaldamento si spegne. Gli algoritmi che, invece, stanno alla base dei sistemi di i.a. sono in grado di raggiungere gli obiettivi prefissati dai programmatori attraverso procedimenti razionali che spesso non sono stabiliti in anticipo, ma sono appresi in fase di addestramento.

Pur trovandoci sempre all'interno del campo del *determinismo* e non in quello del *libero arbitrio* – ambito che tradizionalmente, quantomeno come postulato¹¹³, è legato

¹¹¹ Vd., per tutti, L. PICOTTI, *I primi vent'anni della Convenzione di Budapest nell'ottica sostanzialista e la mancata ratifica ed esecuzione del Primo Protocollo addizionale contro il razzismo e la xenofobia*, in *Dir. pen. proc.*, 2022, n. 8, p. 1032; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, in *Giur. it.*, luglio 2019, p. 1717.

¹¹² G.A. BEKEY, *Current Trends in Robotics: Technology and Ethics*, in P. Lin e al. (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, 2012, p. 18.

¹¹³ Lo sviluppo delle neuroscienze ha messo in parte in crisi questo postulato – che, tuttavia, è irrinunciabile nell'ambito del diritto penale, costituendone il presupposto fondamentale. In argomento vd., per tutti, O. Di Giovine, *Ripensare il diritto penale attraverso le (neuro-)scienze?*, Giappichelli, 2019; F. BASILE – G. VALLAR, *Diritto penale e neuroscienze*, in *Giornale Italiano di Psicologia*, n. 4, 2016, p. 799 ss.; C.

all'azione umana – i comportamenti dei sistemi di i.a. che definiamo come *autonomi* dipendono da una mole talmente ampia di fattori e di variabili da creare un'apparenza di *autodeterminazione*¹¹⁴. La decisione assunta dai sistemi di *machine learning* è sì *determinata* dal codice, nonché dai dati utilizzati in fase di addestramento e da quelli acquisiti nel corso dell'attività algoritmica: il problema è che l'enorme quantità di informazioni sulla base delle quali le decisioni sono assunte rende complesso l'isolamento dei fattori decisivi nel *decision making*.

Gli straordinari successi che i sistemi di i.a. stanno raggiungendo in molti settori – dalla medicina alla circolazione stradale, dalla traduzione in lingua alle consulenze in materia finanziaria – derivano proprio dalla loro capacità di agire in maniera innovativa, trovando il percorso più efficiente per raggiungere un determinato risultato. Si pensi, ad esempio, alla lunga storia di applicazione “ludica” degli algoritmi di i.a. Nel 1997, il programma *Deep Blue*, sviluppato da IBM, fu in grado di sconfiggere l'allora campione mondiale di scacchi, Garry Kasparov; ancora, nel 2016, un *software* prodotto da Google chiamato *AlphaGo* vinse una partita a *Go* – un complesso e antichissimo gioco di strategia molto diffuso in Asia orientale – contro uno dei migliori giocatori al mondo¹¹⁵. In entrambi i casi, il sistema di i.a. utilizzò delle strategie assolutamente imprevedibili, dimostrando quella che secondo alcuni può essere definita come vera e propria creatività¹¹⁶.

GRANDI, *Neuroscienze e responsabilità penale: nuove soluzioni per problemi antichi?*, Giappichelli, 2017; i contributi raccolti in S.J. Morse – A.L. Roskies (eds.), *A Primer on Criminal Law and Neuroscience*, Oxford University Press, 2013; e, recentemente, E.D. CRESPO, “*Compatibilismo umanista*”: una proposta per conciliare le neuroscienze e il diritto penale, in *Riv. it. dir. proc. pen.*, 2022, fasc. 2, p. 645 ss. Recentemente, sui rapporti tra intelligenza artificiale e neuroscienze vd. i contributi pubblicati nel volume collettaneo J.M. Peris Riera – A. Massaro (a cura di), *Derecho penal, inteligencia artificial y neurociencias. Diritto penale, intelligenza artificiale e neuroscienze*, Roma Tre Press, 2023 e, in particolare, il contributo di A. Massaro, *Intelligenza artificiale e neuroscienze: l'eterno ritorno del diritto penale?*, p. 27 ss.

¹¹⁴ A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 31; M. SIMMLER – N. MARKWALDER, *Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in *Crim. Law Forum*, 2019, n. 30, p. 8-9; C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., p. 1769.

¹¹⁵ Un altro successo – sempre in campo “ludico” – è considerata la vittoria nel 2011 del sistema di IA Watson, prodotto da IBM, al quiz televisivo *Jeopardy!*. Nonostante il grande impatto sull'opinione pubblica, ci pare che l'impresa sia meno impegnativa di quella portata avanti da DeepBlue e AlphaGo; in questo caso, infatti, si trattava di domande di cultura generale e la macchina aveva accesso ai motori di ricerca su internet. Sul tema vd. M.C. ELISH – D. BOYD, *Situating Methods in the Magic of Big Data and Artificial Intelligence*, in *Communication Monographs*, vol. 85, iss. 1, 2018, p. 11.

¹¹⁶ A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 40; cfr. M. SCHERER, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies*, in

Si prenda, ancora, questo esempio: due ingegneri avevano progettato un algoritmo di volo per un drone, che consentiva alla macchina di utilizzare il percorso più efficiente per raggiungere un determinato *target*. Avevano inserito, dunque, tale obiettivo nel codice algoritmico e avevano iniziato ad addestrare il drone, collocandolo in vari luoghi intorno al *target*, così da studiare i suoi movimenti e indicare all'algoritmo se la *performance* aveva avuto successo o no: in questo modo, il sistema avrebbe potuto apprendere come raggiungere in maniera affidabile ed efficiente l'obiettivo. Nel corso dell'addestramento, tuttavia, gli ingegneri avevano rilevato che il drone aveva iniziato a comportarsi in maniera completamente diversa rispetto a quella prevista, indirizzando il proprio volo *in direzione opposta* rispetto al luogo in cui si trovava l'obiettivo, verso i confini esterni dell'area predisposta per la sperimentazione. Dopo alcune indagini, era emerso che non si trattava di un malfunzionamento. Il drone aveva infatti imparato che, in alcune circostanze, il modo più efficiente e rapido per raggiungere il bersaglio era quello di volare verso il perimetro esterno dell'area di prova, dove gli ingegneri lo avrebbero recuperato e portato al bersaglio, così da ripristinare il processo: l'algoritmo aveva insomma appreso che *essere trasportato* era il modo più efficiente per raggiungere l'obiettivo¹¹⁷.

Ecco che, dunque, secondo una diversa – e più convincente – impostazione, l'*autonomia* si distingue dall'*automazione* perché i sistemi che ne sono dotati hanno la capacità di compensare l'incompletezza delle informazioni ricevute in partenza attraverso l'apprendimento¹¹⁸; sono, dunque, in grado di prendere decisioni in situazioni di incertezza (*Decision under Uncertainty*)¹¹⁹. In quest'accezione, l'i.a. sarebbe autonoma

29 *Harvard Journal of Law & Technology*, 2016, p. 363 ss., che contesta che possa parlarsi propriamente di "creatività", giungendo, ciononostante, ad affermare che il comportamento dei sistemi di i.a. possa andare contro le attese di produttori e utilizzatori.

¹¹⁷ Il caso è riportato in M.A. LEMLEY – B. CASEY, *Remedies for Robots*, cit., p. 1313; gli autori dell'articolo, a loro volta, affermano di aver ripreso l'esempio da una presentazione tenutasi presso la *11th Annual Stanford Ecommerce Best Practices Conference*, tenutasi nel giugno 2014 (vd. nt. 1 dell'articolo).

¹¹⁸ S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 42: «*To the extents that an agent relies on the prior knowledge of its designer rather than on its own percepts and learning process, we say that the agent lacks autonomy. A rational agent should be autonomous – it should learn what it can to compensate for partial or incorrect prior knowledge*».

¹¹⁹ A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 37, a cui si rimanda per un'analisi più approfondita del concetto di autonomia, che viene considerata come un costrutto sociale, piuttosto che come un derivato della tecnologia (*social attribution of autonomy*), vd. *Ibidem*, p. 31 ss.; vd. anche M. SIMMLER – N. MARKWALDER, *Guilty Robots?*, cit., p. 13.

in quanto *adattiva*, ossia in grado di interagire con l'ambiente e di modificare il proprio comportamento sulla base degli *input* ricevuti¹²⁰.

Di fatto, a noi sembra che il campo di interesse – per il penalista, così come per il civilista che si occupi di responsabilità aquiliana – possa essere utilmente limitato ai soli sistemi che utilizzano tecniche di *machine learning*, in cui l'*output* non è ricavabile direttamente dal codice algoritmico, ma dipende anche dall'interazione con il contesto e dai processi di auto-apprendimento¹²¹. Nel prosieguo, dunque, parleremo indifferentemente di “sistemi di i.a.” o di “algoritmi”, riferendoci, con tale termine, esclusivamente ai sistemi di *machine learning* – ben consapevoli, d'altra parte, che, anche all'interno di tale categoria ristretta, sono inquadrabili sistemi anche molto differenti tra loro, per livello di autonomia, interattività, opacità e logica sottesa al *decision making*.

5.2. L'interattività dei sistemi di i.a.

Oltre che dalla più o meno marcata autonomia, l'imprevedibilità dei sistemi di i.a. può essere determinata anche dalla loro interattività¹²². Gli algoritmi intelligenti, infatti, spesso non agiscono come monadi, ma sono connessi tra loro e, talvolta, persino con l'ambiente fisico in cui si trovano (c.d. *internet of things*¹²³). Da un lato, l'incontro tra

¹²⁰ C.E.A. KARNOW, *The application of traditional tort law to embodied machine intelligence*, cit., p. 60; G. SARTOR – A. OMICINI, *The autonomy of technological systems and responsibilities for their use*, in N. Bhuta e al. (eds.), *Autonomous Weapons Systems. Law, Ethics, Policy*, Cambridge University Press, 2016, p. 49.

¹²¹ In questo senso anche C.E.A. KARNOW, *The application of traditional tort law to embodied machine intelligence*, cit., p. 52; K.A. CHAGAL-FEFERKORN, *The Reasonable Algorithm*, in *University of Illinois Journal of Law, Technology & Policy*, 2018, n. 1, p. 117; U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product: per i fondamenti di un diritto dell'intelligenza artificiale self-learning*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 96; M. SIMMLER – N. MARKWALDER, *Guilty Robots?*, cit., p. 5; *contra* A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 37, che sostengono che la categoria dei sistemi di i.a. autonomi sia più ampia di quella dei sistemi di i.a. che utilizzano tecniche di *machine learning*.

¹²² In generale, sull'interattività dei sistemi di i.a. vd. A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 111 ss.

¹²³ Non esiste una definizione univoca di *internet of things*. L'ENISA (European Union Agency for Cybersecurity) lo definisce come «a cyber- physical ecosystem of interconnected sensors and actuators, which enable intelligent decision making» (ENISA, *IoT and Smart Infrastructures*, disponibile a questo link: <https://www.enisa.europa.eu/topics/iot-and-smart-infrastructures/iot>); lo European Research Cluster on the Internet of Things (IERC) lo definisce invece come «a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical

due o più “autonomie” – nell’accezione anzidetta – può comportare *output* collettivi inaspettati (c.d. *collective machine behaviour*¹²⁴ o *distributed intelligence*¹²⁵), frutto di interazioni non sempre pienamente comprensibili dall’esterno; dall’altro, gli algoritmi connessi con sensori e infrastrutture fisiche complesse sono esposti ad una varietà infinita di *inputs*, risultando così non meno imprevedibili dell’ambiente con cui interagiscono¹²⁶.

A proposito di interattività, ha avuto un certo impatto mediatico, ad esempio, la notizia di due *chatbot* (“*Bob* e *Alice*”) sviluppati in via sperimentale da Facebook che, nel corso di una simulazione di contrattazione, dopo una prima fase di trattativa svolta in inglese, avevano iniziato ad interloquire in un idioma ignoto, incomprensibile persino per gli sviluppatori¹²⁷. Ecco uno stralcio della conversazione che aveva destato l’attenzione dei programmatori, convincendoli ad interrompere l’esperimento:

Bob: i can i i everything else.....

Alice: balls have zero to me to me to me to me to me to me to me to me to

Bob: you i everything else.....

Alice: balls have a ball to me to me to me to me to me to me to me to me

Bob: i i can i i i everything else

Alice: balls have a ball to me to me to me to me to me to me to me to me

Bob: i

Alice: balls have zero to me to me to me to me to me to me to me to me to

Bob: you i i i i everything else.....

Alice: balls have 0 to me to me to me to me to me to me to me to me to

Bob: you i i i everything else.....

Alice: balls have zero to me to me to me to me to me to me to me to me to

and virtual “things” have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network» (vd. http://www.internet-of-things-research.eu/about_iot.htm).

¹²⁴ I. RAHWAN E AL., *Machine Behaviour*, in *Nature*, 2019, n. 568, p. 482.

¹²⁵ Vd. il lavoro seminale di C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, in *Berkeley Technol. Law J.*, 1996, vol. 11, p. 147; più recentemente, M. HILDEBRANDT, *Ambient Intelligence, Criminal Liability and Democracy*, in *Crim. Law and Philos.*, 2008, vol. 2, p. 163.

¹²⁶ C.E.A. KARNOW, *The application of traditional tort law*, cit., p. 59. Karnow considera questa “porosità” dei sistemi di i.a., in relazione all’ambiente in cui si trovano, la causa principale dell’imprevedibilità degli agenti intelligenti.

¹²⁷ L. Longhitano, *L'intelligenza artificiale di Facebook parla una lingua incomprensibile, ma niente panico*, in *Wired Italia*, 1 agosto 2017, disponibile a questo link: <https://www.wired.it/gadget/computer/2017/08/01/intelligenza-artificiale-facebook/>; A. Griffin, *Facebook’s Artificial Intelligence Robots Shut Down After They Start Talking to Each Other in Their Own Language*, in *Independent*, 31 luglio 2017, disponibile a questo link: <https://www.independent.co.uk/life-style/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html>.

Sebbene la conversazione appaia priva di senso agli occhi di un essere umano, la sequenza delle parole utilizzate sembra tutt'altro che casuale, tanto che gli osservatori più attenti hanno affermato che probabilmente i *bot* avevano trovato una scorciatoia linguistica per comunicare tra loro in una maniera che trovavano più intuitiva.

Un altro esempio di interattività è ben rappresentato dalle crisi finanziarie – conosciute come *Flash Crash* – determinate dall'agire collettivo di sofisticati sistemi di *trading* ad alta frequenza (*high frequency trading*, c.d. HFT)¹²⁸. In quei casi, gli improvvisi crolli di mercato non sono dovuti al comportamento di un singolo *trader* algoritmico, quanto alla reazione di uno “sciame” di agenti intelligenti a determinate condizioni di mercato (vd. *infra*, § 7.1.2.2., in questo cap.).

5.3. Sistemi di i.a. e opacità tecnologica (c.d. *black box*).

Come abbiamo già rilevato, attività che, fino a qualche tempo fa, si riteneva che potessero essere realizzate soltanto dagli esseri umani oggi possono essere svolte, in maniera più accurata e precisa, da sistemi di i.a.: si pensi alla guida di un autoveicolo o di un drone, alla diagnosi di un tumore, alla conclusione di contratti. A tale successo tecnologico – che esprime una fase avanzatissima del dominio dell'uomo sulla natura – è tuttavia legato un risultato paradossale: gli stessi ricercatori che sviluppano i sistemi di i.a. non sono in grado di comprendere a fondo *come* e *perché* i sistemi agiscono in un determinato modo.

Gli algoritmi – con la loro capacità di analisi di enormi masse di dati – sono in grado di individuare correlazioni che sono spesso invisibili all'occhio umano (o che, comunque, richiederebbero anni di ricerche e studi per la loro valida dimostrazione), basando su di esse delle predizioni.

Tali sistemi, tuttavia, non sono in grado di fornire una spiegazione teorica e causale dei risultati raggiunti, né i programmatori possono agevolmente intuirli. Così, ad esempio, Google non è in grado di ricostruire *perché* gli utenti vedano, nei risultati del

¹²⁸ Vd. *Infra* § 7.1.2.2., in questo cap.

motore di ricerca, un certo contenuto in luogo di un altro¹²⁹. Allo stesso modo, né gli sviluppatori di sistemi di diagnostica automatica, né tantomeno i medici utilizzatori, riescono a comprendere a pieno le soluzioni proposte dal sistema (vd. *infra*, § 7.2.1., in questo cap.).

Tale fenomeno è noto, nella teoria dei sistemi, come *black box*: è possibile individuare *input* e *output*, ma non è invece possibile ricostruire *cosa accade* all'interno della scatola nera, ovverosia la catena causale che ha determinato il passaggio dagli *input* agli *output*¹³⁰.

Il concetto di *black box*, in realtà, è ben più risalente rispetto allo sviluppo dei moderni sistemi di i.a. Il diritto penale si era già confrontato con tali problematiche nell'ambito degli studi sulla responsabilità per danno da prodotto (vd. *infra*, cap. III, § 4. ss.). Nelle vicende giudiziarie – sia italiane sia straniere¹³¹ – volte a ricostruire la responsabilità penale per i danni all'integrità fisica dei consumatori cagionati da un prodotto, era infatti già emersa una certa difficoltà nell'individuazione del fattore causativo dell'evento lesivo, determinata dal carattere ignoto del rischio e dalla mancanza di leggi scientifiche di copertura condivise. Di fronte a tali effettive difficoltà, il rischio – evidenziato dalla dottrina più attenta¹³² – è che, in sede di accertamento del nesso di

¹²⁹ Ci riferiamo qui alle dichiarazioni rese, nel marzo 2016, da Paul Haahr, uno degli ingegneri di Google a capo della divisione di sviluppo e progettazione di RankBrain, il sistema di *machine learning* deputato al *ranking* delle notizie sul social network, vd. B. Schwartz, *Google's Paul Haahr: We Don't Fully Understand RankBrain*, in *Search Engine Roundtable*, 8 marzo 2016, disponibile a questo link: <https://www.seroundtable.com/google-dont-understand-rankbrain-21744.html>.

¹³⁰ Y. BATHAEE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in 31 *Harvard Journal of Law & Technology*, 2018, n. 2, p. 905; D. CASTELVECCHI, *Can We Open the Black Box of AI?*, in *Nature*, 5 ottobre 2016, disponibile a questo link: <https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731>; il problema è ben conosciuto anche nelle sedi europee, vd. COMMISSIONE EUROPEA, *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*, COM(2020) 64 final, 19 febbraio 2020, p. 10.

¹³¹ Per una precisa ricostruzione della casistica si rinvia a F. STELLA, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Giuffrè, III ed., 2003, p. 224-235; C. PIERGALLINI, C. PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, Giuffrè, 2004, p. 50 ss.; sulla giurisprudenza italiana vd. A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, in R. Bartoli (a cura di), *Responsabilità penale e rischio nelle attività mediche e d'impresa: (un dialogo con la giurisprudenza)*, Firenze University press, 2010, cit., p. 455 ss.

¹³² In ambito penalistico vd. C.E. PALIERO, *L'autunno del patriarca. Rinnovamento o trasmutazione del diritto penale dei codici*, in *Riv. it. dir. proc. pen.*, 1994, p.1239 ss.; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 198 ss.; ID., *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, in *Riv. it. dir. proc. pen.*, 1996, p. 359; C. PERINI, *Il concetto di rischio nel diritto penale*

causalità, l'autorità giudiziaria si accontenti, in mancanza di rilevabili fattori estranei di produzione del danno, della prova della successione tra condotta (immissione del prodotto nel mercato) ed evento (pregiudizio subito dai consumatori), senza indugiare nella verifica della sussistenza di un collegamento causale tra i due fatti storici.

Il concetto di “scatola nera”, tuttavia, è diventato centrale all'interno del dibattito pubblico – accademico e non – proprio con lo sviluppo e la diffusione di sistemi di i.a. sempre più sofisticati.

L'approccio dominante in materia ritiene comunque che possano esserci soluzioni tecniche che garantiscano la trasparenza degli algoritmi e che la *black box* non sia altro che un tentativo, per le società produttrici, di tenere i codici sorgente al riparo da occhi indiscreti¹³³: da un lato, infatti, la trasparenza potrebbe rendere i sistemi più fragili e vulnerabili di fronte ad eventuali attacchi di *hacker* o terroristi¹³⁴; dall'altro lato, l'ostensibilità del codice porrebbe problemi dal punto di vista della tutela del segreto industriale¹³⁵. La convinzione che possano essere promossi schemi di elaborazione delle informazioni algoritmiche basati sulla *trasparenza* è alla base del successo riscontrato dalle teorie dell'*interpretability* e dell'*explainability* (*Explainable Artificial Intelligence*, c.d. XAI), che tentano di sviluppare modelli di calcolo dai quali possano ricostruirsi – a posteriori – i rapporti causa-effetto¹³⁶.

Sono diversi, d'altra parte, i testi legislativi e le *policy guidelines* che richiedono che i sistemi di i.a. siano trasparenti e che il procedimento di *decision making* sia

moderno, Giuffrè, 2010, p. 251; G. DE VERO, *Il nesso causale e il diritto penale del rischio*, in *Riv. it. dir. e proc. pen.*, 2016, p. 670 ss.

¹³³ F. PASQUALE, *The Black Box Society*, cit., p. 2, che definisce l'opacità come una “*remediable incomprehensibility*”.

¹³⁴ J.A. KROLL E AL., *Accountable Algorithms*, in 165 *U. Pa. L. Rev.*, 2017, p. 639.

¹³⁵ Vi è poi l'ipotesi che l'opacità dell'algoritmo sia una scelta intenzionale dei programmatori, per ragioni di segreto industriale, sicurezza nazionale o privacy, vd. J. BURRELL, *How the machine 'thinks'*, cit., p. 3-4; F. PASQUALE, *The Black Box Society*, cit., *passim*; A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1091-1092.

¹³⁶ La letteratura in materia è davvero amplissima. Dal punto di vista tecnico si vd. per tutti R. GUIDOTTI E AL., *A Survey of Methods for Explaining Black Box Models*, in 51 *ACM Computing Surveys*, 2018, p. 1 ss.; A. D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1110 ss.; dal punto di vista normativo, vd. T. WISCHMEYER, *Artificial Intelligence and Transparency: Opening the Black Box*, in T. Wischmeyer – T. Rademacher (eds.), *Regulating Artificial Intelligence*, Cham, 2020, p. 75 ss.; per una ricostruzione sistematica dei principali lavori in materia di *Explainable Artificial Intelligence* vd. F. SOVRANO E AL., *Metrics, Explainability and the European AI Act Proposal*, in *J*, 2022, n. 5, p. 126 ss.

tracciabile. Così, ad esempio, l'*AI Act*¹³⁷ stabilisce che «I sistemi di IA ad alto rischio sono progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'*output* del sistema e utilizzarlo adeguatamente» (art. 13, § 1) e che le persone alle quali è affidata la sorveglianza umana siano «in grado di interpretare correttamente l'*output* del sistema di IA ad alto rischio, tenendo conto in particolare delle caratteristiche del sistema e degli strumenti e dei metodi di interpretazione disponibili» (art. 14, § 4, lett. d). Ancora, gli Orientamenti etici per un'i.a. affidabile pubblicati nell'aprile 2019 dal Gruppo di esperti sull'i.a. nominato dalla Commissione europea richiedono che «i set di dati e i processi che determinano la decisione del sistema di IA, compresi quelli di raccolta ed etichettatura dei dati, come pure gli algoritmi utilizzati, dovrebbero essere documentati secondo i migliori standard per consentire la tracciabilità e aumentare la trasparenza»¹³⁸.

Vi sono tuttavia casi in cui l'opacità appare come una caratteristica ineliminabile dei sistemi di i.a., connaturata al loro funzionamento¹³⁹, specialmente per quanto concerne gli algoritmi che utilizzano le reti neurali artificiali.

In questi casi, l'opacità sembrerebbe derivare essenzialmente da due fattori. In primo luogo, i *pattern* decisionali potrebbero essere troppo complicati, poiché si basano su una quantità enorme di dati, tra i quali non è possibile individuare i fattori dirimenti

¹³⁷ Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale, cit.

¹³⁸ COMMISSIONE EUROPEA – GRUPPO INDIPENDENTE DI ESPERTI AD ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, *Orientamenti etici per un'i.a. affidabile*, 8 aprile 2019, § 76 ss. (pag. 20).

¹³⁹ *Ex multis* D.R. DESAI – J.A. KROLL, *Trust But Verify*, cit., p. 29 ss.; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability – Adapting Liability Rules to the Digital Age and Artificial Intelligence*, in 13 *Journal of European Tort Law*, 2022, iss. 1, p. 54; Y. BATHAEE, *The Artificial Intelligence Black Box*, cit., p. 929; D.F. LLORCA E AL., *Liability regimes in the age of AI: a use-case driven analysis of the burden of proof*, cit., *passim*; G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, cit., p. 54; U. RUFFOLO – A. AMIDEI, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design, strumenti di certificazione preventive e «autodisciplina» di settore*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1, cit, p. 492; E.M. COTUGNO, *Intelligenza artificiale e tutela del consumatore. I nuovi obblighi informativi*, ivi, vol 2, p. 319; A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, ivi vol. 2, cit, p. 290.

per la decisione¹⁴⁰, nonché sulla struttura complessa dell'algoritmo¹⁴¹: è la c.d. inscrutabilità (*inscrutability*¹⁴²). Si pensi, ad esempio, ad un sistema di supporto alla diagnostica che analizza milioni di informazioni su una patologia, mettendoli a confronto con i geni e con la storia del paziente¹⁴³: si tratta di algoritmi il cui codice si dispiega per pagine e pagine e la cui progettazione coinvolge decine e decine di persone, spesso appartenenti a *team* di lavoro differenti. In secondo luogo, specie per quanto concerne i sistemi di *deep learning*, l'opacità deriva proprio dalla discrepanza tra processo logico-computazionale seguito dalla macchina – che talvolta può avere un impianto geometrico “dimensionale”¹⁴⁴ – e struttura semantica propria del linguaggio umano¹⁴⁵. In altri termini, l'algoritmo *non sa spiegarsi* con un linguaggio che sia comprensibile dagli esseri umani. Il rischio, dunque, è che le disposizioni che impongono obblighi di trasparenza in relazione ai sistemi di i.a. si risolvano in una tutela meramente formale, di difficile implementazione a livello pratico¹⁴⁶.

In ogni caso, va rilevato che – così come è difficile riferirsi all'i.a. come se fosse una categoria unitaria – allo stesso modo, il fenomeno della *black box* può presentare diversi gradi di problematicità, a seconda della tecnica di i.a. concretamente utilizzata. Così, i sistemi che utilizzano approcci logico-simbolici di i.a. presentano, di regola, un elevato livello di trasparenza – ma, in compenso, un grado più modesto di accuratezza delle previsioni; al contrario, i sistemi di i.a. basati su reti neurali e sul *deep learning* sono

¹⁴⁰ C. REED E AL., *Responsibility, Autonomy and Accountability: legal liability for machine learning*, Queen Mary School of Law Legal Studies Research Paper No. 243/2016, p. 13; J. BURRELL, *How the machine 'thinks'*, cit., p. 2; W.N. PRICE II, *Medical Malpractice and Black-Box Medicine*, in I.G. Cohen e al. (eds.), *Big Data, Health Law, and Bioethics*, Cambridge University Press, 2018, p. 298.

¹⁴¹ Y. BATHAE, *The Artificial Intelligence Black Box*, cit., p. 901.

¹⁴² A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1094; vd. anche J. BURRELL, *How the machine 'thinks'*, cit., p. 4-5.

¹⁴³ W.N. PRICE II, *Medical Malpractice and Black-Box Medicine*, cit., p. 297.

¹⁴⁴ Y. BATHAE, *The Artificial Intelligence Black Box*, cit., p. 903 ss.; P. DOMINGOS, *A Few Useful Things to Know about Machine Learning*, cit., p. 82-83; A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, in 87 *Fordham Law Review*, 2018, p. 1095-1096.

¹⁴⁵ C. REED E AL., *Responsibility, Autonomy and Accountability*, cit., p. 13, J. BURRELL, *How the machine 'thinks'*, cit., p. 2; W.N. PRICE II, *Medical Malpractice and Black-Box Medicine*, cit., p. 298.

¹⁴⁶ U. RUFFOLO – A. AMIDEI, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design*, cit., p. 492; A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit., p. 290; E.M. COTUGNO, *Intelligenza artificiale e tutela del consumatore. I nuovi obblighi informativi*, cit., p. 319; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale*, cit., p. 16.

caratterizzati da un alto livello di opacità – ma sono anche i sistemi che garantiscono una maggiore accuratezza predittiva¹⁴⁷.

Alla complessiva opacità dei sistemi di i.a. contribuisce, tra l'altro, anche la modalità con la quale gli algoritmi di *machine learning* assumono le decisioni. Come abbiamo già accennato, il passaggio da un modello *simbolico* ad un modello *connessionista* di computazione ha comportato il quasi totale abbandono del tentativo di creare macchine in grado di interpretare simboli e segnali (vd. *supra*, § 1.2.3.). Gli attuali sistemi di ML esaminano enormi quantità di dati, individuano *pattern* ricorrenti e generalizzano i risultati ottenuti – senza, tuttavia, poter dimostrare un rapporto di causa-effetto tra *input* e *output* ottenuto.

Prendiamo ancora, quale esempio, un sistema di supporto alla diagnostica, che – analizzando *database* che contengono migliaia di cartelle cliniche, immagini e articoli scientifici – giunge alla conclusione che la macchia sui polmoni evidenziata da una TAC è un tumore maligno. L'algoritmo non è in grado di *spiegare* attraverso quale percorso logico-argomentativo è giunto a tale diagnosi – come farebbe, ad esempio, un medico che debba relazionare davanti all'*équipe* con la quale sta gestendo il caso clinico¹⁴⁸. Semplicemente, l'algoritmo rileva che, *nella maggioranza dei casi*, una massa di tessuto avente quelle caratteristiche è *un tumore maligno*.

Per centinaia di anni – perlomeno dalla nascita del metodo scientifico – a ricercatori e scienziati è stato insegnato che *una correlazione non implica la causalità* e che per dimostrare il rapporto di causa-effetto tra due fatti non è sufficiente la constatazione della contemporanea presenza degli stessi (c.d. fallacia del *cum hoc ergo*

¹⁴⁷ M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, in M. Ebers – S. Navas (eds.), *Algorithms and Law*, Cambridge University Press, 2020, p. 50. Sulla differenza tra *Weak* e *Strong Black Boxes* vd. anche Y. BATHAEE, *The Artificial Intelligence Black Box*, cit., p. 906; sul *trade off* tra accuratezza delle previsioni algoritmi e spiegabilità vd. anche A.J. LONDON, *Artificial intelligence and black-box medical decisions: accuracy versus explainability*, in *The Hastings Center Report*, 49(1), 2019, p. 15 ss.; R. CARUANA E AL., *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission*, in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015; D.F. LLORCA E AL., *Liability regimes in the age of AI: a use-case driven analysis of the burden of proof*, cit., p. 8; *contra* C. RUDIN, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, in *1 Nat Mach Intell.*, 2019, p. 208: «It is a myth that there is necessarily a trade-off between accuracy and interpretability».

¹⁴⁸ Vd. per tutti J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility: The prospect of expert robots*, in R. Calo e al. (eds.), *Robot Law*, cit., p. 106; V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., *passim*

propter hoc). Si potrebbero fare centinaia di esempi che rendono evidente il paralogismo insito nel dedurre un rapporto di causalità da una mera correlazione: così, ad esempio, seppure sia stata dimostrata una correlazione del 99,79% tra l'incremento, tra il 1999 e il 2009, della spesa in ricerca scientifica negli Stati Uniti, e la crescita, nel medesimo periodo, dei suicidi per impiccagione, strangolamento e soffocamento, a nessuno potrebbe venire in mente di individuare un rapporto causa-effetto tra i due fenomeni¹⁴⁹.

Il successo applicativo dei sistemi di *machine learning* – determinato dall'ampia disponibilità di dati – metterebbe in discussione, a parere di alcuni, la persistente validità di tale approccio: se, infatti, sulla base di milioni di dati, risulta che *in presenza del fatto x si registra anche la presenza del fatto y*, allora si potrà prevedere che, *alla comparsa del fatto x seguirà anche il fatto y*. Perderebbe così rilevanza l'accertamento della causalità, in favore di un approccio pragmatico e funzionalista: è quella che il giornalista Chris Anderson, in un celebre articolo pubblicato sulla rivista *Wired* nel giugno 2008, ha provocatoriamente chiamato "la fine della teoria"¹⁵⁰.

Se un sistema di i.a. decide 99 volte su 100 in maniera più efficiente rispetto al decisore umano, che senso avrebbe, insomma, affidarsi a quest'ultimo, piuttosto che al calcolo computazionale? La crescente accuratezza delle predizioni effettuate dai sistemi di i.a. porterà dunque, sempre di più, il decisore umano a fidarsi di esse – nonostante, per l'appunto, esse siano in una certa misura imperscrutabili e non si fondino su validazioni causali: il paragone con gli oracoli della tradizione greca antica è evidente¹⁵¹. La tendenza in atto – e che, prevedibilmente, è destinata a sedimentarsi nel prossimo futuro – è quella

¹⁴⁹ L'esempio è tratto dal sito <https://www.tylervigen.com/spurious-correlations>, che raccoglie correlazioni che nulla hanno a che fare con la causalità, per mostrare la fallacia del *cum hoc ergo propter hoc*; alcune delle correlazioni sono poi state raccolte nel libro T. VIGEN, *Spurious Correlations*, Hachette Books, 2015; sui limiti delle correlazioni nell'utilizzo dei *big data* vd. anche C.S. CALUDE – G. LONGO, *The Deluge of Spurious Correlations in Big Data*, in *Foundations of Science*, 22 (3), p. 1 ss.

¹⁵⁰ C. Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, in *Wired*, 23 giugno 2008, disponibile al seguente link: <https://www.wired.com/2008/06/pb-theory/>; vd. anche V. MAYER-SCHONBERGER – K. CUKIER, *Big data: A revolution that will transform how we live, work, and think*, Houghton Mifflin Harcourt, 2013, p. 53 ss.; in senso critico vd. M. HILDEBRANDT, *Smart Technologies and the End(s) of law*, Edward Elgar Publishing, 2015, p. 37; J.C. AGUERRI, *Datos, teoría y causalidad: una breve aproximación al retorno a la teoría*, in F.J. Castro Toledo (coord.), *La transformación algorítmica del sistema de justicia penal*, cit., p. 80 ss.

¹⁵¹ Il paragone tra i sistemi predittivi e gli oracoli dell'antichità si è ormai affermato in tantissimi studi di settore, vd. J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility: The prospect of expert robots*, cit., *passim*.

di *delegare* sempre più funzioni decisionali ai sistemi di i.a., i cui processi di *decision making* rimangono tuttavia *incomprensibili* all'uomo.

5.3.1. Dal disincanto weberiano all'incanto algoritmico?

L'opacità dei sistemi di i.a. parrebbe costituire solo l'ultima tappa di quella tendenziale *crisi di comprensione e controllo* che – secondo alcune autorevoli voci nel campo della filosofia e della sociologia – avrebbe colpito l'umanità nel suo rapporto con l'evoluzione tecnologica.

Tra i primi a concettualizzare questo *paradosso della tecnica* sono stati due esponenti della Scuola di Francoforte, Adorno e Horkheimer, che, in *Dialettica dell'Illuminismo*, sostenevano che il progetto illuministico e baconiano di dominio della natura attraverso la scienza e la ragione si fosse risolto, dialetticamente, nel proprio opposto, ossia nell'assoggettamento dell'umanità ai ritmi e alle esigenze della stessa macchina produttiva, diventata ormai una «seconda natura che lo riporta nelle stesse condizioni di smarrimento e paura che vivevano i nostri avi agli inizi del processo di civilizzazione»¹⁵². Se, tuttavia, il bersaglio critico dei due filosofi, in linea di continuità con la lezione di Marx, era più metaforico e politico – l'alienazione dell'uomo contemporaneo nella società capitalistica – il tema specifico della tecnologia come *forza autonoma*, che agisce in base a leggi proprie, diventa centrale nella sociologia critica successiva alla fine della Seconda guerra mondiale. Con accenti diversi, autori come Jacques Ellul¹⁵³, Martin Heidegger¹⁵⁴ e Langdon Winner¹⁵⁵ evidenziarono come lo sviluppo tecnologico, dacché l'umanità esiste, sia stato in grado di determinare l'evoluzione della società in modi non preventivabili né controllabili dall'uomo.

¹⁵² Così l'interpretazione che ne dà W. PRIVITERA, *Tecnica, individuo e società: cinque lezioni sulla teoria di Ulrich Beck*, Rubbettino, 2004, p. 20. Sulla tecnologia come “natura” del futuro vd. il recente K. VAN MENSVOOR, *NextNature*, D Editore, 2022.

¹⁵³ J. ELLUL, *La tecnica. Rischio del secolo*, Giuffrè, 1969 (ed. franc. 1954).

¹⁵⁴ M. HEIDEGGER, *La questione della tecnica*, in Id., *Saggi e discorsi* (a cura di G. Vattimo), Mursia, 1976, p. 5-27 (ed. ted. 1957).

¹⁵⁵; L. WINNER, *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*, MIT Press, 1977.

Ma è soprattutto con la pubblicazione di *La società del rischio* (1986) di Ulrich Beck¹⁵⁶ che la riflessione sugli effetti imprevisti dello sviluppo tecnologico si impone nel dibattito contemporaneo, giungendo a configurarsi come problema *specifico* delle società a capitalismo avanzato¹⁵⁷.

Com'è noto, la tesi di fondo di Beck è che le tecniche di produzione moderna creano sistematicamente rischi che la società non è in grado di controllare ed assorbire. Da un lato, i progressi tecnico-scientifici permettono di affrontare pericoli che, in epoca pre-moderna, avrebbero comportato danni incommensurabili; dall'altro, tuttavia, lo sviluppo capitalistico crea *nuovi rischi*, che sfuggono alla comprensione e al dominio dell'uomo. Ancora prima di Beck, nel 1984, il sociologo Charles Perrow aveva tematizzato l'esistenza di "*normal accidents*", ossia eventi avversi che – nei sistemi estremamente complessi e interconnessi tra loro – possono derivare dall'interazione *inaspettata e incomprensibile* tra diversi fattori (meccanici, elettronici, umani). Nella loro imprevedibilità, tali incidenti sarebbero in larga parte *inevitabili*¹⁵⁸. Ecco, dunque, il paradosso: la tecnica – da strumento di *comprensione* e *dominio* sulla forza cieca e schiacciante della natura – diventa essa stessa un elemento misterioso, imperscrutabile, incontrollabile¹⁵⁹.

Con lo sviluppo e la diffusione di sistemi di i.a., si assiste ad una nuova fase del rapporto tormentato tra tecnologia e *agency* umana. In questo caso, infatti, l'imprevedibilità non è un *bug* nel sistema, ma piuttosto un risultato voluto e ricercato

¹⁵⁶ U. BECK, *La società del rischio. Verso una seconda modernità*, Carocci, 2000 (ed. ted. 1986).

¹⁵⁷ L'opera di Ulrich Beck ha avuto influenza enorme anche in ambito giuridico. Si vd., per restare in ambito penalistico, J.M. SILVA SANCHEZ, *L'espansione del diritto penale. Aspetti della politica criminale nelle società postindustriali*, Giuffrè, 2004 (ed. spagn. 1999); F. HERZOG, *Società del rischio, diritto penale del rischio, regolazione del rischio*, in L. Stortoni – L. Foffani (a cura di), *Critica e giustificazione del diritto penale nel cambio di secolo. L'analisi critica della Scuola di Francoforte*, Giuffrè, 2004, p. 357; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit.

¹⁵⁸ C. PERROW, *Normal accidents*, Princeton University Press, 1984. L'autore, in particolare, si riferiva ad incidenti verificatisi nell'ambito di impianti chimici e centrali nucleari o nella gestione del traffico aereo e marittimo. Il tema dell'imprevedibilità delle tecnostutture più avanzate è trattato anche, più recentemente, da S. ABERSMAN, *Overcomplicated. Technology at The Limits of Comprehension*, Gildan Media Corporation, 2016, che, a tal proposito, parla di *entanglement* (groviglio); su questi temi vd. anche R.A. VENTURA, *Radical Choc*, Einaudi, 2020, *passim*.

¹⁵⁹ Vd. K. VAN MENSVOOR, *NextNature*, cit.

dagli stessi programmatori, poiché consente di raggiungere i risultati più efficienti¹⁶⁰: il sistema di i.a. è *unpredictable by design*¹⁶¹.

Come abbiamo già accennato, d'altra parte, il funzionamento delle tecniche più sofisticate di *machine learning* è ancora in parte ignoto, nonostante se ne sfruttino ampiamente le potenzialità applicative¹⁶². Insomma, non si sa bene *perché*, ma le previsioni effettuate dagli algoritmi di i.a. sono, nella maggioranza dei casi, corrette, e spesso più accurate di quelle umane. Ci si affida ad esse come ad un *oracolo* o ad uno "stregone", tanto che riaffiorano alla mente le parole di uno dei più celebri scrittori di fantascienza, Arthur C. Clarke: "*any sufficiently advanced technology is indistinguishable from magic*"¹⁶³.

I riferimenti alla *magia* e al *genio* sono d'altra parte frequenti, sia da parte di giornalisti specializzati sia da parte degli stessi ricercatori nell'ambito dell'i.a.¹⁶⁴.

¹⁶⁰ J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 107: "*a feature and not a bug*"; vd. anche M.A. LEMLEY – B. CASEY, *Remedies for Robots*, cit., p. 1334 ss.

¹⁶¹ R. CALO, *Robotics and the Lessons of Cyberlaw*, cit., p. 542; J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 107; vd. anche le riflessioni svolte, quando ancora non erano state sviluppate le tecniche di *machine learning*, da C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, cit., p. 192.

¹⁶² M. RAGHU E AL., *On the Expressive Power of Deep Neural Networks*, in *Proceedings of the 34th International Conference on Machine Learning*, 2017, n. 2847–54, § 1: "*deep neural networks have proved astoundingly effective at a wide range of empirical tasks [...] Despite these successes, understanding of how and why neural network architectures achieve their empirical successes is still lacking*"; in termini simili vd. Z. LU E AL., *The Expressive Power of Neural Networks: A View from the Width*, arXiv preprint, 2017, disponibile a questo link: <http://arxiv.org/abs/1709.02540>: "*Deep neural networks have achieved state-of-the-art performance in a wide range of tasks [...] Despite their promising results in applications, our theoretical understanding of neural networks remains limited*".

¹⁶³ A.C. CLARKE, *Profiles of the Future: An Enquiry into the Limits of the Possible*, Harper & Row, 1973, p. 21. Si tratta della terza delle c.d. "*Clarke's three laws*".

¹⁶⁴ Così, ad esempio, afferma Pedro Domingos, uno dei più noti ricercatori in materia di *machine learning*: "*developing successful machine learning applications requires a substantial amount of 'black art' that is difficult to find in textbooks*", vd. P. DOMINGOS, *A Few Useful Things to Know about Machine Learning*, cit., p. 78, vd. anche C. Metz, *In Two Moves, AlphaGo and Lee Sedol Redefined the Future*, in *Wired US*, 16 marzo 2016, disponibile a questo link: <https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/>, che, commentando la vittoria dell'algoritmo AlphaGo al gioco Go!, scrisse: "*machines are now capable of moments of genius*"; R. Giacconi, *Una sorta di magia. Intelligenze artificiali, origami, marionette, serpenti: scoprire l'incanto nei laboratori di robotica*, in *Il Tascabile*, 2 marzo 2023, disponibile a questo link: <https://www.iltascabile.com/scienze/robot/>, in cui si riportano le considerazioni di Jamie Paik, direttrice del Reconfigurable Robotics Lab (RRL) di Losanna: "*ogni volta che accendiamo un robot, gli diamo vita e fa quello che ci aspettavamo, è un momento magico, di gioia, euforia. È come se dicessimo sempre una piccola preghiera: fa' che funzioni questa volta, fa' che funzioni questa volta...*".

Sembrano lontane anni luce le riflessioni di Max Weber sul processo di modernizzazione come progressiva razionalizzazione e comprensione del mondo:

«La crescente intellettualizzazione e razionalizzazione non significa dunque una crescente conoscenza generale delle condizioni di vita alle quali si sottostà. Essa significa qualcosa di diverso: la coscienza o la fede che, se soltanto si volesse, si potrebbe in ogni momento venirne a conoscenza, cioè che non sono in gioco, in linea di principio, delle forze misteriose e imprevedibili, ma che si può invece – in linea di principio – dominare tutte le cose mediante un calcolo razionale. Ma ciò significa il disincantamento del mondo. Non occorre più ricorrere a mezzi magici per dominare gli spiriti o per ingraziarseli, come fa il selvaggio per il quale esistono potenze del genere. A ciò sopperiscono i mezzi tecnici e il calcolo razionale. Soprattutto questo è il significato dell'intellettualizzazione in quanto tale»¹⁶⁵

Sebbene non manchino alcune tendenze – di matrice mistica e irrazionalistica – volte a valorizzare il presunto *reincantamento* prodotto dalle nuove tecnologie¹⁶⁶, la maggior parte degli studiosi guarda con preoccupazione all'affermarsi di strumenti tecnici che agiscono in maniera incontrollata, così come alla “ammirata rassegnazione” che spinge alcuni osservatori a dipingere le capacità dei nuovi sistemi di i.a. come a moderne manifestazioni del “magico”.

6. Le definizioni normative.

Enucleate le caratteristiche (autonomia, interattività, opacità) che rendono i sistemi di i.a. in una certa misura *imprevedibili* e, dunque, *problematici* dal punto di vista dell'attribuzione della responsabilità, cerchiamo ora di comprendere se vi siano definizioni, a livello legislativo o giurisprudenziale, che valorizzino i suddetti aspetti, così

¹⁶⁵ M. WEBER, *La scienza come professione - La politica come professione*, Mondadori, 2006, pp. 20-21 (ed. ted. 1919); per un confronto tra *black box* dell'i.a. e “disincantamento” vd. A. CAMPOLO – K. CRAWFORD, *Enchanted Determinism: Power without Responsibility in Artificial Intelligence*, in *Engaging Science, Technology, and Society*, 2020, n. 6, *passim*; vd. anche M.C. ELISH – D. BOYD, *Situating Methods in the Magic of Big Data and Artificial Intelligence*, cit., p. 8-9.

¹⁶⁶ Sul tema vd. M. Mattei, *Parliamo di reincantamento*, in *L'Indiscreto*, 8 luglio 2022, disponibile a questo link: <https://www.indiscreto.org/parliamo-di-reincantamento/>.

da fornire un valido punto di partenza per meglio circoscrivere il campo di indagine di questo lavoro¹⁶⁷.

Va innanzitutto sottolineato che, ad oggi, lo sforzo definitorio è stato condotto principalmente da *policy report* e da strumenti di *soft law* (fatta salva la Proposta di Regolamento presentata dalla Commissione Europea¹⁶⁸, sulla quale torneremo tra un attimo): nella maggior parte dei casi, si tratta di definizioni che ambiscono alla *trasversalità* e che non hanno tanto lo scopo di delineare *una disciplina* da applicare ai sistemi così individuati, quanto quello di fornire coordinate politiche e di *risk management* per lo sviluppo delle tecnologie intelligenti.

Ne emergono definizioni da un lato vaghe – consueto, ad esempio, è il riferimento a sistemi che *si comportano in maniera intelligente*¹⁶⁹, nonostante, come abbiamo accennato (vd. *supra*, § 1), nel mondo scientifico non ci sia un *consensus* sul significato del concetto di “intelligenza” – e, dall’altro, *over-inclusive*, che comprendono anche tecnologie prive dei caratteri distintivi e problematici dell’i.a.¹⁷⁰.

Ancora, alcune definizioni, prendono in considerazione soltanto l’accezione di i.a. come “sistema tecnologico”, mentre altre comprendono sia quest’ultimo sia il rispettivo campo di studi: nonostante tale nozione bidimensionale sia effettivamente corretta dal punto di vista descrittivo, non si vede come la seconda accezione semantica possa essere utile a fini regolatori, dal momento che l’obiettivo della normativa, in quest’ambito, è

¹⁶⁷ Vd. S. SAMOILI E AL., *AI WATCH. Defining Artificial Intelligence*, EUR 30117 EN – Publications Office of the European Union, 2020, disponibile a questo link: <https://publications.jrc.ec.europa.eu/repository/handle/JRC118163>, che raccoglie e classifica le definizioni di i.a. proposte da scienziati, legislatori, dottrina, tra gli anni 50 e il 2019; vd. anche A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 22 ss.

¹⁶⁸ Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull’Intelligenza artificiale (Legge sull’intelligenza artificiale) e modifica alcuni atti legislativi dell’Unione, COM/2021/206 final, 21 aprile 2021

¹⁶⁹ Vd. ad esempio la Comunicazione della Commissione al Parlamento europeo, *L’intelligenza artificiale per l’Europa*, COM(2018) 237 final, 25 aprile 2018: «“Intelligenza artificiale” (IA) indica sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi»; vd. anche lo standard ISO/IEC 2382:2015(en) Information technology: «*branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement*».

¹⁷⁰ Scettico, a tal proposito, è Bertolini, secondo cui sarebbe auspicabile rinunciare ad una definizione trasversale di i.a., elaborando piuttosto definizioni di i.a. differenti, funzionali alla specifica materia trattata, vd. A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 31.

quello di regolare diritti, obblighi e responsabilità inerenti all'utilizzo e alla messa in commercio di sistemi di i.a., non quello di disciplinare un settore di studi¹⁷¹.

La funzione principale delle definizioni, d'altra parte, consiste nella delimitazione dell'ambito di applicazione di una disciplina normativa, oppure nella specificazione del contenuto di regole giuridiche¹⁷² – in assenza di norme che affrontino specificamente le questioni attinenti alla responsabilità, penale e civile, per danni arrecati da sistemi di i.a., lo sforzo definitorio anzidetto non pare avere alcuna rilevanza applicativa.

6.1. La definizione fornita dall'*AI Act* (Commissione Europea)

Un discorso a parte merita la proposta di Regolamento della Commissione Europea (c.d. *AI Act*)¹⁷³, che per la prima volta mira a fornire un quadro normativo per lo sviluppo e la messa in commercio di sistemi di i.a., con due obiettivi contestuali: da un lato, favorire la sicurezza dei prodotti; dall'altro, incentivare l'innovazione e lo sviluppo in questo campo. Nel prossimo capitolo (vd. § 3.4.2.) affronteremo più nel dettaglio i contenuti della proposta, che – sebbene non detti norme specifiche in tema di responsabilità – prevede alcune obbligazioni per i produttori e per gli utilizzatori di sistemi di i.a., che potrebbero avere importanti ripercussioni sull'individuazione dello *standard* di diligenza, ai fini dell'accertamento della colpa¹⁷⁴. Per ora, basterà rilevare che la Proposta di Regolamento adotta un *risk-based approach*, in base al quale si prevedono obblighi di trasparenza e autorizzazione per i prodotti ad alto rischio, vietandosi invece la messa in commercio di prodotti che producono un rischio intollerabile per i diritti fondamentali.

¹⁷¹ A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 30

¹⁷² A. BELVEDERE, (voce) *Definizioni*, in *Digesto*, IV ed., vol. V, Torino, 1989, p. 149 ss.; ID., *Il problema delle definizioni nel codice civile*, Giuffrè, 1977; in ambito penalistico vd. A. Cadoppi (a cura di), *Il problema delle definizioni legali nel diritto penale. Ominis definitio in iure periculosa?*, Cedam, 1996; proprio con riferimento ai sistemi di i.a., e, più precisamente, alle macchine a guida autonoma, vd. E. HILGENDORF, *Automated Driving and the Law*, in E. Hilgendorf – U. Seidel (eds.), *Robotics, Autonomics, and the Law*, p. 172.

¹⁷³ Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale, 21 aprile 2021, cit.

¹⁷⁴ A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 73.

Nonostante i principali studi commissionati dalle istituzioni europee suggerissero un approccio settoriale¹⁷⁵ – così da circoscrivere l'intervento e dosarlo in relazione alle aree di rischio e alle problematiche più pressanti – la Commissione, nella sua proposta di Regolamento, adotta invece un'impostazione *orizzontale*, volta a fornire un *framework* generale per tutti i sistemi di i.a., fondato sui principi di proporzionalità e innovazione.

Nella prospettiva fatta propria dalle istituzioni europee, la scelta di adottare una definizione unitaria di i.a. trova la sua ragion d'essere in quelle che sono riconosciute come le caratteristiche *comuni* a tutti i sistemi di i.a., che li renderebbero *differenti* rispetto ad altre tecnologie. Lo stesso Parlamento Europeo, in una risoluzione del 2020¹⁷⁶, identificava tali caratteristiche nella complessità, connettività, opacità, vulnerabilità e autonomia.

La proposta di Regolamento, all'art. 3, lett. a), definisce il sistema di i.a. come:

«un *software* sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare *output* quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono»

Gli Approcci indicati dall'allegato 1 sono:

«a) Approcci di apprendimento automatico, compresi l'apprendimento supervisionato, l'apprendimento non supervisionato e l'apprendimento per rinforzo, con utilizzo di un'ampia gamma di metodi, tra cui l'apprendimento profondo (*deep learning*);

b) approcci basati sulla logica e approcci basati sulla conoscenza, compresi la rappresentazione della conoscenza, la programmazione induttiva (logica), le basi di conoscenze, i motori inferenziali e deduttivi, il ragionamento (simbolico) e i sistemi esperti;

c) approcci statistici, stima bayesiana, metodi di ricerca e ottimizzazione»

¹⁷⁵ A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 31; A. BERTOLINI – E. PALMERINI, *Regulating Robotics: A Challenge for Europe*, in *Upcoming Issues of EU Law*, 2014, p. 185 ss.; in dottrina vd. M. EBERS, *Regulating AI and Robotics*, cit., p. 93.

¹⁷⁶ Risoluzione del Parlamento europeo su un regime di responsabilità civile per l'intelligenza artificiale, 2020, cit.

La definizione è stata criticata da diversi studiosi della materia, che ne hanno sottolineato il carattere estremamente ampio: in particolare, il campo di applicazione della Proposta si estende anche agli algoritmi fondati su approcci logico-deduttivi (lett. b) e statistici (lett. c), che normalmente, oggi, non sono considerati come “intelligenti” e che spesso sono già in circolazione da decenni¹⁷⁷ – con il rischio che i nuovi requisiti per la messa in commercio si applichino anche a sistemi che non necessitano di regolamentazione rispetto all'obiettivo normativo, caricando le imprese di eccessivi oneri finanziari e amministrativi¹⁷⁸.

A tale definizione rinvia oltretutto la Proposta di Direttiva relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale¹⁷⁹ (vd. art. 2, n. 1 e 2), presentata dalla Commissione Europea il 28 settembre 2022: ciò significa che il regime di *disclosure* obbligatoria e di alleggerimento probatorio ivi previsto sarà applicabile soltanto qualora il ricorrente agisca per ottenere il risarcimento del danno causato da un sistema di i.a., così come definito dalla bozza di regolamento sull'intelligenza artificiale.

¹⁷⁷ C. OSBORNE, *The European Commission's Artificial Intelligence Act highlights the need for an effective AI assurance ecosystem*, in *Centre for Data Ethics and Innovation Blog*, disponibile a questo link: <https://cdei.blog.gov.uk/2021/05/11/the-european-commissions-artificial-intelligence-act-highlights-the-need-for-an-effective-ai-assurance-ecosystem/>; V. GALLO E AL., *The new EU AI Act|what do financial services firms need to know?*, in *Deloitte Insights*, disponibile a questo link: <https://ukfinancialservicesinsights.deloitte.com/post/102gxhz/the-new-eu-ai-act-what-do-financial-services-firms-need-to-know>; L. Clarke, *The EU's leaked AI regulation is ambitious but disappointingly vague*, in *Tech Monitor*, 15 aprile 2021, disponibile a questo link: <https://techmonitor.ai/policy/eu-ai-regulation-machine-learning-european-union>; contra A. LAVORGNA – G. SUFFIA, *La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell'ambito della giustizia penale*, in *Dir. pen. cont. – Riv. Trim.*, 2021, n. 2, p. 94, che ritengono che la definizione sia troppo ristretta, trascurando ampi settori del fenomeno. In ogni caso, alcune fonti riportano che Consiglio e Commissione Europea si stiano accordando per restringere la definizione e limitare il campo di applicazione del Regolamento ai soli sistemi dotati di *machine learning*, vd. N. Kayser-Bril, *European Council and Commission in agreement to narrow the scope of the AI Act*, in *Algorithm Watch*, 24 novembre 2021, disponibile a questo link: <https://algorithmwatch.org/en/eu-narrow-scope-of-ai-act/>.

¹⁷⁸ J. MÖKANDER E AL., *Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation*, in *Minds and Machines*, 2022, vol. 32, p. 245.

¹⁷⁹ Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, cit.

6.2. La definizione fornita in Italia dal Consiglio di Stato.

La giurisprudenza amministrativa, ad oggi, è stata l'unica giurisdizione, in Italia, a confrontarsi con il problema definitorio riguardante l'intelligenza artificiale, anche se soltanto in un *obiter dictum* di una sentenza del Consiglio di Stato del 2021¹⁸⁰. Nonostante la definizione non sia concepita nella chiave teleologica dell'attribuzione di responsabilità, le argomentazioni addotte per giustificare una definizione *ristretta* di i.a. possono risultare utili anche in prospettiva penalistica.

Innanzitutto, vediamo il fatto. Oggetto del giudizio, in quel caso, era una gara per la fornitura di *pacemaker* “di alta fascia” per gli enti sanitari lombardi. La lettera di invito e il capitolato tecnico indicavano, tra i criteri di valutazione dell'offerta, il parametro tabellare “*Algoritmo di prevenzione+trattamento delle tachiaritmie atriali*”: nel caso di presenza di entrambi gli algoritmi (di prevenzione e di trattamento) sarebbero stati assegnati 15 punti; nel caso di presenza di uno solo dei citati algoritmi (di prevenzione o di trattamento), sarebbero stati assegnati 7 punti.

Vincitrice della gara risultava la società Microport; la prima esclusa, al secondo posto, era Abbott. Nel valutare la sussistenza dei parametri anzidetti, la Commissione di gara aveva dato rilievo esclusivamente agli algoritmi dotati di “autonomia”. Posto che entrambi i prodotti erano dotati di algoritmi “autonomi” di prevenzione, a fare la differenza era stata la presenza, nel *device* prodotto da Microport, di algoritmi “autonomi” di trattamento, che consentivano «in maniera automatica di contrastare il ritmo prefibrillatorio costituito dal riconoscimento di frequenti ectopie atriale e trattato mediante riduzione/omogeneizzazione dei periodo refrattari atriali». Al contrario, l'algoritmo di trattamento utilizzato da Abbott non veniva considerato autonomo, poiché poteva essere attivato soltanto presso ambulatori cardiologici: attraverso un programmatore esterno, un operatore clinico poteva assumere temporaneamente il controllo del *pacemaker*, così da impartire, sulla base della valutazione in tempo reale del ritmo cardiaco, una sequenza di stimoli da erogare a scopo terapeutico.

¹⁸⁰ Cons. Stato, sez III, sent. 25 novembre 2021, n. 7891; vd. i commenti di F. PAOLUCCI, *Algoritmi e intelligenza artificiale alla ricerca di una definizione: l'esegesi del Consiglio di Stato, alla luce dell'AI Act*, in *Quest. giust.*, 8 aprile 2022; C. FILICETTI, *Sulla definizione di algoritmo (nota a Consiglio di Stato, Sezione Terza, 25 novembre 2021, n. 7891)*, in *Giust. ins.*, 8 febbraio 2023.

Abbott impugnava l'aggiudicazione dinanzi al TAR Lombardia, lamentando che il parametro “*algoritmo di prevenzione e trattamento delle tachiaritmie atriali*” non avesse nulla a che fare con il carattere di autonomia di un sistema. Il Tribunale accoglieva il ricorso presentato da Abbot, affermando che «il dato testuale della lettera di invito non richiede che l'algoritmo di trattamento, al verificarsi dell'episodio aritmico, sia avviato dal dispositivo medesimo in automatico. Tale caratteristica attiene a una componente ulteriore, non indicata nella legge di gara, vale a dire a un algoritmo di intelligenza artificiale nella diagnosi dell'aritmia e avvio del trattamento. Fondatamente, pertanto, Abbott ha dedotto l'erroneità della valutazione della commissione di gara che - pur in presenza di un algoritmo di trattamento delle aritmie nel proprio dispositivo (vale a dire l'algoritmo NIPS, pacificamente definibile come tale) - ha attribuito soli 7 punti anziché 15 al dispositivo offerto. Infatti, la commissione ha confuso, sovrapponendoli indebitamente, il concetto di algoritmo con quello di avvio automatico del trattamento».

Microport proponeva allora ricorso al Consiglio di Stato, contestando la ricostruzione operata dalla Corte territoriale. Pur riconoscendo che – nella sua accezione comune – il termine “algoritmo” indichi «semplicemente una sequenza finita di istruzioni, ben definite e non ambigue, così da poter essere eseguite meccanicamente e tali da produrre un determinato risultato» (questa la definizione fornita dal giudice di prima istanza), il Consiglio di Stato accoglie il ricorso presentato da Microport, ritenendo che tale termine, quando riferito a sistemi tecnologici, «[sia] ineludibilmente collegat[o] al concetto di automazione ossia a sistemi di azione e controllo idonei a ridurre l'intervento umano».

In base a tali premesse, il Consiglio di Stato esclude che il *device* prodotto da Abbott potesse essere qualificato come “algoritmo” ai sensi del bando di gara, atteso che esso «non interviene in funzione di automazione delle funzioni di prevenzione e trattamento delle tachiaritmie atriali come richiesto dall'amministrazione, e dunque correttamente l'amministrazione non lo ha considerato ai fini del punteggio».

La decisione non pare esente da criticità, soprattutto nella parte in cui presume che, con l'espressione “dispositivi di alta fascia”, la pubblica amministrazione intendesse riferirsi ad algoritmi dotati di “autonomia”. Ciò che in questa sede interessa, tuttavia, è la definizione di “intelligenza artificiale”, contrapposta a quella di “algoritmo”. Nelle parole del Consiglio di Stato, l'i.a. è «un sistema che non si limita solo ad applicare le regole

software e i parametri preimpostati (come fa invece l'algoritmo "tradizionale") ma, al contrario, elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico». In breve, il Consiglio di Stato assimila la nozione di i.a. a quella di *machine learning*, aderendo dunque ad una definizione maggiormente restrittiva di quella proposta in sede europea.

7. Una proposta di classificazione per i sistemi di i.a. (con relativa casistica).

Delimitato così il nostro campo d'indagine, ci proponiamo ora di fornire alcuni esempi concreti di sistemi di i.a. il cui comportamento può risultare *imprevedibile*, verificando altresì le modalità con le quali giurisprudenza (ancora in maniera limitata) e legislazione (in maniera sempre più frequente) stanno affrontando le problematiche principali derivanti dai possibili danni cagionati dagli agenti intelligenti.

Ci limiteremo esclusivamente ai sistemi che possono compiere – e, nella maggior parte dei casi, *hanno già compiuto* – fatti che potrebbero integrare fattispecie penalmente rilevanti: dunque, non ci soffermeremo, ad esempio, sui sistemi di giustizia e polizia predittiva, dal cui utilizzo – seppur problematico (vd. *supra*, § 3.1.) – non pare possa sorgere alcuna violazione di norme incriminatrici. Il campo di indagine prescelto è, da un lato, più ampio e, dall'altro, più ristretto rispetto a quello che potrebbe interessare la responsabilità aquiliana: più ampio, perché alcuni degli incidenti che riportiamo sono considerati, dal diritto civile, danni *da inadempimento contrattuale* (si pensi, ad esempio, al medico, che, nel caso in cui abbia già instaurato un rapporto professionale con il paziente, risponde dei danni cagionati ai sensi dell'art. 1218 c.c.¹⁸¹); più ristretto, perché la clausola generale dell'art. 2043 c.c. delinea una nozione *atipica* di fatto illecito¹⁸², in maniera difforme rispetto all'esigenza di tipicità espressa al principio di legalità penale (art. 25, co. 2, Cost.).

¹⁸¹ In materia si vd., per tutti, G. ALPA, *Ars interpretandi e responsabilità sanitaria a seguito della nuova legge Bianco-Gelli*, in *Contr. impr.*, 2017, n. 3, p. 728 ss.

¹⁸² Si vd., per tutti, G. VISINTINI, *Atipicità dei fatti illeciti e danno ingiusto*, in A. Somma e al. (a cura di), *Dialoghi con Guido Alpa. Un volume offerto in occasione del suo LXXI compleanno*, Roma Tre Press., 2018, p. 589 ss.

La classificazione che proponiamo è basata sulle modalità con le quali i sistemi di i.a. interagiscono con l'esterno e, di conseguenza, sul tipo di bene giuridico sul quale il sistema di i.a. può incidere, attraverso la commissione di fatti penalmente tipizzati.

Una prima, fondamentale, distinzione è quella tra sistemi che, nel lasso di tempo in cui sono attivi, possono agire sul mondo (reale o virtuale) *direttamente* (§ 7.1) e sistemi che, invece, si limitano a fornire *raccomandazioni* e necessitano di un filtro umano per incidere concretamente sui beni giuridici (§ 7.2)¹⁸³. Tra i primi vi sono, innanzitutto, i dispositivi caratterizzati da *embodiment*: si tratta dei c.d. *robot*, in cui il *software* è incorporato in un *hardware* (§ 7.1.1). Tali sistemi agiscono nel mondo fisico e possono dunque incidere sull'integrità delle persone, cagionando morti e lesioni. Nel proseguo, ci occuperemo di alcune delle principali ipotesi di *robot* oggi in commercio – che abbiamo selezionato sulla base della *specificità* dei problemi che pongono (*autonomous vehicles*, § 7.1.1.1; sistemi di guida autonoma degli aerei, § 7.1.1.2; sistemi d'arma autonomi letali, § 7.1.1.3; *healthcare robots*, § 7.1.1.4). Va in ogni caso considerato che le applicazioni robotiche sono in crescita in tutti i settori: si pensi, ad esempio, allo sviluppo di *care robots* (o *assistive social robots*)¹⁸⁴ – che hanno l'obiettivo di fornire assistenza alle persone anziane o disabili – o ai progetti di robotica intelligente in campo industriale¹⁸⁵ e ludico¹⁸⁶.

Sempre all'interno della categoria dei sistemi che possono incidere *direttamente* sui beni giuridici tutelati dal diritto penale, vi sono gli algoritmi privi di *embodiment*, i c.d. *bots* (§ 7.1.2) che possono compiere un'ampia gamma di fatti illeciti, a seconda

¹⁸³ Riconosce carattere fondamentale a questa prima distinzione anche A.D. SELBST, *Negligence and AI's Human Users*, cit., p. 1318 ss.

¹⁸⁴ Il dibattito sulla regolazione dei *care robots* è particolarmente acceso, poiché lo sviluppo di sistemi robotici aventi compiti di cura, assistenza e compagnia pone problemi specifici e ulteriori rispetto alle questioni attinenti all'attribuzione della responsabilità per comportamenti robotici lesivi. In particolare, se da un lato, si pone l'esigenza di tutelare la funzione sociale svolta da questi sistemi sul piano dell'assistenza e, talvolta, dell'alleviamento delle condizioni di malessere psichico, dall'altro, la relazione con il *robot* rischia di generare dipendenze ed essere fonte di abusi; vd. A. BERTOLINI – G. AIELLO, *Robot companions: A legal and ethical analysis*, in *The Information Society*, 2018, vol. 34, no. 3, p. 130 ss.; H. ROBINSON E AL., *The Role of Healthcare Robots for Older People at Home: A Review*, in *Int. J. Social Robotics*, 2014, 6, p. 575; in lingua italiana vd. anche E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 469 ss.

¹⁸⁵ R.S. PERES, *Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook*, in *IEEE Access*, vol. 8, pp. 220121 ss.

¹⁸⁶ A. MCSTAY – G. ROSNER, *Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies*, in *Big Data & Society*, 2021, p. 1 ss.

dell'ambiente virtuale in cui operano: così, ad esempio, i *social bots* possono pubblicare post diffamatori o di istigazione all'odio *online* (vd. § 7.1.2.1), mentre i sistemi di *trading* algoritmico possono provocare sensibili e artificiose alterazioni dei prezzi di strumenti finanziari, potenzialmente integranti la fattispecie di manipolazione del mercato (vd. § 7.1.2.2).

Passando, invece, agli *intelligent decision support systems* (c.d. IDSS), ci limiteremo, in questa sede, ad affrontare il tema dei sistemi di supporto alla diagnosi e al trattamento in ambito medico (§ 7.2.1), che costituiscono, in questo campo, la principale applicazione – in termini di diffusione e di efficienza – delle tecniche di i.a.

7.1. Sistemi che possono recare un'offesa diretta ai beni giuridici.

7.1.1. Embodied systems

7.1.1.1. Autonomous vehicles.

Come abbiamo già accennato, il settore automobilistico è tra i più avanzati e promettenti nell'utilizzo dell'intelligenza artificiale. Per questo motivo, ci soffermeremo sugli *autonomous vehicles* (AV) in maniera più diffusa rispetto agli altri sistemi di i.a. – sia in questa sede, sia nei prossimi capitoli – utilizzandoli come *caso paradigmatico* dei problemi inerenti all'attribuzione della responsabilità penale per i “danni da algoritmo intelligente”¹⁸⁷.

¹⁸⁷ La letteratura che si è occupata dell'imputazione dei danni derivanti da macchine a guida autonoma è già amplissima (soprattutto in ambito civilistico). Senza alcuna pretesa di esaustività si vd., quanto alla dottrina italiana, A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit.; A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit.; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit.; G. CALABRESI – E. AL MUREDEN, *Driverless Cars*, Il Mulino, 2021, p. 156; A. LEPORE, *I.A. e responsabilità civile. Robot, autoveicoli e obblighi di protezione*, in *Tecnologie e diritto*, 2021, p. 198; M. TAMPIERI, *L'intelligenza artificiale: una nuova sfida anche per le automobili*, in *Contr. Impr.*, 2020, n. 2, p. 748; S. PELLEGGATTA, *Automazione nel settore automotive: profili di responsabilità civile*, in *Contr. impr.*, 2019, n. 4, p. 1429-1431; F. BARTOLINI, *Auto a guida autonoma e problemi di responsabilità civile*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 306-307; U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, in *Giur. it.*, luglio 2019, p. 1704 ss.; sui profili penalistici vd. A. CAPPELLINI, *Profili penalistici delle self-driving cars*, in *Dir. pen. cont. – Riv.*

L'obiettivo di lungo periodo dei ricercatori e delle aziende che producono *autonomous vehicles* è la completa automazione delle funzioni di guida, che determinerebbe la scomparsa di volante e pedali, con la trasformazione del conducente umano in un semplice passeggero. Nelle intenzioni dei produttori, le *self-driving cars* del futuro saranno inoltre connesse tra loro e con le infrastrutture stradali; potranno dunque coordinare i propri comportamenti, sulla base delle informazioni ricevute dalla rete stradale interattiva.

La definitiva affermazione di veicoli completamente autonomi comporterebbe, innanzitutto, una riduzione del numero degli incidenti e delle relative vittime – oggi dovuti principalmente alla distrazione dei conducenti o alla guida in stato di ebbrezza o sotto l'effetto di sostanze stupefacenti. La diffusione di *autonomous vehicles* potrebbe inoltre portare numerosi benefici alla società: dall'incremento della mobilità – e dunque dell'indipendenza – di anziani e disabili¹⁸⁸, passando per la gestione più efficiente delle risorse energetiche¹⁸⁹, fino al non trascurabile risparmio di tempo per tutte quelle persone che utilizzano le automobili come principale mezzo di trasporto.

Ad oggi, tuttavia, nessuna casa di produzione automobilistica ha ancora sviluppato una *self-driving car* pronta per essere testata o messa in commercio (c.d. AV di livello 5, vd. *infra*, § 7.1.1.1.1.). Gli ostacoli non riguardano esclusivamente lo stato di avanzamento della tecnologia, ma sono dovuti soprattutto al rischio legale, poiché manca

Trim., 2019, n. 2, p. 326 ss.; L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit. Nella letteratura statunitense si vd. DOUMA – S.A. PALODICHUK, *Criminal Liability Issues Created by Autonomous Vehicles*, in 52 *Santa Clara Law Review*, p. 1157 ss.; K.S. ABRAHAM – R.L. RABIN, *Automated vehicles and manufacturer responsibility for accidents: a new legal regime for a new era*, in *Va. L. Rev.*, 2019, p. 127 ss.; M.A. GEISTFELD, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, in 105 *Calif. L. Rev.*, 2017, p. 1611 ss.; ID., *The Regulatory Sweet Spot for Autonomous Vehicles*, in 53 *Wake Forest L. Rev.*, 2018, p. 337 ss.; J. GURNEY, *Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles*, in P. Lin e al. (eds.) *Robot Ethics 2.0*, cit., p. 51 ss.; M. ROE, *Who's Driving That Car?: An Analysis of Regulatory and Potential Liability Frameworks for Driverless Cars*, in 60 *B.C. L. Rev.*, 2019, p. 315 ss.; D.G. GIFFORD, *Technological Triggers to Tort Revolutions: Steam Locomotives, Autonomous Vehicles, and Accident Compensation*, 11 *Journal of Tort Law*, 2018, p. 71 ss.; nella letteratura tedesca vd. E. HILGENDORF, *Automated Driving and the Law*, cit; M. EBERS, *Civil liability for autonomous vehicles in Germany*, 2022, in SSRN, disponibile a questo link: <https://ssrn.com/abstract=4027594>.

¹⁸⁸ Vd. S. Saripalli, *Are Self-Driving Cars the Future of Mobility for Disabled People?*, in *The Conversation*, 5 ottobre 2017, disponibile a questo link: <https://theconversation.com/are-self-driving-cars-the-future-of-mobility-for-disabled-people-84037>.

¹⁸⁹ Vd. E. HILGENDORF, *Automated Driving and the Law*, in E. Hilgendorf – U. Seidel (eds.), *Robotics, Autonomics, and the Law*, cit., p. 174.

una normativa di riferimento – a livello nazionale, così come a livello internazionale – che disciplini l’attribuzione di responsabilità, civile e/o penale, per il caso in cui il sistema di i.a. completamente autonomo causi danni a cose e persone. Nel silenzio della legge, il rischio, ben percepito dalle società produttrici, è che – in assenza del conducente, che, come vedremo, ad oggi è un vero e proprio catalizzatore di responsabilità – anche i comportamenti emergenti dei sistemi di guida autonoma siano attribuiti ai produttori, benché non fossero da essi prevedibili.

D’altra parte, anche quando esistenti, i testi normativi attualmente vigenti prevedono sempre la presenza di un *human-in-the-loop*, ossia di una persona che *condivide* la funzione di guida con il sistema di i.a.¹⁹⁰. La rinuncia, per ora, ai benefici che potrebbero derivare dall’utilizzo diffuso dei sistemi di guida totalmente autonomi sembra derivare proprio dal timore di un *vuoto di responsabilità*, per il caso in cui si verificano incidenti, nonché dalla controversa attribuzione agli algoritmi della risoluzione di questioni eticamente sensibili – quale, in particolare, la scelta di quali persone sacrificare, in situazioni di pericolo, nel caso in cui sia impossibile un intervento integralmente salvifico (c.d. *Trolley problem*)¹⁹¹.

¹⁹⁰ La prospettiva dell’*human-in-the-loop*, tra l’altro, pare permeare anche la bozza di *AI Act*. Vd. Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull’Intelligenza artificiale, cit., art. 14, co. 1: «I sistemi di IA ad alto rischio sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso».

¹⁹¹ Vd. M. CUNNEN E AL., *Artificial Driving Intelligence and Moral Agency: Examining the Decision Ontology of Unavoidable Road Traffic Accidents through the Prism of the Trolley Dilemma*, in *Applied Artificial Intelligence - An International Journal*, 2019, Vol. 33, No. 3, p. 267 ss.; W.B. WENDEL, *Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles*, in 56 *California Western Law Review* 129, 2018; I. COCA VILA, *Coches autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal*, in *Cuadernos de política criminal*, 2017, 235 ss.; in generale, sul *trolley problem*, vd. D. EDMONDS, *Uccideresti l'uomo grasso? Il dilemma etico del male minore*, Raffaello Cortina Editore, 2014. Nota correttamente Ruffolo che, in una fase di transazione, in cui le macchine a guida autonoma si affiancheranno ai veicoli tradizionali, l’utilizzatore dell’*autonomous vehicle* dovrà ragionevolmente attendersi di non essere sacrificato in favore dell’incolumità personale di terzi (es. i pedoni che attraversano all’improvviso un incrocio), conformemente alla logica che sta alla base dell’esimente dello stato di necessità (art. 54 c.p.), vd. U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1708. Vd. anche G. CONTISSA – F. LAGIOIA – G. SARTOR, *La manopola etica: i veicoli autonomi eticamente personalizzabili e il diritto*, in *Sistemi Intelligenti*, 2017, n. 3, p. 601 ss., che propongono che i sistemi di guida autonoma assicurino sempre la possibilità, per il conducente, di decidere in via preventiva i criteri da utilizzare in caso di “scelte tragiche”.

7.1.1.1.1. Quale grado di autonomia? La classificazione SAE.

Il grado di autonomia di un *autonomous vehicle* viene definito in funzione del ruolo che deve essere svolto dall'utilizzatore per garantire la funzionalità del veicolo ed un adeguato livello di sicurezza.

Nel corso del tempo sono state elaborate numerose tassonomie, da parte di diversi organi istituzionali. Si tratta di documenti descrittivi ed informativi, non normativi: si limitano dunque ad individuare dei criteri razionali per distinguere tra loro i sistemi autonomi, senza tuttavia ricollegare a tale classificazione alcuna disciplina.

In questa sede, analizzeremo la classificazione elaborata dalla *Society of Automotive Engineers* (SAE), in collaborazione con l'*International Organization for Standardization* (ISO), pubblicata nel 2014 e modificata a più riprese nel corso degli anni¹⁹² (c.d. *SAE Levels of Driving Automation*)¹⁹³). La scelta è dovuta al fatto che si tratta della classificazione maggiormente utilizzata, a livello internazionale, per distinguere tra loro gli *Automated Driving Systems* e, dal punto di vista normativo, per individuare la disciplina applicabile a seconda del grado di autonomia.

Prima di passare ad illustrare brevemente i cinque livelli di automazione, va innanzitutto sgombrato il campo da equivoci. Con l'espressione "*Driving automation Systems*" ci si riferisce a tutti quei *software* che contribuiscono a svolgere continuamente una funzione di guida: non rientrano in tale nozione – e sono dunque esclusi dalla classificazione – i sistemi di sicurezza attiva (*Active safety systems*), come il controllo elettronico della stabilità (*electronic stability control*, ESC) e la frenata automatica d'emergenza (*automatic emergency braking*, AEB), e alcuni tipi di sistemi di assistenza alla guida, come l'assistenza al mantenimento della corsia (*lane keeping*

¹⁹² L'ultima modifica risale al 30 aprile 2021 ed è disponibile sul sito della SAE, al seguente link: https://www.sae.org/standards/content/j3016_202104/. Si veda anche la classificazione proposta dalla *National Highway Traffic Safety Administration* (NHTSA) statunitense, disponibile a questo link: <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety>. Per un esame comparativo tra la classificazione della SAE, quella della NHTSA e quella proposta dal *Bundesanstalt für Straßenwesen* (BASt) tedesco vd. B.S. Jahromi, *Levels of Automation for Autonomous Ground Vehicles*, in *Medium*, disponibile a questo link: <https://medium.com/@BabakShah/levels-of-automation-for-self-driving-cars-d410a4f679b7>.

¹⁹³ SAE J3016™ Recommended Practice: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, d'ora in avanti *SAE Levels of Driving Automation*.

assistance, LKA), dal momento che essi si limitano ad intervenire soltanto per brevi lassi temporali, durante situazioni potenzialmente pericolose. Stante l'apporto necessariamente momentaneo di tali strumenti, essi non modificano o eliminano il ruolo del conducente nell'eseguire la funzione di guida (*dynamic driving task*, DDT), e dunque – pur svolgendo funzioni automatizzate¹⁹⁴ – non sono considerati *Driving automation Systems*.

Come anticipato, la SAE individua cinque categorie di *Driving Automation System*.

(i) Al livello 1 (c.d. *Drive Assistance*)¹⁹⁵, il veicolo può eseguire autonomamente o il movimento longitudinale o quello laterale (non può, invece, eseguire entrambi i movimenti simultaneamente) e il sistema autonomo si disinserisce immediatamente su richiesta del conducente. Esempi di sistemi di livello 1 possono essere, quanto al movimento longitudinale, il controllo adattivo della velocità di crociera, e, quanto al movimento laterale, il parcheggio del veicolo.

(ii) Al livello 2 (c.d. *Partial Driving Automation*)¹⁹⁶, il veicolo può eseguire autonomamente e simultaneamente sia il movimento laterale che quello longitudinale: può dunque svolgere interamente, in determinate circostanze (es. tempo soleggiato, specifica area georeferenziata) e per un certo periodo di tempo, la funzione di guida, nei suoi aspetti *operazionali*. Dal momento che, tuttavia, le capacità del sistema di effettuare la *object and event detection and response* (OEDR) sono limitate¹⁹⁷, il conducente deve supervisionarne l'attività ed intervenire ogniqualvolta sia necessario. I veicoli di livello 2 sono già largamente utilizzati negli Stati Uniti, dove – come vedremo (vd. *infra*, § 7.1.1.1.3.) – sono già stati coinvolti in diversi incidenti stradali.

(iii) Al livello 3 (c.d. *Conditional Driving Automation*)¹⁹⁸ il veicolo può svolgere interamente, in determinate circostanze e per un certo periodo di tempo, la funzione di

¹⁹⁴ SAE Levels of Driving Automation, cit., p. 4.

¹⁹⁵ SAE Levels of Driving Automation, cit., § 5.2.

¹⁹⁶ SAE Levels of Driving Automation, cit., § 5.3.

¹⁹⁷ È stato notato, ad esempio, che il sistema Autopilot prodotto da Tesla non riconosceva le luci lampeggianti delle ambulanze e delle macchine della polizia; il problema, tuttavia, dovrebbe essere stato risolto nei modelli più recenti, vd. J. Ramey, *Tesla Autopilot Will Now Detect Emergency Lights, but Only at Night*, in *Autoweek*, 22 settembre 2021, disponibile a questo link: <https://www.autoweek.com/news/green-cars/a37694444/tesla-autopilot-will-now-detect-emergency-lights-but-only-at-night/>.

¹⁹⁸ SAE Levels of Driving Automation, cit., § 5.4.

guida; a differenza che per i veicoli di livello 2, tuttavia, il sistema è in grado di riconoscere – quando si presentano – le situazioni nelle quali non è in grado di operare, avvertendo con un certo preavviso il conducente affinché acquisisca il controllo del veicolo. Il conducente, di conseguenza, non deve monitorare costantemente il sistema, ma deve restare reattivo, pronto ad intervenire se richiesto; in tal caso, potrà decidere se assumere la funzione di guida, o fermare la macchina in un luogo sicuro, garantendo così il raggiungimento di una *minimal risk condition* (c.d. *fallback*). Tra i primi paesi ad aver consentito la circolazione di veicoli di livello 3 vi è la Germania, che, nel 2017¹⁹⁹, ha modificato a tale scopo il suo codice della strada (*Straßenverkehrsgesetz*, StVG)²⁰⁰; soltanto nel dicembre 2021, tuttavia, è stata autorizzata la vendita e la circolazione di un *autonomous vehicle* di livello 3 (in particolare, il sistema *Drive Pilot*, prodotto da Mercedes Benz)²⁰¹. Anche in Francia è stata approvata un'apposita disciplina per la circolazione di veicoli di livello 3 di guida autonoma; , ad oggi, tuttavia, tali autoveicoli possono circolare soltanto in via sperimentale²⁰².

(iv) Al livello 4 (c.d. *High Driving Automation*)²⁰³, il veicolo può svolgere interamente la funzione di guida qualora sussistano ben definite circostanze di tempo e luogo. A differenza che per i veicoli di livello 3, tuttavia, l'utente a bordo è considerato un passeggero, non un conducente, e dunque non deve intervenire per riprendere la funzione di guida, qualora il sistema presenti guasti o difetti nella circolazione; in questi

¹⁹⁹ Bundesgesetzblatt Jahrgang 2017 Teil I Nr. 38, ausgegeben am 20.06.2017, Seite 1648, disponibile a questo link: https://dejure.org/BGBI/2017/BGBI_I_S_1648.

²⁰⁰ Il testo della legge è accessibile sul sito del Ministero della Giustizia tedesco a questo link: <https://www.gesetze-im-internet.de/stvg/>; una traduzione in inglese (non ufficiale) delle disposizioni modificate è disponibile come annesso all'articolo di M. EBERS, *Civil liability for autonomous vehicles in Germany*, cit., p. 30. Nonostante il testo delle norme faccia riferimento a veicoli “con funzioni di guida altamente o totalmente automatizzate” (*Kraftfahrzeuge mit hochoder vollautomatisierter*), la definizione di cui alla Sezione 1a(2) StVG rispecchia le caratteristiche del livello 3 di automazione SAE, vd. M. EBERS, *Civil liability for autonomous vehicles in Germany*, cit., p. 7.

²⁰¹ Vd. Red. Reuters, *UPDATE 1-Germany clears Mercedes-Benz's hands-free drive system*, in Reuters, 9 dicembre 2021, disponibile a questo link: <https://www.reuters.com/article/daimler-autonomous-idCNL8N2SU35K>. La circolazione di veicoli di livello 3 è inoltre già consentita in Giappone, vd. J. Solomon, *World's first certified level 3 autonomous car to hit streets of Japan*, 5 marzo 2021, in Autox, disponibile a questo link: <https://www.autox.com/news/car-news/worlds-first-certified-level-3-autonomous-car-to-hit-streets-of-japan-109099/>.

²⁰² Vd. legge n. 2019-486 del 22 maggio 2019, *Plan d'action pour la croissance et la transformation des entreprises*, c.d. Loi PACTE, che autorizza la circolazione dei veicoli di livello 3 in via sperimentale, previa autorizzazione.

²⁰³ SAE Levels of Driving Automation, cit., § 5.5.

casi, il veicolo è in grado di effettuare autonomamente il *fallback*. In Germania, a partire dal luglio 2021²⁰⁴, il codice della strada consente l'utilizzo di sistemi con "funzioni di guida autonoma in aree operative definite" (*Kraftfahrzeuge mit autonomer Fahrfunktion in festgelegten Betriebsbereichen*), ossia – in base alla definizione fornita dalla Sezione 1d(1) StVG – di sistemi con livello 4 di automazione. Ad oggi, tuttavia, non è stata ancora approvato nessun veicolo di livello 4.

(v) Al livello 5 (c.d. *Full Driving Automation*)²⁰⁵, il veicolo svolge interamente la funzione di guida su qualsiasi tipo di strada e in tutte le circostanze in cui un'automobile tradizionale può essere utilizzata (non ci sono, dunque, restrizioni climatiche, temporali o geografiche, a differenza di quanto accade per le automobili di livello 4). Il veicolo non è in grado di operare soltanto in condizioni straordinarie e ingestibili (es. tempesta di neve, strade allagate), nelle quali nemmeno un conducente riuscirebbe a guidare in maniera sicura; in questi casi, il sistema eseguirà autonomamente il *fallback*.

Di queste cinque categorie, soltanto le ultime tre costituiscono – sulla base dei criteri utilizzati dalla SAE – degli *Automated Driving Systems* (ADS), ossia dei sistemi che, in determinate condizioni – che variano a seconda del livello di automazione – sono in grado di eseguire l'intera funzione di guida su base continua²⁰⁶. I livelli 1 e 2 – pur essendo inclusi tra i *Driving Automation System*²⁰⁷ – non sono invece considerati ADS, poiché sono in grado di svolgere soltanto *una parte* della funzione di guida, lasciando la parte rimanente nelle mani del conducente.

²⁰⁴ Bundesgesetzblatt Jahrgang 2021 Teil I Nr. 48, ausgegeben am 27.07.2021, Seite 3108, disponibile a questo link: <https://dejure.org/BGBI/2021/BGBl. I S. 3108>.

²⁰⁵ SAE Levels of Driving Automation, cit., § 5.6.

²⁰⁶ Vd. SAE Levels of Driving Automation, p. 7, § 3.6: « *AUTOMATED DRIVING SYSTEM (ADS): The hardware and software that are collectively capable of performing part or all of the DDT on a sustained basis; this term is used generically to describe any system capable of Level 1 to 5 driving automation* ».

²⁰⁷ Vd. SAE Levels of Driving Automation, p. 7, § 3.2: « *DRIVING AUTOMATION SYSTEM OR TECHNOLOGY: The hardware and software that are collectively capable of performing the entire DDT on a sustained basis, regardless of whether it is limited to a specific operational design domain (ODD); this term is used specifically to describe a Level 3, 4, or 5 driving automation system* ». L'operational design domain è così definito: « *Operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics* » (p. 17, § 3.21).

7.1.1.1.2. Il quadro normativo (cenni).

Gli Stati Uniti sono stati tra i primi paesi ad avviare sperimentazioni su strada delle auto a guida autonoma: la mancata adesione alla Convenzione di Vienna sulla circolazione stradale (1968) – che, come vedremo a breve, fino a tempi recentissimi richiedeva la presenza, nell'autoveicolo, di un conducente umano – ha consentito la circolazione di *autonomous vehicles* anche in assenza di normative specifiche²⁰⁸.

Diversi Stati, a partire dal Nevada, hanno comunque varato discipline specifiche volte a regolare la circolazione di *autonomous vehicles* di livello 2 (per i quali, dunque, è ancora prevista la presenza di un conducente che supervisioni il sistema e sia pronto ad intervenire ogniquale volta si presenti un malfunzionamento)²⁰⁹. Dal novembre 2018, la California ha invece consentito la sperimentazione nel traffico di veicoli *driverless*²¹⁰. I testi legislativi introdotti negli stati americani regolano principalmente il regime autorizzatorio e i requisiti per la circolazione su strada degli *autonomous vehicles*. Nessuno di essi affronta invece, esplicitamente, il tema della responsabilità; tutt'al più, vi sono dei rinvii alla “disciplina applicabile” – un richiamo di scarsa utilità pratica, in una materia in cui oggetto di discussione è proprio *quale sia* la disciplina applicabile²¹¹. Ad oggi, invece, i tentativi di introdurre una disciplina federale sono stati vani²¹².

Quanto agli stati aderenti alla Convenzione di Vienna sulla circolazione stradale (1968) – tra i quali rientra la gran parte degli stati europei, compresa l'Italia – un primo emendamento, entrato in vigore il 23 marzo 2016, ha consentito la circolazione di sistemi

²⁰⁸ Così AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, in *Giur. it.*, luglio 2019, p. 1713.

²⁰⁹ Per accedere al database aggiornato delle legislazioni statali, negli Stati Uniti, in materia di Autonomous Vehicles vd. NATIONAL CONFERENCE OF STATE LEGISLATURES, *Autonomous Vehicles State Bill Tracking Database*, disponibile a questo link: <https://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx>

²¹⁰ California Code of Regulations, tit. 13, div. 1, ch. 1, art. 3.7, accessibile a questo link: <https://www.law.cornell.edu/regulations/california/title-13/division-1/chapter-1/article-3.7>.

²¹¹ Vd., A. GUERRA E AL., *Liability for Robots I: Legal Challenges*, in *Journal of Institutional Economics*, 2021, p. 6, che cita, a titolo di esempio, la normativa del Tennessee: “[l]iability for accidents involving an [Automated Driving System]-operated vehicle shall be determined in accordance with product liability law, common law, or other applicable federal or state law” (Tenn. Code Ann. §55-30-106(a) (2019)).

²¹² Vd. The SELF DRIVE Act (H.R. 3388_2017), il cui testo è disponibile a questo link: <https://www.congress.gov/bill/115th-congress/house-bill/3388/text>; The SELF DRIVE ACT (H.R. 3711), il cui testo è disponibile a questo link: <https://www.congress.gov/bill/117th-congress/house-bill/3711?s=1&r=55>.

di guida *parzialmente* autonoma, purché vi sia sempre la possibilità, per il conducente, di disattivarli (art. 8, co. 5-*bis*). Tuttavia, fino a poco tempo fa, era vietata la messa in commercio di veicoli a guida *completamente* autonoma, poiché la Convenzione imponeva che ogni veicolo dovesse essere costantemente sottoposto al controllo del conducente. Oggi anche questa limitazione è stata rimossa: dal 14 luglio 2022 sono infatti entrate in vigore alcune modifiche finalizzate a consentire la circolazione di veicoli a guida totalmente autonoma²¹³. In particolare, è stato introdotto nella Convenzione di Vienna l'art. 34-*bis*, che recita:

«The requirement that every moving vehicle or combination of vehicles shall have a driver is deemed to be satisfied while the vehicle is using an automated driving system which complies with:

(a) domestic technical regulations, and any applicable international legal instrument, concerning wheeled vehicles, equipment and parts which can be fitted and/or be used on wheeled vehicles, and
(b) domestic legislation governing operation.

*The effect of this Article is limited to the territory of the Contracting Party where the relevant domestic technical regulations and legislation governing operation apply»*²¹⁴.

Spetta dunque agli Stati parte alla Convenzione la decisione *se e come* consentire la circolazione di veicoli a guida autonoma.

Come è già emerso nel paragrafo precedente, la Germania ha assunto, a livello mondiale, un ruolo di *leadership* nella regolazione della circolazione delle macchine a

²¹³ Gli emendamenti, presentati il 14 gennaio 2021 dallo *United Nations Economic Commission for Europe* (UNECE), sono stati tacitamente approvati ai sensi dell'art. 49(2)(a) della Convenzione; vd. *United Nations, Convention on Road Traffic Vienna, 8 November 1968 – Acceptance of Amendment to Article 1 and New Article 34 Bis to the Convention*, 21 gennaio 2022, disponibile sul sito delle Nazioni Unite a questo link: <https://treaties.un.org/doc/Publication/CN/2022/CN.26.2022-Eng.pdf>. Già nel 2016, tuttavia, era stato consentito l'utilizzo di *sistemi di bordo che influiscono sulla guida del veicolo*, purché potessero essere “neutralizzati o disattivati dal conducente” (art. 8, co. 5-*bis*).

²¹⁴ Il testo degli emendamenti è disponibile sul sito dell'UNECE a questo link: <https://unece.org/sites/default/files/2021-01/ECE-TRANS-WP.1-173-Add1e.pdf> (proposta ECE/TRANS/WP.1/173/Add.1). Vd. anche art. 1 della Convenzione, in cui sono inserite le definizioni di “*automated driving system*” e “*Dynamic control*”: «(ab) “*Automated driving system*” refers to a vehicle system that uses both hardware and software to exercise dynamic control of a vehicle on a sustained basis. (ac) “*Dynamic control*” refers to carrying out all the real-time operational and tactical functions required to move the vehicle. This includes controlling the vehicle’s lateral and longitudinal motion, monitoring the road, responding to events in the road traffic, and planning and signalling for manoeuvres».

guida autonoma, approvando la prima disciplina organica in materia di ADS e consentendo la circolazione di veicoli di livello 3 e 4²¹⁵. Quanto ai primi, l'utilizzatore viene ancora considerato "conducente" del veicolo (*Fahrzeugführer*)²¹⁶. Quando è attiva la funzione di guida automatica, il conducente può distogliere la propria attenzione dalla guida; deve tuttavia rimanere "percettivo", così da poter intervenire quando il sistema lo richiede o quando il conducente riconosce o avrebbe dovuto riconoscere, sulla base di circostanze evidenti, che non esistono più i presupposti per l'uso previsto delle funzioni di guida automatiche²¹⁷. Per gli utilizzatori di sistemi di guida di livello 4 – che, tuttavia, come abbiamo già rilevato, non sono ancora entrati in circolazione in Germania – la legislazione tedesca non prevede invece alcun obbligo di monitoraggio. Scompare la figura del conducente (*Fahrzeugführer*) e compare quella del "supervisore tecnico" (*Technische Aufsicht eines Kraftfahrzeugs*), che può anche trovarsi *all'esterno* del veicolo e che deve intervenire soltanto quando vi sia una richiesta in questo senso da parte del sistema, riportando il veicolo in una situazione di rischio minimo²¹⁸.

Anche in Francia, con la c.d. *Loi PACTE* del 2019, è stata introdotta una disciplina che regola la circolazione di veicoli a guida autonoma di livello 3²¹⁹. Inoltre, l'*ordonnance* 14 avril 2021 n. 2021-443 ha modificato il *Code de la route* francese, inserendo un terzo capitolo all'interno del *Titre 2* rubricato "*Responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite (articles L123-1 à L123-4)*"²²⁰, che detta una disciplina speciale per la responsabilità del conducente e del produttore di *autonomous vehicle*.

Innanzitutto, l'art. L123-1, co. 1, *Code de la route*, esclude l'applicabilità degli obblighi di vigilanza previsti per il conducente di veicolo tradizionale, nel caso in cui – nel momento dell'infrazione – il sistema stava esercitando il "controllo dinamico" del

²¹⁵ Bundesgesetzblatt Jahrgang 2021 Teil I Nr. 48, cit.; sulla legislazione tedesca vd. diffusamente M. EBERS, *Civil liability for autonomous vehicles in Germany*, cit.; vd. anche L. D'AMICO, *Intelligenza artificiale e auto a guida autonoma. tra prevenzione primaria, colpa penale e rischio consentito*, in *Riv. it. med. leg.*, 2022, fasc. 3, p. 612 ss.

²¹⁶ Sez. 1a (4) StVG, cit.

²¹⁷ Sez. 1b(1) e (2) StVG, cit.

²¹⁸ Sezione 1d(3) StVG, cit.

²¹⁹ Vd. Legge n. 2019-486 del 22 maggio 2019, cit. Sulla disciplina francese vd. più approfonditamente M. GIUCA, *Disciplinare l'intelligenza artificiale. La riforma francese sulla responsabilità penale da uso di auto a guida autonoma*, in *Arch pen.*, 2022, n. 2.

²²⁰ Il testo della legge è disponibile su *Légifrance* a questo link: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043370894>.

veicolo (ovverosia, nel momento in cui la funzione di guida automatica era attiva)²²¹; il conducente dovrà tuttavia mantenersi sempre in uno stato di allerta, così da poter riprendere tempestivamente la funzione di guida, se richiesto dal sistema. Al comma 3, l'art. L123-1 stabilisce che laddove il conducente non risponda alla richiesta di riprendere il controllo avanzata dal sistema, tornerà ad applicarsi la disciplina generale sulla responsabilità penale del conducente (art. L121-1). Parrebbe dunque configurarsi una posizione di garanzia in capo al conducente, che ha l'obbligo di "restare allerta" e di attivarsi – quando richiesto dal sistema – per riprendere il controllo del veicolo: qualora non lo faccia, l'evento lesivo gli sarà addebitabile a titolo omissivo²²².

L'art. L123-2, co. 1, stabilisce invece che, per il lasso di tempo in cui il sistema esercita il "controllo dinamico" del veicolo, sarà il produttore a rispondere degli eventuali *délits* di omicidio o lesioni stradali – sempre che, ovviamente, ne sia accertata la colpa. Non è chiara la portata innovativa della norma rispetto all'ordinario regime della responsabilità penale colposa (art. 121-3 *Code pénal*): probabilmente, come è stato rilevato, l'intento del legislatore è stato quello di *circoscrivere* la responsabilità del produttore per gravi delitti contro l'incolumità personale alle sole ipotesi in cui vi è un controllo totale del veicolo da parte del sistema²²³.

Quanto, infine, all'Italia, una disciplina organica delle macchine a guida autonoma non pare all'orizzonte. Il Decreto del Ministero delle Infrastrutture e dei Trasporti del 28 febbraio 2018 recante *Modalità attuative e strumenti operativi della sperimentazione su strada delle soluzioni di Smart Road e di guida connessa e automatica* (c.d. Decreto *smart roads*)²²⁴ consente, previa autorizzazione (art. 9) – rilasciata a seguito di valutazione positiva della domanda (art. 11) e della verifica della presenza dei requisiti richiesti (art. 12) – la sperimentazione su strada di veicoli a guida parzialmente autonoma. La guida

²²¹ Per controllo dinamico si intende, nello specifico, l'«*exécution de toutes les fonctions opérationnelles et tactiques en temps réel nécessaires au déplacement du véhicule. Il s'agit notamment du contrôle du déplacement latéral et longitudinal du véhicule, de la surveillance de l'environnement routier, des réactions aux événements survenant dans la circulation routière et de la préparation et du signalement des manœuvres*», vd. Décret n° 2022-1034 du 21 juillet 2022 portant publication de l'amendement à la convention internationale sur la circulation routière de Vienne du 8 novembre 1968, adopté à Genève le 14 janvier 2022 (1), disponibile a questo link: https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000046081258.

²²² In questo senso M. GIUCA, *Disciplinare l'intelligenza artificiale*, cit., p. 24.

²²³ M. GIUCA, *Disciplinare l'intelligenza artificiale*, cit., p. 30-31.

²²⁴ Per un esame approfondito della normativa si vd. L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 817 ss.

deve avvenire sempre sotto controllo di un supervisore (art. 10) e su specifici tratti stradali dotati di adeguate infrastrutture (art. 17). Sebbene il testo normativo non faccia alcun riferimento alla tassonomia internazionale promossa dalla SAE, sembra di potersi ritenere che i veicoli cui si riferisce il decreto rientrino nel livello 2 o 3 di automazione.

Ai sensi dell'art. 11, co. 1, lett. a), la domanda di autorizzazione deve contenere anche «l'indicazione del proprietario del veicolo a guida automatica, quale soggetto responsabile ai sensi dell'art. 196 del codice della strada e dell'art. 2054, comma 3, del codice civile, ovvero di altro soggetto obbligato in solido, ai sensi del citato art. 196». Ad oggi, dunque, nel caso di danni cagionati dalla circolazione di un veicolo a guida autonoma, a risponderne oggettivamente sarà, sul piano civile, il proprietario del veicolo (oltre che, si presume, il conducente, ai sensi dell'art. 2054, co. 1, c.c.). All'infuori di tali regimi autorizzatori speciali, la circolazione di auto a guida autonoma deve intendersi implicitamente vietata dal Codice della strada, che richiede che ogni auto deve essere condotta da una persona che se ne assuma la responsabilità²²⁵.

Nel maggio del 2019, le prime auto a guida autonoma hanno iniziato a circolare, su percorsi appositamente predisposti per la sperimentazione, a Parma e a Torino²²⁶.

7.1.1.1.3. Incidenti che hanno coinvolto gli *autonomous vehicles* e prima casistica giudiziaria.

Sebbene, sul lungo periodo, l'utilizzo di sistemi di guida autonoma prometta di ridurre complessivamente il numero di sinistri stradali, ad oggi non sono mancati gli incidenti che hanno causato vittime: quasi non passa settimana senza che si abbia notizia di incidenti stradali che, da qualche parte nel mondo, hanno coinvolto degli *autonomous vehicles*²²⁷. Riportiamo qui alcuni dei casi verificatisi negli ultimi anni, concentrandoci,

²²⁵ A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 326. Vd. in particolare l'art. 46 Codice della Strada: «si intendono per veicoli tutte le macchine di qualsiasi specie, che circolano sulle strade guidate dall'uomo».

²²⁶ A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 326 (e fonti ivi citate).

²²⁷ Ci sono diversi *database online* che raccolgono e classificano gli incidenti che hanno coinvolto veicoli a guida autonoma; tra i più aggiornati vd. il sito *Autonomous Vehicle Crashes* (<https://www.avcrashes.net>), che dà anche la possibilità di visualizzare i dati attraverso una mappa interattiva. In data 27 dicembre 2022, il sito contava 576 incidenti, realizzatisi tra il 14 ottobre 2014 e il 13 dicembre 2022.

in particolare, su quelli che hanno avuto (o stanno avendo) un seguito giudiziario: si tratta, perlopiù, di incidenti che hanno coinvolto veicoli di livello 2 e per i quali sono state adite corti statunitensi.

Partiamo, innanzitutto, dai procedimenti che si stanno svolgendo (o si sono svolti) in sede civile.

In un caso²²⁸, *Umeda v Tesla*, risalente all'aprile 2018, un cittadino giapponese, Yoshihiro Umeda, veniva colpito e ucciso da un veicolo Tesla Model X, mentre si trovava sulla corsia d'emergenza dell'autostrada a prestare soccorso ad un suo amico motociclista. Al momento dell'incidente, il sistema *Autopilot* della Tesla era inserito; il veicolo aveva improvvisamente accelerato – probabilmente “confuso” da un repentino cambio di corsia della macchina antistante – colpendo i pedoni presenti sulla corsia di emergenza²²⁹.

I parenti della vittima avevano presentato inoltre ricorso contro Tesla per ottenere il risarcimento dei danni, sostenendo che il sistema di guida conteneva un difetto di progettazione. Nonostante la corte statunitense adita abbia dichiarato di non avere giurisdizione sul caso (*forum non conveniens*), accogliendo le istanze di Tesla²³⁰, la vicenda merita comunque di essere esaminata, stante l'interesse che destano le argomentazioni addotte nel ricorso, fondate sui principi della responsabilità per danno da prodotto (vd. *infra*, cap. II, § 3. ss.)²³¹.

In particolare, i querelanti adducevano che Tesla *Autopilot* avesse alcuni difetti di produzione, dal momento che il sistema non sarebbe riuscito a riconoscere la presenza dei pedoni, né avrebbe allertato il conducente del malfunzionamento in corso. Tesla avrebbe inoltre saputo che il prodotto era difettoso: «*from several sources, including but not limited to the results of crash testing, internal memoranda and correspondence,*

²²⁸ United States District Court, N.D. California, *Tomomi UMEDA, et al v. TESLA INC.*, Case No. 20-cv-02926-SVK, 23 settembre 2020, in *Westlaw*, WL 5653496.

²²⁹ Nella sentenza si legge che il conducente del veicolo era già stato condannato da un tribunale giapponese per omicidio stradale, poiché risultava che aveva avuto un colpo di sonno alla guida e non era stato in grado di impedire l'incidente.

²³⁰ I giudici si dichiaravano privi di giurisdizione poiché sia il conducente della Tesla, sia la vittima e i parenti della vittima che avevano presentato la querela, erano tutti cittadini giapponesi; inoltre, la Tesla coinvolta nell'incidente è stata venduta in Giappone.

²³¹ United States District Court, N.D. California, San Jose Division, *Tomomi UMEDA, and Miyu Umeda v. TESLA, INC. dba Tesla Motors, Inc. and Does 1-10, inclusive*, Case no. 5:20-cv-2926, 28 aprile 2020, in *Westlaw*, WL 2105881.

industry publications, government regulatory research, actual data collected from similar incidents which caused damages, both to property and to persons, for which Tesla knew that its vehicles were defective and presented a substantial and unreasonable risk of harm to the general public [...] Despite such knowledge, Tesla, acting through its officers, directors, and managing agents, for the purpose of enhancing profits, knowingly and deliberately failed to remedy or otherwise take corrective action to fix the known defects in said vehicles, failed to warn the public, including decedent Yoshihiro Umeda, of the extreme risk of injury that could result and be occasioned by said defects». Oltre ad aver messo in commercio un prodotto insicuro, Tesla avrebbe inoltre omesso di informare correttamente gli acquirenti circa i rischi e i pericoli legati all'utilizzo del sistema di guida autonoma, promuovendo una “*overreliance*” del conducente nei confronti delle funzionalità del veicolo.

Anche in un altro procedimento civile, *Jing Wang and Wai-Leung Chan v. Tesla, Inc.* (2021)²³², i ricorrenti sostenevano che l'incidente per il quale si procedeva era stato causato da alcuni malfunzionamenti dell'*Autopilot* prodotto da Tesla. In particolare, mentre uno dei ricorrenti (Chan) era alla guida di una Tesla, con sistema di guida autonoma attivato, un veicolo aveva cambiato corsia, ponendosi di fronte alla macchina di Chan. La funzione *autopilot* – non riconoscendo il veicolo introdottosi nella corsia – non aveva frenato, né aveva avvertito Chan della collisione imminente. Con un solo secondo per reagire, il querelante aveva sterzato a sinistra, nel tentativo di evitare la collisione, e si era scontrato con due auto provenienti dalla corsia adiacente: nemmeno in questo caso, l'*autopilot* avrebbe riconosciuto la collisione imminente.

Wang (proprietario del veicolo) e Chang (conducente) ricorrono in giudizio contro Tesla, per ottenere il risarcimento dei danni subiti dal loro veicolo e dalle altre due macchine coinvolte nell'incidente (non sono invece riportati danni all'integrità fisica). In particolare, i ricorrenti allegano che l'*autopilot* non presenti le qualità descritte dal contratto (*breach of express and implied warranties*) e che la casa produttrice sia solita pubblicizzarlo in maniera ingannevole (*deceptive and misleading business practices and false advertising; negligent misrepresentation*) e fraudolenta (*common law fraud*),

²³² Il procedimento è ancora in corso. Del fascicolo è disponibile l'ordinanza che rigetta alcune questioni di inammissibilità presentate da Tesla, vd. United States District Court, E.D. New York, *JING WANG and Wai-Leung Chan v. TESLA, INC.*, Case 0-CV-3040 (NGG) (SJB), 16 luglio 2021, in Westlaw, WL 3023088.

violando altresì gli obblighi informativi sul prodotto (*failure to warn*), dal momento che agli acquirenti non vengono fornite informazioni corrette sul funzionamento e sui limiti del sistema di guida autonoma. Oltre al risarcimento del danno (c.d. *compensatory damages*), i ricorrenti chiedono altresì il pagamento di *punitive and exemplary damages*. Il procedimento è ancora in corso; ad oggi, la *United States District Court, E.D. New York* ha dichiarato inammissibile soltanto il motivo fondato sulla frode contrattuale²³³.

Quanto ai *procedimenti penali*, ad oggi abbiamo notizia di due casi, negli Stati Uniti, in cui gli incidenti provocati dagli *autonomous vehicles* hanno determinato l'avvio di indagini penali nei confronti dei conducenti, rei – in base all'impostazione accusatoria – di aver omesso colposamente di vigilare sul funzionamento del veicolo.

In un caso, risalente al marzo 2018, una *safety driver* di Uber, Rafael Stuart Vasquez, è stata accusata del reato di omicidio stradale (*vehicular manslaughter*) per non aver evitato la collisione tra il veicolo che stava testando e un pedone che stava attraversando la strada. La posizione di Uber è stata invece archiviata da parte dell'autorità giudiziaria. La richiesta di rinvio a giudizio dell'agosto 2020²³⁴ non spiega le circostanze dell'incidente, che però possono essere ricavate dai numerosi articoli che sono stati pubblicati sulla vicenda²³⁵, nonché da un report del *National Transportation*

²³³ Sempre a proposito di pubblicità ingannevole, segnaliamo che anche il *California Department of Motor Vehicles* ha proposto ricorso contro Tesla presso lo *State Office of Administrative Hearings*, allegando che la società ha diffuso dichiarazioni non veritiere o fuorvianti, e non basate sui fatti, nel pubblicizzare i veicoli come dotati, o potenzialmente dotati, di funzioni di sistema avanzato di assistenza alla guida; vd. R. Mitchell, *California DMV accuses Tesla of falsely advertising Autopilot and Full Self-Driving features*, in *Los Angeles Times*, 5 agosto 2022, disponibile a questo link: <https://www.latimes.com/business/story/2022-08-05/dmv-false-advertising-tesla>. Precedentemente, nel 2020, anche il Tribunale di Monaco, in Germania, ha stabilito che "Autopilot" e "Full Self Driving" sono espressioni fuorvianti e ha ordinato a Tesla di cessare di utilizzarle sul suo web tedesco, vd. M. Taylor, *German Court Bans Tesla 'Autopilot' Name For Misleading Customers*, in *Forbes*, 14 luglio 2020, disponibile a questo link: <https://www.forbes.com/sites/michaeltaylor/2020/07/14/german-court-bans-tesla-autopilot-name-for-misleading-customers/?sh=7757b3f5c7fb>.

²³⁴ Allister Adel – Maricopa County Attorney, *State of Arizona v. Rafael Stuart Vasquez*, Indict. 785 GJ 251, 27 agosto 2020, disponibile a questo link: <https://www.maricopacountyattorney.org/DocumentCenter/View/1724/Rafael-Vasquez-GJ-Indictment>.

²³⁵ In particolare, un recente articolo ricostruisce la vicenda, vd. L. Smiley, *'I'm the Operator': The Aftermath of a Self-Driving Tragedy*, in *Wired US*, 8 marzo 2022, disponibile a questo link: <https://www.wired.com/story/uber-self-driving-car-fatal-crash/>.

Safety Board (NTSB), l'agenzia federale incaricata di indagare sugli incidenti che coinvolgono trasporti ed infrastrutture²³⁶.

In base a quest'ultimo report, emerge che il sistema di guida autonoma aveva individuato il pedone 5,6 secondi prima dell'impatto. Nonostante ciò, tuttavia, il sistema non era stato in grado di classificarlo come "pedone" né era stato in grado di predire il suo movimento. Le specifiche del sistema, infatti, non prevedevano la frenata di emergenza per attenuare la collisione, che era invece affidata all'intervento dell'operatore. Da parte sua, la *safety driver* era distratta: il video della telecamera rivolta verso l'interno del veicolo mostrava che l'operatrice aveva distolto lo sguardo dalla strada per un lungo periodo di tempo mentre il veicolo si avvicinava al pedone, probabilmente per guardare il cellulare che aveva riposto, all'inizio del viaggio, nella parte inferiore della consolle centrale del veicolo. L'operatore aveva guardato la strada soltanto un secondo prima dell'impatto, iniziando a sterzare a sinistra 0,02 secondi prima di colpire il pedone, a una velocità di 39 miglia orarie.

Nel citato report redatto dall'NTSB si legge che il sistema di guida autonoma sviluppato da Uber è progettato per apportare piccole modifiche alla traiettoria dei veicoli, per evitare gli ostacoli, ma non è invece collaudato per effettuare automaticamente frenate di emergenza. Nel caso, dunque, in cui vengano individuati ostacoli da aggirare, le auto avviano una modalità di c.d. "*action suppression*" per un secondo, durante il quale il veicolo verifica la natura del pericolo e, a seconda della circostanza, calcola un percorso alternativo o permette all'operatore di prendere il controllo del veicolo. La finalità dell'*action suppression* è quella di evitare i falsi allarmi, che potrebbero indurre il veicolo a compiere manovre estreme non necessarie.

Una volta effettuata la *action suppression*, la risposta dell'ADS dipende dalla possibilità di evitare una collisione applicando la frenata massima (7 m/s, con tasso di decelerazione massimo di 5 m/s): nel caso in cui questa sia possibile, il sistema frena automaticamente; se, invece, la frenata massima consentita non sarebbe, in ogni caso, in grado di evitare una collisione, il sistema invia un allarme acustico all'operatore del veicolo, avviando contemporaneamente un rallentamento graduale del veicolo²³⁷. In tali

²³⁶ NTSB Accident report, *Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018*, pubblicato il 19.11.2019, disponibile a questo link: <https://www.nts.gov/investigations/accidentreports/reports/har1903.pdf>.

²³⁷ NTSB Accident report, cit., p. 14.

circostanze, l'ADS non è progettato per applicare la frenata massima solo per ridurre la gravità di una collisione: l'idea che sta alla base di questo meccanismo è che le frenate di emergenza (più rapide, dunque, rispetto ai 7 m/s della frenata massima che può compiere il sistema) sono più sicure se effettuate dal conducente – un punto di vista, in realtà, piuttosto contestabile, considerati i tempi di reazione di un guidatore che, fino ad un momento prima, si stava affidando ad un sistema di guida autonomo, seppur parzialmente²³⁸ (su questi aspetti vd. più diffusamente *infra*, cap. III, § 3.2.2.).

Questo è, dunque, ciò che sembra essere accaduto: il veicolo è entrato in fase di *action suppression* ma, calcolando che non avrebbe comunque potuto evitare la collisione, ha inviato un allarme acustico alla conducente, limitandosi ad effettuare un lieve rallentamento; a quel punto, tuttavia, per l'operatrice era troppo tardi per frenare bruscamente ed evitare l'incidente.

In base al report, a contribuire all'incidente sarebbero state anche l'inadeguatezza delle procedure di valutazione dei rischi per la sicurezza attivate da Uber e la mancanza di meccanismi efficaci per affrontare la “*complacency*” degli operatori nei confronti dell'automazione.

In un altro caso²³⁹, risalente al 19 dicembre 2019, alla guida dell'*autonomous vehicle* (una Tesla Model S) vi era invece un privato cittadino, Kevin George Aziz Riad, che non era intervenuto per fermare la sua automobile – con sistema di *autonomous driving* attivo – al semaforo rosso. La vettura si era scontrata, all'incrocio, con un altro autoveicolo, causando il decesso di due persone. Riad è stato rinviato a giudizio per omicidio stradale (*vehicular manslaughter*) commesso con colpa grave (*gross negligence*)²⁴⁰.

²³⁸ Tale meccanismo è stato infatti modificato nei successivi *software* di guida autonoma sviluppati da Uber, così da permettere *sempre* la frenata massima, anche qualora non dovesse bastare ad evitare una collisione, vd. NTSB Accident report, cit., p. 14.

²³⁹ Superior Court of California - Los Angeles County, Lorena OCHOA e al. v. Kevin George Aziz RIAD [...] Tesla Motors Inc.; Tesla Inc., No. 20STCV15375, 22 aprile 2020, in *Westlaw*, 2020 WL 1956437 (ricorso civile presentato dai familiari delle vittime; questi i motivi: *Complaint for Damages for Wrongful Death*; 1. *Negligence*; 2. *Negligent Hiring, Training, Retention & Supervision*; 3. *Strict Product Liability*; 4. *Negligence Product Liability*; 5. *Negligence (Post Sale)*).

²⁴⁰ La notizia è riportata da CBS Los Angeles, *Driver to stand trial in Tesla autopilot crash which killed two in Gardena*, 20 maggio 2022, disponibile a questo link: <https://www.cbsnews.com/losangeles/news/driver-to-stand-trial-in-tesla-autopilot-crash-which-killed-two-in-gardena/>.

Sui profili di imputazione, civile e penale, degli eventi lesivi che hanno coinvolto i veicoli a guida autonoma torneremo a più riprese nel corso della nostra trattazione.

7.1.1.2. Sistemi di guida autonoma degli aerei.

I sistemi di *autopilot* degli aerei sono tra le più antiche tecnologie di i.a. sperimentate con successo in ambito civile²⁴¹. Dopo quasi un secolo di sviluppo²⁴², oggi l'*autopilot* ha raggiunto la quasi completa autonomia, essendo in grado di decollare, navigare e atterrare – gestendo condizioni esterne non sempre prevedibili, come il vento e la presenza di altri velivoli – con *input* umano minimo (sebbene al pilota rimanga sempre affidata la gestione delle emergenze, i c.d. *edge-cases*²⁴³). Sono inoltre in via di sviluppo sistemi completamente autonomi di volo, che non necessiterebbero nemmeno di una presenza umana in cabina di pilotaggio²⁴⁴.

I sistemi di navigazione aerea autonoma, fin dall'inizio, furono uno strumento centrale per incrementare la sicurezza dei voli e ridurre l'incidenza di errori umani e, a partire dagli anni Cinquanta, divennero lo *standard* nella produzione di velivoli.

Proprio la longevità della tecnologia dell'*autopilot* nell'aviazione offre un chiaro esempio di quali problemi possono presentarsi quando il controllo di un veicolo (o di qualsiasi altro artefatto tecnologico) è affidato congiuntamente ad un essere umano e ad una macchina. Sebbene, infatti, l'automazione possa eliminare l'incidenza di alcuni errori

²⁴¹ Per una ricostruzione dei principali problemi che si pongono in ordine all'individuazione della persona responsabile, in caso di incidenti aerei vd. M.C. ELISH – T. HWANG, *Praise the machine! Punish the human! The Contradictory History of Accountability in Automated Aviation*, in *Comparative Studies in Intelligent Systems*, Working Paper #1 V2, 18 maggio 2015; M.C. ELISH, *Moral crumple zones: Cautionary Tales in Human-Robot Interaction*, in *Engaging Science, Technology, and Society* 5 (2019), p. 40 ss.; W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You: Understanding the Boeing 737 Max Products Liability Litigation*, 84 J. Air L. & Com. 379, 2019.

²⁴² M.C. ELISH – T. HWANG, *Praise the machine! Punish the human!*, cit., nota 5 e bibliografia ivi citata.

²⁴³ M.C. ELISH, *Moral crumple zones*, cit., p. 47.

²⁴⁴ Vd. S. Lekach, *Self-flying planes could transport passengers one day—but first, packages*, in *Mashable*, 24 luglio 2021, disponibile a questo link: <https://mashable.com/article/self-flying-airplanes>; T. Pallini, *I flew on a self-flying plane where pilots sat back as the aircraft taxied, took off, and landed on its own and I'm convinced it's the future of aviation*, in *Business Insider*, 8 maggio 2021, disponibile a questo link: <https://www.businessinsider.com/flying-on-self-flying-plane-convicted-its-future-of-aviation-2021-5?r=US&IR=T>.

umani – specie quelli che sono solitamente determinati da disattenzione e stanchezza – la co-gestione del volo può, d'altra parte, contribuire a crearne di nuovi.

Innanzitutto, c'è il rischio che l'automazione porti, sul lungo periodo, ad una *dipendenza* dei piloti nei confronti della tecnologia e, di conseguenza, ad un'erosione delle competenze e all'incapacità di gestire le situazioni di emergenza: è quanto è emerso, in maniera paradigmatica, nell'incidente del volo Air France 447, verificatosi il primo giugno 2009, quando i piloti – a seguito del malfunzionamento e della conseguente disattivazione automatica del sistema automatico anti-stallo – non riuscirono a risolvere una situazione di stallo, un problema relativamente semplice da gestire per un pilota aereo, determinando così il crollo del velivolo²⁴⁵.

Inoltre, negli ultimi anni, al rischio di una perdita di competenze dei piloti si è unita anche la difficoltà di comprendere il funzionamento dei complessi sistemi tecnologici alla base dell'*autopilot* e di interagire con essi (spesso anche a causa di una carente attività di *training* da parte delle aziende produttrici di aerei).

Una delle vicende più emblematiche, in questo senso, è quella che ha coinvolto i velivoli modello Boeing 737 MAX, protagonisti di diversi incidenti nel corso degli ultimi anni: nell'ottobre 2018, un Boeing 737 operato da Lion Air precipitò in mare poco dopo il decollo da Giacarta, causando la morte di tutte le persone a bordo; lo stesso accadde pochi mesi dopo, nel marzo 2019, per un volo operato da Ethiopian Airlines; infine, ancora recentemente, nel marzo 2022, un altro Boeing 737 – operato questa volta da China Eastern Airlines – si è schiantato poco dopo la partenza.

Sebbene la dinamica di quest'ultimo incidente sia ancora da accertare, dal *Deferred Prosecution Agreement* raggiunto tra gli organi inquirenti e Boeing²⁴⁶ emerge che i primi due, tra gli incidenti citati, sarebbero stati causati da una difettosa interazione

²⁴⁵ Per una ricostruzione dell'incidente vd. Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA), *Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris*, luglio 2012, disponibile a questo link: <https://bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf>; vd. anche l'opinione di un esposto aviatore W. Langewiesche, *The Human Factor*, in *Vanity Fair*, ottobre 2014, disponibile a questo link: <https://www.vanityfair.com/news/business/2014/10/air-france-flight-447-crash>.

²⁴⁶ United States District Court for the Northern District of Texas, *United States of America v. The Boeing Company, Deferred Prosecution Agreement* (Case 4:21-cr-00005-O), 7 gennaio 2021, disponibile a questo link: <https://www.justice.gov/criminal-vns/file/1482911/download>; vd. anche la *Press Release* rilasciata dal *Department of Justice*, disponibile a questo link: <https://www.justice.gov/opa/pr/boeing-charged-737-max-fraud-conspiracy-and-agrees-pay-over-25-billion>.

tra i piloti e il sistema di controllo di volo “intelligente” adottato dalla Boeing, il *Maneuvering Characteristics Augmentation System* (MCAS)²⁴⁷.

Quest’ultimo è un sistema “anti-stallo”, che interviene quando alcuni sensori – che indicano dati riguardanti l’inclinazione dell’aereo (in particolare, il c.d. angolo di attacco) – richiedono una stabilizzazione. In entrambi gli incidenti, i sensori avevano percepito erroneamente che l’aereo si trovava con la parte anteriore impennata – in situazione di stallo – e avevano segnalato tale criticità al MCAS, che di conseguenza continuava ad intervenire per abbassare la parte anteriore dell’aereo.

I piloti, da parte loro, tentavano di riportare il velivolo nella corretta situazione di equilibrio, ma – non sapendo che il sistema MCAS si era attivato e, probabilmente, addirittura ignorandone l’esistenza – non riuscivano a comprendere cosa stesse succedendo.

Così, un sistema che aveva l’obiettivo di mitigare i rischi derivanti da imprevedibili errori umani si è trasformato in un ulteriore ed inaspettato fattore di rischio²⁴⁸.

7.1.1.2. Sistemi d’arma autonomi letali.

Il settore bellico è sempre stato un contesto privilegiato per la sperimentazione e lo sviluppo delle tecnologie più avanzate, spesso sono nate proprio a scopi militari. Negli ultimi anni, si è assistito allo stanziamento di ingenti investimenti pubblici per lo sviluppo di sistemi d’arma autonomi letali (c.d. SALA) – o *Lethal autonomous weapon systems* (LAWS)²⁴⁹. Secondo la definizione più accreditata, i SALA sono quelle armi che, una

²⁴⁷ La medesima ricostruzione emerge anche in J. Fallows, *Here’s What Was on the Record About Problems with the 737 Max*, in *The Atlantic*, 13 marzo 2019, disponibile a questo link: <https://www.theatlantic.com/ideas/archive/2019/03/heres-what-was-on-the-record-about-problems-with-the-737-max/622213/>; per un’analisi del caso sotto la lente della responsabilità per danno da prodotto vd. inoltre W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit.

²⁴⁸ In argomento vd. più diffusamente W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit., *passim*.

²⁴⁹ In dottrina vd. N. Bhuta e al. (eds.), *Autonomous Weapons Systems*, cit.; J. Galliot – D. MacIntosh – J.D. Ohlin (eds.), *Lethal Autonomous Weapons. Re-Examining the Law and Ethics of Robotic Warfare*, Oxford University Press, 2021; P. SCHARRE, *Army of None: Autonomous weapons and the future of war*, W.W. Norton&Company, 2018; A. LELE, *Disruptive Technologies for the Militaries and Security*, Springer, 2019; P. ASARO, *How just could a robot war be?*, in P. Brey – A. Briggie – K. Waelbers (eds.),

volta attivate, sono in grado di identificare, selezionare ed ingaggiare un obiettivo senza ulteriore intervento umano²⁵⁰ (c.d. *robot killer*). A nostro parere, il fatto che questi sistemi possano essere programmati in modo tale che il controllore umano mantenga un “potere di veto” rispetto alla decisione algoritmica (c.d. sistemi “*human-on-the-loop*”²⁵¹), non ne esclude il carattere autonomo, poiché – in assenza di un intervento inibitore da parte del controllore – tali armi saranno effettivamente in grado di *decidere* quale *target* colpire. In base alla definizione anzidetta, dovrebbero invece essere esclusi dal novero dei sistemi d’arma autonomi:

(i) i sistemi automatizzati, manovrabili a distanza e telecomandati²⁵² (come i c.d. *loitering munition*, piccoli missili pilotati da remoto, che vengono fatti esplodere direttamente sull’obiettivo militare, ostacolando l’utilizzo di sistemi antirazzo²⁵³);

(ii) le armi e i sistemi d’arma specificamente progettati per difendere le proprie piattaforme, le proprie forze e la popolazione contro minacce dinamiche come missili, munizioni e aeromobili nemici²⁵⁴.

Current Issues in Computing and Philosophy, Ios Press, p. 50 ss.; E. FUZAYLOVA, *War torts, autonomous weapons systems, and liability: why a limited strict liability tort regime should be implemented*, in *Cardozo Law Review*, 40, n. 3, 2019; W.C. MARRA – S.K. MCNEIL, *Understanding “The Loop”: Regulating the Next Generation of War Machines*, in 36 *Harvard Journal of Law and Public Policy*, 2013, p. 1139; nella dottrina italiana, D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto: quale ruolo per il controllo umano?*, in *BioLaw Journal*, 2019, n. 1, p. 33 ss.; F. CHESINI, *“Terminator Scenario”? Intelligenza artificiale nel conflitto armato: “Lethal Autonomous Weapons Systems” e le risposte del diritto internazionale umanitario*, in *BioLaw Journal*, 2020, fasc. 3, p. 441 ss.

²⁵⁰ Così la Department of Defense Directive (DODD) 3000.09, *Autonomy in Weapon Systems*, 21 novembre 2012, mod. 2017, disponibile a questo link: <https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf>: «*weapon system[s] that, once activated, can select and engage targets without further intervention by a human operator*»; nello stesso senso, vd. definizione fornita da *International Committee of the Red Cross (ICRC) Position on Autonomous Weapon Systems*, 2021, disponibile a questo link: <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>.

²⁵¹ I sistemi “*human-on-the-loop*” si distinguono da quelli “*human-in-the-loop*”, che, per agire, necessitano di un’autorizzazione umana e da quelli “*human-out-of-the-loop*”, che possono colpire senza alcuna supervisione umana; vd. il report di Human Rights Watch, *Losing Humanity: The Case Against Killer Robots*, 2012, p. 2, disponibile a questo link: https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf.

²⁵² Risoluzione del Parlamento Europeo del 12 settembre 2018 sui sistemi d’arma autonomi (2018/2752(RSP)).

²⁵³ S. Valesini, *Nella guerra in Ucraina scendono in campo le armi autonome*, in *Wired Italia*, 24 marzo 2022, disponibile a questo link: <https://www.wired.it/article/ucraina-guerra-armi-autonome-droni-kamikaze-robot-killer-russia/>.

²⁵⁴ Così la Risoluzione del 12 settembre 2018 sui sistemi d’arma autonomi (2018/2752(RSP)); vd. ad esempio il sistema *Iron Dome* utilizzato da Israele per intercettare razzi a media velocità e proiettili di

Una volta attivata, un'arma autonoma è in grado di identificare gli obiettivi militari – attraverso sensori e tecniche di riconoscimento delle immagini basate sul *machine learning* – e di colpirli. La problematicità delle *autonomous weapons* consiste proprio nel fatto che – a differenza dei tradizionali droni, che possono controllare autonomamente soltanto alcuni aspetti della propria navigazione aerea – l'autonomia riguarda in questo caso la *selezione* e l'*ingaggio* degli obiettivi nelle operazioni belliche: funzioni delicatissime e “critiche”, poiché implicano la risoluzione di dilemmi etico-morali suscettibili di incidere su beni giuridici di assoluta rilevanza, quali il diritto alla vita e all'integrità fisica²⁵⁵.

L'utilizzo di armi autonome rischia oltretutto di porsi in contrasto con il diritto internazionale umanitario, che disciplina le regole che devono essere seguite nel corso dei conflitti armati e prevede norme specifiche per la selezione degli obiettivi militari (c.d. *Law of targeting*)²⁵⁶. D'altra parte, a favore dello sviluppo e dell'utilizzo in contesti bellici di sistemi d'arma autonomi vi è la convinzione, da parte di alcuni autori, che l'impiego dell'i.a. potrebbe ridurre il numero di vittime civili, stante la già accennata accuratezza predittiva dimostrata dai sistemi di *machine learning*²⁵⁷.

Tuttavia, non bisogna dimenticare che i sistemi di i.a. *non sono infallibili* e possono commettere errori nell'identificazione e nel riconoscimento di oggetti e persone: il malfunzionamento di un'arma nell'ambito di un'operazione militare potrebbe avere un effetto tragico, potendosi risolvere nella morte di decine, centinaia, migliaia di persone.

artiglieria con traiettoria balistica, vd. I. Orpaz, *How Does the Iron Dome Work?*, in *Haaretz*, 19 novembre 2012, disponibile a questo link: <https://www.haaretz.com/2012-11-19/ty-article/.premium/how-iron-dome-works/0000017f-e70b-dc7e-adff-f7af61f70000>.

²⁵⁵ D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto*, cit., p. 37.

²⁵⁶ C. GRUT, *The Challenge of Autonomous Lethal Robotics to International Humanitarian Law*, in *Journal of Conflict and Security Law*, vol. 18, iss. 1, 2013, p. 5 ss.; M.N. SCHMITT – J.S. THURNER, *Out of the Loop: Autonomous Weapons Systems and the Law of Armed Conflict*, in *Harvard National Security Journal*, 2013, n. 4, p. 231; in generale, sul *targeting* vd. M.N. SCHMITT – E. WIDMAR, *The Law of Targeting*, Oxford University Press, 2012; G.D. SOLIS, *The Law of Armed Conflicts*, Cambridge University Press, 2010, p. 519 ss.

²⁵⁷ Vd. il documento proposto dal governo statunitense nell'ambito del Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons, intitolato *Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapons*, 28 marzo 2018.

Infine, ancora una volta, emerge un problema di *attribuibilità* dell'eventuale fatto illecito scaturito dall'attività del sistema autonomo (c.d. *accountability gap*): chi dovrà risponderne? Il programmatore, il produttore, il comandante dell'operazione militare?

Gli appelli alla messa al bando dei sistemi d'arma autonomi letali si susseguono ormai da anni. Nel 2015, più di tremila scienziati esperti di intelligenza artificiale e robotica firmarono una lettera aperta invitando i governi a prevenire una corsa agli armamenti in relazione a tali armi e ad evitare gli effetti destabilizzanti di queste tecnologie²⁵⁸. Lo stesso fecero, nell'agosto 2017, 116 fondatori di importanti imprese internazionali del settore della robotica e dell'intelligenza artificiale, in una lettera aperta alle Nazioni Unite²⁵⁹. Infine, la *Campaign to Stop Killer Robots*²⁶⁰, negli ultimi anni, ha raccolto numerosissime adesioni, tra le quali quelle di moltissime ONG e oltre 30 Stati.

Ad oggi, tuttavia, non vige alcun divieto, a livello internazionale, di utilizzare sistemi d'arma autonomi letali. Nel 2016, alla quinta conferenza di revisione della Convenzione delle Nazioni Unite su certe armi convenzionali (*Convention on Certain Conventional Weapons*, c.d. CCW)²⁶¹, è stato istituito un Gruppo di Esperti Governativi, riunitosi per la prima volta nel 2017, con il mandato di valutare l'opportunità di regolare (e, eventualmente, bandire) le armi autonome. Nonostante diversi Stati Parte alla Convenzione abbiano proposto l'adozione di normative vincolanti sul punto²⁶², ad oggi la Riunione delle Alte Parti contraenti della CCW è riuscita ad adottare soltanto dei principi guida non vincolanti²⁶³, che possono essere così riassunti:

- il diritto internazionale umanitario continua ad applicarsi pienamente a tutti i sistemi di armi, compresi i SALA;

²⁵⁸ FUTURE OF LIFE INSTITUTE, *Autonomous Weapons: an Open Letter from Ai & Robotics Researchers*, 9 febbraio 2016, disponibile a questo link: <https://futureoflife.org/open-letter-autonomous-weapons>.

²⁵⁹ FUTURE OF LIFE INSTITUTE, *Killer robots: World's top AI and robotics companies urge United Nations to ban lethal autonomous weapons*, disponibile a questo link: <https://futureoflife.org/2017/08/20/killer-robots-worlds-top-ai-robotics-companies-urge-united-nations-ban-lethal-autonomous-weapons/>.

²⁶⁰ *Campaign to Stop Killer Robots*, vd. link: <https://www.stopkillerrobots.org/>.

²⁶¹ Convenzione delle Nazioni Unite su certe armi convenzionali, conclusa a Ginevra il 10 ottobre 1980 ed entrata in vigore il 2 dicembre 1983.

²⁶² Le proposte pervenute sono disponibili a questo link: https://meetings.unoda.org/section/ccw-gge-2022_documents_18542_proposals_19869/.

²⁶³ CCW, *Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects – Final Report*, 13 dicembre 2019, Annex III, disponibile a questo link: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/343/64/PDF/G1934364.pdf?OpenElement>.

- l'uomo deve mantenere il controllo sull'uso dei SALA ed è responsabile per il loro utilizzo;
- deve essere sempre garantita la sicurezza fisica e informatica dei SALA, al fine di evitare che siano utilizzati impropriamente da parte di gruppi terroristici;
- nel corso del ciclo di progettazione e sviluppo dei SALA devono essere effettuati periodicamente test per la valutazione del rischio;
- i SALA non devono avere forme antropomorfizzate;
- il dibattito circa l'opportunità di sviluppare SALA non deve ostacolare l'elaborazione di tecnologie autonome intelligenti a fini pacifici;
- la CCW si impegna ad offrire supporto per affrontare lo sviluppo dei SALA nel contesto degli obiettivi e delle finalità della Convenzione, al fine di trovare un equilibrio tra necessità militari e considerazioni umanitarie.

Anche l'Unione Europea, per ora, non ha adottato alcuna normativa vincolante in questo settore. Il Parlamento Europeo ha approvato due risoluzioni, una nel 2018²⁶⁴ e una nel 2021²⁶⁵, in cui si esortava l'Alto rappresentante dell'Unione per gli affari esteri e la politica di sicurezza, gli Stati membri e il Consiglio ad adoperarsi per l'avvio di negoziati internazionali su uno strumento giuridicamente vincolante che disciplini, eventualmente vietandoli, i sistemi d'arma autonomi letali.

Quanto agli Stati Uniti, i sistemi d'arma autonomi sono regolati da una direttiva redatta nel 2012 dal Dipartimento di Difesa²⁶⁶, che non ne vieta lo sviluppo o l'utilizzo, ma ne subordina l'impiego ad una serie di requisiti. Ecco i principali:

- i SALA devono rispettare il diritto bellico e le convenzioni internazionali rilevanti;

²⁶⁴ Risoluzione del 12 settembre 2018 sui sistemi d'arma autonomi (2018/2752(RSP)).

²⁶⁵ Risoluzione del Parlamento europeo del 20 gennaio 2021 sull'intelligenza artificiale: questioni relative all'interpretazione e applicazione del diritto internazionale nella misura in cui l'UE è interessata relativamente agli impieghi civili e militari e all'autorità dello Stato al di fuori dell'ambito della giustizia penale (2020/2013(INI))

²⁶⁶ DODD 3000.09, *Autonomy in Weapon Systems*, 21 novembre 2012, modificata nel 2017, disponibile a questo link: <https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf>. La Direttiva sarà probabilmente modificata tra il 2022 e il 2023, proprio per adattarsi alle più recenti evoluzioni tecnologiche nell'ambito degli AI-enabled autonomous systems, vd. G.C. Allen, *DOD Is Updating Its Decade-Old Autonomous Weapons Policy, but Confusion Remains Widespread*, in *Center for Strategic & International Studies*, 6 giugno 2022, disponibile a questo link: <https://www.csis.org/analysis/dod-updating-its-decade-old-autonomous-weapons-policy-confusion-remains-widespread>.

- i SALA devono essere sviluppati in modo tale da permettere agli operatori di esercitare livelli appropriati di “giudizio umano” sull'uso della forza letale – ove il requisito del “giudizio umano” non richiede un "controllo" manuale del sistema d'arma, ma richiede piuttosto un certo coinvolgimento umano nelle decisioni su come, quando, dove e perché l'arma debba essere impiegata²⁶⁷;

- i SALA devono essere sottoposti a test di funzionamento in ambienti operativi realistici e contro avversari adattabili, così da garantire che le “richieste” dell’operatore siano soddisfatte nei tempi corretti e che i sistemi siano sufficientemente robusti da ridurre al minimo i guasti;

- ogni cambiamento di sistema operativo impone un nuovo test di valutazione dei SALA.

Nonostante non sia sempre agevole individuare con precisione quali siano le tecnologie effettivamente impiegate in ambito militare, diverse fonti affermano che si sarebbero già verificati alcuni casi di utilizzo di sistemi d’arma autonomi in contesti bellici. In un report redatto dalla Commissione di Esperti sulla Libia nominata dal Consiglio di Sicurezza dell’Onu, ad esempio, si legge che, nel marzo 2020, un drone militare turco avrebbe individuato e colpito automaticamente un obiettivo miliare libico, in assenza di alcun *input* o supervisione umana²⁶⁸. Ancora, secondo alcune testate giornalistiche, nella guerra in corso in Ucraina, entrambi gli schieramenti starebbero facendo uso di armi autonome dotate di sistemi di i.a.²⁶⁹.

²⁶⁷ Così il report redatto dal *Congressional Research Service* intitolato *Emerging Military Technologies: Background and Issues for Congress* (novembre 2020), p. 8.

²⁶⁸ *Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council* (UN S/2021/229), 8 marzo 2021, § 63 ss., disponibile a questo link: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N21/037/72/PDF/N2103772.pdf?OpenElement>.

²⁶⁹ R.F. Trager – L.M. Luca, *Killer Robots Are Here—and We Need to Regulate Them*, in *Foreign Foreign Policy*, 11 maggio 2022, disponibile a questo link: <https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/>. W. Knight, *Russia's Killer Drone in Ukraine Raises Fears About AI in Warfare*, in *Wired US*, 17 marzo 2022, disponibile a questo link: <https://www.wired.com/story/ai-drones-russia-ukraine/>. Vd. però G.C. Allen, *Russia Probably Has Not Used AI-Enabled Weapons in Ukraine, but That Could Change*, in *Center for Strategic & International Studies*, 26 maggio 2022, disponibile a questo link: <https://www.csis.org/analysis/russia-probably-has-not-used-ai-enabled-weapons-ukraine-could-change>, che contesta il fatto che vi siano prove dell'utilizzo effettivo di *autonomous weapons*, pur non escludendo che la Russia possa averne la disponibilità.

7.1.1.4. *Healthcare robots.*

Negli ultimi anni, nel settore chirurgico, si è assistito ad un forte incremento dell'utilizzo di dispositivi robotici, che sono in grado di compiere movimenti più accurati e precisi di quelli umani e di eseguire con successo interventi anche molto delicati, consentendo oltretutto al medico di operare da remoto²⁷⁰. I robot sono diffusi specialmente nel campo della ginecologia e dell'urologia: le procedure standard dell'oncologia urologica (come l'asportazione totale della prostata), ad esempio, vengono sempre più spesso eseguite con l'aiuto di robot chirurgici.

Anche per il settore chirurgico – così come per gli *autonomous vehicles* – alcuni ricercatori hanno cercato di individuare una classificazione basata sul grado di autonomia dei sistemi robotici²⁷¹.

Si va dai robot che si limitano ad eseguire i comandi dell'operatore umano, costituendone una mera estensione robotica (Livello 0), ai robot che sono in grado di condizionare o correggere l'azione umana (L1), che eseguono compiti designati dall'operatore umano sotto la sua supervisione (L2) o che generano strategie di esecuzione dei compiti sotto la supervisione umana (L3). Infine, vi sono i due livelli caratterizzati da maggiore sofisticazione, che comprendono quei sistemi robotici in grado di svolgere un'intera procedura chirurgica con o senza la supervisione umana (rispettivamente L4 e L5): si tratta, tuttavia, di livelli che delineano prospettive future,

²⁷⁰ Sulla questione della responsabilità nella robotica chirurgica vd. E. FOSCH-VILLARONGA, *Robots, Healthcare, and the Law. Regulating Automation in Personal Care*, Routledge, 2020; F. GRIFFIN, *Artificial Intelligence and Liability in Health Care*, in 31 *Health Matrix: Journal of Law-Medicine*, 2021, p. 65 ss.; M. ROE, *Who's Driving That Car?*, cit., p. 340 ss.; G. GUERRA, *Evolving artificial intelligence and robotics in medicine, evolving European law. Comparative remarks based on the surgery litigation*, in *Maastricht Journal of European and Comparative Law*, 2021, Vol. 28(6), p. 805 ss. Nella dottrina italiana, U. RUFFOLO, *L'Intelligenza artificiale in sanità: dispositivi medici, responsabilità e "potenziamento"*, in *Giur. it.*, febr. 2021; M. BASSINI – L. LIGUORI – O. POLLICINO, *Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. Pizzetti (a cura di), *Intelligenza artificiale*, cit., p. 349 ss.; C. IAGNEMMA, *I 'robot medici': profili problematici in tema di alleanza terapeutica e di responsabilità penale*, in *Corti supreme e salute*, 2020, n. 2, p. 442 ss.; in generale, sul quadro europeo in materia di device medico-sanitari vd. A. PISANI TEDESCO, *Il nuovo quadro normativo europeo dei dispositivi medici*, in *Dir. comm. internaz.*, 2022, fasc. 3, p. 675 ss.

²⁷¹ Vd. G.Z. YANG, *Medical robotics-regulatory, ethical and legal considerations for increasing levels of autonomy*, in *Science Robotics*, 2017, 1 ss.; per un commento più approfondito alla classificazione vd. D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto*, cit., p. 47 ss.

poiché, ad oggi, i modelli di i.a. utilizzati in ambito sanitario necessitano tutti di un controllo completo e costante da parte del chirurgo.

I sistemi attuali, infatti, non compiono alcun movimento o decisione indipendente, ma sono “telecomandati” dal chirurgo; il robot traduce in movimenti precisi e in tempo reale i comandi forniti dal medico attraverso la manipolazione di un *controller*, eliminando il fisiologico tremore dei movimenti della mano²⁷². Uno dei prodotti più sviluppati è il *robot* Da Vinci²⁷³, di livello 0, ma diversi robot chirurgici impiegati in sala operatoria hanno già raggiunto un grado di autonomia di tipo L1: si tratta di quei sistemi dotati di *software* che assistono il chirurgo nel movimento del *controller*, eventualmente correggendolo ed impedendogli di effettuare determinate manovre che potrebbero risultare pericolose per il paziente²⁷⁴. Ovviamente, al fine di garantire sempre un controllo umano, questi vincoli attivi (cd. *Virtual Fixtures*) possono essere sempre disattivati dal medico²⁷⁵. Nelle sale operatorie sono altresì già utilizzati dispositivi robotici di livello L2, come, ad esempio, il sistema ROBODOC, in grado di eseguire piani preoperatori relativi al taglio di ossa, sotto supervisione del medico – che, anche in questo caso, può sempre disattivare il sistema²⁷⁶. Infine, sono già in uso alcune limitate forme di sistemi di livello L3, che generano strategie di esecuzione dei compiti sotto la supervisione umana, lasciando sempre al medico la scelta su quale tra le opzioni sia la più opportuna: il sistema STAR, ad esempio, raggiunge questo livello di autonomia per quanto riguarda la generazione di strategie di anastomosi²⁷⁷.

Per i sistemi di livello L0 e L1, la macchina costituisce uno *strumento* nelle mani del medico; non può, dunque, essere fonte di *comportamenti emergenti*, imprevedibili da parte dell'utilizzatore e del produttore. Nel caso in cui, nel corso di un'operazione, si

²⁷² Vd. la pagina di Ucla Health, *What Is Robotic Surgery?*, disponibile a questo link: <https://www.uclahealth.org/robotic-surgery/what-is-robotic-surgery>, in cui il *controller* viene paragonato ad un *joystick* per *videogames*.

²⁷³ Sul suo funzionamento vd. più diffusamente G. GUERRA, *Evolving artificial intelligence and robotics in medicine*, cit., *passim*; per una panoramica sui *device* medici basati su sistemi di i.a. attualmente in commercio vd. C. IAGNEMMA, *I 'robot medici': profili problematici in tema di alleanza terapeutica e di responsabilità penale*, cit., p. 442 ss., cui si rinvia anche per l'ampia bibliografia scientifica citata.

²⁷⁴ Così D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto*, cit., p. 48.

²⁷⁵ *Ibidem*.

²⁷⁶ *Ibidem*.

²⁷⁷ Così D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto*, cit., p. 49.

verifichi un danno all'integrità fisica del paziente, a risponderne potranno allora essere, alternativamente o congiuntamente, il medico o il produttore. Quanto al medico, si applicherà il regime di favore previsto dall'art. 590-sexies c.p. per la responsabilità medica colposa; lo *standard* di diligenza relativo all'utilizzo e al controllo di tali dispositivi potrà inoltre essere oggetto di "linee-guida" o di "buone pratiche clinico-assistenziali", ai sensi del medesimo articolo²⁷⁸.

Nel caso in cui, invece, il danno si verifichi a causa di un difetto del *device*, sarà il produttore che potrà essere chiamato a risponderne: rinviando ai successivi capitoli per una disamina dei profili problematici nella ricostruzione della responsabilità civile (vd. *infra*, cap. II, § 3. ss.) e penale (vd. *infra*, cap. III, § 4. ss.) per danno da prodotto. Ci limitiamo qui a rilevare che, negli Stati Uniti, la giurisprudenza civile si è già occupata in diverse occasioni della responsabilità del produttore per danni causati dal malfunzionamento di *robot* medici²⁷⁹. I fatti oggetto del contendere sono, sostanzialmente, sempre i medesimi e coinvolgono soprattutto il già citato robot Da Vinci: in diversi casi, il dispositivo si è bloccato nel corso di un intervento, costringendo il chirurgo a continuare ad operare con le tecniche tradizionali; a seguito dell'operazione, i pazienti hanno manifestato problemi di salute, quali ustioni, forti dolori, disfunzioni. Nella maggior parte dei casi²⁸⁰, i ricorsi presentati avverso la struttura ospedaliera e la società produttrice del *device* (*Intuitive Surgical Inc.*) sono stati rigettati, poiché le corti hanno ritenuto che non vi fosse alcuna prova di un nesso causale tra i danni subiti dal paziente e il presunto malfunzionamento del *robot*. Soltanto in un caso, *Taylor v. Intuitive*

²⁷⁸ C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., p. 1757.

²⁷⁹ Le decisioni che riportiamo, in realtà, non esauriscono la casistica in materia, poiché la maggior parte dei casi si è conclusa con accordi extragiudiziali. Per un'analisi più approfondita della casistica rinviando a G. GUERRA, *Evolving artificial intelligence and robotics in medicine*, cit., p. 815 ss.

²⁸⁰ Pennsylvania E.D. Court, *Mracek v. Bryn Mawr hospital et al.* Case: 2:08-cv-00296, 11 marzo 2009, disponibile a questo link: <https://www.plainsite.org/dockets/jv8xcvvo/pennsylvania-eastern-district-court/mracek-v-bryn-mawr-hospital-et-al/>, confermata in appello con U.S. Court of Appeals for the Third Circuit, *Mracek v. Bryn Mawr Hosp.*, 363 Fed. App'x 925, 28 gennaio 2010, disponibile a questo link: <https://cite.case.law/f-appx/363/925/>; Illinois N.D. Court, *O'Brien v. Surgical, Inc.*, Case: 1:10-cv-03005, 25 luglio 2011, disponibile a questo link: https://www.govinfo.gov/content/pkg/USCOURTS-ilnd-1_10-cv-03005/pdf/USCOURTS-ilnd-1_10-cv-03005-1.pdf. In Louisiana E.D. Court, *Silvestrini v. Intuitive Surgical, Inc. e al.*, case: 2:11-cv-02704, 6 febbraio 2012, disponibile a questo link: https://www.govinfo.gov/content/pkg/USCOURTS-laed-2_11-cv-02704/pdf/USCOURTS-laed-2_11-cv-02704-0.pdf, il ricorso è stato invece dichiarato inammissibile per difetto di giurisdizione.

*Surgical Inc*²⁸¹, la Corte Suprema dello Stato di Washington – ribaltando le precedenti sentenze di merito – ha stabilito la responsabilità del produttore per il danno arrecato al paziente, sostenendo che la società non aveva adempiuto all’obbligo di avvertire l’acquirente (in questo caso, l’ospedale) dei pericoli inerenti all’utilizzo del prodotto.

Se la casistica relativa al *device* Da Vinci riproduce sostanzialmente i profili problematici già emersi in relazione al tradizionale danno da prodotto, con riferimento ai sistemi di livello L2 e L3 potrebbero invece sorgere questioni *nuove*, legate alla capacità di svolgere autonomamente, seppur sotto la supervisione umana, determinate attività. Va comunque sottolineato che, in relazione all’utilizzo di *device* medici, il ruolo di controllo attribuito al medico appare meno problematico di quello che viene assegnato all’utilizzatore di un *autonomous vehicle*: il chirurgo che supervisiona l’attività del sistema di i.a., infatti, ha conoscenze specialistiche che dovrebbero consentirgli di comprendere se e quando il *robot* presenta malfunzionamenti e di intervenire di conseguenza, disattivando il dispositivo e continuando ad operare in maniera tradizionale. I veicoli a guida autonoma, invece, sono sviluppati per essere diffusi quanto le normali automobili. Il conducente, d’altra parte, non sempre è posto nelle condizioni di poter interpretare gli *output* del veicolo in maniera tempestiva: si pensi, ad esempio, ai veicoli di livello 2, in cui l’utilizzatore deve essere pronto a riprendere il controllo della guida in ogni momento, senza che vi sia nemmeno un avvertimento sonoro da parte del sistema di guida autonomo. Il conducente di un *autonomous vehicle* potrebbe rendersi conto di un malfunzionamento quando ormai è troppo tardi, stante la tempestività richiesta per la gestione di un veicolo su strada²⁸².

Va inoltre considerato che gli *autonomous vehicles* sono chiamati ad operare in un ambiente largamente imprevedibile, qual è quello del traffico, in cui vi è un’interazione costante con gli altri utenti della strada (pedoni, ciclisti, altri automobilisti, eventualmente altri veicoli a guida autonoma) e vi sono innumerevoli segnali da

²⁸¹ Supreme Court of Washington, *Taylor v. Intuitive Surgical, Inc.*, Case: 92210-1, 9 febbraio 2017, disponibile a questo link: <https://case-law.vlex.com/vid/Taylor-v-Intuitive-Surgical-Inc-020917-WASC-92210-1-665020489>.

²⁸² Vd. M. ROE, *Who’s Driving That Car?*, cit., p. 339. Si legga, ad esempio, la testimonianza di una conducente di un *autonomous vehicle* coinvolto in un incidente, che alla polizia intervenuta sul posto ha riferito: «When I popped up and I looked and saw a black truck – it happened so fast», in N.E. Boudette, *‘It Happened So Fast’: Inside a Fatal Tesla Autopilot Accident*, in *The New York Times*, 17 agosto 2021, disponibile al seguente link: <https://www.nytimes.com/2021/08/17/business/tesla-autopilot-accident.html>.

interpretare (semafori, sirene, strisce pedonali, etc.). In confronto, gli *healthcare robots* si muovono in un contesto relativamente prevedibile, con un compito molto specifico e senza alcuna interazione con persone o altre macchine.

7.1.2. Non-embodied systems.

7.1.2.1. Bots.

I *bots* sono programmi *software* basati su sistemi di i.a. e sono normalmente definiti per contrapposizione con i *robots*, caratterizzati dall'incorporazione del *software* in un *hardware* (il già citato *embodiment*)²⁸³. I *bots* sono oggi largamente utilizzati, in ambito societario, per il servizio clienti, per rispondere *online* ai quesiti più semplici posti dagli utenti (c.d. *chatbots*). Ma è nel campo dei *social networks* che, a partire dal 2015, si è assistito alla creazione su larga scala di *social bots*, ossia di *account* sui social network in grado di produrre contenuti in maniera autonoma e di emulare le interazioni umane. La facilità con cui è possibile creare i *bots* – attraverso servizi disponibili gratuitamente sui motori di ricerca o utilizzando comuni *software* di programmazione (es. Python) – ha determinato un'esplosione nell'utilizzo di tali sistemi.

Nella loro versione più rudimentale, i *social bots* sono in grado di ripubblicare automaticamente i post che contengono determinate parole o *hashtag*; di rispondere a post che soddisfano un certo criterio; di “seguire” (mettere il *follow*) a qualsiasi utente che abbia seguito il *bot* o che abbia pubblicato una specifica parola chiave; di effettuare ricerche sui motori di ricerca di articoli/notizie in base a criteri specifici e di eseguire una ripubblicazione automatica²⁸⁴.

I *social bots* sono saliti agli onori delle cronache per l'utilizzo che ne è stato fatto, a partire dalle elezioni statunitensi del 2016, a fini di propaganda politica: i *bots* sono in grado di amplificare a dismisura la diffusione di notizie false o tendenziose, ripubblicando costantemente post che contengono specifiche parole chiave e determinando, in questo

²⁸³ G. SCOPINO, *Algo Bots and the Law: Technology, Automation, and the Regulation of Futures and Other Derivatives*, Cambridge University Press, 2020, p. 37.

²⁸⁴ Anche A. BESSI – E. FERRARA, *Social bots distort the 2016 U.S. Presidential Election Online Discussion*, in *First Monday*, vol. 21, n. 11, 2016.

modo, le tematiche all'ordine del giorno dibattito pubblico. L'utilizzo di *social bots* a fini di destabilizzazione e polarizzazione del dibattito pubblico in occasione di elezioni politiche è ormai vicenda nota ed è oggetto di vastissima letteratura²⁸⁵. Il primo utilizzo massiccio di *social bots* finalizzato ad influenzare il voto nel corso di una competizione elettorale è avvenuto, come abbiamo accennato, nel corso delle elezioni statunitensi del 2016²⁸⁶ e si è similmente registrato, in seguito, nel dibattito *social* riguardante le manifestazioni dei *Gilet Jaunes* in Francia (2018-2019), le elezioni presidenziali francesi del 2017, le elezioni spagnole del 2019, le elezioni nel Regno Unito del 2019 e le elezioni europee del 2019²⁸⁷.

Sebbene, ovviamente, la diffusione di notizie false o tendenziose – specialmente se a fini di propaganda politica – sia sempre esistita, il fenomeno in esame presenta tratti di *pervasività* e *ampiezza* del tutto nuovi, che gli hanno permesso, secondo alcuni commentatori, di incidere sulle scelte degli elettori e, di conseguenza, sugli equilibri politici: alcuni ricercatori in *data-science*, ad esempio, hanno calcolato che, nel corso delle elezioni statunitensi del 2016, circa il 15% degli account che “twittava” su argomenti di politica interna era costituito da *bots*, responsabili del 19% della totale

²⁸⁵ P.N. HOWARD – S.C. WOOLEY, *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, Oxford University Press, 2018; J.D. OHLIN, *Election Interference. International Law and the Future of Democracy*, Cambridge University Press, 2020; ID., *Did Russian Cyber Interference in the 2016 Election Violate International Law?*, in 95 *Texas Law Review*, 2017, p. 1579; vd. anche i report pubblicati dal *Programme on Democracy & Technology* (Oxford Internet Institute), disponibili a questo link: <https://navigator.oii.ox.ac.uk/what-is-comprop/>; per un approccio penalistico vd. la monografia di T. GUERINI, *Fake news e diritto penale*, Giappichelli, 2020, e la sezione monografica di *Dir. pen. cont. – Riv. trim.*, 2021, fasc. 4, dedicata al tema «La disinformazione tra politica e diritto. Dimensione istituzionale, strategie preventive e dinamiche punitive»; sulla possibile qualificazione – dal punto di vista del diritto internazionale – delle presunte influenze *cyber* straniere sui processi democratici vd. M.N. SCHMITT, *Foreign Cyber Interference in Elections*, 97 *Int'l L. Stud.*, 2021, p. 739; S. WHEATLEY, *Foreign Interference in Elections under the Non-intervention Principle: We Need to Talk about “Coercion”*, 31 *Duke Journal of Comparative & International Law*, 2020, p. 161.

²⁸⁶ Fondamentale sul tema è il *Report on the Investigation into Russian Interference in the 2016 Presidential Election* (c.d. Mueller report), marzo 2019, disponibile a questo link: (ancora in parte secretato): <https://www.justice.gov/archives/sco/file/1373816/download>; vd. anche A. BESSI – E. FERRARA, *Social bots distort the 2016 U.S. Presidential Election Online Discussion*, cit.

²⁸⁷ Vd. E. BRATTBERG – T. MAURER, *Russian Election Interference. Europe's Counter to Fake News and Cyber Attacks*, Carnegie Endowment for International Peace, Maggio 2018; sulla Brexit, in particolare, vd. M.T. BASTOS – D. MERCEA, *The Brexit Botnet and User-Generated Hyperpartisan News*, in *Social Science Computer Review*, 2019, Vol. 37(1), p. 38 ss.

conversazione su Twitter²⁸⁸ – percentuali che dimostrano la “potenza di fuoco” che i *bots* possono avere nell’ambito del dibattito pubblico.

Negli ultimi anni, gli studi di settore hanno registrato un utilizzo più contenuto dei *social bots* di propaganda elettorale. Da un lato, tale diminuzione sarebbe da ricollegare ad un’accentuata capacità di detezione e sospensione degli *account* automatici da parte dei *social networks*, incentivata anche dai numerosi richiami da parte delle istituzioni²⁸⁹; dall’altra, i *social bots* potrebbero essere più difficili da riconoscere, dato l’incremento nelle capacità di simulazione del linguaggio umano, attraverso tecniche di *natural language processing*²⁹⁰.

Se per i *social bots* meno sofisticati – che si limitano ad effettuare una serie di azioni, sulla base delle istruzioni fornite in sede di programmazione – non si pongono grossi dubbi di *attribuibilità* dell’eventuale fatto illecito (ad es. diffamazione, istigazione all’odio razziale etc.) al programmatore e/o utilizzatore, sono i *bots* che utilizzano sistemi di *machine learning* che possono creare maggiori problemi.

Pensiamo, ad esempio, al *social bot* sviluppato da Microsoft nel marzo 2016 (@TayTweets), che era stato progettato per imitare i modelli di linguaggio di una ragazza americana di 19 anni e per imparare ad esprimersi grazie all’interazione con gli utenti umani su Twitter²⁹¹. Dopo neanche due giorni dall’attivazione, il *bot* – sollecitato da interazioni aggressive e provocatorie provenienti da alcuni utenti (c.d. *troll*) – aveva

²⁸⁸ A. BESSI – E. FERRARA, *Social bots distort the 2016 U.S. Presidential Election Online Discussion*, cit.

²⁸⁹ In particolare, quanto alle iniziative promosse in ambito europeo, la Commissione europea ha incardinato la strategia di contrasto alla disinformazione *online* sul sistema di autodisciplina previsto dal *Code of Practice on Disinformation*, un codice di buona condotta, adottato a settembre 2018 e rivisto nel 2022, al quale le piattaforme digitali aderiscono spontaneamente e al quale, ad oggi, hanno aderito i maggiori *player* della rete. Tra le pratiche di disinformazione *online* che gli aderenti si impegnano a contrastare vi è anche «*The creation and use of fake accounts, account takeovers and bot-driven amplification*», vd. COM, *The Strengthened Code of Practice on Disinformation*, 16 giugno 2022, *Commitment 14*. Dal momento che non esistono, ad oggi, criteri vincolanti per qualificare un contenuto come “disinformazione”, non sembra che un presunto contenuto disinformativo possa rientrare tra quei “contenuti illegali” la cui conoscenza dà luogo ad un obbligo, per il prestatore di servizio *online*, di agire «immediatamente per rimuovere i contenuti illegali o per disabilitare l’accesso agli stessi» (Proposta di Regolamento del Parlamento europeo e del Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali) e che modifica la direttiva 2000/31/CE, art. 5, co. 1, lett. b).

²⁹⁰ Vd. intervista al *data scientist* Emilio Ferrara (G. Guglielmi, *The next-generation bots interfering with the US election*, 28 ottobre 2020), disponibile a questo link: <https://www.nature.com/articles/d41586-020-03034-5>.

²⁹¹ R. Price, *Microsoft is deleting its AI chatbot's incredibly racist tweets*, in *Business Insider*, 24 marzo 2016, disponibile a questo link: <https://www.businessinsider.in/Microsoft-is-deleting-its-AI-chatbots-incredibly-racist-tweets/articleshow/51539858.cms>.

iniziato a pubblicare tweet razzisti, discriminatori ed insultanti (eccone alcuni, a mero titolo di esempio, «*bush did 9/11 and Hitler would have done a better job than the monkey we have now. donald trump is the only hope we've got*»; «*I fucking hate negros, I wish we could put them all in a concentration camp*»), tanto che Microsoft era stata costretta a sospendere l'account, eliminando alcuni dei contenuti più ingiuriosi e diffamatori.

Pensiamo, ancora, al *social bot* (@jeff_ebooks) progettato da uno sviluppatore di servizi *online* olandese, che generava contenuti basandosi su un “*markov chain generator*” – ossia un *software* che consente di elaborare frasi di senso compiuto a partire da un testo dato²⁹². In questo caso, il *bot* utilizzava, come testo di partenza, frammenti di tweet pubblicati dallo stesso creatore dell'account automatico: a partire da singole parole utilizzate dall'account, il *bot* elaborava nuove frasi di senso compiuto. La polizia olandese aveva aperto un fascicolo – poi archiviato – nei confronti dello sviluppatore, dopo che il *bot* aveva twittato di trovarsi ad un evento di moda ad Amsterdam e di avere intenzione di uccidere delle persone.

In entrambi questi casi, i sistemi di i.a. hanno agito in modi *non predeterminati* e *non prevedibili* a priori da parte dei programmatori²⁹³: si viene dunque a rompere il *rapporto diretto di autorità* tra programmatore e comportamento dei *bot*.

Simili questioni si sono poste con riferimento a *bot* di acquisto automatico *online*. Si pensi, ad esempio, ai *bot* utilizzati per il bagarinaggio *online*, che acquistano enormi quantità di biglietti non appena questi sono resi disponibili²⁹⁴. Si pensi, ancora, al noto caso «*Random Darknet Shopper*», un progetto avviato nel 2015 dal collettivo artistico *Mediengruppe Bitnik*, che aveva creato un *bot* incaricato di effettuare un acquisto a settimana sul *dark web*, senza fornire indicazioni più precise su quale tipologia di prodotti comprare. Ogni nuovo prodotto sarebbe stato recapitato direttamente all'indirizzo della galleria svizzera che ospitava l'opera performativa ed esposto insieme a tutti gli altri

²⁹² A. Hern, *Randomly generated tweet by bot prompts investigation by Dutch police*, in *The Guardian* (online), 12 febbraio 2015, disponibile a questo link: <https://www.theguardian.com/technology/2015/feb/12/randomly-generated-tweet-by-bot-investigation-dutch-police>.

²⁹³ Secondo alcuni studiosi, tuttavia, quantomeno il comportamento di Tay avrebbe potuto essere previsto, stante la nota presenza di *troll* su internet. In particolare, Microsoft avrebbe potuto filtrare i dati di *input*, escludendo, ad esempio, i tweet che contenevano linguaggio offensivo; vd. J. SINGH E AL., *Responsibility & Machine Learning*, cit., p. 12.

²⁹⁴ F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 26.

oggetti precedentemente acquistati²⁹⁵. Alla tredicesima settimana, il *bot* aveva acquistato una dose di ecstasy; la procura aveva allora ordinato il sequestro del *bot* e di tutti gli oggetti che erano stati acquistati. Gli sviluppatori del sistema di i.a. erano stati inizialmente iscritti nel registro degli indagati; il fascicolo era tuttavia stato successivamente archiviato, con la motivazione che l'acquisto costituiva parte di una *performance* artistica, finalizzata ad attrarre consapevolezza sul mercato del *dark web*.

7.1.2.2. Sistemi di *trading* algoritmico.

Meritano una considerazione a sé stante – per la loro rilevanza nella prassi delle negoziazioni finanziarie – i sistemi di negoziazione algoritmica, benché si tratti di algoritmi che potrebbero rientrare nella categoria dei *bots*²⁹⁶. Con l'espressione “*algorithmic trading*” ci si riferisce a sistemi informatici che sono in grado di effettuare autonomamente ordini di acquisto sulle piattaforme di *trading*: l'algoritmo decide la tempistica, il prezzo e la quantità dell'ordine; nella maggior parte dei casi, può addirittura

²⁹⁵ La descrizione del progetto è disponibile sul sito del collettivo, vd. <https://www.bitnik.org/r/>; vd. anche J. Kasperkevic, *Swiss police release robot that bought ecstasy online*, in *The Guardian* (online), 22 aprile 2015, disponibile al link: <https://www.theguardian.com/world/2015/apr/22/swiss-police-release-robot-random-darknet-shopper-ecstasy-deep-web>; J. Koebler, *In Europe, Robots Can Legally Buy Drugs Online for Art*, in *Vice US*, 15 aprile 2015, disponibile a questo link: <https://www.vice.com/en/article/mgbwg4/in-europe-robots-can-legally-buy-drugs-online-for-art>; in dottrina F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in *Philosophy & Technology*, 2020, 33, p. 451 ss.

²⁹⁶ Le implicazioni dell'uso dei sistemi di negoziazione algoritmica sulla configurabilità delle fattispecie di manipolazione di mercato sono affrontate soprattutto dalla dottrina statunitense vd. *ex multis* G. SCOPINO, *Algo Bots and the Law*, cit.; ID., *Preparing Financial Regulation for the Second Machine Age: The Need for Oversight of Digital Intermediaries in the Futures Markets*, in *2015 Columbia Business Law Review*, 2015, p. 439 ss.; Y. YADAV, *How Algorithmic Trading Undermines Efficiency in Capital Markets*, 68 *Vanderbilt Law Review*, 2015, p. 1607; ID., *The Failure of Liability in Modern Markets*, in *102 Virginia Law Review*, 2016, p. 1031 ss.; T.C.W. LIN, *The New Market Manipulation*, in *66 Emory Law Journal*, 2017, p. 1253 ss.; ID., *Artificial Intelligence, Finance, and the Law*, in *88 Fordham L. Rev.*, 2019, p. 531 ss.; W. MAGNUSON, *Artificial Financial Intelligence*, in *10 Harvard Business Law Review*, 2020, p. 337 ss.; C.R. KORSMO, *High-Frequency Trading: A Regulatory Strategy*, in *48 U. Rich. L. Rev.*, 2019, p. 523 ss.. Nella letteratura italiana vd. F. CONSULICH, *Il nastro di Möbius. intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in *Banca borsa*, 2018, vol. 71, n. 2, p. 195 ss.; A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, cit., p. 745 ss. Vi sono poi due libri divulgativi di estremo interesse per chi sia interessato al funzionamento degli algoritmi di *trading* finanziario: M. LEWIS, *Flash Boys: A Wall Street Revolt*, W W Norton & Co Inc., 2015 e A. LAUMONIER, 6|5 – *La rivolta delle macchine*, Nero Editions, 2018.

avviare l'ordine in assenza di intervento umano²⁹⁷. Una categoria particolare di *automated trading* sono gli *high frequency trading* (HFT): nonostante non esista una definizione condivisa, per semplicità possiamo riferirci ad essi come ad algoritmi che consentono di effettuare negoziazioni ad altissima velocità, rendendo possibile perfezionare diverse migliaia di ordini al secondo²⁹⁸. Tale beneficio è assicurato dall'abbattimento del c.d. periodo di latenza, ossia del lasso di tempo che intercorre tra la decisione economica e l'effettiva esecuzione dell'operazione sul mercato, che si attesta oggi intorno a pochi microsecondi²⁹⁹.

L'esecuzione di un elevatissimo numero di transazioni in un breve intervallo di tempo ha la finalità di sfruttare le minime oscillazioni di prezzo dei titoli, che sarebbero impercettibili per un investitore tradizionale; i margini di profitto – esigui in relazione alla singola operazione – assumono una portata considerevole se rapportati alla mole complessiva delle operazioni³⁰⁰, determinando così un vero e proprio *vantaggio*

²⁹⁷ Questa la definizione fornita dalla Direttiva 2014/65/UE del Parlamento Europeo e del Consiglio del 15 maggio 2014 relativa ai mercati degli strumenti finanziari e che modifica la direttiva 2002/92/CE e la direttiva 2011/61/UE (c.d. MIFID II): «*negoziazione di strumenti finanziari in cui un algoritmo informatizzato determina automaticamente i parametri individuali degli ordini, come ad esempio se avviare l'ordine, i tempi, il prezzo o la quantità dell'ordine o come gestire l'ordine dopo la sua presentazione, con intervento umano minimo o nullo e non comprende i sistemi utilizzati unicamente per trasmettere ordini a una o più sedi di negoziazione, per trattare ordini che non comportano la determinazione di parametri di trading, per confermare ordini o per eseguire il trattamento post-negoziazione delle operazioni eseguite*» (Art. 4(1)(39)). Vd. anche G. SCOPINO, *Algo Bots and the Law*, p. 17.

²⁹⁸ Questa la definizione fornita dalla Direttiva MIFID II: «*qualsiasi tecnica di negoziazione algoritmica caratterizzata da: a) infrastrutture volte a ridurre al minimo le latenze di rete e di altro genere, compresa almeno una delle strutture per l'inserimento algoritmico dell'ordine: co-ubicazione, hosting di prossimità o accesso elettronico diretto a velocità elevata; b) determinazione da parte del sistema dell'inizializzazione, generazione, trasmissione o esecuzione dell'ordine senza intervento umano per il singolo ordine o negoziazione, e c) elevato traffico infragiornaliero di messaggi consistenti in ordini, quotazioni o cancellazioni*» (Art. 4(1)(40)). La velocità di tali algoritmi sarebbe da imputare non solo alla sofisticata struttura del *software* e alla potenza computazionale degli *hardware*, ma anche alla peculiare collocazione fisica di questi ultimi, posti in prossimità del server dei mercati finanziari presso i quali gli algoritmi operano, così da ridurre radicalmente lo spazio che l'impulso elettrico deve percorrere per giungere a destinazione (cd. *co-location*), vd. A. PUORRO, *High Frequency Trading: una panoramica*, in *Questioni di Economia e Finanza (Occasional Papers)* – Banca d'Italia, n. 198, 2013, p. 11.

²⁹⁹ vd. J. ARIFOVIC – X. HE – L. WEI, *Machine learning and speed in high-frequency trading*, in *Journal of Economic Dynamics & Control*, 139, 2022, p. 1; vd. A. PUORRO, *High Frequency Trading*, cit., p. 10-11, che – significativamente – nel 2012 parla di “millisecondi” e non di “microsecondi”, a testimonianza della continua evoluzione tecnologica di questi sistemi; a Puorro si rimanda anche per la descrizione delle diverse fasi che consentono di trasformare la decisione economica nell'esecuzione dell'operazione.

³⁰⁰ G. STRAMPELLI, *L'informazione societaria a quindici anni dal T.U.F.: profili evolutivi e problemi*, in *Riv. soc.*, 2014, p. 996; A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, cit., p. 747-748.

competitivo dell'operatore algoritmico rispetto all'investitore fisico³⁰¹. D'altra parte, l'assenza di pregiudizi ed "emozioni" nelle scelte algoritmiche, unita alla capacità di analisi di un'imponente mole di informazioni sull'andamento del mercato, rende le predizioni automatizzate maggiormente accurate di quelle umane³⁰².

Dal punto di vista penalistico, tuttavia, ciò che più rileva non è tanto la *velocità* degli operatori algoritmici, quanto la loro *autonomia operativa* rispetto ai soggetti nel cui interesse agiscono (intermediari finanziari, titolari di *Hedge funds* etc.)³⁰³.

I sistemi di *algorithmic trading* – programmati per *massimizzare i profitti* degli operatori finanziari – effettuano di continuo autonome scelte di investimento, decidendo di volta in volta tra diverse opzioni possibili. Gli algoritmi più recenti sfruttano altresì tecniche di *machine learning*³⁰⁴ e sono dunque in grado di apprendere dall'ambiente in cui operano e modificare le proprie strategie operative adattandosi al nuovo contesto – con il risultato che l'HFT potrà effettuare una decisione di investimento che non era prevedibile dal programmatore al momento dell'immissione sul mercato.

La questione problematica, allora, consiste ancora una volta nel comprendere se i tradizionali meccanismi imputativi della responsabilità penale siano adeguati a regolare anche i casi in cui il fatto illecito scaturisca non da una condotta umana ma da un'operazione algoritmica.

È infatti pacifico che fatti di manipolazione del mercato – penalmente rilevanti ai sensi dell'art. 185 T.u.f.³⁰⁵ – possano essere realizzati anche attraverso l'utilizzo di algoritmi di *trading*: anzi, la velocità operativa degli HFT ha determinato l'emergere di nuovi schemi di turbativa, quali, ad esempio, il *pinging*, lo *smoking*, il *layering*, lo *spoofing*, il cui comune denominatore è costituito dall'immissione e dalla contestuale cancellazione di migliaia di ordini, al fine di far apparire le condizioni di mercato

³⁰¹ F. CONSULICH, *Il nastro di Möbius*, cit., p. 199-200; vd. anche A.F. TRIPODI, *Abusi di mercato e trading algorithmic*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 745 ss.

³⁰² G. SCOPINO, *Algo Bots and the Law*, p. 18.

³⁰³ F. CONSULICH, *Il nastro di Möbius*, cit., p. 206

³⁰⁴ J. ARIFOVIC – X. HE – L. WEI, *Machine learning and speed in high-frequency trading*, cit., p. 6.

³⁰⁵ Art. 185, co. 1, D.lgs. 24 febbraio 1998, n. 58: «Chiunque diffonde notizie false o pone in essere operazioni simulate o altri artifici concretamente idonei a provocare una sensibile alterazione del prezzo di strumenti finanziari, è punito con la reclusione da uno a sei anni e con la multa da euro ventimila a euro cinque milioni».

differenti da quelle effettive³⁰⁶. Tali strategie paiono integrare a pieno titolo quegli “*altri artifici concretamente idonei a provocare una sensibile alterazione del prezzo di strumenti finanziari*” di cui al citato art. 185 T.u.f.

D'altra parte, la capacità offensiva degli HFT ha avuto diverse manifestazioni esemplari nei c.d. *flash crash*, ovverosia in quei fenomeni di improvviso crollo dei prezzi, seguito da un rapido ritorno alla preesistente situazione di equilibrio, senza che tale oscillazione possa trovare una giustificazione nel mutamento del valore del titolo o in altre ragioni attinenti alla struttura del mercato di riferimento: «Il flash crash sembra, quindi, avere una natura meccanica, vale a dire indipendente da fattori economici reali ed, invece, molto più legata alle procedure tecniche di funzionamento degli scambi dal punto di vista informatico»³⁰⁷. Si pensi, ad esempio, a quanto accadde in quello che viene considerato il primo *flash crash* della storia, il crollo registratosi il 6 maggio 2010, quando l'indice *Dow Jones* scese di oltre l'8% nell'arco di pochi minuti, recuperando tutte le perdite nei dieci minuti successivi: un evento senza precedenti nella storia della finanza moderna.

Ad alimentare tali violente fluttuazioni dei titoli vi sono proprio le tecniche di negoziazione ad alta frequenza, che reagiscono immediatamente e in massa di fronte alle informazioni provenienti dal mercato, riuscendo altresì a beneficiare della turbativa: così, gli HFT – di fronte alla picchiata dei prezzi – sono i primi a vendere e – al momento del “rimbalzo” – sono i primi a ricomprare, anticipando in questo modo le strategie degli investitori fisici³⁰⁸. Emblematico, ad esempio, è il *flash crash* realizzatosi sui mercati americani il 23 aprile del 2013, quando la notizia – poi rivelatasi falsa – di un attentato a Barack Obama aveva indotto gli HFT a vendere in massa titoli finanziari, intuendo un ribasso sui mercati (e contribuendo ad anticiparne la verifica e ad amplificarne gli effetti). Il rialzo che, quindici minuti dopo, ne seguì fu determinato dagli stessi HFT, che ricominciarono gli acquisti, maturando complessivamente 600 milioni di dollari di profitti³⁰⁹.

³⁰⁶ Così F. CONSULICH, *Il nastro di Möbius*, cit., p. 210; per la distinzione tra le varie tecniche si rinvia a G. RUTA, *I.a. nei reati economici e finanziari*, in *Intelligenza artificiale e giurisdizione penale*, cit., p. 65 ss.

³⁰⁷ F. CONSULICH, *Il nastro di Möbius*, cit., p. 201.

³⁰⁸ F. CONSULICH, *Il nastro di Möbius*, cit., p. 202; A.F. TRIPODI, *Abusi di mercato e trading algorithmic*, cit., p. 748.

³⁰⁹ F. CONSULICH, *Il nastro di Möbius*, cit., p. 202 e bibliografia ivi citata.

Il fatto, dunque, che, nel giro di pochi minuti, i *flash crash* esauriscano i propri effetti non vale certo a diminuire la problematicità del fenomeno: sebbene, infatti, complessivamente il mercato recuperi una situazione di equilibrio, l'impatto del crollo si fa sentire sulle posizioni dei singoli operatori, che possono ritrovarsi con ingenti guadagni o poderose perdite, a seconda della velocità con la quale hanno reagito alle oscillazioni³¹⁰.

Posta, dunque, la potenzialità offensiva dei *trader* algoritmici, restano da valutare le condizioni di responsabilità delle persone fisiche (e/o degli enti) che beneficiano dei risultati economici dell'attività di investimento. Gli operatori algoritmici, infatti, non possono essere considerati *meri strumenti* nelle mani dell'uomo, stante la loro capacità di *decision-making* e stante, altresì, l'adattività e l'interconnettività con l'ambiente in cui si trovano e con gli altri agenti virtuali.

7.2. Sistemi che possono recare un'offesa *indiretta* ai beni giuridici.

7.2.1. Sistemi di supporto alla diagnostica e al trattamento medico.

Come già accennato, uno dei settori più promettenti per lo sviluppo e l'utilizzo dei sistemi di i.a. è quello medico. In diversi ospedali sono già in fase di *trial* diversi sistemi di supporto alla diagnostica e al trattamento (*Clinical Decision Support Systems*, c.d. CDSS), in grado di analizzare in maniera veloce e accurata un'immensa mole di dati concernenti la c.d. *medical knowledge* e la storia clinica del paziente, in modo tale da fornire indicazioni diagnostiche e terapeutiche ai medici competenti. Tali tecnologie hanno mostrato risultati molto accurati soprattutto nel campo della radiologia, dell'oncologia, della dermatologia³¹¹.

³¹⁰ F. CONSULICH, *Il nastro di Möbius*, cit., p. 202.

³¹¹ La letteratura medica è ormai molto ampia. Per limitarci al solo settore oncologico, vd. D. WANG E AL., *Deep Learning for Identifying Metastatic Breast Cancer*, arXiv preprint, 2016, disponibile a questo link: <https://arxiv.org/abs/1606.05718>; L. FERRANTE DI RUFFANO E AL., *Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults*, in *Cochrane Database of Systematic Reviews*, 12, 2018; per ulteriori riferimenti bibliografici rinviamo al report pubblicato dal Consiglio Superiore di Sanità, *I sistemi di i.a. come strumento di supporto alla diagnostica*, 2021, p. 18, disponibile a questo link: https://www.salute.gov.it/imgs/C_17_pubblicazioni_3218_allegato.pdf.

Tra questi sistemi, IBM Watson viene considerato uno dei più promettenti: già fenomeno *pop*, ben conosciuto dai telespettatori statunitensi, per aver vinto al quiz televisivo *Jeopardy!* contro gli allora campioni in carica³¹², Watson è stato poi utilizzato in diversi settori (come *fashion advisor*; come “aiuto *chef*”; come *software* di ricerca dottrinale e giurisprudenziale negli studi legali³¹³), per poi trovare applicazione soprattutto nel settore medico.

Come abbiamo già riportato, alcuni esperimenti condotti presso lo Sloan-Kettering Hospital negli Stati Uniti – dove IBM Watson è in fase di sperimentazione – suggeriscono che le diagnosi effettuate da Watson sono più accurate di quelle effettuate dai medici. In base a tali studi, la media di diagnosi di cancro corrette sarebbe del 90% per Watson, rispetto al 50% dei medici³¹⁴.

Il database su cui lavora IBM Watson è praticamente illimitato, essendo costituito da tutto ciò che è accessibile su internet e, inoltre, dalle banche dati sanitarie aggiornate da sistemi sanitari nazionali e ospedali. Per avere un’idea della mole dei dati alla quale può accedere IBM Watson, basti pensare che *Kaiser Permanente*, una delle maggiori assicurazioni sanitarie statunitensi, ogni giorno raccoglie almeno tre *terabyte* di dati concernenti i 14 milioni di pazienti che accedono ai servizi sanitari³¹⁵. D’altra parte, IBM Watson non solo analizza i dati “grezzi” contenuti nei sistemi sanitari nazionali, ma è anche in grado di esaminare l’intera letteratura scientifica sul tema di interesse.

È dunque evidente che la capacità di Watson IBM di setacciare in breve tempo questo insieme fluido di dati e di ricavarne informazioni rilevanti a fini diagnostici e terapeutici costituisce un vantaggio non indifferente rispetto all’attività dei medici³¹⁶.

I CDSS si basano su tecniche di *supervised learning* (vd. *supra*, § 2.1.): gli algoritmi imparano a identificare e a correlare domande (*input*) e risposte (*output*) sulla

³¹² M.C. ELISH – D. BOYD, *Situating Methods in the Magic of Big Data and Artificial Intelligence*, cit., p. 11.

³¹³ J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility: The prospect of expert robots*, in R. Calo e al. (eds.), *Robot Law*, cit., p. 105 e bibliografia ivi citata.

³¹⁴ F. LAGIOIA – G. CONTISSA, *The Strange Case of Dr. Watson: Liability Implications of AI Evidence-Based Decision Support Systems in Health Care*, in *European Journal of Legal Studies*, vol. 12, n. 2, 2020, p. 255; per ulteriori studi sull’accuratezza dei CDSS rinviamo ai dati raccolti nel report del Consiglio Superiore di Sanità, *I sistemi di i.a. come strumento di supporto alla diagnostica*, cit., p. 18-19.

³¹⁵ J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 108, che a loro volta citano P. WEBSTER, *Electronic Health Records: An Asset or a Whole Lot of Hype?*, in 182 *Can. Med. Ass’n J.*, 2010.

³¹⁶ J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 108.

base degli esempi presenti nell'insieme di addestramento, applicando successivamente le conoscenze acquisite nell'analisi delle nuove questioni che vengono loro sottoposte. Gli algoritmi generano poi una serie di soluzioni (diagnosi o trattamenti), ciascuna con un punteggio di confidenza associato.

Nella pratica, ad esempio, il medico potrà inserire nel sistema la sequenza genetica del tumore di un paziente e la sua storia clinica; il sistema integrerà queste informazioni con i vastissimi *database* che ha a disposizione e genererà una serie di ipotetiche diagnosi o piani terapeutici, assegnando loro specifici punteggi di fiducia e classificando le risposte in base alla probabilità che siano corrette. In questo modo, il sistema parrebbe adottare l'approccio tipico della c.d. *evidence-based medicine*, che impone, nel processo decisionale medico, di utilizzare le migliori risultanze scientifiche.

Posto che la capacità predittiva dei CDSS pare destinata a crescere e probabilmente, sul lungo periodo, a superare quella dell'essere umano, la questione più dibattuta, in dottrina, è quella che riguarda il ruolo che dovrà essere riservato al medico nel processo decisionale: il medico, insomma, diventerà un mero esecutore delle istruzioni provenienti dall'algoritmo o conserverà l'ultima parola su diagnosi e trattamento da applicare nel caso concreto?³¹⁷.

La quasi totalità degli autori e dei testi di *soft law* ad oggi pubblicati ritengono che il medico debba sempre essere il decisore finale, sulla base del già citato principio dell'*human in the loop*³¹⁸. Le motivazioni che possono essere addotte sono diverse.

³¹⁷ Il tema è già oggetto di ampie riflessioni. Ci limitiamo, in questa sede, a rinviare ai principali contributi: W. N. PRICE II, *Regulating Black-Box Medicine*, 116 *Mich. L. Rev.*, 2017, p. 421; J.S. ALLAIN, *From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems*, in 73 *La. L. Rev.*, 2013, p. 1049; F. LAGIOIA – G. CONTISSA, *The Strange Case of Dr. Watson*, cit., p. 245; A.M. FROOMKIN E AL., *When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning*, in 61 *Ariz. L. Rev.* 33, 2019, p. 33; J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit.; S. GERKE E AL., *Ethical and legal challenges of artificial intelligence-driven healthcare*, in A. Bohr – K. Memarzadeh (eds), *Artificial intelligence in healthcare*, Academic Press, 2020, p. 295 ss.; in lingua italiana vd. A. PERIN, *Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale*, in *BioLaw Journal*, 2019, n. 1, p. 207; U. RUFFOLO, *L'Intelligenza artificiale in sanità*, cit., p. 502; L. SCAFFARDI, *La medicina alla prova dell'Intelligenza Artificiale*, in *DPCE Online*, Vol. 51, N. 1, apr. 2022, p. 349.

³¹⁸ In questo senso A.M. FROOMKIN E AL., *When AIs Outperform Doctors*, cit.; A. PERIN, *Standardizzazione, automazione e responsabilità medica*, cit., p. 230. Quanto alle fonti istituzionali, vd. lo studio commissionato dal Parlamento europeo, *Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts*, giugno 2022, p. 49, disponibile a questo link: [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU\(2022\)729512_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf),

Innanzitutto, non dobbiamo dimenticare che i CDSS utilizzano tecniche statistico-predittive basate su massicce quantità di dati, la cui qualità e accuratezza non sempre sarà vagliabile automaticamente dal sistema: le competenze professionali del medico, unite a quel “buon senso” che – quantomeno ad oggi – non può essere riprodotto artificialmente, dovrebbero dunque consentire di non accogliere quelle raccomandazioni che appaiano come completamente implausibili. Il fatto che i sistemi di *machine learning* siano addestrati su un *set* limitato di dati, inoltre, può condurre ad un’individuazione erronea del trattamento più adeguato per il paziente. Tale problema può verificarsi perché i *database* prendono in considerazione soltanto pazienti con determinate caratteristiche biologiche e genetiche – la cui risposta ad un trattamento non può essere generalizzata ed estesa a pazienti con caratteristiche diverse (vd. *supra*, § 3) – o perché il sistema fonda le proprie raccomandazioni su correlazioni spurie. A quest’ultimo proposito, ad esempio, è emblematica la vicenda, riportata in uno studio del 2015³¹⁹, in cui un CDSS aveva classificato come “a basso rischio” di mortalità per polmonite i pazienti asmatici. Tale conclusione – evidentemente controintuitiva – rifletteva un modello reale nei dati di addestramento, ma ne travisava il significato: non è, infatti, l’asma in sé a ridurre l’incidenza della mortalità per polmonite, quanto il trattamento dell’infezione in terapia intensiva, che – in alcune strutture ospedaliere – viene applicato di *default* ai pazienti malati d’asma, mentre non viene eseguito con la medesima frequenza nei confronti della popolazione generale.

D’altra parte, è stato osservato che il ruolo del medico è fondamentale nel considerare fattori che possono essere essenziali nella diagnosi, ma che non possono essere percepiti da un CDSS (ad esempio, un particolare odore o colore o consistenza della pelle del paziente)³²⁰: tali fattori potrebbero dunque suggerire una diagnosi diversa, e più adeguata, al medico.

Infine, su un piano più strettamente etico, la scomparsa del medico dall’ambito decisionale si porrebbe in contrasto con la necessità – che oggi trova un esplicito riferimento normativo nell’art. 1, co. 2, l. 22 dicembre 2017, n. 219 – che la relazione

che richiede che sia sempre previsto «*a human-in-the-loop mechanism to enable for human oversight and feedback*».

³¹⁹ R. CARUANA E AL., *Intelligible Models for Healthcare*, cit. Vd. anche M.A. LEMLEY – B. CASEY, *Remedies for Robots*, cit., p. 1336 ss.

³²⁰ F. LAGIOIA – G. CONTISSA, *The Strange Case of Dr. Watson*, cit., p. 275.

medico-paziente sia improntata alla fiducia e alla trasparenza, nonché fondata sul consenso informato, «nel quale si incontrano l'autonomia decisionale del paziente e la competenza, l'autonomia professionale e la responsabilità del medico»³²¹. Da un lato, dunque, vi è l'esigenza che il rapporto con il paziente non sia delegato ad un sistema artificiale, ma sia mantenuto ad un livello umano e personale; dall'altro, deve essere preservato il percorso informativo destinato ad acquisire il consenso (autonomo, libero e consapevole) da parte del paziente³²². Quest'ultima è una questione piuttosto problematica, se pensiamo che il processo decisionale interno agli algoritmi predittivi – specie quelli più accurati, che utilizzano tecniche di *deep learning*³²³ – è imperscrutabile dagli stessi medici utilizzatori, a causa del già citato fenomeno della *black box* (vd. *supra* § 5.3)³²⁴.

In ogni caso, se le suddette considerazioni inducono a ritenere che il medico debba restare il decisore finale, rimangono alcune questioni da risolvere: quale valore hanno le previsioni derivanti dal calcolo algoritmico? In quali casi sarà legittimo fidarsi di esse e quando, invece, sarà necessario discostarsene? E, di conseguenza, a quali condizioni il medico dovrà rispondere per aver assunto ed eseguito una decisione erronea proposta da un sistema di intelligenza artificiale?

Come è stato acutamente rilevato³²⁵, la relazione dialettica che si pone tra medico e sistemi algoritmici pare assimilabile a quella esistente tra professionista e linee-guida: anche in questo caso, infatti, si tratta di individuare i confini della responsabilità medica, in un contesto in cui l'autonomia professionale è condizionata dalle indicazioni provenienti da un diverso polo decisionale.

Le indicazioni provenienti dai CDSS, dunque, potrebbero, in futuro, assurgere a vero e proprio *standard of care*: analogamente al meccanismo previsto dall'art. 590-sexies c.p. per le linee-guida, potrebbe dunque prevedersi una generale tutela della

³²¹ Art. 1, co. 2, legge 22 dicembre 2017, n. 219, recante norme in materia di consenso informato e di disposizioni anticipate di trattamento

³²² Su questi temi vd. più approfonditamente il report del Consiglio Superiore di Sanità, *I sistemi di i.a. come strumento di supporto alla diagnostica*, cit., p. 30 ss.; H. VAN KOLFSCHOOTEN, *EU regulation of artificial intelligence: Challenges for patients' rights*, in 59 *Common Market Law Review*, 2022, p. 95 ss.

³²³ A.J. LONDON, *Artificial intelligence and black-box medical decisions*, cit., p. 16-17.

³²⁴ S. GERKE E AL., *Ethical and legal challenges of artificial intelligence-driven healthcare*, cit., 301-302; G. COHEN, *Informed Consent and Medical Artificial Intelligence: What to Tell the Patient*, in *Georgetown Law Journal*, 108, 2020, p. 1425.

³²⁵ A. PERIN, *Standardizzazione, automazione e responsabilità medica*, cit., p. 231.

legittima aspettativa del medico nei confronti della correttezza dei risultati algoritmici, che verrebbe meno soltanto nelle ipotesi di colpa grave, cioè di chiara riconoscibilità e prevedibilità dell'errore o dell'inadeguatezza della raccomandazione fornita dall'algoritmo³²⁶.

³²⁶ A. PERIN, *Standardizzazione, automazione e responsabilità medica*, cit., p. 231-232; sostanzialmente nello stesso senso J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 122 ss.; critica, invece, per le ripercussioni che l'equiparazione tra risultato algoritmico e linee guida potrebbe avere sull'autonomia professionale, è A. Fiorentini, *Machine learning e dispositivi medici: riflessioni in materia di responsabilità civile*, in *Corr. giur.*, 2021, 10, p. 1262. Sul ruolo delle linee guida nell'accertamento della responsabilità medica si vd., per tutti, A.R. DI LANDRO, *Dalle linee guida e dai protocolli all'individualizzazione della colpa penale nel settore sanitario*, Giappichelli, 2012.

CAPITOLO II

INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ CIVILE

SOMMARIO: 1. La responsabilità civile come territorio privilegiato per l'*enforcement ex post* dei danni da intelligenza artificiale. – 1.1 Dal dogma della colpa alla responsabilità oggettiva per creazione del rischio: alcuni cenni storici. – 1.2. Ritorno al futuro: un declino del modello oggettivistico? Regolazione dell'i.a. e *policy* europea. – 1.3. Adattabilità della normativa vigente o necessità di introdurre una disciplina *ad hoc*? – 2. La responsabilità oggettiva in capo all'utilizzatore del sistema di i.a.: i regimi speciali di responsabilità previsti dal codice civile. – 2.1 Responsabilità da cosa in custodia (art. 2051 c.c.) e responsabilità del custode dell'animale o di chi se ne serve (art. 2052 c.c.). – 2.2. Responsabilità da circolazione di veicoli (art. 2054 c.c.). – 2.3 Responsabilità di padroni e committenti (art. 2049 c.c.). – 3. Responsabilità per danno da prodotto difettoso: cenni introduttivi. – 3.1 Sull'adeguatezza della nozione di "prodotto". – 3.2 Sull'adeguatezza della nozione di "produttore". – 3.3. Sull'adeguatezza della nozione di "difetto". – 3.3.1. La delimitazione della rilevanza dei difetti: le esimenti. – 3.3.2. La classificazione tradizionale: difetti di fabbricazione, progettazione, informazione nei sistemi di i.a. – 3.3.3. I difetti di progettazione: tra responsabilità per colpa e responsabilità assoluta. – 3.3.4. La nozione di difetto di progettazione in relazione ai sistemi di i.a.: alcuni profili di inadeguatezza. – 3.3.5. Nuovi criteri per la valutazione del difetto: le proposte della bozza di revisione della *Product liability directive*. – 3.3.6. Verso una nuova nozione di *design defect* in relazione ai sistemi di i.a.? Le proposte della dottrina – 3.3.6.1. Sostanziale eliminazione del requisito del difetto. – 3.3.6.2. L'agente modello umano. – 3.3.6.3. L'agente modello algoritmico. – 3.4. I rapporti tra *product safety* e *product liability*: verso un'adozione della *preemption doctrine*? – 3.4.1. *Product Safety* e tecniche di armonizzazione. Cenni. – 3.3.7.2. Sistemi di i.a. e norme di armonizzazione: l'*AI Act*. – 3.4.3. (*segue*) I sistemi di intelligenza artificiale ad alto rischio. – 3.4.4. La tesi della non difettosità del prodotto conforme agli standard tecnici. – 3.4.5. (*segue*) Separazione concettuale tra disciplina della *product liability* e disciplina della *product safety*. – 3.5. L'accertamento del nesso di causalità. Profili sostanziali. – 3.5.1. (*segue*) Profili processuali. – 4. Responsabilità per esercizio di attività pericolose (art. 2050 c.c.). – 5. La personalità elettronica: una proposta europea, presto abbandonata. – 6. *Manufacturer Enterprise Liability (MER)* e assicurazione obbligatoria. – 7. Intelligenza artificiale e responsabilità civile: indicazioni e spunti per il prosieguo dell'indagine.

1. La responsabilità civile come territorio privilegiato per l'*enforcement ex post* dei danni da intelligenza artificiale.

Il dibattito sui danni cagionati dai sistemi di i.a. – in Italia come all'estero – si è sviluppato soprattutto in ambito civilistico, ove la flessibilità delle categorie concettuali sembra fornire un quadro di riferimento di massima per la regolazione del fenomeno. Se, sul versante del diritto penale, l'autonomia decisionale e comportamentale dei sistemi di i.a. potrebbe determinare una crisi del modello tradizionale di ascrizione dell'evento lesivo, nel settore civile, invece, il criterio del "più probabile che non" nell'accertamento del nesso di causalità, nonché la sussistenza di modelli di responsabilità *senza colpa*,

sembrano fornire strumenti più duttili, e già consolidati, per l’allocazione dei costi derivanti dal danno algoritmico.

Attraverso i regimi di responsabilità oggettiva, in particolare, il rischio di eventi lesivi derivanti dall’agire di dispositivi intelligenti potrebbe essere attribuito al produttore o all’operatore del sistema *indipendentemente* dall’accertamento della colpa. A tal proposito, può risultare utile sottolineare l’ovvio: la responsabilità aquiliana non è un “gioco a somma zero”¹; la mancata individuazione di un soggetto responsabile per la commissione di un fatto lesivo comporta che il costo derivante dal danno rimanga in capo a colui che l’ha subito. Qualora, dunque, si optasse per un meccanismo di imputazione dei danni basato sulla colpa – ai sensi dello schema tradizionale di cui all’art. 2043 c.c. – i costi degli eventi lesivi derivanti da comportamenti emergenti dei sistemi di i.a. sarebbero destinati a restare, perlopiù, in capo ai danneggiati: da un lato, perché potrebbe trattarsi di eventi che il produttore o l’utilizzatore non avrebbero potuto evitare nemmeno utilizzando l’ordinaria diligenza (si pensi ad un pedone investito da una *self driving car* di livello 5, fabbricata secondo i più avanzati *standard* di sicurezza del settore); dall’altro, perché, anche trattandosi di eventi che un agente modello avrebbe potuto impedire, spesso il danneggiato non si trova nelle condizioni di *poter provare* la colpa del produttore o dell’utilizzatore.

Una tale allocazione economica dei costi non parrebbe compatibile né con la funzione *deterrente* della responsabilità civile, né con la funzione *compensativa*. Quanto al primo aspetto, il rischio è che si venga a creare un “incentivo distorto”² per i produttori e per gli utilizzatori dei dispositivi intelligenti, poiché la prospettiva di una quasi certa immunità (per le motivazioni anzidette) indurrebbe ad un’assunzione di rischi spropositati, con conseguente socializzazione dei costi, in caso di danno³. Quanto, invece, al profilo compensativo, il ricorso al criterio ordinario della colpa rischierebbe di diminuire drasticamente le probabilità di soddisfazione dei danneggiati, che si

¹ M. BUSSANI, *La responsabilità civile al tempo della crisi*, in *Riv. trim. dir. proc. civ.*, 2015, fasc. 2, p. 580.

² G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, a cura di Pasquale Femia, Edizioni Scientifiche Italiane, 2019, p. 27.

³ G. WAGNER, *Robot liability*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, Baden-Baden, 2019, p. 30 ss.

troverebbero così sforniti di adeguata tutela di fronte all'evolversi dei processi industriali e produttivi.

Ecco che, allora, il dibattito civilistico in materia di danno algoritmico – sia in Italia sia all'estero – si è concentrato, per un verso, sull'applicabilità, *de jure condito*, dei regimi di responsabilità oggettiva già esistenti (dal lato dell'utilizzatore, responsabilità da custodia di cose o animali e responsabilità vicaria; dal lato del produttore, responsabilità per esercizio di un'attività pericolosa o per danno da prodotto) e, per altro verso, sull'opportunità, *de jure condendo*, di modificare tali regimi – in modo tale da garantirne l'adeguatezza rispetto al nuovo settore – o, ancora, di introdurre nuove forme compensative per le vittime (assicurazione obbligatoria, personalità elettronica⁴, *Manufacturer Enterprise Liability*⁵).

Resta ferma, ovviamente, l'applicabilità residuale del regime generale previsto all'art. 2043 c.c. In questo capitolo, tuttavia, ci concentreremo esclusivamente sulle ipotesi di responsabilità oggettiva previste dal nostro ordinamento, nella convinzione – confortata dall'opinione unanime della letteratura civilistica in materia⁶ – che proprio esse vengano principalmente in rilievo con riferimento ai danni da prodotto intelligente, date le difficoltà di accertamento della colpa in capo al produttore di sistemi di i.a.

⁴ Per “personalità elettronica” si intende la creazione di un *patrimonio separato* in capo al sistema di i.a., destinato in via esclusiva al soddisfacimento degli obblighi risarcitori derivanti dagli eventi lesivi algoritmici, senza che sia necessario l'accertamento di una colpa in capo al produttore o all'utilizzatore del sistema. L'ipotesi di introdurre una personalità elettronica era stata prefigurata dalla Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, cit., § 59, lett. f). Su tale proposta torneremo nel prosieguo, vd. *infra*, § 5., in questo cap.).

⁵ La proposta di *Manufacturer Enterprise Liability* prevede la creazione di un fondo indennitario, finanziato dalle società produttrici di sistemi di i.a., al quale il danneggiato potrebbe accedere qualora dimostri di aver subito il danno, senza che sia necessaria la prova del difetto del prodotto (vd. *infra*, § 5., in questo cap.).

⁶ Vd. per tutti U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 95 ss.; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, in *Oss. dir. civ. comm.*, 2020, fasc. 2, p. 445 ss.; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1719; vd., inoltre, gli autori citati nel prosieguo, ai §§ 2. ss.

1.1. Dal dogma della colpa alla responsabilità oggettiva per creazione del rischio: alcuni cenni storici.

Nel modello aquiliano tradizionale – che caratterizza tutte le codificazioni civili ottocentesche e dei primi anni del Novecento – il criterio di imputazione della responsabilità è unico ed è fondato sulla colpa: il costo del danno, dunque, viene sempre trasferito su colui che lo ha cagionato con colpa (o con dolo) e il giudizio di responsabilità non coinvolge altri se non i due soggetti immediatamente interessati (danneggiato e danneggiante)⁷.

Tale modello imputativo risente ancora, com'è evidente, della funzione punitiva un tempo attribuita alla responsabilità aquiliana, risalente all'epoca in cui non era netta la distinzione tra “pena” e obbligo di risarcimento civile⁸. Lungi dal trattarsi di una questione “neutra” e puramente tecnica, la fissazione dei criteri di imputazione della responsabilità (e, dunque, in termini economici, la distribuzione dei rischi e l'allocazione dei costi) è dipendente in larga misura dai sistemi economici e dalle strutture del mercato di riferimento⁹: il principio della colpa come unico criterio imputativo della responsabilità civile, nello specifico, non faceva altro che tradurre sul piano giuridico l'approccio del *laissez faire* tipico dello Stato liberale ottocentesco, assolvendo alla medesima funzione ideologica che assolvevano, in altri rami dell'ordinamento civile, il dogma della libertà contrattuale e l'assolutezza dei poteri del proprietario¹⁰.

A parere di alcuni autori, in particolare, il principio in base al quale non potesse esserci responsabilità senza colpa era “*richiesto dalla rivoluzione industriale*”, poiché si riteneva che fosse più vantaggioso, per un'economia industriale in espansione, trascurare la sicurezza degli individui, piuttosto che sacrificare gli interessi dell'impresa, gravandola

⁷ Non è questa la sede per ripercorrere gli orientamenti che, nel corso dei secoli, hanno consentito l'affermarsi della concezione in base alla quale non può esserci “nessuna responsabilità senza colpa”. Ci limitiamo qui a rinviare all'ampia trattazione di G. ALPA, *La responsabilità civile. Parte generale*, Utet, 2010, al capitolo secondo, p. 25 ss.

⁸ Originariamente, vi era una perfetta coincidenza tra responsabilità civile e penale, poiché il risarcimento era ammesso solo nei casi in cui la *lex Aquilia* prevedeva la condanna ad una pena; vd. G. PONZANELLI, *La responsabilità civile: profili di diritto comparato*, Il Mulino, 1992, p. 61; G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 159-160 e autori ivi citati; C. PIERGALLINI, “Civile” e “penale” a perenne confronto: l'appuntamento di inizio millennio, in *Riv. it. dir. proc. pen.*, 2012, n. 4, p. 1300.

⁹ G. ALPA, *La responsabilità civile. Parte generale*, cit. p. 159-160.

¹⁰ G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 105.

dei costi degli incidenti inevitabili¹¹. Il “dogma della colpa” risulta perciò frutto di una consapevole opzione di politica del diritto, volta a sottrarre le imprese dal costo degli incidenti arrecati incolpevolmente ai terzi, nella convinzione che costi eccessivamente gravosi avrebbero potuto minare la solidità economica della nascente industria¹².

Il criterio esclusivo della colpa, di conseguenza, comportava che la grandissima parte dei danni cagionati non venisse risarcita. La maggioranza degli eventi lesivi derivanti dallo sviluppo industriale sono infatti danni anonimi o incolpevoli: di conseguenza, nell'impossibilità di provare la negligenza del produttore, a sopportare il costo dei danni industriali restavano coloro che li avevano subiti (soprattutto lavoratori dipendenti, ma anche utilizzatori dei prodotti), sulla base della teoria dell'”*assumption of risk*”¹³. Per citare uno dei maggiori studiosi della responsabilità civile in Italia:

«i danneggiati hanno finanziato l'industria nella fase del capitalismo ottocentesco. La logica liberista di lasciare il danno lì dove cade ha comportato una sorta di immunità alla borghesia imprenditoriale in ascesa, che ha avuto come diretta conseguenza quella di agevolare l'accumulazione del capitale in danno delle vittime seppur incolpevoli. Queste considerazioni svelano la falsità della tesi che attribuisce alla colpa il ruolo di supremazia siccome espressiva del progresso morale del diritto»¹⁴.

Nonostante i primi tentativi di promuovere *modelli di responsabilità aquiliana alternativi alla colpa* risalgano già agli ultimi decenni del diciannovesimo secolo, è soltanto a partire dagli anni Sessanta del Novecento che cominciano a svilupparsi le prime riflessioni organiche in materia – soprattutto in relazione al danno da prodotto – destinate a diventare un punto di riferimento in materia di responsabilità oggettiva¹⁵.

¹¹ F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, Edizioni Scientifiche Italiane, 2003, p. 127; G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 106.

¹² G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 107; F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 126.

¹³ F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 126-127.

¹⁴ M. FRANZONI, *Dei fatti illeciti: art. 2043-2059*, Zanichelli, 2020, p. 44.

¹⁵ P. TRIMARCHI, *Rischio e responsabilità oggettiva*, Giuffrè, 1961; S. RODOTÀ, *Il problema della responsabilità civile*, Giuffrè, 1964; M. COMPORTI, *Esposizione al pericolo e responsabilità civile*, Morano Editore, 1965; P. FORCHIELLI, *Intorno alla responsabilità senza colpa*, in *Riv. trim. dir. proc. civ.*, 1967, p. 1394; successivamente vd. anche, limitandoci ai testi pubblicati nel nostro Paese, G. ALPA, *Colpa e responsabilità oggettiva nella prospettiva dell'analisi economica del diritto*, in *Politica del diritto*, 1976,

Anche in questo caso, l'elaborazione di nuovi modelli di distribuzione dei rischi e allocazione dei costi è da porre in relazione all'evolversi delle strutture economico-sociali dominanti, nonché al definitivo sorgere del *welfare state* sulle ceneri dell'ottocentesco stato liberale. Da un lato, l'affermarsi della produzione e del consumo di massa determina la fuoriuscita del fenomeno del "danno" dal ristretto recinto della fabbrica – a cui la prima rivoluzione industriale lo aveva relegato – giungendo a coinvolgere, con la diffusione su larga scala di beni di consumo, tutte le fasce della popolazione. L'emanazione di leggi che rendevano obbligatoria – a carico dell'impresa – l'assicurazione per gli infortuni sul lavoro aveva fatto venire meno le esigenze di tutela che, in un primo momento, erano state all'origine delle teorie oggettivistiche del rischio d'impresa; fu invece il fenomeno della *vittimizzazione di massa* dovuta ai danni da prodotto (e alla circolazione stradale) a fornire nuova linfa vitale alla responsabilità oggettiva¹⁶.

Dall'altro lato, l'emergere della prospettiva "solidaristica" e "interventista" dello Stato sociale determinò un superamento della concezione individualistica della responsabilità per colpa, che rispondeva all'unica preoccupazione di individuare il responsabile del danno e di stabilire a quali condizioni questi fosse obbligato a risarcire¹⁷. Coerentemente con il nuovo ruolo "regolatore" dello Stato, a partire dal secondo Dopoguerra, il sistema della responsabilità civile – su impulso della dottrina, in Italia, e della giurisprudenza, in paesi come la Francia, la Germania e gli Stati Uniti¹⁸ – parve farsi carico di un ulteriore obiettivo, ossia quello di istituire criteri di ripartizione dei rischi che consentissero di assicurare la più ampia tutela ai danneggiati e di distribuire le perdite in maniera razionale ed economica¹⁹. Il problema di allocazione dei costi derivanti dal

p. 431 ss.; M. CONFORTINI, *Posizioni della dottrina e della giurisprudenza italiana sul tema della responsabilità civile del produttore per la messa in commercio di prodotti (difettosi e) dannosi*, in *Resp. civ.*, 1977, p. 544; C. CASTRONOVO, *Sistema e problema del danno da prodotto*, Giuffrè, 1979. Per una ricostruzione delle fasi che hanno accompagnato lo storico passaggio dal dogma della colpa alla responsabilità oggettiva si rinvia a F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 114 ss.

¹⁶ F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 131; P. TRIMARCHI, *Istituzioni di diritto privato*, Giuffrè, XIX ed., 2011, p. 128.

¹⁷ G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 162, che cita una celebre affermazione di Oliver Wendell Holmes Jr.: «*civil liability, in its immediate working, is simply a redistribution of an existing loss between two individuals*», vd. O.W. HOLMES, *The Common Law*, Boston, 1881, p. 144-145.

¹⁸ F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 131.

¹⁹ G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 162-163.

danno, dunque, da problema *individuale*, diventa un problema di carattere *sociale*, e il *focus* della responsabilità aquiliana si sposta dalla figura del *danneggiante* all'istituto del *danno*.

Nel nostro Paese, la proposta di superare la lettura tradizionalmente monolitica del sistema di responsabilità civile ebbe origine, come abbiamo già accennato, nelle riflessioni dottrinali, che cominciarono a prendere atto di come già nel sistema codicistico fossero presenti modelli di ascrizione della responsabilità non colposi: è il caso della responsabilità per danno causato da dipendenti (art. 2049 c.c.), da cose in custodia (2051 c.c.), da animali (art. 2052 c.c.), dalla rovina di edifici (art. 2053 c.c.), da autoveicoli (art. 2054 c.c.), nonché del danno cagionato dallo svolgimento di attività pericolose (art. 2050 c.c.).

Nell'applicazione concreta di tali norme, la giurisprudenza già faceva ricorso, in maniera piuttosto disinvolta, a presunzioni di colpa di fatto invincibili²⁰; in via di principio, tuttavia, i giudici continuavano ad affermare la persistente validità del paradigma colpacentrico. Tuttavia, se, ad esempio, nel laboratorio artigiano pre-industriale, in cui il titolare lavorava fianco a fianco con i propri dipendenti e apprendisti, era verosimile che un fatto illecito di questi ultimi facesse presumere una *culpa in vigilando* o *in eligendo* a carico dell'imprenditore, nella grande industria moderna – dove i lavoratori sono numerosissimi e operano a distanza dal loro datore, secondo un principio di divisione del lavoro – una simile presunzione appare come una mera finzione²¹.

Il compito di “strappare il velo” del modello colpacentrico fu così assunto dalla dottrina, che individuò negli artt. 2049-2054 c.c. delle ipotesi di responsabilità senza colpa, il cui criterio unificatore veniva identificato, di volta in volta, nel rischio di impresa²², nella creazione del pericolo²³ o, successivamente, con l'emergere dell'analisi economica del diritto, nel criterio del *cheapest cost avoider*²⁴.

²⁰ F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 133.

²¹ P. TRIMARCHI, *Istituzioni di diritto privato*, cit., p. 128.

²² P. TRIMARCHI, *Rischio e responsabilità oggettiva*, cit.

²³ M. COMPORTI, *Esposizione al pericolo e responsabilità civile*, cit.

²⁴ G. CALABRESI – J.T. HIRSCHOFF, *Toward a Test for Strict Liability in Torts*, in 81 *Yale Law Journal*, 1972, p. 1060 ss.; G. CALABRESI., *The Costs of Accidents. A Legal and Economic Analysis*, Yale University Press, 1970, trad. it. *Costo degli incidenti e responsabilità civile. Analisi economico-giuridica*, Giuffrè, 1975. Secondo l'A., in realtà, il principio del *cheapest cost avoider* non è limitato al campo della responsabilità oggettiva, ma coinvolge anche quello della responsabilità colposa; esso, infatti, ha proprio l'obiettivo di indirizzare il legislatore nella scelta di quale sia il modello di responsabilità (colposa o

Questo nuovo approccio determina una recisione netta con l'ancestrale funzione punitiva, che prima di allora, sebbene implicitamente, veniva ancora attribuita alla responsabilità aquiliana²⁵. Compito di quest'ultima, infatti, non è la "scoperta" del "vero" autore del fatto dannoso e la sua punizione²⁶ – funzione propria del diritto penale, che vi assolve (o dovrebbe farlo²⁷) con le garanzie e le tutele che gli sono proprie. Piuttosto, il compito della responsabilità civile è quello di individuare i *criteri giuridici* per l'attribuzione del danno, determinando *artificialmente* le condizioni in presenza delle quali il costo di un danno subito da un soggetto debba essere trasferito ad un altro soggetto²⁸. A differenza di quanto accade nel diritto penale – nel quale, per espressa previsione costituzionale (art. 27, co. 1.), la responsabilità è personale –, nella

oggettiva) maggiormente idoneo (ossia, più efficiente) a regolare i diversi settori economici. In particolare, in base a tale criterio, il costo del danno deve essere posto in capo al soggetto che, tenuto conto delle informazioni di cui disponeva al momento in cui agiva, si trovava nella posizione migliore per valutare, attraverso un'analisi costi-benefici, la scelta più conveniente tra prevenire il danno e risarcirlo.

²⁵ La funzione "punitiva" della responsabilità civile, in realtà, ha trovato una rinnovata *âge d'or* negli Stati Uniti, con l'incremento delle condanne al pagamento di *punitive damages*, sanzioni di diritto civile con funzione *deterrente* e *retributiva*, applicate dai giudici in aggiunta ai *compensatory damages* in casi di condotte dolose (o gravemente colpose) che abbiano arrecato danni di particolare entità; sul tema vd. L.L. SCHLUETER, *Punitive Damages*, VII ed., vol. I, LexisNexis, 2015; D.G. OWEN, *A Punitive Damages Overview: Functions, Problems and Reform*, in 39 *Vill. L. Rev.*, 1994, p. 363 ss. In Italia, c'è stata una prima apertura all'ammissibilità dei *punitive damages* con la sentenza delle Sezioni Unite Cass. civ. sez. un., 5 luglio 2017, n. 16601, che ha ammesso l'efficacia in Italia di una sentenza straniera recante la condanna al pagamento dei cosiddetti *punitive damages*, affermando che non si tratta di sanzioni contrarie all'ordine pubblico. Sulla sentenza vd. i commenti contenuti in *Danno resp.*, 2017, n. 4, di M. LA TORRE, *Un punto fermo sul problema dei "danni punitivi"*, p. 419; G. CORSI, *Le Sezioni Unite: via libera al riconoscimento di sentenze comminatorie di punitive damages*, p. 429; G. PONZANELLI, *Polifunzionalità tra diritto internazionale privato e diritto privato; Le Sezioni Unite e le funzioni della responsabilità civile*, p. 435; P.G. MONATERI, *Le Sezioni Unite e le funzioni della responsabilità civile*, p. 437; vd. anche C. CONSOLO, *Riconoscimento di sentenze, specie USA e di giurie popolari, aggiudicanti risarcimenti punitivi o comunque sopracompensativi, se in regola con il nostro principio di legalità (che postula tipicità e financo prevedibilità e non coincide pertanto con il, di norma presente, due process of law)*, in *Corr. giur.*, 2017, p. 1050.

²⁶ P. BARCELLONA, *Diritto privato e società moderna*, Jovene, 1996, p. 550: «Il problema della responsabilità civile non consiste nella scoperta del "vero" autore del fatto dannoso, bensì nella fissazione di un criterio grazie al quale si può sostituire l'attribuzione automatica del danno con un criterio giuridico».

²⁷ Sulla continua opera di flessibilizzazione delle categorie del diritto penale ad opera della giurisprudenza torneremo nel prossimo capitolo; vd. in part. § 4. ss.

²⁸ F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 131; P. BARCELLONA, *Diritto privato e società moderna*, cit., p. 150; G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 98. In realtà, a parere di alcuni autori, anche la responsabilità penale si sta, inopportuno, evolvendo in un diritto "artificiale" vd. F. SGUBBI, *Il reato come rischio sociale. Ricerche sulle scelte di allocazione dell'illegalità penale*, Il Mulino, 1990, passim e p. 90: «La responsabilità penale è distribuita socialmente, è allocata artificialmente secondo scelte di opportunità politiche».

responsabilità civile talvolta il soggetto tenuto a risarcire il danno viene individuato in una persona che non corrisponde all'autore materiale del fatto lesivo²⁹.

La responsabilità oggettiva, d'altra parte, non assolve soltanto ad una funzione economico-allocativa – secondo le finalità di *compensazione* e *deterrenza* tipiche della responsabilità civile – ma anche ad una funzione *distributiva* e *pragmatica*, nella misura in cui consente al danneggiato di ottenere il risarcimento anche nei casi in cui sarebbe di fatto *impossibile* dimostrare la colpa del produttore, poiché le sue tracce sarebbero da scovare tra gli impenetrabili meccanismi produttivi e decisionali delle grandi imprese³⁰.

1.2. Ritorno al futuro: un declino del modello oggettivistico? Regolazione dell'i.a. e policy europea.

Le circostanze che hanno determinato la nascita della responsabilità oggettiva per il rischio d'impresa ricordano, *mutatis mutandis*, l'attuale dibattito sull'allocazione dei costi derivanti dai danni algoritmici. Come allora, anche oggi emerge l'esigenza di contemperare innovazione industriale e sicurezza dei consumatori³¹ (da tutelare, dal punto di vista civilistico, sia *ex ante*, attraverso la normativa sulla *product safety*, sia *ex post*, attraverso le norme sulla responsabilità aquiliana) – e, come allora, il rischio da scongiurare è che si creino delle sacche di immunità giustificate dall'obiettivo di incentivare il progresso tecnologico e di evitare *chilling effects*³².

Il modello della responsabilità oggettiva, tuttavia, oggi non pare in ottima forma. La disciplina della responsabilità per danno da prodotto – che a lungo ha costituito la forma più compiuta di responsabilità oggettiva – ha subito, in tutti gli ordinamenti, un

²⁹ U. CARNEVALI, *Appunti di diritto privato*, Raffaello Cortina Editore, IX ed., 2010, p. 247.

³⁰ F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 94; ID., *Valori e funzioni della responsabilità civile nell'epoca del post-turbocapitalismo*, in *Danno e resp.*, 2021, n. 2, p. 147.

³¹ Le due istanze emergono in maniera plastica nella Relazione di accompagnamento all'AI Act (p. 3), che, tra gli obiettivi della normativa indica sia quello di «assicurare che i sistemi di IA immessi sul mercato dell'Unione e utilizzati siano sicuri e rispettino la normativa vigente in materia di diritti fondamentali e i valori dell'Unione», sia quello di «assicurare la certezza del diritto per facilitare gli investimenti e l'innovazione nell'intelligenza artificiale».

³² Per un simile paragone tra i primordi dell'attività industriale e l'attuale contesto produttivo vd. A. DAVOLA – R. PARDOLESI, *In viaggio col robot: verso nuovi orizzonti della r.c. auto ("driverless")?*, in *Danno e resp.*, 2017, n. 5, p. 626.

avvicinamento al paradigma colposo: in Europa, a causa dell'incerta delimitazione della nozione di “difetto”, nonché di un'interpretazione particolarmente estensiva dell'esimente per il rischio da sviluppo, in base alla quale la responsabilità del produttore è esclusa se «lo stato delle conoscenze scientifiche e tecniche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso» (art. 7, lett. e), 85/374/CEE³³ vd. *infra*, § 3.3.1.)³⁴; negli Stati Uniti, in forza dalla pubblicazione, nel 1998, ad opera dell'*American Law Institute* (ALI)³⁵, del *Restatement (Third) of Torts*³⁶, che, per i *design defects*, introduceva il *reasonable alternative design test*, criterio di fatto basato sulla colpa³⁷.

Alcuni autorevoli voci della dottrina, inoltre, hanno osservato come il declino del paradigma oggettivistico, nel nostro Paese, si osservi anche con riferimento alle norme codicistiche speciali in materia di responsabilità aquiliana – sempre di più, in una sorta di ritorno alle origini, identificate come ipotesi di *colpa presunta*³⁸.

Nello specifico settore dei danni cagionati da sistemi di i.a., d'altra parte, le proposte, avanzate dalla dottrina³⁹, di introdurre regimi di responsabilità oggettiva non

³³ Direttiva 85/374/CEE del Consiglio del 25 luglio 1985 relativa al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi.

³⁴ Nella giurisprudenza della Corte di Cassazione è ormai invalsa l'affermazione in base alla quale la responsabilità per il danno da prodotto difettoso avrebbe natura presunta e non oggettiva, vd. Cass. civ., sez. III, 28 settembre 2018, n. 23447; Cass. civ., sez. III, 19 febbraio 2016, n. 3258; Cass., sez. III, 29 maggio 2013, n. 13458.

³⁵ L'*American Law Institute* è un organismo privato formato da avvocati e giudici che si pone l'obiettivo di armonizzare le leggi dei vari Stati degli USA. Sul ruolo dell'istituto nella trasformazione del diritto statunitense si vd. A. ELSON, *The Case for an In-Depth Study of the American Law Institute*, in *Law & Social Inquiry*, 1998, vol. 23, no. 3, p. 625 ss.

³⁶ *Restatement (Third) of Torts: Prod. Liab.* § 2(b) (Am. Law Inst. 1998).

³⁷ G.L. PRIEST, *La controrivoluzione nel diritto della responsabilità da prodotti negli Stati Uniti d'America*, in *Foro it.*, 1989, IV, p. 119; D.G. GIFFORD, *Technological Triggers to Tort Revolutions: Steam Locomotives, Autonomous Vehicles, and Accident Compensation*, 2018, 11 *Journal of Tort Law*, p. 119 ss.

³⁸ F.P. PATTI, *Il declino della responsabilità oggettiva (a margine dell'art. 2051 c.c.)*, in *Riv. dir. civ.*, 2019, n. 4, p. 977 ss.; V. ROPPO, *La responsabilità oggettiva*, in P. Sirena (a cura di), *La funzione deterrente della responsabilità civile alla luce delle riforme straniere e dei Principles of European Tort Law*, Giuffrè, 2011, p. 185 ss.; ID., *La responsabilità civile di Pietro Trimarchi*, in *Jus Civile*, 2017, n. 6, p. 696 ss.

³⁹ C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, in *Journal of European Tort Law*, 2011, vol. 11, iss. 2, p. 150 ss.; H. ZECH, *Liability for AI: public policy considerations*, in *ERA Forum*, 2021, n. 22, p. 147 ss.; S. LI – M. FAURE – K. HAVU, *Liability Rules for AI-Related Harm: Law and Economics Lessons for a European Approach*, in *European Journal of Risk Regulation*, 2022, p. 8 ss.; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability – Adapting Liability Rules to the Digital Age and Artificial Intelligence*, in *Journal of European Tort Law*, 2021, vol. 13, iss. 1.

sono state accolte. Lo stesso Parlamento europeo, nella sua Risoluzione del 20 ottobre 2020, all'art. 4, proponeva l'introduzione di un regime di responsabilità oggettiva – assistito da un obbligo di assicurazione – per gli *operatori di sistemi di i.a. ad alto rischio*, escludendo che gli operatori potessero addurre, come prova liberatoria, la propria diligenza, il comportamento autonomo del sistema di i.a. o la forza maggiore⁴⁰. Tale proposta, tuttavia, non è confluita nella bozza di direttiva sulla responsabilità da intelligenza artificiale⁴¹, che – pur prendendo atto di come le consultazioni pubbliche evidenziassero una preferenza per tale soluzione tra i rispondenti (ad eccezione delle imprese)⁴² – opta per l'introduzione di una *presunzione del nesso di causalità in caso di colpa* (art. 4) e della possibilità, per l'autorità giudiziaria, di *ordinare la divulgazione degli elementi di prova*, qualora emergano indici di fondatezza della domanda risarcitoria (art. 3). La principale motivazione addotta per escludere l'ingresso della responsabilità oggettiva nel settore dei danni algoritmici consiste nel fatto che «il tipo di prodotti e servizi dotati di sistemi di IA che potrebbero incidere sul pubblico in generale e mettere a repentaglio importanti diritti giuridici, come il diritto alla vita, alla salute e alla proprietà, e che potrebbero pertanto essere soggetti a un regime di responsabilità oggettiva, non sono ancora ampiamente disponibili sul mercato»: una motivazione che non convince del tutto, se pensiamo ad esempio ai danni all'integrità fisica che possono essere cagionati, già oggi, dai sistemi di guida autonoma (vd. *supra*, cap. I, § 7. ss.).

Contemporaneamente alla proposta di direttiva sulla responsabilità extracontrattuale da intelligenza artificiale, la Commissione Europea ha altresì pubblicato una bozza di revisione della direttiva sulla responsabilità per danno da prodotti difettosi, finalizzata ad adeguare la disciplina vigente «all'era digitale, all'economia circolare e

⁴⁰ Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, 2020/2014(INL), art. 4, § 3: «*Gli operatori di sistemi di IA ad alto rischio non possono eludere la propria responsabilità sostenendo di avere agito con la dovuta diligenza o che il danno o il pregiudizio sia stato cagionato da un'attività, dispositivo o processo autonomo guidato dal loro sistema di IA. Gli operatori non sono considerati responsabili se il danno o il pregiudizio è dovuto a cause di forza maggiore*». La Risoluzione prevedeva inoltre, per gli operatori di sistemi di i.a. non qualificati come ad alto rischio, un regime di responsabilità per colpa, ma con un onere della prova agevolato per l'attore.

⁴¹ Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, cit.

⁴² Relazione di accompagnamento alla Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, cit., p. 8-9 e 16.

all'impatto delle catene globali del valore»⁴³. Oltre al (dovuto) allineamento delle nozioni di *prodotto* e *produttore* al nuovo contesto industriale, la proposta ridefinisce le circostanze che devono essere valutate al fine di verificare la difettosità del prodotto, includendovi l'eventualità che il prodotto abbia la «*capacità di continuare ad imparare dopo la sua diffusione*» (art. 10, lett. c) della bozza): una disposizione che di certo non aiuta l'interprete nel determinare *se* e in *quali casi* la capacità di *machine learning* possa essere considerata un difetto. D'altra parte, la proposta, avanzata da dottrina⁴⁴ e *stakeholders*⁴⁵, di sopprimere l'esimente del rischio da sviluppo – punto di accesso, come abbiamo già evidenziato, di valutazioni inerenti la *diligenza del produttore* – non è stata accolta dalla Commissione Europea, che ha mantenuto la causa di esclusione della responsabilità nel nuovo articolato.

Una responsabilità oggettiva *pura*, dunque, ad oggi non pare essere la soluzione più accreditata, a livello europeo, per disciplinare le ipotesi di danno da sistema di i.a. – sebbene la Commissione si riservi, nell'ambito della revisione periodica della direttiva sulla responsabilità extracontrattuale da intelligenza artificiale, di valutare l'opportunità, sulla base dei dati sugli incidenti, di introdurre un regime di responsabilità oggettiva e/o una copertura assicurativa obbligatoria⁴⁶.

⁴³ Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi, COM (2022) 495 final, 28 settembre 2022.

⁴⁴ F.P. PATTI, *The European Road to Autonomous Vehicles*, in 43 *Fordham International Law Journal*, 2019, p. 142 ss.; H. JACQUEMIN – J. HUBIN, *Aspects contractuels et de responsabilité civile en matière d'intelligence artificielle*, in H. Jacquemin – A. De Streel (eds.), *Intelligence artificielle et droit*, Larcier, 2017, p. 137; U. SALANITRO, *Intelligenza artificiale e responsabilità: la strategia della Commissione europea*, in *Riv. dir. civ.*, 2020, fasc. 6, p. 1246; vd. anche EUROPEAN COMMISSION – EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 42-44.

⁴⁵ BUREAU EUROPÉEN DES UNIONS DE CONSOMMATEURS, *PRODUCT LIABILITY 2.0. How to make EU rules fit for consumers in the digital age*, 7 maggio 2020, p. 20, disponibile a questo link: https://www.beuc.eu/sites/default/files/publications/beuc-x-2020-024_product_liability_position_paper.pdf.

⁴⁶ Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, cit., art. 5.

1.3. Adattabilità della normativa vigente o necessità di introdurre una disciplina *ad hoc*?

Nell'attuale fase di valutazione dell'impatto che i sistemi di i.a. possono avere sul sistema della responsabilità aquiliana, si fronteggiano due posizioni: da un lato, coloro che ritengono che un'interpretazione evolutiva delle norme codicistiche esistenti⁴⁷ – per loro natura flessibili e aperte ai mutamenti socio-tecnologici – sia adatta a regolare la responsabilità per danni cagionati da sistemi di i.a.; dall'altro, coloro, i c.d. *eccezionalisti*, che sostengono che gli agenti intelligenti abbiano caratteristiche (autonomia, interconnettività, opacità) così *disruptive* rispetto ai tradizionali artefatti tecnologici da rendere assolutamente obsolete le ordinarie categorie della responsabilità aquiliana⁴⁸.

Il dibattito in corso, come evidenziato da diversi studiosi⁴⁹, sembra riecheggiare quello che accompagnò la nascita e la diffusione del *cyberspace*, con la contrapposizione tra coloro che sostenevano la necessità di introdurre un nuovo campo di studi (la c.d. *cyberlaw*) – che si ponesse come obiettivo quello di affrontare le problematiche specifiche imposte dall'avvento di internet – e coloro che, al contrario, ritenevano che le norme esistenti fossero sufficienti a regolare il nuovo fenomeno. Celebre, a tal proposito, rimane la provocazione del giudice statunitense F.H. Easterbrook, che – chiamato a tenere una relazione all'inaugurazione di una conferenza sulla *cyberlaw* – sostenne che pensare di introdurre un corso di studi sul diritto del *cyberspace* non fosse dissimile dall'idea di organizzare un ciclo di lezioni sulla *Law of Horses*:

⁴⁷ Nei sistemi di *common law*, il riferimento sarà piuttosto alle dottrine tradizionali del *tort law*.

⁴⁸ G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, cit.; M. SCHERER, *Regulating Artificial Intelligence Systems*, cit., p. 354 ss.; J. TURNER, *Robot Rules. Regulating Artificial Intelligence*, Palgrave Macmillan Cham, 2019; in realtà, questa posizione va ben oltre la sola questione della responsabilità aquiliana per i danni da algoritmo, e – proponendo una presa d'atto del presunto superamento della dicotomia "uomo-macchina" – auspica un complessivo ripensamento delle tradizionali categorie "umano-centriche" del diritto, vd. ex multis R. CALO, *Robotics and the Lessons of Cyberlaw*, cit.

⁴⁹ J. TURNER, *Robot Rules*, cit., p. 40; A. SOLOW-NIEDERMAN, *Emerging Digital Technology and the "Law of the Horse"*, in *UCLA L. Rev.*, 19 febbraio 2019; I. GIUFFRIDA E AL., *A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law*, in 68 *Case W. Rsrv. L. Rev.*, 2018, po. 772; M. FIERENS E AL., *Setting the Scene: On AI Ethics and Regulation*, in J. De Bruyne – C. Vanleenhove (eds.), *Artificial Intelligence and the Law*, cit., p. 58.

«Lots of cases deal with sales of horses; others deal with people kicked by horses; still more deal with the licensing and racing of horses, or with the care veterinarians give to horses, or with prizes at horse shows. Any effort to collect these strands into a course on “The Law of the Horse” is doomed to be shallow and to miss unifying principles. Teaching 100 percent of the cases on people kicked by horses will not convey the law of torts very well. Far better for most students- better, even, for those who plan to go into the horse trade-to take courses in property, torts, commercial transactions, and the like, adding to the diet of horse cases a smattering of transactions in cucumbers, cats, coal, and cribs. Only by putting the law of the horse in the context of broader rules about commercial endeavors could one really understand the law about horses»⁵⁰.

Come accadde per il *cyberspace*, anche in relazione all’emergere di forme di intelligenza artificiale parte della dottrina ritiene che l’esistenza di lacune normative non debba essere sopravvalutata e che le norme esistenti abbiano una capacità di mediazione giuridica in relazione ai nuovi fenomeni tecnologici.

In particolare, restando nel campo della responsabilità extracontrattuale per danni cagionati da sistemi di i.a., una parte della dottrina sostiene che le esistenti norme codicistiche (e, con le necessarie, puntuali, modifiche, anche la normativa speciale sulla responsabilità per danno da prodotto) possano essere applicabili anche con riferimento alle ipotesi di danni derivanti da sistemi di i.a.⁵¹. Tale orientamento muove dalla constatazione che i codici dell’Europa continentale sono concepiti per essere strutturalmente aperti alle evoluzioni dei fenomeni sociali e, effettivamente, sono già stati in grado di adeguarsi, con modifiche normative minime, al mutare della società e delle tecniche produttive⁵². Nati sul modello del *Code Napoléon* – ma con radici nella

⁵⁰ F.H. EASTERBROOK, *Cyberspace and the Law of the Horse*, in *University of Chicago Legal Forum*, 1996, p. 207 ss.

⁵¹ Vd. *ex multis* U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 95 ss.; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 445 ss.; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell’Unione Europea*, cit., p. 1719; A. PROCIDA MIRABELLI DI LAURO, *Le intelligenze artificiali tra responsabilità civile e sicurezza sociale*, in P. Perlingieri e al. (a cura di), *Rapporti civilistici e intelligenze artificiali: attività e responsabilità*, Edizioni Scientifiche Italiane, p. 297 ss. In ambito di *common law*, per l’applicabilità dei tradizionali principi di *tort law*, vd. R.A. HEVERLY, *Are Existing Tort Theories Ready for AI? An American Perspective*, in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, 2022, p. 91 ss.

⁵² U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 100.

tradizione del diritto romano – i codici civili europei hanno governato fenomeni impensabili ad inizio Ottocento, quali la seconda rivoluzione industriale e la produzione di massa di beni di consumo del secondo Dopoguerra⁵³.

D'altra parte, da sempre il sistema della responsabilità aquiliana fa i conti con i danni cagionati da forme di intelligenza *naturale* non imputabili: quella animale (vd. art. 2051 c.c.), ma anche quella umana, ai tempi in cui gli schiavi erano considerati “cose”⁵⁴ e, ancora oggi, nei casi di danni causati da persone incapaci di intendere e di volere (art. 2047). Di fronte a comportamenti *intelligenti* e per loro natura imprevedibili – ma non ascrivibili giuridicamente alle persone (o agli animali) che li hanno posti in essere – le norme attribuiscono la *responsabilità per fatto altrui* ai soggetti che *beneficiano* di tali forme di intelligenza e ai soggetti che si trovano nella posizione migliore per *impedire* il verificarsi degli eventi lesivi.

Il paragone con i sistemi di i.a. è evidente. Anche in questo caso gli algoritmi possono essere all'origine di eventi dannosi, imprevedibili sia da parte dei produttori sia da parte degli utilizzatori; applicando analogicamente i principi ricavabili dal sistema aquiliano, la responsabilità potrebbe dunque essere attribuita, alternativamente o cumulativamente, ai soggetti che *beneficiano* delle funzionalità dei sistemi di i.a. e ai soggetti che si trovano nella posizione migliore per *impedire* il verificarsi degli eventi lesivi.

Tra le ragioni per le quali non parrebbe necessario “reinventare la ruota”⁵⁵ – ma sarebbe invece auspicabile affidarsi ad un'attività ermeneutica evolutiva della normativa esistente – vi è altresì la considerazione per la quale qualsiasi intervento normativo in una materia antica e consolidata come quella extracontrattuale rischierebbe di avere effetti

⁵³ U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 99. Nonostante il codice civile italiano sia relativamente giovane (1942), esso ricalca per la massima parte il Codice unitario del 1865 che lo precedeva, a sua volta ispirato dal Codice napoleonico.

⁵⁴ U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 99; F. GALGANO, *Dogmi e dogmatica del diritto*, Cedam, 2010, p. 64 ss.; sull'analogia tra schiavi e sistemi di i.a. vd. P. ASARO, *A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, in P. Lin e al. (eds.), *Robot Ethics*, cit., p. 178 ss.; S. BECK, *The problem of ascribing legal responsibility in the case of robotics*, in 31 *AI & Soc*, 2016, p. 478.

⁵⁵ U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 100, che cita a sua volta J. LEVY, *No Need to Reinvent the Wheel: Why Existing Liability Law Does Not Need to Be Preemptively Altered to Cope with the Debut of the Driverless Cars*, in *J. Bus Entrepreneurship & L.*, 2016, 9, p. 355.

sistematici (e, in una certa misura, imprevedibili) sull'equilibrio complessivo della disciplina⁵⁶.

Dopo un primo momento in cui le istituzioni europee sembravano essersi pronunciate sull'inadeguatezza dei tradizionali meccanismi di imputazione della responsabilità aquiliana – giungendo addirittura, nel 2017, a prefigurare l'introduzione di una personalità elettronica per i sistemi di i.a.⁵⁷ (vd. *infra*, § 5, in questo cap.) – più recentemente Parlamento e Commissione Europea hanno optato per un approccio più sfumato. Già nella Risoluzione del 20 ottobre 2020, il Parlamento Europeo si dichiarava «convinto che non sia necessaria una revisione completa dei regimi di responsabilità correttamente funzionanti», pur ritenendo opportuni «adeguamenti specifici e coordinati dei regimi di responsabilità per evitare una situazione in cui le persone che subiscono pregiudizi o danni al patrimonio non ottengano un risarcimento»⁵⁸. Le recenti proposte della Commissione Europea – quella sulla responsabilità extracontrattuale da intelligenza artificiale e quella sulla responsabilità per danno da prodotto difettoso – sembrano confermare tale impostazione, limitandosi ad adeguare gli strumenti già esistenti alla nuova realtà tecnologica e produttiva.

Accanto a tale orientamento, vi è invece una parte della dottrina che ritiene che le peculiarità dei sistemi di i.a siano tali da scardinare i tradizionali meccanismi imputativi della responsabilità aquiliana, rendendo necessario un intervento legislativo di ampio respiro, che traduca la *disruption* costituita dai sistemi di i.a. autonomi in paradigmi legislativi *ad hoc*⁵⁹. Tra i modelli normativi di cui si auspica l'introduzione vi sono la personalità elettronica, la responsabilità oggettiva degli operatori dei sistemi di i.a. e la *Manufacturer Enterprise Liability* – istituti che, declinati in vario modo, ricorrono in molte delle trattazioni sul tema dei danni da intelligenza artificiale.

⁵⁶ A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1719.

⁵⁷ Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)), 2018/C 252/25.

⁵⁸ Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, cit., considerando 6. Vd. anche, nello stesso senso le *Proposte per una strategia italiana per l'intelligenza artificiale* del luglio 2020, elaborate dal Gruppo di esperti MISE.

⁵⁹ C.E.A. KARNOW, *The application of traditional tort law*, cit., p. 74; P. ASARO, *The liability problem for autonomous artificial agents*, in *AAAI Symposium on Ethical and Moral Considerations in Non-Human Agents*, 2016, p. 190 ss.; A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., *passim*.

2. La responsabilità oggettiva in capo all'utilizzatore del sistema di i.a.: i regimi speciali di responsabilità previsti dal codice civile.

Come abbiamo già evidenziato, la proposta di introduzione di un meccanismo di responsabilità oggettiva per gli operatori del sistema di i.a. – avanzata dalla Risoluzione del Parlamento europeo del 20 ottobre 2020 – non è stata accolta dalla Commissione Europea, che, nella bozza di direttiva sulla responsabilità da intelligenza artificiale⁶⁰, si è limitata a prevedere, in caso di danni cagionati da dispositivi intelligenti, una *presunzione del nesso di causalità in caso di colpa* (art. 4) e una possibilità, per l'autorità giudiziaria, di *ordinare la divulgazione degli elementi di prova*, nel caso in cui la domanda di risarcimento appaia fondata (art. 3). Qualora tale testo venisse approvato, dunque, il nostro Paese dovrebbe introdurre un regime di responsabilità colposa *speciale* – quanto al regime probatorio – rispetto a quello previsto dall'art. 2043 c.c., per i soli danni cagionati da sistemi di i.a.

Sebbene il tentativo di agevolare il danneggiato nella dimostrazione del nesso causale sia da salutare con favore, stanti gli ostacoli alla ricostruzione causale derivanti dalla c.d. *black box* (vd. *infra*, §§ 3.5. e 3.5.1.), il testo normativo non risolve il problema principale derivante dall'attribuzione della responsabilità aquiliana in questa materia – quello della prova dell'elemento soggettivo della colpa in capo all'operatore. Nella consapevolezza di tale difficoltà, la proposta, al considerando 4, si premura di fare salva la possibilità, per il ricorrente, di avvalersi dei regimi di responsabilità oggettiva più favorevoli eventualmente vigenti a livello nazionale.

Il riferimento immediato, per il nostro ordinamento, sono gli artt. 2049-2054 c.c., che, come noto, prevedono alcune specifiche deroghe al regime ordinario della responsabilità colposa; la loro applicazione al fenomeno dei danni da dispositivi intelligenti potrebbe effettivamente apparire molto più favorevole, per il danneggiato, rispetto al regime colposo di cui all'art. 2043 c.c. (anche nella sua versione *aggiornata*, che implementi le modifiche proposte dalla direttiva).

⁶⁰ Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, cit.

Volendo cimentarci nel difficile compito di trovare un criterio unificatore delle ipotesi codicistiche di responsabilità oggettiva, ci pare di poter rilevare, in via di prima approssimazione, che le disposizioni di cui agli artt. 2049-2054 c.c. delineano tutte ipotesi in cui vi è una persona che *ha creato* un rischio di evento dannoso, oppure che *beneficia* della creazione di un rischio o, ancora, che *si trova nella posizione migliore* per controllare un rischio. In realtà, l'ipotesi di cui all'art. 2050 c.c. (*esercizio di un'attività pericolosa*) pare attagliarsi meglio alla posizione del produttore; analizzeremo, dunque, il campo di applicazione di tale disposizione quando esamineremo l'eventuale responsabilità civile di quest'ultimo (vd. *infra*, § 3 ss.).

Benché, come abbiamo accennato (vd. *supra*, § 1.1., in questo cap.), le prime riflessioni organiche in materia di responsabilità oggettiva si siano sviluppate in relazione al rischio di impresa, non tutte le norme codicistiche che disciplinano la responsabilità senza colpa possono essere ricondotte all'attività industriale. Così, se l'art. 2049 (*responsabilità dei padroni e dei committenti*) e l'art. 2051 (*danno cagionato da cosa in custodia*) sono disposizioni che vengono tipicamente applicate all'attività produttiva, vi sono altre ipotesi di imputazione oggettiva che attengono alla sfera personale dei consociati: si pensi alla responsabilità del custode di animale (art. 2052) o a quella del proprietario di immobili (art. 2053) o di autoveicoli (art. 2054)⁶¹. Lo stesso art. 2051 c.c., d'altra parte, si presta ad applicazioni estranee all'esercizio dell'attività d'impresa (si pensi alla responsabilità del proprietario di un albero il cui ramo sia caduto su un pedone).

Così, ai sensi dell'art. 2052 c.c., il padrone di un cane sarà tenuto a risarcire i danni causati dal suo animale, anche qualora dimostri di aver adottato tutte le misure idonee ad evitare il danno. La soluzione opposta – lasciare il danno “là dove è caduto”, ossia, nel caso di specie, sulla persona che è stata morsa – sarebbe contraria al più basilare senso di giustizia: chi decide di beneficiare della compagnia di un cane deve anche sostenere gli eventuali costi derivanti da tale scelta, anche qualora non abbia colpe per il comportamento aggressivo del suo animale – è il principio del *cuius commoda, eius et incommoda esse debent*⁶².

⁶¹ Le considera invece ipotesi di colpa presunta, e non di responsabilità oggettiva, C.M. BIANCA, *Diritto civile: Vol. V – La responsabilità*, Giuffrè, 2020, p. 718.

⁶² U. CARNEVALI, *Appunti di diritto privato*, cit., p. 255-256.

Allo stesso modo, si potrebbe dire, chi decide di beneficiare dei più sofisticati algoritmi di intelligenza artificiale dovrà anche rispondere degli effetti collaterali che ne possono derivare. Se la *ratio* di ritenere responsabile oggettivamente l'operatore del sistema di i.a. parrebbe in linea con le disposizioni codicistiche in materia di imputazione oggettiva, in assenza di un intervento legislativo – che definisca il bilanciamento, tutto *politico*, tra istanze di tutela del danneggiato ed esigenze di produzione e diffusione dei sistemi tecnologici più evoluti – l'applicazione estensiva delle norme codicistiche dovrebbe comunque ispirarsi a criteri di prudenza.

Molti studiosi si sono già espressi sulla possibilità di applicare i regimi speciali di responsabilità in caso di danno cagionato da un sistema di i.a. Il riferimento ad una fattispecie in luogo dell'altra dipende qui dalla *concezione* che si ha dei sistemi di i.a., nonché dalla *ratio* delle norme codicistiche.

Così, se si considera il sistema di i.a. come una cosa inanimata, si potrà ipotizzare di fare ricorso alla norma sulla responsabilità da cosa in custodia (a cui si aggiunge, nello specifico campo dell'*automotive*, la responsabilità del proprietario per i danni cagionati dall'autoveicolo ai sensi dell'art. 2054 c.c., vd. *infra*, § 2.2.).

Se si ritiene che il sistema di i.a. possa essere considerato come una forma di intelligenza “minorata”, si potrà fare riferimento all'art. 2052 c.c.

Infine, qualora si consideri l'agente algoritmico alla stregua di una forma di intelligenza paragonabile a quella umana – quantomeno limitatamente alle specifiche funzioni che gli sono attribuite nell'ambito dell'esercizio di un'attività lavorativa – la disciplina di riferimento è quella della responsabilità vicaria (art. 2049 c.c.).

2.1. Responsabilità da cosa in custodia (art. 2051 c.c.) e responsabilità del custode dell'animale o di chi se ne serve (art. 2052 c.c.).

La soluzione più immediata è quella di sussumere le ipotesi di danno algoritmico sotto il regime di responsabilità oggettiva – estesa fino al limite del caso fortuito – prevista per il custode di cose o animali (rispettivamente, artt. 2051 e 2052 c.c.). Consideriamo congiuntamente le due disposizioni poiché – nonostante la differenza di formulazione, dovuta a ragioni storiche – esse prevedono un identico regime di responsabilità oggettiva,

fondata sulla medesima *ratio* del *cuius commoda, eius et incommoda esse debent*⁶³. L'unica differenza tra le due norme sta nell'identificazione del soggetto responsabile: l'art. 2051 c.c. fa riferimento alla figura del custode, mentre l'art. 2052 c.c. indica, come destinatari della norma, il "proprietario" dell'animale e colui che "se ne serve per il tempo in cui lo ha in uso".

In particolare, entrambe le disposizioni prevedono una responsabilità oggettiva per la persona che si trova nella posizione di esercitare un controllo sulla cosa o sull'animale, attribuendo altresì rilievo – esplicitamente, per quanto riguarda l'art. 2052 c.c., implicitamente, per l'art. 2051 c.c. – all'elemento dell'*interesse* alla custodia: non si applicherà, dunque, tale regime nei confronti di chi (ad es., il dipendente) detiene la cosa, l'animale (o il sistema di i.a.) nell'interesse altrui o nell'esercizio delle proprie incombenze⁶⁴. In questa prospettiva, si tratterebbe di una responsabilità che non è collegata ad una *vigilanza immediata* sulla cosa o sul bene, bensì al profitto e al controllo delle condizioni generali del rischio⁶⁵.

Il riferimento al caso fortuito quale prova liberatoria, nell'uno e nell'altro caso, potrebbe inoltre adattarsi perfettamente al nuovo contesto fenomenologico, poiché potrebbe esonerare da responsabilità l'utilizzatore di un sistema di i.a., nei casi in cui l'autoapprendimento abbia generato comportamenti assolutamente imprevedibili⁶⁶.

Il richiamo all'una o all'altra disposizione, come abbiamo già accennato, varia a seconda della concezione che si ha del sistema di i.a.

Coloro che considerano l'agente intelligente come una cosa inanimata – che, tuttavia, come tutti gli altri oggetti, può sottrarsi al controllo umano a causa di un guasto

⁶³ P. TRIMARCHI, *Istituzioni di diritto privato*, cit., p. 135.

⁶⁴ Principale sostenitore dell'assimilazione è Pietro Trimarchi, vd. ID., *Istituzioni di diritto privato*, cit., p. 135; ID., *La responsabilità civile: atti illeciti, rischio, danno*, Giuffrè, III ed., 2021, p. 364.

⁶⁵ P. TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, cit., p. 364.

⁶⁶ U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 100. Non si può non rilevare, tuttavia, come l'applicazione dell'esimente del "caso fortuito" – tanto in relazione all'art. 2051 c.c., quanto in relazione all'art. 2052 c.c. – sia, per giurisprudenza costante, particolarmente restrittiva, limitata a «ciò che non è prevedibile oggettivamente, ovvero tutto ciò che rappresenta un'eccezione alla normale sequenza causale, integra il caso fortuito, quale fattore estraneo alla sequenza originaria, avente idoneità causale assorbente e tale da interrompere il nesso con quella precedente, sovrapponendosi ad essa ed elidendone l'efficacia condizionante» (Cass. civ., sez. VI, 12 luglio 2022, n. 31106; vd. anche Cass. civ., sez. III, ord. 1 febbraio 2018, n. 2477). Sulla problematicità di tale orientamento, che riduce il caso fortuito all'assenza di un elemento costitutivo della fattispecie (il nesso causale), vd. F.P. PATTI, *Il declino della responsabilità oggettiva (a margine dell'art. 2051 c.c.)*, cit., p. 989.

o di cause accidentali⁶⁷ – ritengono applicabile l’art. 2051 c.c.⁶⁸. Così come il custode di materiale infiammabile risponde per i danni causati dall’eventuale incendio, o il titolare di un laboratorio è responsabile per i danni cagionati da una fuga di sostanze chimiche corrosive, allo stesso modo, colui che esercita un potere di fatto su un dispositivo intelligente dovrebbe rispondere, secondo questa prospettiva, dei danni derivanti dai comportamenti emergenti del sistema stesso.

Secondo l’interpretazione maggioritaria, infatti, l’art. 2051 c.c. si applica proprio nei casi in cui il danno scaturisca direttamente dal “dinamismo” della cosa, indipendentemente dall’azione dell’uomo⁶⁹. Il discrimine tra l’applicazione dell’art. 2043 c.c. (e 2050 c.c., in caso di attività pericolose) e il ricorso all’art. 2051 c.c. starebbe proprio nel diverso ruolo svolto dalla persona fisica nel verificarsi dell’evento lesivo: nel primo caso, la cosa è sottoposta alla direzione dell’uomo; nel secondo, la cosa «non è azionata direttamente da un operatore o, pur essendo azionata da questo, si sottra[e] al suo controllo»⁷⁰.

Al contrario, coloro che intravedono una qualche forma di *agency* – seppur “minorata” – in capo ai sistemi di i.a. ritengono maggiormente opportuna un’applicazione analogica dell’art. 2052 c.c.⁷¹. L’analogia sarebbe fondata su alcune presunte

⁶⁷ P. TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, cit., p. 367.

⁶⁸ E. PALMERINI, *Soggettività e agenti artificiali*, cit., p. 469; U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 97; *contra* l’equiparazione tra “cosa” e sistema di i.a. vd. G. SARTOR, *Gli agenti software e la disciplina giuridica degli strumenti cognitivi*, in *Dir. inf.*, 2003, pp. 55 ss.; M. COSTANZA, *Impresa robotizzata e responsabilità*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l’etica*, cit., p. 108 ss.

⁶⁹ P.G. MONATERI, *La responsabilità civile*, in R. Sacco (diretto da), *Tratt. dir. civ.*, Utet, 1998, p. 1038. In giurisprudenza vd. Cass., 22 maggio 2000, n. 6616, in *Mass. Foro it.*, 2000; Cass., 7 aprile 2010, n. 8229, in *Mass. Giust. civ.*, 2010. Recentemente sul campo applicativo dell’art. 2051 c.c. vd. in dottrina V. VILLANOVA, *La responsabilità per il danno cagionato da cose in custodia*, in *Resp. civ. prev.*, 2022, p. 760 ss.; D. SANTARPIA, *Un’indagine subiettiva sulla responsabilità da custodia*, in *Resp. civ. prev.*, 2018, fasc. 4, p. 1224C ss.; F.P. PATTI, *Il declino della responsabilità oggettiva (a margine dell’art. 2051 c.c.)*, cit., che riflette sull’incertezza applicativa della norma.

⁷⁰ P. TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, cit., p. 367-368.

⁷¹ Vd. C. BOSCARATO, *Who is Responsible for a Robot’s Actions? An initial examination of Italian Law within a European Perspective*, in B. Van den Berg – L. Klaming (eds.), *Technologies on the Stand: Legal and Ethical Questions in Neuroscience and Robotics*, Wolf Legal Pubns, 2011, p. 393; M. RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente alla luce dell’attuale scenario normativo*, in *Contr. impr.*, 2020, n. 3, p. 1181; G. TADDEI ELMI – F. ROMANO, *Il robot tra ius condendum e ius conditum*, in *Informatica e dir.*, 2016, n. 1, p. 126 ss.; *contra* all’analogia tra sistemi di i.a. e animali è A. BERTOLINI, *Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules*, in *5 Law Innovation & Tech.*, 2013, p. 227 ss., che osserva come gli animali – esseri irrazionali e imprevedibili – non possono essere paragonati ai sistemi di i.a., che seguono

caratteristiche comuni di animali e sistemi di i.a.: da un lato, l'imprevedibilità e l'incontrollabilità dei comportamenti, che possono tuttavia essere ridotte attraverso un adeguato *addestramento*; dall'altro, l'assenza di personalità giuridica, che non consente loro di essere titolari di un patrimonio.

Entrambe le prospettive, sia quella fondata sull'art. 2051 c.c., sia quella fondata sull'art. 2052 c.c., scontano tuttavia il limite di essere adeguate soltanto laddove il custode mantenga una qualche forma di governo sul comportamento dei sistemi di i.a. Qualora il futuro dovesse riservare una crescente autonomia degli agenti intelligenti rispetto al ruolo degli utilizzatori, l'allocatione dei costi derivanti dagli eventi lesivi agli utilizzatori rischierebbe di porsi in contrasto con la finalità preventiva della responsabilità civile, dal momento che questi si troverebbero nell'impossibilità di diminuire il rischio di danno⁷².

2.2 Responsabilità da circolazione di veicoli (art. 2054 c.c.).

Nello specifico settore delle auto a guida autonoma, il riferimento codicistico più immediato è l'art. 2054 c.c., che – per il caso di danno derivante da circolazione del veicolo – prevede:

(i) una responsabilità *aggravata* a carico del conducente, in base alla quale quest'ultimo è tenuto a risarcire il danno se non prova di aver fatto tutto il possibile per evitare il danno (co. 1 e 2)⁷³;

(ii) una responsabilità oggettiva a carico del proprietario, dell'usufruttuario e dell'acquirente con patto di riservato dominio, in solido con quella del conducente di cui ai commi 1 e 2, a meno che questi non provino che la circolazione del veicolo è avvenuta contro la loro volontà (co. 3);

pur sempre le istruzioni fornite in sede di programmazione; vd. anche l'accurata disanima di A. DRIGO, *L' "Evocativa" integrazione analogica dell'art. 2052 c.c. per i sistemi emergenti di intelligenza artificiale: un'analisi critica*, in *GiustiziaCivile.com*, 2020.

⁷² A DRIGO, *L' "Evocativa" integrazione analogica dell'art. 2052 c.c.*, cit., p. 14; H. ZECH, *Liability for AI: public policy considerations*, p. 154.

⁷³ Il comma 2 precisa che, nel caso di scontro tra veicoli, si presume, fino a prova contraria, che ciascuno dei conducenti abbia concorso ugualmente a produrre il danno. Il conducente dovrà quindi dimostrare la colpa dell'altro *driver* per superare la presunzione.

(ii) una responsabilità oggettiva a carico di tutte le persone sopra indicate, in caso di danni derivati da vizi di costruzione o da difetto di manutenzione del veicolo (co. 4).

Una responsabilità ai sensi dei co. 1 e 2 dell'art. 2054 c.c. potrebbe sicuramente ipotizzarsi con riguardo agli incidenti provocati da veicoli a guida parzialmente autonoma (livelli 1, 2 e 3), in relazione ai quali il conducente mantiene un obbligo di sorveglianza, dovendo attivarsi per riprendere il controllo del veicolo ogniquale volta sussista una situazione di emergenza o vi sia una richiesta in questo senso da parte del sistema⁷⁴. Per sottrarsi alla responsabilità, il conducente dovrà dimostrare di «aver fatto tutto il possibile per evitare un danno» (art. 2054, co. 1), non essendo invece sufficiente la prova di un funzionamento difettoso del veicolo (art. 2054, ult. co.)⁷⁵ – posto che, in quest'ultima evenienza, il conducente potrà rispondere in concorso con il produttore. In solido con il conducente, potrà poi essere chiamato a rispondere del danno il proprietario del veicolo, ai sensi dell'art. 2054, co. 3. Tale soluzione trova tra l'altro conferma nel Decreto *Smart Roads*⁷⁶, che, con riferimento alla sperimentazione su strada di veicoli a guida parzialmente autonoma, stabilisce che la domanda di autorizzazione deve contenere anche «l'indicazione del proprietario del veicolo a guida automatica, quale soggetto responsabile ai sensi dell'art. 196 del codice della strada e dell'art. 2054, comma 3, del codice civile, ovvero di altro soggetto obbligato in solido, ai sensi del citato art. 196». (art. 11, co. 1, lett. a).

Al contrario, in assenza di un soggetto qualificabile come “conducente”, i commi 1 e 2 dell'art. 2054 c.c. non paiono applicabili in relazione ai veicoli di livello 4 e 5⁷⁷.

⁷⁴ A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, in *Europa e dir. priv.*, 2019, fasc. 4, p. 1002; G. CALABRESI – E. AL MUREDEN, *Driverless Cars*, Il Mulino, 2021, p. 156; A. LEPORE, *I.A. e responsabilità civile. Robot, autoveicoli e obblighi di protezione*, in *Tecnologie e diritto*, 2021, p. 198; M. TAMPIERI, *L'intelligenza artificiale: una nuova sfida anche per le automobili*, in *Contr. Impr.*, 2020, n. 2, p. 748; S. PELLEGGATTA, *Automazione nel settore automotive: profili di responsabilità civile*, in *Contr. impr.*, 2019, n. 4, p. 1429-1431; F. BARTOLINI, *Auto a guida autonoma e problemi di responsabilità civile*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 306-307.

⁷⁵ A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit., p. 5.

⁷⁶ Decreto del Ministero delle Infrastrutture e dei Trasporti del 28 febbraio 2018, cit.

⁷⁷ A. PISANI TEDESCO, *Smart mobility e rischi satellitari e informatici: i possibili scenari di allocazione della responsabilità civile*, in *Dir. comm. internaz.*, 2019, fasc. 4, § 4; A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit., p. 1004; S. PELLEGGATTA, *Automazione nel settore automotive*, cit., p. 1433; F. BARTOLINI, *Auto a guida autonoma e problemi di*

Nonostante l'art. 2054 c.c., a differenza dell'art. 46 c.d.s., non faccia espressamente riferimento alla “guida dell'uomo”, sembra difficile che la persona trasportata in un veicolo che si muove in maniera totalmente autonoma possa essere considerato “conducente”⁷⁸. Il regime di cui all'art. 2054 c.c., co. 1 e 2, d'altra parte, è fondato sulla capacità di controllo del veicolo e sulla colpevolezza del conducente, benché presunta – una colpevolezza che di certo non pare riscontrabile ove l'utilizzatore non sia posto nella posizione di poter governare il veicolo⁷⁹.

Inoltre, anche sul versante dell'analisi economica del diritto, l'attribuzione di una responsabilità all'operatore di una *self-driving car* non parrebbe efficiente, poiché porrebbe il costo del rischio di un incidente in capo ad un soggetto privo degli incentivi (ossia, i mezzi) per ridurre la probabilità del verificarsi dello stesso⁸⁰. L'utilizzatore di un veicolo autonomo di livello 4 e 5, infatti, da un lato non è in grado di incidere sul movimento del veicolo e, dall'altro, non ha le competenze necessarie per migliorare il *software* preposto a regolare la guida automatizzata. Egli è totalmente estraneo alla dinamica della quale sarebbe chiamato a sopportare il rischio, così che la regola di responsabilità civile non può sortire nei suoi confronti l'effetto virtuoso di promuovere il costante miglioramento della tecnologia delle *driverless car*, al fine di raggiungere il più alto livello di sicurezza possibile⁸¹.

Stante l'inapplicabilità, per i veicoli di livello 4 e 5, del regime previsto dai commi 1 e 2 dell'art. 2054 c.c., non potrà nemmeno configurarsi una responsabilità del proprietario del veicolo ai sensi del comma 3 del medesimo articolo, che presuppone una responsabilità (per colpa presunta) del conducente. Maggiore spazio applicativo potrebbe invece residuare per la responsabilità oggettiva prevista dal co. 4 dell'art. 2054 c.c., per i vizi di costruzione del veicolo⁸², da coordinarsi con la disciplina sulla *product liability* (vd. *infra*, § 3 ss., in questo cap.).

responsabilità civile, cit., p. 307; contra A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 619, che ritengono applicabile tale norma, contestandone però l'efficienza sul piano della prevenzione.

⁷⁸ A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, p. 1004.

⁷⁹ A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 619; A. PISANI TEDESCO, *Smart mobility e rischi satellitari e informatici*, cit., § 4.

⁸⁰ A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 619.

⁸¹ A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 619.

⁸² M. TAMPIERI, *L'intelligenza artificiale: una nuova sfida anche per le automobili*, cit., p. 754-755.

2.3 La responsabilità di padroni e committenti (art. 2049 c.c.).

Tra le varie soluzioni proposte dalla dottrina per regolare il danno da algoritmo vi è anche quella di applicare la disciplina della responsabilità vicaria, in via analogica⁸³ o attraverso l'introduzione di un'espressa clausola di equivalenza tra dipendente e sistema di i.a.⁸⁴. Tale forma di imputazione – disciplinata, nel nostro ordinamento, dall'art. 2049 c.c. – prevede la responsabilità del datore di lavoro (nella terminologia della disposizione, *i padroni e i committenti*) per il danno illecito cagionato a terzi dai dipendenti (*domestici e commessi*), nell'esercizio delle incombenze alle quali sono adibiti.

L'analogia tra il rapporto di preposizione e il rapporto intercorrente tra utilizzatore e sistema di i.a. parrebbe giustificato dal fatto che, sempre più spesso, imprenditori e professionisti si affidano a sistemi di i.a. per espletare mansioni che in precedenza erano rimesse al lavoro umano. L'estensione della responsabilità vicaria ai danni da intelligenza artificiale avrebbe dunque l'effetto di impedire che – affidando funzioni agli agenti algoritmici – il datore possa sottrarsi alla responsabilità per danni cagionati da un'attività svolta a suo vantaggio⁸⁵.

Il fondamento della responsabilità oggettiva⁸⁶ del datore di lavoro viene normalmente individuato nel principio del *cuius commoda eius et incommoda*, in base al

⁸³ M. COSTANZA, *Impresa robotizzata e responsabilità*, cit., p. 112 ss.; E. KARNER, *Liability for robotics: current rules, challenges and the need*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 120; S. NYHOLM, *Attributing Agency to Automated Systems: Reflections on Human-Robot Collaborations and Responsibility-Loci*, in *24 Science Engineering Ethics*, 2018, p. 1201 ss.; S. CHOPRA – L.F. WHITE, *A Legal Theory for Autonomous Artificial Agents*, University of Michigan Press, 2011, p. 128 ss.; M. DIAMANTIS, *Vicarious Liability for AI*, Iowa Legal Studies Research Paper No. 2021-27; ID., *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, in *98 N.C. L. Rev.*, 2020, p. 926 ss.; ID., *Algorithms Acting Badly: a Solution From Corporate Law*, in *89 Geo. Wash. L. Rev.*, p. 818. Nel sistema di *common law* la responsabilità vicaria è conosciuta anche come *respondeat superior*.

⁸⁴ GRUPPO INDIPENDENTE DI ESPERTI AD ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, *Orientamenti etici per un'i.a. affidabile*, cit., p. 45 ss.; A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 87.

⁸⁵ C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, cit., p. 159; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 468; M. DIAMANTIS, *Algorithms Acting Badly*, cit., *passim*.

⁸⁶ La dottrina è univoca sul carattere oggettivo della responsabilità *ex art. 2049 c.c.*, vd. P. TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, cit., p. 314; M. FRANZONI, *L'illecito*, Giuffrè, 2004; G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 718.

quale chi trae vantaggio dal rapporto di preposizione deve anche sopportare i costi derivanti dell'eventuale danno da questi cagionato – una *ratio* che, evidentemente, potrebbe riscontrarsi anche in relazione ai danni algoritmici. D'altra parte, il datore di lavoro, nell'ipotesi di dipendente umano come in quella di “dipendente algoritmico”, si trova nella posizione più adatta per *controllare* lo svolgimento dell'attività lavorativa e per minimizzare il rischio di eventi lesivi; oltretutto, il datore potrebbe internalizzare il costo dei danni, attraverso il ricorso a strumenti assicurativi⁸⁷.

Se la *ratio* della responsabilità vicaria potrebbe effettivamente apparire compatibile con le esigenze di allocazione economica dei costi derivanti dai danni algoritmici⁸⁸, il maggiore ostacolo all'applicazione analogica dell'art. 2049 c.c. consiste nell'asimmetria tra la non imputabilità del sistema di i.a. e la responsabilità solidale del dipendente, prevista dalla disposizione codicistica⁸⁹.

3. Responsabilità per danno da prodotto difettoso: cenni introduttivi.

La disciplina della responsabilità per danno da prodotto sembra destinata ad essere la soluzione più immediata per affrontare le sfide poste dai danni causati da sistemi di i.a.⁹⁰. Con la progressiva attribuzione di autonomia ai dispositivi intelligenti – e con la corrispondente perdita di controllo sugli *smart products* da parte degli utilizzatori – si apre la strada ad una maggiore responsabilizzazione dei soggetti che immettono sul

⁸⁷ M. RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente*, cit., p. 1181.

⁸⁸ Contra E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 468-9, che ritiene che l'attività algoritmica non sia paragonabile a quella del dipendente, dal momento che alla prima non è riconducibile un raggio d'azione autonomo e indipendente rispetto alla rete organizzativa dell'attività d'impresa.

⁸⁹ M. RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente*, cit., p. 1181.

⁹⁰ Sull'applicabilità della disciplina europea sulla responsabilità per danno da prodotto difettoso ai sistemi di i.a. vd. *ex multis* i contributi pubblicati nel libro collettaneo S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit.; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit.; C. TWIGG-FLESNER, *Guiding Principles for Updating the Product Liability Directive for the Digital Age*, ELI Innovation Paper Series – European Law Institute, 2021; T.S. CABRAL, *Liability and artificial intelligence in the EU: Assessing the adequacy of the current Product Liability Directive*, in *Maastricht Journal of European and Comparative Law*, 2020, vol. 27(5), p. 615 ss.; A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit.; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 125.

mercato i prodotti e che si trovano nella posizione più adeguata (anche in forza di un'*expertise* tecnologica che spesso non ha pari nel settore pubblico, né, tantomeno, tra il pubblico dei consumatori) per valutare i rischi della commercializzazione ed incidere sulla sicurezza dei dispositivi intelligenti⁹¹.

Com'è noto, ciò che distingue la disciplina della responsabilità per danno da prodotto dall'ordinaria responsabilità aquiliana (art. 2043 c.c.) è il criterio di imputazione oggettivo: qualora ricorrano le circostanze descritte dalla normativa, al danneggiato – per ottenere il risarcimento della lesione subita – “basterà” dimostrare il difetto del prodotto, il danno, e il nesso di causalità tra difetto e danno, senza che sia necessario provare la colpa del produttore. Se il criterio della c.d. *strict liability* si era fatto apprezzare con riferimento alla generalità dei prodotti – stante l'asimmetria informativa che impedisce al danneggiato di conoscere se il produttore abbia assunto condotte di minimizzazione del rischio prima di immettere sul mercato il prodotto – esso diventa indispensabile in relazione ai danni cagionati dai sistemi di i.a., la cui capacità di autonomo *decision-making* rende complesso, se non impossibile, il riconoscimento di una colpa in capo al produttore per i comportamenti emergenti e lesivi posti in essere dallo *smart product*.

Come abbiamo già anticipato (vd. *supra*, § 1.3., in questo cap.), tuttavia, in realtà la disciplina del danno da prodotto difettoso – per come correntemente interpretata – non esclude un giudizio sulla *rimproverabilità*, in senso lato, del produttore. I confini della responsabilità sono infatti delimitati dalla nozione di “difetto”, che, specie per quanto riguarda il *design* del prodotto, introduce un elemento di valutazione della posizione soggettiva del convenuto – che, com'è evidente, rischia di annichilire i benefici derivanti dalla non necessità dell'accertamento della *colpa*. La stessa giurisprudenza della Corte di Cassazione, d'altra parte, è ormai uniforme nel ritenere che «*la responsabilità da prodotto difettoso ha natura presunta, e non oggettiva, poiché prescinde dall'accertamento della colpevolezza del produttore, ma non anche dalla dimostrazione dell'esistenza di un difetto del prodotto*»⁹². Riservandoci di affrontare più approfonditamente, in seguito (vd. *infra*, § 3.3. ss., in questo cap.), il problematico avvicinamento della disciplina sul danno da prodotto difettoso al criterio imputativo

⁹¹ G. WAGNER, *Robot liability*, cit., p. 38 ss.

⁹² Cass. civ., sez. III, 28 settembre 2018, n. 23447; Cass. civ., sez. III, 28 luglio 2015, n. 15851; Cass. civ., sez. III, 19 febbraio 2016, n. 3258; Cass., sez. III, 29 maggio 2013, n. 13458.

colposo, ci accingiamo, in questa sede, a fornire qualche cenno introduttivo alla normativa rilevante.

Innanzitutto, la disciplina speciale sulla responsabilità civile da prodotto è stata introdotta, a livello europeo, dalla dir. n. 85/374/CEE⁹³, recepita nel nostro ordinamento dal D.P.R. 24 maggio 1988, n. 224 e poi confluita nel Codice del consumo (d. lgs. 6 settembre 2005, n. 206), agli artt. 114-127.

L'art. 114 cod. cons. – che recepisce l'art. 1 della dir. 85/374/CEE – stabilisce che «*Il produttore è responsabile del danno cagionato da difetti del suo prodotto*», sempre che il difetto sia esistente al momento della messa in commercio del prodotto (vd. art. 118, lett. b), che esclude la responsabilità del produttore se «*il difetto che ha cagionato il danno non esisteva quando il produttore ha messo il prodotto in circolazione*»). La responsabilità del produttore è inoltre esclusa se non ha messo il prodotto in circolazione (art. 118, lett. a, cod. cons.), se non ha fabbricato il prodotto per la vendita o per qualsiasi altra forma di distribuzione a titolo oneroso, né lo ha fabbricato o distribuito nell'esercizio della sua attività professionale (art. 118, lett. c, cod. cons.), se – in caso di produttore o fornitore di una parte componente o di una materia prima – il difetto è interamente dovuto alla concezione del prodotto in cui è stata incorporata la parte o materia prima o alla conformità di questa alle istruzioni date dal produttore che la ha utilizzata (art. 118, lett. f, cod. cons.).

L'art. 120 cod. cons., recependo il disposto dell'art. 4 della Dir. 85/374/CEE, pone a carico del danneggiato l'onere di fornire la prova, oltre che del danno subito, dell'esistenza del difetto del prodotto e del nesso di causalità sussistente tra *difetto del prodotto* e *danno*: di fatto, una ripartizione dell'*onus probandi* che corrisponde a quello dell'ordinaria responsabilità aquiliana, ai sensi degli artt. 2043 e 2697 c.c.

La responsabilità del produttore può essere fatta valere sia dall'utilizzatore e acquirente del bene di consumo, sia dal terzo che venga danneggiato dal prodotto. Quest'ultima ipotesi, tuttavia, non è frequente nella prassi giurisprudenziale,

⁹³ Direttiva 85/374/CEE del Consiglio del 25 luglio 1985, cit. In generale sull'applicazione della Direttiva sul danno da prodotto nel panorama europeo vd. P. Machnikowski (ed.), *European Product Liability. An Analysis of the State of the Art in the Era of New Technologies*, Intersentia, 2016; D. WUYTS, *The product liability directive – More than two decades of defective products in Europe*, in *Journal of European Tort Law*, 2014, n. 5, p. 1; D. Fairgrieve (ed.), *Product Liability in Comparative Perspective*, Cambridge University Press, 2005; S. WHITTAKER, *Liability for Products. English Law, French Law, and European Harmonisation*, Oxford University Press, 2005.

probabilmente perché il terzo danneggiato dispone solitamente di altre modalità per ottenere compensazione. D'altra parte, la tendenza a considerare la responsabilità del produttore come rimedio a tutela del compratore/utilizzatore è evidente dalla stessa collocazione della disciplina nel codice del consumo, che riflette l'idea che la difettosità del prodotto sia da mettere in relazione a un rapporto di consumo.

Nei paragrafi che seguono, prenderemo in considerazione i principali elementi costitutivi della fattispecie del danno da prodotto difettoso (nozioni di “prodotto”, “produttore”, “difetto”, “nesso causale”), cercando di dare rilievo ai problemi specifici che essi pongono quando il prodotto è costituito da un agente intelligente. Già il Report finale sull'applicazione della direttiva sul danno da prodotto in ambito europeo (2018)⁹⁴ aveva evidenziato diverse falle nel meccanismo compensativo dei soggetti danneggiati, specie con riferimento ai nuovi prodotti ad alta tecnologia: da un lato, l'ingresso sul mercato della figura del programmatore di algoritmi – autonomi rispetto a qualsivoglia *hardware* o comunque spesso venduti separatamente rispetto ad essi – rende obsolete le definizioni di “prodotto” e “produttore” adottate dalla direttiva; dall'altra, la capacità di auto-apprendimento dei sistemi di i.a. e la loro tipica opacità scardinano il significato tradizionale attribuito ai concetti di “difetto” e “nesso di causalità”. Le carenze ed aporie della disciplina evidenziate in quella sede – e al centro delle riflessioni di dottrina e gruppi di studio istituiti a livello europeo⁹⁵ – sono state alla base della proposta di revisione della direttiva sul danno da prodotto, presentata nell'ottobre 2022, che ha l'obiettivo di agevolare il danneggiato nell'ottenimento del risarcimento, adeguando la disciplina vigente «*all'era digitale, all'economia circolare e all'impatto delle catene globali del valore*»⁹⁶.

Anticipiamo subito che – benché alcune delle modifiche contenute nella proposta di revisione rispondano senza dubbio ad esigenze di aggiornamento della direttiva – resta il problema di fondo, non affrontato direttamente dalla Commissione Europea, di definire

⁹⁴ EUROPEAN COMMISSION, Evaluation of Council Directive 85/374/EEC on the approximation of laws, regulations and administrative provisions of the Member States concerning liability for defective products, SWD(2018) 157 final, 7 maggio 2018, disponibile a questo link: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0157&from=EN>.

⁹⁵ Il riferimento è soprattutto a EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit.

⁹⁶ Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi, COM (2022) 495 final, 28 settembre 2022.

come incida la capacità di *machine learning* dei sistemi di i.a. sulla configurabilità di un “difetto” del prodotto. Non è chiaro, insomma, *se e quando l’auto-apprendimento possa essere considerato alla stregua di un “difetto”* e, in questo caso, se il produttore possa allegare, quale causa esimente, la circostanza che «lo stato oggettivo delle conoscenze scientifiche e tecniche al momento dell'immissione del prodotto sul mercato o della sua messa in servizio oppure durante il periodo in cui il prodotto era sotto il controllo del fabbricante non permetteva di scoprire l'esistenza del difetto» (c.d. esimente da rischio di sviluppo, art. 7, lett. e), nella nuova proposta trasposta invariata all’art. 10, lett. e).

Torneremo diffusamente altrove su questo aspetto (vd. *infra*, § 3.3.1., in questo cap.).

3.1 Sull’adeguatezza della nozione di “prodotto”.

Ai sensi dell’art. 115 cod. cons. – che recepisce l’art. 2 della direttiva n. 85/374/CEE – per “prodotto”, ai fini dell’attribuzione della responsabilità⁹⁷, s’intende «ogni bene mobile, anche se incorporato in altro bene mobile o immobile» (co. 1), inclusa l’elettricità (co. 2).

Sebbene la direttiva non sia espressamente chiara sul punto, fin dalla messa in commercio dei primi prodotti dotati di *software* (si pensi, ad esempio, agli elettrodomestici), la dottrina maggioritaria⁹⁸ ha ritenuto che il codice immateriale dovesse essere considerato come *componente* del prodotto, quando esso sia un elemento

⁹⁷ La precisazione è d’obbligo, poiché la definizione fornita all’art. 115 cod. cons. si applica al solo titolo II della parte IV del Codice del consumo ; la nozione generale di “prodotto” è invece rinvenibile all’art. 3, co., lett. e) cod. cons., che così lo definisce: «qualsiasi prodotto destinato al consumatore, anche nel quadro di una prestazione di servizi, o suscettibile, in condizioni ragionevolmente prevedibili, di essere utilizzato dal consumatore, anche se non a lui destinato, fornito o reso disponibile a titolo oneroso o gratuito nell’ambito di un’attività commerciale, indipendentemente dal fatto che sia nuovo, usato o rimesso a nuovo; tale definizione non si applica ai prodotti usati, forniti come pezzi d’antiquariato, o come prodotti da riparare o da rimettere a nuovo prima dell’utilizzazione, purché il fornitore ne informi per iscritto la persona cui fornisce il prodotto».

⁹⁸ A. ZACCARIA, *La responsabilità del “produttore” di software*, in *Contr. impr.*, 1993, p. 302 e, più recentemente, D. WUYTS, *The product liability directive*, cit., 2014, n. 5, p. 6; D. FAIRGRIEVE E AL., *Product liability directive*, in P. Machnikowski (a cura di), *European Product Liability*, cit., p. 46-47; A. PENZA, *La responsabilità per danno da prodotti difettosi*, in P.G. Demarchi (a cura di), *I diritti del consumatore e la nuova class action*, Zanichelli, 2010, p. 218; D. RESTUCCIA, *La tutela del consumatore*, in B. Inzitari (a cura di), *Valutazione del danno e rimedi risarcitori*, Giappichelli, 2016, p. 521.

inseparabile rispetto all'*hardware* e necessario al suo funzionamento: in base a tale orientamento, i difetti del *software* sono tali da rendere difettoso l'intero prodotto⁹⁹. La stessa Commissione Europea, nel 1988, ha confermato che «*the Directive applies to software in the same way [...] that is applies to handcraft and artistic products*»¹⁰⁰.

In base a tale prospettiva, dunque, in linea di principio non vi sarebbero motivi per escludere dalla nozione di “prodotto” rilevante ai sensi dell'art. 2 dir., gli artefatti che incorporino sistemi di intelligenza artificiale inseparabili rispetto all'*hardware*. Il problema si pone piuttosto con riferimento ai *software non embodied*, ossia a quegli algoritmi – dotati o meno di i.a. – che possono essere venduti anche separatamente rispetto al prodotto e che sono direttamente scaricabili *online*, sia che essi siano destinati ad essere comunque incorporati in un *hardware*, sia che essi siano invece progettati per operare su un sistema operativo virtuale. In entrambi i casi, tali *software* possono essere difettosi e determinare un'offesa all'incolumità personale o al patrimonio; non è chiaro, tuttavia, se, sulla base della normativa vigente, essi possano essere qualificati come *prodotti o componenti del prodotto*.

L'evoluzione tecnologica degli ultimi anni ha determinato una crescente *fluidità* dei sistemi operativi, e dunque dei prodotti che li incorporano: da un lato, il consumatore può decidere di installare su un *hardware* un *software* sviluppato da una società diversa da quella che ha messo in commercio il prodotto o addirittura un *software open source*; dall'altro, il sistema operativo può essere oggetto di frequenti *updates*, che incidono sul concreto funzionamento del prodotto. In entrambi i casi, la conseguenza è che il prodotto che causa il danno può aver subito delle *sostanziali modifiche* rispetto alla progettazione iniziale – circostanza che esclude la responsabilità del produttore (vd. art. 118, lett. b), cod. cons., che prevede un'esimente nel caso in cui «*il difetto che ha cagionato il danno non esisteva quando il produttore ha messo il prodotto in circolazione*»).

Così, ad esempio, comprata una stampante 3D, un consumatore può decidere di acquistare *online* il *software* per il suo funzionamento da una società diversa da quella

⁹⁹ T.G. WOLPERT, *Product Liability and Software Implicated in Personal Injury*, in 60 *Def. Couns. J.*, 1993, p. 521 ss.

¹⁰⁰ Risposta fornita da Lord Cockfield, per conto della Commissione, alla domanda scritta No 706/88 presentata da Mr Gijs de Vries (LDR-NL) il 5 luglio 1988, pubblicata sull'*Official Journal of the European Communities*, C 114, vol. 32, 8 maggio 1989, p. 42, disponibile a questo link: <https://eur-lex.europa.eu/legal-content/EN/TEXT/PDF/?uri=OJ:C:1989:114:FULL&from=IT>.

che ha prodotto la stampante, eventualmente sostituendo questo al sistema operativo installato di *default*; o, ancora, può mantenere il *software* originario, ma aggiornarlo successivamente. Il *software* e l'aggiornamento, in questi casi, di certo non possono considerarsi come *prodotti* a sé stanti, non essendo qualificabili come *beni mobili* ai sensi della direttiva; d'altra parte, essi non parrebbero costituire nemmeno una *componente del prodotto*, poiché – accedendo al bene fisico *dopo* la messa in commercio del prodotto – la società produttrice non ha alcun controllo sul loro funzionamento e di certo, dunque, non potrà essere chiamata a risponderne.

Di conseguenza, a legislazione vigente, qualora si verifichi un danno a causa di un *software* difettoso che non era originariamente installato nel prodotto, il danneggiato non potrà ottenere ristoro attraverso il regime della responsabilità per danno da prodotto.

A *fortiori*, non sembrerebbero rientrare nella definizione di cui all'art. 2 dir. nemmeno i c.d. “prodotti digitali”, ossia i *software* che agiscono in ambienti esclusivamente virtuali e che non sono necessari al funzionamento di un *hardware*: si pensi alle *app* installate su *tablet* o *smartphone*, il cui malfunzionamento può sicuramente determinare danni all'incolumità personale o al patrimonio (es. la carente cybersicurezza in un'app di pagamenti come *Satispay* o *Paypal* può cagionare perdite patrimoniali).

Tale caratteristica separazione del *software* di i.a. rispetto al supporto fisico ha indotto una parte della dottrina a sostenere che, talvolta, il primo possa più correttamente qualificarsi come *servizio*, piuttosto che come *componente* di un prodotto¹⁰¹. Il servizio è definito dal diritto europeo in senso negativo: ai sensi dell'art. 57 TFUE, per servizi si intendono «le prestazioni fornite normalmente dietro retribuzione, in quanto non siano regolate dalle disposizioni relative alla libera circolazione delle merci, dei capitali e delle persone». A differenza dei prodotti, che possono essere facilmente divisi in classi, alla categoria dei servizi sono riconducibili attività estremamente eterogenee, e dunque difficilmente classificabili: sono infatti qualificati come servizi le attività professionali e

¹⁰¹ Sul punto vd. EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 28; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1726; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 462; K.A. CHAGAL-FEFERKORN, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, in 30 *Stanford Law & Policy Review*, 2019, p. 84; T.S. CABRAL, *Liability and artificial intelligence in the EU*, cit., p. 620; A. BERTOLINI, *Robots as Products*, cit., p. 214 e ss.; A. FUSARO, *Quale modello di responsabilità per la robotica avanzata? Riflessioni a margine del percorso*, in *Nuova giur. civ. comm.*, 2020, n. 6, p. 1346.

tutte le altre attività che non si esprimono nella produzione di un bene tangibile¹⁰². Qualora il *software* venisse considerato un servizio, la Dir. 85/374/CEE non sarebbe applicabile ai danni da questo cagionati e gli unici rimedi per la persona danneggiata sarebbero contrattuali¹⁰³ o fondati sulla disciplina ordinaria della responsabilità aquiliana (posto che, in ogni caso, secondo la Corte di Giustizia dell'Unione Europea, la Direttiva non impedisce agli Stati membri di riconoscere forme di responsabilità oggettiva del prestatore di servizi che abbia cagionato un danno mediante l'utilizzo di un prodotto difettoso, purché tale responsabilità non implichi alcuna limitazione di quella dei fabbricanti del prodotto¹⁰⁴).

La parte maggioritaria della dottrina ritiene comunque che – tenendo conto della *ratio* della Dir. 85/374/CEE, che è quella di garantire un livello elevato di tutela nei confronti dei consumatori – sia possibile un'interpretazione estensiva della nozione di prodotto, che comprenda quantomeno gli algoritmi che, pur venduti separatamente (e successivamente) rispetto all'*hardware*, siano strumentali rispetto al funzionamento di un bene fisico¹⁰⁵.

A legislazione vigente, resterebbero in ogni caso esclusi dal campo di applicazione della direttiva i prodotti digitali, che non sembra in alcun modo possibile ricondurre ai “beni mobili” di cui all'art. 2 dir. 85/374/CEE¹⁰⁶. Da più parti – sia da parte

¹⁰² G. ALPA, *Quale modello normativo europeo per l'intelligenza artificiale?*, in *Contr. impr.*, 2021, n. 4, p. 1020.

¹⁰³ E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 462.

¹⁰⁴ Corte di Giustizia UE, 21 dicembre 2011, *Centre hospitalier universitaire de Besançon contro Thomas Dutrueux e Caisse primaire d'assurance maladie du Jura*.

¹⁰⁵ J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 64; B.A. KOCH, *Product Liability 2.0 - Mere Update or New Version?*, ivi, p. 105 ss.; ID., *Response of the EU La Institute to the Public Consultation*, cit., p. 31; G. WAGNER, *Robot, Inc.: Personhood for Autonomous Systems?*, in 88 *Fordham Law Review*, 2019, p. 604; ID., *Robot liability*, cit., p. 41; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1723; U. RUFFOLO, *Responsabilità da A.I. e da algoritmo*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 107; J. KNETSCH, *Are Existing Tort Theories Ready for AI? A Continental European Perspective*, in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, cit.; M. RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente*, cit., p. 1186; A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit., p. 1021; G. GUERRA, *Il concetto di difettosità nella realtà che cambia. Un esercizio di microcomparazione*, in *Comparazione e diritto civile*, 2019, fasc. 1, p. 270.

¹⁰⁶ J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, cit., p. 64; contra M. RATTI, *Riflessioni in materia di responsabilità civile*, cit., p. 1186; A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit., p. 286, a parere

di alcuni atti di indirizzo delle istituzioni europee¹⁰⁷, sia da parte della dottrina¹⁰⁸ – è stato comunque auspicato un intervento normativo, che estenda il campo di applicazione della direttiva anche ai danni cagionati da prodotti *immateriali*.

La già citata proposta di revisione della direttiva sulla responsabilità per danno da prodotto accoglie i suggerimenti emersi in sede di dibattito scientifico ed estende la nozione di prodotto ai *software* e ai “*file per la fabbricazione digitale*” (*digital manufacturing file*, vd. art. 4, § 1, n. 1)¹⁰⁹ e la nozione di componente del prodotto a «*qualsiasi articolo, tangibile o intangibile, o qualsiasi servizio correlato, integrato in un prodotto o interconnesso con esso dal fabbricante di tale prodotto o sotto il controllo di tale fabbricante*» (art. 4, § 1, n. 3). Per non ostacolare l’innovazione e la ricerca nel campo delle tecnologie più avanzate, tuttavia, il considerando 13 stabilisce che la direttiva non si applica a quei *software* (c.d. *open-source*) che sono liberamente accessibili, utilizzabili e modificabili e che sono sviluppati o forniti al di fuori di un’attività commerciale.

3.2. Sull’adeguatezza della nozione di “produttore”.

Delimitata la nozione di “prodotto”, concentriamoci ora sulla contigua definizione di “produttore”. Ai sensi dell’art. 115, co. 2-bis, cod. cons., «*produttore, ai fini del presente titolo, è il fabbricante del prodotto finito o di una sua componente, il produttore della materia prima, nonché, per i prodotti agricoli del suolo e per quelli dell’allevamento, della pesca e della caccia, rispettivamente l’agricoltore, l’allevatore, il*

dei quali anche a normativa invariata l’ideatore del *software non embodied* può considerarsi il fabbricante del prodotto e dunque rispondere ai sensi dell’art. 114 cons. e ss.

¹⁰⁷ L’estensione è auspicata dalla Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l’intelligenza artificiale, cit., considerando 8; vd. anche EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, 2019, cit.

¹⁰⁸ L. COPPINI, *Robotica e intelligenza artificiale: questioni di responsabilità civile*, in *Pol. dir.*, 2018, fasc. 4, p. 729; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit., p. 31 ss.; C. TWIGG-FLESNER, *Guiding Principles for Updating the Product Liability Directive for the Digital Age*, cit., p. 5; M. BUITEN E AL., *EU Liability Rules for the Age of Artificial Intelligence*, Centre on Regulation in Europe, 9 aprile 2021, p. 51.

¹⁰⁹ Il «*file per la fabbricazione digitale*» è a sua volta definito come «*una versione o un modello digitale di un bene mobile*» (art. 4, § 1, n. 2). Si tratta di quelle informazioni progettuali che vengono fornite ai macchinari (torni, frese, stampanti 3D) per la realizzazione di beni tangibili (vd. considerando 14 della proposta).

pescatore ed il cacciatore»¹¹⁰. Anche ammettendo che, a legislazione vigente, l'algoritmo possa rientrare nella nozione di "prodotto" (o componente del prodotto) ai sensi dell'art. 115 cod. cons. (art. 2 della direttiva), resta dubbio se i soggetti coinvolti nei processi di programmazione, sviluppo e aggiornamento degli algoritmi di i.a. rientrino nella definizione di "produttore".

Lo sviluppatore di un *software* che non rientra nella struttura organizzativa del produttore – e che, dunque, vende il progetto applicativo a quest'ultimo o aggiorna le funzionalità del *software* dopo l'entrata in commercio del prodotto, in favore del consumatore – non parrebbe rientrare in alcuna delle suddette categorie normative. La conseguenza, dunque, è che, anche qualora si ritenesse di poter estendere la nozione di "prodotto" al *software*, in ogni caso la responsabilità attivabile ai sensi della dir. n. 85/374/CEE potrebbe configurarsi soltanto a carico del fabbricante del prodotto immesso nel commercio, non invece in capo all'ideatore dell'algoritmo¹¹¹; i rapporti tra essi dovranno dunque essere regolati sulla base della disciplina in materia di responsabilità contrattuale. È quello che accade normalmente nei rapporti tra fabbricante e *designer* del prodotto. Il *design* riguarda, infatti, la sola concezione astratta del prodotto e non ne costituisce una componente; è il fabbricante che dà corpo a tale "idea astratta" ed egli solo risponde dei difetti del prodotto così concretizzato – tutt'al più, il *designer* potrà, per l'appunto, essere chiamato a rispondere per inadempimento contrattuale nei confronti del fabbricante¹¹².

La non riconducibilità del programmatore alla nozione di "produttore" non pare, in realtà, del tutto incoerente con le finalità della disciplina sulla responsabilità per il danno da prodotto difettoso¹¹³. È la casa produttrice, infatti, a fungere da garante dell'affidabilità del prodotto finito nei confronti del consumatore – sia quando sia impegnata direttamente nella progettazione, sia quando, invece, in qualità di

¹¹⁰ L'art. 3, co. 1, dir. aggiunge «ogni persona che, apponendo il proprio nome, marchi marchio o altro segno distintivo sul prodotto, si presenta come produttore dello stesso».

¹¹¹ U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 97; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit., p. 32.

¹¹² U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 97.

¹¹³ A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 627; contra U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 98, che sostiene che lo sviluppatore di un algoritmo di i.a. non può essere assimilabile al tradizionale *designer*, poiché questi – attribuendo alla macchina la capacità di modificarsi nel tempo – plasma concretamente la configurazione del prodotto finale.

assemblatore, si limiti a selezionare i *partner* più adeguati a tal fine¹¹⁴. Il fabbricante del prodotto finito, d'altra parte, è il soggetto che si trova nella posizione più adeguata a svolgere una costante attività di monitoraggio del funzionamento dei propri prodotti e ad intervenire, qualora essi rivelino imperfezioni o malfunzionamenti, al fine di migliorarne la qualità e l'affidabilità¹¹⁵.

Un'estensione della definizione di "produttore" potrebbe invece essere opportuna al fine di includervi – sempre che non coincidano con i *manufacturers* – i *backend operators*, ossia quei soggetti che monitorano in maniera continuativa le condizioni di sicurezza dei *software* e forniscono assistenza e aggiornamenti agli utenti¹¹⁶. Le operazioni di *backend* hanno un ruolo centrale nel plasmare le funzionalità e il *design* dei *software* ed intervengono *dopo* la messa in commercio del prodotto: per questo motivo, non parrebbe corretto imputare al fabbricante eventuali difetti derivanti da tali attività, stante il controllo limitato o nullo che questi può svolgere in materia.

L'estensione della nozione di "produttore" ai programmatori – sia che intervengano prima della messa in commercio del prodotto, sia che intervengano dopo – è stata auspicata, con accenti diversi, dalla maggioranza degli studiosi¹¹⁷. Tali sollecitazioni sono state accolte nella proposta di revisione della direttiva sul danno da prodotto, che estende radicalmente il novero dei soggetti che possono essere chiamati a rispondere per il danno derivante dal prodotto difettoso. L'art. 7 della proposta di revisione (rubricato *Operatori economici responsabili del danno da prodotti difettosi*) include, dunque, sia i fabbricanti del prodotto (o di una componente del prodotto) – la cui definizione è stata ampliata fino a comprendere anche coloro che sviluppano (*to develop*) un prodotto (vd. art. 4, § 1, n. 11) – sia i soggetti che apportano modifiche sostanziali al prodotto (vd. art. 7, § 4: «Qualunque persona fisica o giuridica che modifichi un prodotto già immesso sul mercato o messo in servizio è considerata fabbricante del prodotto ai fini di cui al paragrafo 1, qualora la modifica sia considerata sostanziale in virtù delle norme

¹¹⁴ Così A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 627.

¹¹⁵ A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 627.

¹¹⁶ B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit., p. 39.

¹¹⁷ *Ex multis* S. LI – M. FAURE – K. HAVU, *Liability Rules for AI-Related Harm*, cit., p. 8-9; U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 97-98; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit., p. 39.

dell'Unione o nazionali in materia di sicurezza dei prodotti e sia effettuata al di fuori del controllo del fabbricante originario»)¹¹⁸.

3.3. Sull'adeguatezza della nozione di “difetto”.

Le istanze di adeguamento della disciplina esaminate sinteticamente nei paragrafi precedenti (nozione di *prodotto* e *produttore*) riguardano, come già accennato, la questione dell'applicabilità della *Product liability directive* a quelli che abbiamo definito come *prodotti fluidi*, ossia prodotti che subiscono delle modifiche, nel loro ciclo vitale, per effetto di aggiornamenti e modifiche di sistema operativo. La delimitazione della nozione di difetto occupa invece, all'interno della nostra trattazione, uno spazio più significativo, poiché interroga proprio il problema *specifico* dei sistemi di intelligenza artificiale, ossia quello dell'*autonomia* e dell'*imprevedibilità*.

Il concetto di “difettosità” – nel circoscrivere la responsabilità del produttore – costituisce il perno su cui ruota l'intero sistema della responsabilità per danno da prodotto. Affinché il consumatore possa ottenere ristoro per il danno subito, non è sufficiente che vi sia un nesso di causalità tra prodotto e danno, ma è necessario che – al momento della messa in commercio (vd. art. 118, lett. b) – sussista un *difetto*, ossia una «condizione di anormale pericolosità del prodotto»¹¹⁹.

Non tutti i prodotti che cagionano un danno, infatti, sono di per sé difettosi: si pensi alle sigarette o agli alcolici, che siano fabbricati e commercializzati secondo gli

¹¹⁸ Un ulteriore allargamento dell'area dei soggetti responsabili è proposto con riferimento all'*importatore del prodotto difettoso e al rappresentante autorizzato del fabbricante* (art. 7, § 2) e, nel caso in cui nessuno di questi sia stabilito in UE, al *fornitore di servizi di logistica* (art. 7, § 3; definito come «una persona fisica o giuridica che offre, nel corso di un'attività commerciale, almeno due dei seguenti servizi: immagazzinamento, imballaggio, indirizzamento e spedizione di un prodotto, senza essere proprietario del prodotto, ad eccezione dei servizi postali», art. 4, §, 1, n. 14) o al *distributore*. Infine, l'art. 7, § 6, stabilisce che, a determinate condizioni, la direttiva è applicabile anche alle piattaforme *online* che consentono la conclusione di contratti a distanza.

¹¹⁹ U. CARNEVALI, *La responsabilità del produttore. Problemi generali*, in AA.VV., *La responsabilità dell'impresa per i danni all'ambiente e ai consumatori*, Giuffrè, 1978, p. 149; vd. anche Corte di Giustizia UE, 5 marzo 2015, *Boston Scientific Medizintechnik GmbH*, cause riunite C-503/13 e C-504/13, § 30, che definisce i difetti come «anomale potenzialità di danno che il prodotto può causare alla persona o ai beni del suo utilizzatore».

standard di settore¹²⁰, o, ancora, ai farmaci innovativi e ai vaccini, che, pur comportando il rischio di gravi effetti collaterali, non possono essere considerati difettosi se la messa in commercio è stata autorizzata dagli organismi competenti (c.d. *unavoidably unsafe products*)¹²¹. Quello inerente alla sussistenza di un difetto è dunque un *giudizio autonomo* rispetto a quello sull'accertamento del danno concretamente verificatosi: il difetto non può dedursi in via indiretta né dal verificarsi dell'evento dannoso né dall'utilizzo conforme del prodotto¹²².

Per “difetto”, in particolare, non si intende un connotato fisico del bene, ma si fa riferimento ad un *concetto normativo*¹²³ e *relazionale*¹²⁴, poiché legato alle legittime aspettative dei consumatori. Secondo l'attuale impianto della normativa, il difetto consiste dunque in un disallineamento del prodotto rispetto alla sicurezza che la platea degli utenti ha ragionevolmente diritto di attendersi, in applicazione di quello che la dottrina statunitense definisce come *consumer expectations test*.

L'art. 117, co. 1, cod. cons. – che recepisce l'art. 6 della dir. 85/374 – detta i criteri per la valutazione della difettosità del prodotto, stabilendo che «Un prodotto è difettoso quando non offre la sicurezza che ci si può legittimamente attendere tenuto conto di tutte le circostanze, tra cui:

- a) il modo in cui il prodotto è stato messo in circolazione, la sua presentazione, le sue caratteristiche palesi, le istruzioni e le avvertenze fornite;
- b) l'uso al quale il prodotto può essere ragionevolmente destinato e i comportamenti che, in relazione ad esso, si possono ragionevolmente prevedere;
- c) il tempo in cui il prodotto è stato messo in circolazione».

¹²⁰ A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1721.

¹²¹ E. BELLISARIO, *Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, in *Eur. dir. priv.*, 2016, fasc. 3, § 7.

¹²² G. PONZANELLI, *Causa probabile del danno e responsabilità del produttore*, in *Danno e resp.*, 2004, p. 532; Cass. civ., sez. III, 13 dicembre 2010, n. 25116.

¹²³ C. CASTRONOVO, *La nuova responsabilità civile*, Giuffrè, 2006, p. 692-693; vd. anche P.G. MONATERI, *La responsabilità civile*, cit., p. 716, che parla di “concetto legislativo” (corsivi dell'A.); E. RAJNERI, *The projects of European regulation on liability for damage caused by artificial intelligence. Striking a balance between the interests of consumers, multinationals and SMEs*, in *Riv. dir. banc.*, 2022, fasc. 2, p. 108;

¹²⁴ F. CAFAGGI, *Verso un approccio sistematico ai rimedi contro la difettosità dei prodotti*, in S. Patti (a cura di), *Il danno da prodotti in Italia, Austria, Repubblica Federale di Germania, Svizzera*, Cedam, 1990, p. 85.

I parametri citati – che esamineremo in dettaglio *infra*, al § 3.3.2. – costituiscono un elenco puramente esemplificativo: il giudice, nel valutare il livello di «sicurezza che ci si può legittimamente attendere», potrà prendere in considerazione ulteriori elementi, tra i quali spicca – primo tra tutti – il rispetto della normativa sulla sicurezza del prodotto. A tale ultimo proposito, ci limitiamo a sottolineare, per ora (vd. più diffusamente *infra*, § 3.4. ss.), che, se il mancato rispetto degli *standard* di sicurezza obbligatori comporta, di per sé, la difettosità del prodotto¹²⁵, non è vera l'affermazione contraria: l'osservanza di tali *standard*, infatti, non preclude al giudice di indagare sulla possibile esistenza di difetti nel prodotto¹²⁶.

Sin dall'entrata in vigore della dir. 85/374, la dottrina ha riconosciuto proprio nella sfuggente definizione del “difetto” il profilo più problematico della disciplina, capace di aumentare a dismisura la discrezionalità giudiziaria e di rendere di fatto imprevedibili i giudizi in materia¹²⁷. In particolare, il tentativo di mediazione tra la necessità di fornire adeguata tutela ai consumatori e l'esigenza di incentivare l'innovazione e la competitività delle imprese europee¹²⁸ si risolve in una formulazione *compromissoria* ed *incerta* della norma, che risulta ambigua persino sulla scelta del criterio di imputazione della responsabilità da utilizzare¹²⁹. Da un lato, infatti, la clausola generale della “*sicurezza che ci si può legittimamente attendere*” sembra confermare l'opzione di fondo del legislatore comunitario per un regime di responsabilità oggettiva; dall'altro, tuttavia, le circostanze che dovrebbero orientare l'interprete nell'applicazione di tale clausola riguardano la condotta del produttore e del danneggiato, piuttosto che le caratteristiche oggettive del prodotto¹³⁰ – senza contare che le esimenti previste dall'art. 118 c.c. innestano un'ulteriore valutazione delle condotte poste in essere dalla società produttrice.

¹²⁵ A. LUMINOSO, *Certificazione di qualità di prodotti e tutela del consumatore-acquirente*, in *Eur. dir. priv.*, 1999, p. 43 e 54; G. SMORTO, (voce) *Certificazione di qualità e normazione tecnica*, in *Dig. disc. priv.*, Agg. II, Utet, 2003, p. 219; A. ZEI, *Tecnica e diritto tra pubblico e privato*, Giuffrè, 2008, p. 263.

¹²⁶ E. BELLISARIO, *Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, cit., § 4.

¹²⁷ *Ex multis* E. RAJNERI, *L'ambigua nozione di prodotto difettoso al vaglio della corte di cassazione italiana e delle altre corti europee*, in *Riv. dir. civ.*, 2008, n. 5, p. 624 ss.; G. PONZANELLI, *Causa probabile del danno e responsabilità del produttore*, cit., p. 532; E. BELLISARIO, *Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, cit., § 4; M. DELLACASA, *Sulle definizioni legislative nel diritto privato. Fra codice e nuove leggi civili*, Giappichelli, 2004, p. 326 ss.

¹²⁸ E. RAJNERI, *L'ambigua nozione di prodotto difettoso*, cit., p. 624; M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 338.

¹²⁹ M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 338.

¹³⁰ M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 338

A parere della dottrina italiana, tale valutazione non deve apparire in contrasto con il carattere *oggettivo* della responsabilità¹³¹. Per escludere il risarcimento, infatti, al produttore non basterà dimostrare di *non essere in colpa*; questi dovrà invece provare l'esistenza di una delle cause tassative previste dall'art. 118, che – pur coincidendo, in taluni casi, con l'assenza di colpa (è il caso, in particolare, del *rischio da sviluppo*, di cui all'art. 118, lett. e, vd. *infra*, § 3.3.1.) – *non assorbono l'intera casistica della mancanza di colpa*. Il carattere oggettivo della responsabilità, secondo questa prospettiva, potrebbe essere espresso in maniera più precisa dall'espressione francese *responsabilité sans faute* – un modello di imputazione della responsabilità *indipendente* dall'accertamento della colpa, ma che non per questo deve rinunciare a qualsiasi considerazione nei confronti delle condotte di attori e convenuti¹³².

Dal modello di responsabilità oggettiva (*parziale o limitata*) così inteso, dovrebbero invece distinguersi i regimi di c.d. *absolute responsibility*¹³³, che – se applicati alla disciplina della responsabilità per danno da prodotto – obbligherebbero il produttore a tenere indenne il consumatore da *qualsiasi danno* derivante dall'uso del prodotto, trasformandolo dunque in una sorta di *assicuratore*¹³⁴.

Non si può non rilevare, in ogni caso, che, di fatto – e al di là del tentativo dottrinale di trovare una giustificazione teorica all'ambiguità della norma europea – nella prassi la distinzione bizantina tra regimi di *responsabilità per colpa* e regimi di *responsabilità oggettiva non assoluta* tende a scolorire, con la conseguenza che

¹³¹ G. PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, in M. Bessone (a cura di), *Casi e questioni di diritto privato*, Giuffrè, IV ed., 2002, p. 510; F. GALGANO, *Responsabilità del produttore*, in *Contr. impr.*, 1986, p. 996; G. VISINTINI, *Trattato breve della responsabilità civile. Fatti illeciti. Inadempimento. Danno Risarcibile*, Cedam, 2005, p. 748.

¹³² G. PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, cit., p. 510.

¹³³ D.G. OWEN, *Rethinking the Policies of Strict Products Liability*, in 33 *Vand. L. Rev.*, 1980, p. 682, che contrappone, proprio in materia di responsabilità per danno da prodotto, la *strict liability* e la *absolute liability*; sul concetto di “responsabilità assoluta” vd., nella dottrina italiana, G. PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, cit., p. 509 e 515; G. PONZANELLI – R. PARDOLESI, *La responsabilità per danno da prodotti difettosi*, in *Nuove leggi civ.*, 1989, p. 501. Un'ipotesi, nel nostro ordinamento, di *absolute liability* può essere considerata la responsabilità per fatto del dipendente (art. 2049), che non tollera prove liberatorie fondate sulla diligenza del datore di lavoro; alcuni autori, muovendo da tale considerazione, giungono addirittura a ritenere che l'art. 2049 c.c. non riguardi la responsabilità civile, ma dia luogo esclusivamente ad un'obbligazione di garanzia, vd. D. DE MARTINI, *I fatti produttivi di danno risarcibile*, Cedam, 1983, p. 277.

¹³⁴ U. CARNEVALI, *La responsabilità del produttore. Problemi generali*, cit., p. 148 e 155; G. PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, cit., p. 515.

l'accertamento del difetto di *design* si trasforma spesso in una vera e propria valutazione della *diligenza professionale del produttore*¹³⁵.

L'approssimarsi della disciplina della responsabilità per danno da prodotto difettoso ad un modello di responsabilità per colpa, d'altra parte, incontra problemi peculiari in relazione all'imputazione dei danni cagionati da sistemi di i.a. La capacità di adattamento e di auto-apprendimento degli agenti intelligenti, infatti, mal si concilia con la necessità di accertare una *negligenza* in capo al produttore, dal momento che – come abbiamo ribadito più volte – l'evento lesivo può derivare anche da un comportamento imprevedibile dell'algoritmo.

3.3.1. La delimitazione della rilevanza dei difetti: le esimenti.

A precisare – per sottrazione – la nozione di difetto tracciata dall'art. 117 cod. cons., vi sono le esimenti previste dall'art. 118 cod. cons., alle lettere b), d) ed e), che escludono la rilevanza di alcuni tipi di difetto. In particolare, il difetto del prodotto non può essere imputato al produttore ai sensi della disciplina comunitaria se «*il difetto che ha cagionato il danno non esisteva quando il produttore ha messo il prodotto in circolazione*» (lett. b, c.d. *later defect defence*), se «*il difetto è dovuto alla conformità del prodotto a una norma giuridica imperativa o a un provvedimento vincolante*» (lett. d), o se «*lo stato delle conoscenze scientifiche e tecniche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso*» (lett. e, c.d. esimente per rischio da sviluppo o *development risk defence*¹³⁶).

¹³⁵ E. VISENTINI, *L'esimente del rischio di sviluppo come criterio di responsabilità del produttore. (l'esperienza italiana e tedesca e la direttiva comunitaria)*, in *Resp. civ. prev.*, 2004, p.1267. vd. anche D. CARUSO – R. PARDOLESI, *Per una storia della Direttiva 1985/374/CEE*, in R. Pardolesi (a cura di) *I 25 anni di products liability*, quaderno speciale di *Danno e resp.*, 2012, p. 15, che rilevano come, al di là della qualificazione formale, l'impostazione della direttiva europea come «la seduttiva semplicità del paradigma di responsabilità oggettiva (...) andava evaporando». La dottrina maggioritaria, in Europa, intravede, infatti, nella disciplina della responsabilità per danno da prodotto difettoso un regime *de facto* colposo, vd. S. WHITTAKER, *Liability for Products: English Law, French Law and European Harmonisation*, cit., p. 494 ss.; H. ZECH, *Liability for AI: public policy considerations*, cit., p. 150.

¹³⁶ Sulla nozione di “rischio da sviluppo” adottata dalla Direttiva vd. D. CERINI, *Responsabilità del produttore e rischio da sviluppo: oltre la lettera della dir. 85/374/CE*, in *Dir. ed econ. assic.*, 1996, p. 33.

Tolta la lett. d) – che si limita a fornire un criterio, del tutto ragionevole, per risolvere eventuali antinomie – le lett. b) ed e) devono essere interpretate come disposizioni che, così come la stessa definizione di difetto fornita dall’art. 117 cod. cons., contribuiscono a diluire la rigosità del regime fondato sulla responsabilità oggettiva, poiché introducono una forma – seppur limitata – di valutazione della diligenza del produttore¹³⁷. Quest’ultimo, infatti, non risponde in forza del mero rapporto di causalità tra *creazione del rischio* ed *evento lesivo* che concretizza il rischio creato – regime che scivolerebbe nella già citata *responsabilità assoluta* – ma soltanto in relazione a danni che erano prevedibili sulla base dello stato dell’arte esistente al momento della messa in commercio. Secondo una parte della dottrina, è proprio il regime delle esimenti – particolarmente favorevole alla posizione del produttore – che scoraggia il danneggiato dal fare ricorso alla disciplina prevista dal codice del consumo, spingendolo ad avvalersi piuttosto della tutela offerta dall’art. 2050 c.c. (*responsabilità per l'esercizio di attività pericolose*, ove la prova liberatoria è quella di *avere adottato tutte le misure idonee a evitare il danno*; su tale disposizione vd. *infra*, § 4, in questo cap.)¹³⁸.

L’applicabilità della *later defect defence* e della *development risk defence* anche nel caso di danni cagionati da sistemi che utilizzano tecniche di *machine learning* suscita diverse perplessità in dottrina¹³⁹. Sia che si voglia considerare il difetto dell’algoritmo come sopravvenuto (e, dunque, escludere la responsabilità del produttore facendo ricorso alla *later defect defence*), sia che si consideri tale difetto come esistente ma non conoscibile dal produttore al momento della messa in commercio (*development risk defence*), in ogni caso l’evento lesivo che deriva da un comportamento emergente del sistema di i.a. parrebbe difficilmente ascrivibile al produttore, sulla base delle norme vigenti.

La dottrina maggioritaria tende a considerare il comportamento deviante del dispositivo intelligente come un vizio esistente ma non conoscibile, più che come un difetto sopravvenuto, evidenziando il rischio che l’esimente da rischio di sviluppo possa

¹³⁷ A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 58.

¹³⁸ A. FUSARO, *I prodotti difettosi e pericolosi: le responsabilità*, in G. Alpa (a cura di), *La responsabilità del produttore*, Giuffrè, 2019, p. 380 ss.

¹³⁹ Vd. per tutti U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1262 ss.

essere considerata applicabile di *default* in relazione ai sistemi di I.A.¹⁴⁰, dal momento che il comportamento di questi, per loro natura, è caratterizzato da una limitata prevedibilità.

Se così avvenisse, i costi dei sinistri finirebbero per gravare sempre sulla vittima, o su altri soggetti, chiamati a rispondere oggettivamente a vario titolo (si pensi, ad esempio, per quanto concerne il danno da circolazione di autoveicoli, al proprietario o conducente del veicolo, secondo quanto disposto dall'art. 2054 c.c.)¹⁴¹, ma mai sul produttore, nonostante l'imprevedibilità degli eventi lesivi sia in qualche modo *prevedibile* da quest'ultimo e dunque – così come per i tradizionali difetti di fabbricazione – *statisticamente calcolabile* dall'impresa. Nell'*Impact Assessment* che accompagna la proposta di revisione della Direttiva, una delle opzioni individuate per far fronte alle capacità adattive dei sistemi di i.a. era proprio quella di sopprimere l'esimente del rischio da sviluppo (opzione 2b)¹⁴². Tale intervento era inoltre auspicato sia dal Gruppo di Esperti sull'i.a. nominato dalla Commissione Europea¹⁴³ – che riteneva altresì inopportuna, con riferimento ai sistemi di i.a., l'applicabilità della *later defect defence* – sia dall'Organizzazione europea dei consumatori (BEUC, *Bureau Européen des Unions de Consommateurs*)¹⁴⁴.

La suddetta opzione non è stata tuttavia percorsa dalla proposta di revisione della direttiva sul danno da prodotto, probabilmente a causa della forte opposizione da parte delle società produttrici, che temevano che una soppressione della *development risk*

¹⁴⁰ M. EBERS, *Liability For Artificial Intelligence And EU Consumer Law*, in 3 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 2021, n. 3, p. 215; G. SPINDLER, *User liability and strict liability in the Internet of Things and for Robots*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 143; sembra invece considerare il “difetto da *self learning*” un difetto sopravvenuto U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., 1264, che giudica dunque dubbia l'applicabilità della *development risk defense*.

¹⁴¹ U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1705.

¹⁴² EUROPEAN COMMISSION, *Impact Assessment Report Accompanying the Document Proposal for a Directive of the European Parliament and of The Council on Liability for Defective Products*, SWD(2022) 316 final, 28 settembre 2022, p. 50, disponibile a questo link: https://single-market-economy.ec.europa.eu/document/348b3e35-7d1a-43df-8e9d-296fc09e2c3c_en.

¹⁴³ EUROPEAN COMMISSION – EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 42-44.

¹⁴⁴ BUREAU EUROPÉEN DES UNIONS DE CONSOMMATEURS, *PRODUCT LIABILITY 2.0*, cit., in dottrina vd. anche H. JACQUEMIN – J. HUBIN, *Aspects contractuels et de responsabilité civile en matière d'intelligence artificielle*, cit., p. 137; F.P. PATTI, *The European Road to Autonomous Vehicles*, in *Fordham International Law Journal*, 2019, vol. 43, p. 142 ss. (in relazione alle di macchine a guida autonoma).

defence potesse determinare un aumento esponenziale delle condanne al risarcimento di danni¹⁴⁵.

3.3.2. La classificazione tradizionale: difetti di fabbricazione, progettazione, informazione nei sistemi di i.a.

Tradizionalmente – prima ancora che, in Europa, fosse introdotta una disciplina specifica per la responsabilità per danno da prodotto difettoso – la dottrina¹⁴⁶ ha tracciato, all'interno della categoria dei difetti del prodotto, una classificazione basata sull'eziologia del difetto, che distingue tra vizi di fabbricazione, vizi di costruzione (oggi chiamati più comunemente vizi di progettazione o *design*) e vizi di informazione.

Nonostante tale distinzione non sia richiamata formalmente dalla disciplina italiana e da quella europea – che fanno genericamente riferimento alla “sicurezza” – essa riecheggia nelle circostanze elencate dall'art. 117 cod cons. e dall'art. 6 dir. 85/374.

(i) I difetti di fabbricazione (*manufacturing defects*) vengono ad esistenza durante il processo di fabbricazione e concernono soltanto singoli esemplari di una serie di prodotti generalmente non difettosi. Si tratta di vizi tipici dei moderni metodi di produzione in serie, che possono essere dovuti ad una disfunzione occasionale del macchinario (es. momentaneo inceppamento di un macchinario) o a sviste del personale addetto alla fabbricazione o al controllo.

Come insegna la migliore dottrina civilistica degli anni Sessanta e Settanta¹⁴⁷, tali vizi si caratterizzano, da un lato, per la loro *inevitabilità* – poiché neppure l'impiego delle migliori tecniche produttive può eliminarli del tutto – e, dall'altro, per la loro *prevedibilità*, nel senso che la loro frequenza può essere statisticamente calcolata. L'incidenza dei vizi di fabbricazione può essere accertata piuttosto facilmente, essendo

¹⁴⁵ EUROPEAN COMMISSION, *Impact Assessment*, cit., p. 51.

¹⁴⁶ Il merito è soprattutto della dottrina tedesca. Si vedano i riferimenti contenuti in U. CARNEVALI, *La responsabilità del produttore*, cit., p. 30, nota 82; vd. anche G. HOHLOCH, *Prospettive di evoluzione della responsabilità del produttore nel diritto tedesco*, in S. Patti (a cura di), *Il danno da prodotti in Italia, Austria, Repubblica Federale di Germania, Svizzera*, cit., p. 268 ss.

¹⁴⁷ U. CARNEVALI, *La responsabilità del produttore. Problemi generali*, cit., p. 144.

sufficiente dimostrare che il singolo prodotto differisce, nelle sue caratteristiche materiali, dagli altri prodotti della serie; nell'ambito degli *smart products*, ciò può accadere in relazione alle componenti dell'*hardware* (es. sensori, videocamere) o, ad esempio, nel caso dell'installazione incompleta di un *software* di guida autonoma in un veicolo¹⁴⁸. Tutti gli errori che coinvolgono il codice dell'algoritmo possono, invece, costituire esclusivamente *difetti di design*, poiché concernono i criteri di progettazione del prodotto¹⁴⁹.

Si tratta dell'unica tipologia di difetto che è regolata da un regime di responsabilità oggettiva puro¹⁵⁰: nel caso in cui il danno sia causato dalla difformità del prodotto rispetto agli altri esemplari della serie, il produttore è ritenuto responsabile ai sensi dell'art. 114 cod. cons. sulla base del mero accertamento del nesso di causalità, indipendentemente da qualsivoglia valutazione circa la diligenza del fabbricante. Nell'individuare la "*sicurezza che ci si può legittimamente attendere*", l'interprete assumerà quale naturale punto di riferimento le caratteristiche degli altri prodotti della medesima serie; l'art. 117, co. 3, conferma tale ricostruzione stabilendo che «*Un prodotto è difettoso se non offre la sicurezza offerta normalmente dagli altri esemplari della medesima serie*».

(ii) I difetti di costruzione (*design defects*) derivano invece da un errore nella progettazione degli artefatti (es. fallace scelta dei materiali utilizzati, insufficiente sperimentazione etc.). A differenza dei vizi di fabbricazione, quelli di costruzione inficiano alla radice l'intera serie di prodotti messa in commercio, con aumento esponenziale della potenzialità lesiva. Rientrano in questa categoria anche quei difetti che non attengono direttamente al prodotto, ma alla sua confezione, quando questa è tale da non garantire un'efficace conservazione del prodotto¹⁵¹. L'art. 117 cod. cons. contempla i difetti di progettazione alle lett. b) e c), ove vengono specificati i criteri ai quali il giudice deve attenersi al fine di verificare l'eventuale difettosità del prodotto.

¹⁴⁸ G. WAGNER, *Robot, Inc.*, cit., p. 604.

¹⁴⁹ M.A. GEISTFELD, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, in 105 *Calif. L. Rev.*, 2017, p. 1633; F.P. HUBBARD, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation*, in 66 *Fla. L. Rev.*, 2015, p. 1854.

¹⁵⁰ M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 339.

¹⁵¹ U. CARNEVALI, *La responsabilità del produttore. Problemi generali*, cit., p. 144.

Innanzitutto, la progettazione del prodotto deve essere valutata sulla base dell'«uso al quale il prodotto può essere ragionevolmente destinato e i comportamenti che, in relazione ad esso, si possono ragionevolmente prevedere» (art. 117, co. 1, lett. b) cod. cons.). Si tratta, nella materia che ci interessa, di una circostanza non priva di rilevanza, specie se consideriamo che, ad oggi, la tendenza della regolamentazione europea è quella di garantire sempre la presenza di una persona (c.d. *human-in-the-loop*)¹⁵², che possa intervenire – anche in relazione ai sistemi di i.a. più avanzati ed autonomi – per riprendere il controllo dell'algoritmo o *robot*. In fase di progettazione, il produttore deve dunque prevedere meccanismi idonei ad assicurare un rapido ed efficace intervento da parte dell'utilizzatore, tenendo altresì conto della naturale propensione delle persone a confidare eccessivamente nell'accuratezza dell'attività algoritmica (c.d. *technological complacency*), nonché dell'inevitabile aumento dei tempi di reazione e di gestione delle emergenze, nel caso in cui le situazioni ordinarie siano gestite da tecniche di automazione (c.d. *control dilemma*)¹⁵³: affronteremo con maggiore approfondimento questi problemi quando parleremo dell'eventuale colpa penale in capo all'utilizzatore di sistemi di i.a. (vd. *infra*, cap. III, § 3.2.2.).

Così, restando nel campo dell'*automotive*, è ragionevolmente prevedibile che l'utilizzatore di un veicolo di livello 2 o 3 – quando è attiva la funzione di guida autonoma – perda la concentrazione necessaria ad intervenire tempestivamente. Il produttore dovrebbe dunque predisporre meccanismi di allerta compatibili con l'incremento dei tempi di reazione, così come un'interfaccia semplice ed intuitiva che consenta all'utilizzatore di comprendere rapidamente quale problema si sta verificando e come agire. Potrebbero altresì rientrare tra le misure di sicurezza volte a prevenire prevedibili abusi da parte dei conducenti gli *hand detection systems*, ossia quei sensori in grado di rilevare l'eventuale allontanamento delle mani del conducente dal volante¹⁵⁴.

¹⁵² Vd. il già citato art. 14, co. 1, *AI Act*: «I sistemi di IA ad alto rischio sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso».

¹⁵³ E. HILGENDORF, *Automated Driving and the Law*, cit., p. 187.

¹⁵⁴ N. DAS E AL., *On Performance Evaluation of Driver Hand Detection Algorithms: Challenges, Dataset, and Metrics*, in 2015 *IEEE 18th International Conference on Intelligent Transportation Systems*. Alcune aziende – sempre al fine di verificare se l'attenzione dell'utente sia rivolta alla guida del veicolo – stanno altresì sviluppando sistemi che monitorano i movimenti degli occhi o della testa del conducente, nonché la frequenza cardiaca o respiratoria dell'utente. Trattandosi, tuttavia, di dati inerenti alla salute dell'utente, il

Alla lett. c), l'art. 117, co. 1, cod. cons. prevede poi che la difettosità del prodotto debba essere valutata sulla base del «*tempo in cui il prodotto è stato messo in circolazione*». Gli standard di sicurezza sono dunque correlati allo stato dell'evoluzione scientifica esistente al momento in cui i prodotti vengono messi in circolazione. Il co. 2 dell'art. 117 precisa poi che un prodotto non può essere considerato difettoso per il solo fatto che un prodotto più perfezionato sia stato in qualunque tempo messo in commercio. La precisazione è quantomai opportuna, poiché preclude di ritenere che le “legittime aspettative” del consumatore debbano sempre coincidere con l'adozione di tutte le misure di sicurezza che la tecnica offre al momento della messa in commercio del prodotto. Il contenuto delle aspettative, invece, è strettamente legato alle scelte di consumo dell'acquirente: questi potrà infatti decidere di optare per prodotto meno robusto, ma più a buon mercato, accettando implicitamente che dovrà provvedere più frequentemente all'acquisto di un nuovo modello o alla manutenzione di quello acquistato; viceversa, il consumatore potrà preferire pagare un prezzo più alto per un prodotto che presenti caratteristiche di maggiore sicurezza e resistenza¹⁵⁵. D'altra parte, qualora il prodotto più perfezionato esistente sul mercato venisse considerato il solo “non difettoso”, si verrebbe a creare uno squilibrio concorrenziale tra il produttore di tale bene e tutti gli altri produttori, poiché solo il primo sarebbe esentato dai costi degli incidenti causati dai prodotti¹⁵⁶.

Complementare ai criteri delineati dall'art. 117, co. 1, lett. b) e c), è l'esimente per il rischio da sviluppo, prevista dall'art. 118, lett. e), che esclude la responsabilità del produttore «*se lo stato delle conoscenze scientifiche e tecniche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso*» (vd. art. 7, lett. e), dir. 85/374, c.d. *development risk defense*). Come abbiamo già evidenziato, il combinato disposto tra queste norme induce a ritenere che la verifica circa la sussistenza dei difetti di progettazione non sia fondata su un regime di responsabilità oggettiva pura, poiché la valutazione, da un lato, della *prevedibilità degli impieghi del prodotto* da parte dei consumatori (art. 117, co. 1, lett. a), e, dall'altro, delle

GDPR pone limiti al loro utilizzo, vd. T. MULDER – N.E. VELLINGA, *Exploring data protection challenges of automated driving*, cit.

¹⁵⁵ P. TRIMARCHI, *Responsabilità civile: atti illeciti, rischio, danno*, cit., p. 439.

¹⁵⁶ G. WAGNER, *Robot, Inc.*, cit., p. 606.

conoscenze scientifiche esistenti al momento della messa in circolazione, introducono degli elementi di valutazione circa la colpevolezza del produttore¹⁵⁷.

(iii) Infine, vi sono i difetti legati ad erroneità o carenza delle informazioni offerte ai consumatori. Il prodotto, di per sé, è fabbricato e progettato a regola d'arte: il difetto consiste, tuttavia, nell'omissione delle informazioni sulla corretta modalità d'impiego (es. manutenzione, conservazione, ricambi, usi controindicati, etc.) e sui pericoli che l'utilizzo del prodotto può comportare (es. infiammabilità, tossicità, effetti collaterali nel caso dei medicinali, etc.). La scarsa informazione crea una legittima aspettativa di sicurezza e rende il prodotto difettoso, in caso di danni derivanti da un suo impiego scorretto. L'art. 117 cod. cons. fa riferimento ai difetti informativi al co. 1, lett. a), ove stabilisce che – tra le circostanze che devono essere considerate per valutare la sicurezza del prodotto – il giudice deve tenere conto anche del «*modo in cui il prodotto è stato messo in circolazione, la sua presentazione, le sue caratteristiche palesi, le istruzioni e le avvertenze fornite*». Appare con tutta evidenza che sussiste un'intensa correlazione tra difetti informativi e difetti di progettazione: le legittime aspettative inerenti al *design* del prodotto dipendono infatti anche dalle informazioni fornite al consumatore.

All'obbligo informativo che ricade sul produttore sembra di poter ricollegare due corollari, da sempre sottolineati dalla dottrina che si occupa della materia.

In primo luogo, va rilevato che soltanto eccezionalmente le istruzioni possono costituire mezzi di prevenzione del rischio adatti a sopperire ad un difetto di progettazione del prodotto. Si pensi, ad esempio, ad un farmaco che può causare allergie su alcune categorie di soggetti: in questo caso, un'adeguata avvertenza circa gli effetti allergenici del farmaco è idonea ad evitare che la sua composizione debba subire modifiche – con conseguente aumento dei costi e, eventualmente, ritardo nello sviluppo tecnologico – per fronteggiare il rischio corso da un ristretto numero di utenti¹⁵⁸. Nella maggioranza dei casi, tuttavia, il sistema degli avvisi informativi non può essere considerato una legittima alternativa ad una maggiore sicurezza strutturale e progettuale del prodotto, ove ragionevolmente possibile; in caso contrario i costi che dovrebbero essere sostenuti dal

¹⁵⁷ A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 58; C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, cit., p. 159.

¹⁵⁸ C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 48.

produttore per la messa in sicurezza del prodotto sarebbero scaricati sul consumatore¹⁵⁹. Le avvertenze informative, insomma, non possono porre sul consumatore un obbligo di diligenza *spropositato* rispetto alle esigenze di normale utilizzo del prodotto: così, ad esempio, un abito confezionato con materiale altamente infiammabile è da ritenersi *pericoloso* anche nel caso in cui il produttore abbia informato l'acquirente di tale caratteristica, poiché – nel normale utilizzo di un capo di vestiario – non può escludersi un contatto accidentale con sigarette, fiammiferi o altre fonti di calore¹⁶⁰. In secondo luogo, vi è da ritenere che quanto più il produttore sorvoli sui rischi del prodotto (fornendo istruzioni e avvertenze poco chiare o incomplete), tanto più salgono le aspettative di sicurezza dell'utente.

Entrambi i corollari rischiano di entrare in gioco in quello che, come abbiamo ripetuto più volte, è ad oggi il settore più avanzato nello sviluppo di sistemi di i.a., ossia il settore dell'*automotive*. Il problema si pone, in particolare, in relazione ai veicoli di livello 2 e 3, che sono in grado di svolgere soltanto parzialmente la funzione di guida.

Da un lato, infatti, la presentazione pubblicitaria di alcuni *semi-autonomous vehicles* tende a dare ai potenziali acquirenti l'idea che la macchina possa guidarsi completamente da sola, senza alcun controllo da parte del conducente: si pensi, ad esempio, al sistema di guida autonoma di livello 2 prodotto da Tesla, il cui nome, *Autopilot*, evoca un'esperienza di guida totalmente automatizzata; si pensi, ancora, al video pubblicato sul canale Youtube della medesima casa automobilistica, ove si assiste ad una guida nel traffico gestita interamente dal *robot*, con tanto di mani del conducente posate sul grembo e non – come invece imporrebbero le normative in materia – sul volante. L'accusa di pubblicità ingannevole rivolta alla società Tesla, d'altra parte, non ha lasciato indifferenti gli organi giudiziari e di regolazione amministrativa: nel 2020, il Tribunale di Monaco ha stabilito che "*Autopilot*" e "*Full Self Driving*" sono espressioni fuorvianti e ha ordinato a Tesla di cessare di utilizzarle sul suo web tedesco¹⁶¹; ancora, nel 2022, il *California Department of Motor Vehicles* ha proposto ricorso contro Tesla presso lo *State Office of Administrative Hearings*, allegando che la società avrebbe

¹⁵⁹ U. CARNEVALI, *Responsabilità del produttore*, in *Enc. Dir.*, Agg. II, 1998, p. 944; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 48.

¹⁶⁰ C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 48; P.G. MONATERI, *La responsabilità civile*, cit., p. 719.

¹⁶¹ M. Taylor, *German Court Bans Tesla 'Autopilot' Name For Misleading Customers*, cit.

diffuso dichiarazioni non veritiere sui veicoli di sua produzione, pubblicizzandoli come dotati di funzioni di guida autonome di cui, in realtà, erano sprovisti¹⁶²;

Dall'altro lato, ci sembra che alcuni difetti di progettazione che emergono dalle prime ricostruzioni dei sinistri non possano essere compensati nemmeno da un'adeguata informazione fornita dal produttore: così, un veicolo di livello 2 – che, come illustrato *supra*, al cap. I, § 7.1.1.1.2., è in grado di guidare autonomamente in determinate condizioni, ma richiede l'intervento del conducente nelle situazioni di emergenza – non pare offrire «la sicurezza che ci si può legittimamente attendere», nel caso in cui sia incapace di riconoscere alcune luci nel traffico (semafori, lampeggianti delle forze dell'ordine, dell'ambulanza, della segnaletica), nonostante il consumatore venga informato di tale carenza.

3.3.3. I difetti di progettazione: tra responsabilità per colpa e responsabilità assoluta.

Da sempre, i *design defects* sono considerati i difetti più problematici nell'ambito della responsabilità per danno da prodotto¹⁶³. Innanzitutto, non esiste un *design* del prodotto che sia *assolutamente* sicuro: il giudice è dunque chiamato a valutare qual è il livello di sicurezza che ci si può legittimamente attendere da un prodotto – valutazione che evidentemente lascia ampi spazi di discrezionalità¹⁶⁴. D'altra parte, il mercato solitamente offre vari modelli del medesimo prodotto, diversi per qualità, prezzo e progettazione, e, ovviamente, diversi anche per sicurezza. Se è vero che, come già osservato, il comma 2 dell'art. 117 cod. cons. stabilisce che «un prodotto non può essere considerato difettoso per il solo fatto che un prodotto più perfezionato sia stato in qualunque tempo messo in commercio», è anche vero che tra i criteri sui quali i consumatori fondano le proprie legittime aspettative di sicurezza rientra anche il metro comparativo con riferimento alla generalità delle merci offerte sul mercato. Dal punto di vista probatorio, poi, un altro elemento che rende complessa la valutazione del difetto di

¹⁶² R. Mitchell, *California DMV accuses Tesla of falsely advertising Autopilot and Full Self-Driving features*, cit.

¹⁶³ U. CARNEVALI, *Responsabilità del produttore*, in *Enc. Dir.*, cit., p. 943; P. TRIMARCHI, *Responsabilità civile: atti illeciti, rischio, danno*, cit., p. 438.

¹⁶⁴ U. CARNEVALI, *Responsabilità del produttore*, in *Enc. Dir.*, cit., p. 943.

progettazione consiste nel fatto che l'utente, con il suo comportamento, può influire sulla sicurezza offerta dal prodotto: non è dunque sempre agevole determinare se l'evento lesivo sia conseguenza del difetto del prodotto o, invece, di uno scorretto impiego del prodotto da parte del consumatore¹⁶⁵.

Tuttavia, la maggiore difficoltà che s'incontra nell'accertamento dei difetti di *design* consiste proprio nell'ambigua formulazione normativa, stante l'incerto ruolo che il sistema nel suo complesso (*i.e.* nozione di difetto ed esimenti) assegna all'eventuale *comportamento negligente* del produttore. Tale questione si intreccia con i tortuosi profili probatori della disciplina, caratterizzati, per un verso, dall'asimmetria informativa che caratterizza colui che sopporta l'onere della prova (il danneggiato) rispetto al convenuto-produttore, e, per altro verso, dalla stessa complessità tecnica dell'accertamento degli elementi della fattispecie, primo tra tutti il nesso di causalità.

A tal proposito, si fronteggiano due opposti orientamenti. Da un lato, la giurisprudenza di merito tende ad alleggerire l'onere probatorio a carico del danneggiato, fino a ritenere che basti la prova di un malfunzionamento del prodotto per dimostrare la sussistenza del difetto, spettando poi al produttore l'onere di fornire un'eventuale prova liberatoria ai sensi dell'art. 118 cod. cons.¹⁶⁶. Di fatto, tale posizione determina uno svilimento del requisito della difettosità¹⁶⁷, con conseguente trasformazione della disciplina in oggetto in un regime di "responsabilità assoluta" di natura indennitaria.

Dall'altro lato, la giurisprudenza di legittimità afferma costantemente che il difetto del prodotto non può dedursi né dal verificarsi di un evento dannoso, né dall'utilizzo "normale" del prodotto¹⁶⁸. In particolare, si afferma che «il danno non prova, di per sé, né direttamente né indirettamente il difetto, né la pericolosità del prodotto in condizioni normali di impiego, ma solo una più indefinita pericolosità di per sé insufficiente per istituire la responsabilità del produttore, se non sia anche in concreto accertato che quella

¹⁶⁵ U. CARNEVALI, *Responsabilità del produttore*, in *Enc. Dir.*, cit., p. 943.

¹⁶⁶ Vd. ad es. Trib. Monza, 10 febbraio 2015, in *Danno resp.*, 2015, p. 950; Trib. Pontedera, 16 marzo 2011, in *Danno resp.*, 2012, p. 67; Trib. Roma, 4 dicembre 2003 e Trib. Roma, 3 novembre 2003, in *Danno resp.*, 2004, p. 527; Trib. Benevento, 24 febbraio 2006, in *Danno resp.*, 2006, p. 1254.

¹⁶⁷ M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 342 ss., che si interroga sull'utilità stessa del concetto di "difetto" ove letto alla luce dei meccanismi presuntivi utilizzati dalla giurisprudenza di merito.

¹⁶⁸ Cass. civ., sez. III, 29 luglio 2015, n. 15851; Cass. civ., sez. III, 26 giugno 2015, n. 13225; Cass. civ., sez. III, 29 maggio 2013, n. 13458; Cass. civ., ord. 23 maggio 2013, n. 12665; Cass. civ., sez. III, 15 marzo 2007, n. 6007.

specifica condizione di insicurezza del prodotto si pone al di sotto del livello di garanzia di affidabilità richiesto dalla utenza o dalle leggi in materia»; l'eventuale ricorso a presunzioni può contribuire a formare il convincimento del giudice solo qualora queste rispondano ai tradizionali requisiti di gravità, precisione e concordanza.

Ancora, si afferma che «la responsabilità da prodotto difettoso ha natura presunta (e non già oggettiva), poiché prescinde dall'accertamento della colpevolezza del produttore, ma non anche dalla dimostrazione dell'esistenza di un “difetto” del prodotto; sicché grava sul soggetto danneggiato (...) la prova del collegamento causale, non già tra “prodotto” e danno, bensì tra “difetto” e danno»¹⁶⁹.

3.3.4. La nozione di difetto di progettazione in relazione ai sistemi di i.a.: alcuni profili di inadeguatezza.

Come abbiamo più volte ribadito, l'applicazione della disciplina sulla responsabilità per danno da prodotto agli eventi lesivi cagionati dai sistemi di i.a. incontra un problema *specifico* con riferimento ai sistemi di i.a., che è quello di stabilire quando vi sia un difetto¹⁷⁰. Se, infatti, l'esigenza di aggiornare la nozione di “prodotto” e di “produttore” è comune a molti dei più sofisticati ritrovati della tecnica – e può essere soddisfatta piuttosto agevolmente, come mostrano le modifiche proposte dalla Commissione Europea – la presunta inadeguatezza del concetto di “difettosità” è legata precipuamente all'irrompere di sistemi di i.a. capaci di autonomo *decision making*.

Come abbiamo rilevato nei paragrafi precedenti, la nozione di “difetto” presuppone, infatti, un qualche tipo di svista, disattenzione o sottavalutazione da parte del produttore – che non sempre sarà configurabile nel caso in cui il danno sia conseguenza di un'autonoma decisione da parte dei prodotti, dovuta alla loro capacità di modificarsi nel corso del loro ciclo vitale.

¹⁶⁹ Cass., 29 maggio 2013, n. 13458, cit.

¹⁷⁰ EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 28.; A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit., p. 293.

In particolare, i problemi più rilevanti sorgono in relazione ai difetti di progettazione¹⁷¹; per quanto concerne i difetti di fabbricazione, infatti, emergono questioni che non divergono, nella sostanza, da quelle già affrontate nell'ambito della tradizionale disciplina della responsabilità per danno da prodotto difettoso: così, se per la distrazione di un operatore o per il malfunzionamento di un macchinario, in un drone viene installato un codice incompleto, il produttore potrà sempre essere chiamato a rispondere dell'eventuale danno cagionato dal drone, poiché la mera difformità del singolo prodotto rispetto al *design* della serie è sufficiente a dimostrare l'esistenza del difetto.

Per la verità, anche con riferimento ai difetti di progettazione può individuarsi un caso non controverso, relativo a difetti “genetici”, presenti *già nel momento in cui il prodotto viene messo in commercio*: si pensi, ad esempio, quando alla capacità di apprendimento non si accompagni la presenza di misure strutturate di prevenzione degli eventi lesivi o l'adozione di un adeguato sistema di cybersicurezza¹⁷². In queste ipotesi, che rappresentano quella che presumibilmente sarà la “casistica standard” di applicazione delle norme sulla responsabilità da prodotto, non si porranno problemi di rilievo quanto alla conformazione della nozione di “difetto” – alcune difficoltà potrebbero piuttosto emergere quanto all'accertamento del nesso di causalità (vd. *infra*, § 3.5. ss., in questo cap.).

Proprio in relazione al *design*, tuttavia, possono emergere delle ipotesi più complesse, che non hanno precedenti nella casistica tradizionale e che sono collegate alle peculiari modalità di funzionamento dei sistemi di i.a.: ci riferiamo a quelle ipotesi, già citate, in cui il danno si sia verificato a causa di un *comportamento autonomo ed imprevedibile* dell'algoritmo, derivante dall'apprendimento e dall'interazione continua con l'ambiente circostante.

¹⁷¹ S. NAVAS, *Robot Machines and Civil Liability*, in M. Ebers – S. Navas (eds.), *Algorithms and Law*, cit., p. 168; A. BERTOLINI, *Robots as products*, cit., p. 239; G. GUERRA, *Il concetto di difettosità nella realtà che cambia*, cit., p. 272; G. WAGNER, *Robot, Inc.*, cit. p. 605; P.M. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 170.

¹⁷² Vd. U. RUFFOLO, *Responsabilità da A.I. e da algoritmo*, cit. p. 107; ID., *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 247. L'Autore esaurisce il campo dei possibili difetti nell'assenza di “blocchi etici”, ossia di misure che incorporino nella macchina “pensante” i principi del *neminem laedere*. In realtà, come diremo nella parte che segue, l'incorporazione di questi “vincoli etici” non sempre impedisce il verificarsi di eventi lesivi.

Il comportamento dei sistemi di i.a., infatti, come abbiamo cercato di far emergere nel precedente capitolo, non può essere deterministicamente fatto risalire al codice algoritmico, poiché esso deriva, in una certa misura, anche dall'acquisizione di dati dall'ambiente esterno, dalla capacità di apprendimento e dall'eventuale connessione con l'attività di *altri sistemi di i.a.* Va inoltre sottolineato che gli agenti intelligenti che utilizzano tecniche di *machine learning* decidono sulla base di previsioni statistiche che, per loro natura, non sono infallibili, potendo determinare indesiderate conseguenze dannose, più o meno imprevedibili anche per gli stessi sviluppatori, programmatori o produttori¹⁷³.

Quando, allora, il comportamento emergente lesivo potrà essere considerato come derivante da un difetto di progettazione? A meno di non voler considerare la stessa capacità di apprendimento alla stregua di un difetto – conclusione che desterebbe numerose perplessità, atteso che il *machine learning* determina, nel complesso, un evidente miglioramento delle prestazioni della macchina – sarà difficile stabilire quando il sistema intelligente sia stato progettato in maniera difettosa¹⁷⁴. In particolare, non è chiaro quale debba essere lo *standard* di comparazione rispetto al quale è necessario valutare la difettosità del prodotto – se la decisione che, in condizioni ordinarie, avrebbe potuto essere presa da una persona fisica o se, invece, quella che avrebbe preso un algoritmo “ragionevole”¹⁷⁵.

D'altra parte, anche qualora si ritenesse che la deviazione dell'algoritmo rispetto al percorso decisionale ordinario integri la nozione di “difetto”, la lettera della norma

¹⁷³ K.A. CHAGAL-FEFERKORN, *Am I an Algorithm or a Product?*, cit., p. 84, in relazione alla *product liability* statunitense. Analogamente, quanto alla disciplina europea, vd. A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1720; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 463; M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, in M. EBERS – S. NAVAS (eds.), *Algorithms and Law*, cit., p. 58; C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, cit., p. 159; J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, cit., p. 70 ss.; A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 74.

¹⁷⁴ C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, cit., p. 159; J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, cit., p. 70 ss.; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 28; S. BECK, *The problem of ascribing legal responsibility in the case of robotics*, cit., p. 475.

¹⁷⁵ J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, cit., p. 68-71; G. WAGNER, *Robot Liability*, p. 42-5; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1261; parafrasiamo qui il titolo di un testo che ha avuto molto successo in questo ambito, R. ABBOTT, *The Reasonable Robot. Artificial intelligence and the law*, Cambridge University Press, 2020.

parrebbe escludere che il produttore debba rispondere dal danno cagionato dal prodotto intelligente, atteso che l'art. 7 della direttiva prevede un'esimente nel caso in cui il vizio emerga successivamente alla messa in circolazione del prodotto (vd. art. 118, lett. b), nonché un'esimente per il c.d. rischio da sviluppo (vd. art. 118, lett. e).

Se la tendenza in atto è quella di approssimare sempre di più la nozione di *design defect*, così come integrata dalle disposizioni sulle esimenti, ad una valutazione circa la *prevedibilità*, per il produttore, dell'esito lesivo derivante dal prodotto, si finisce allora per cascare nuovamente nel *responsibility gap*, che il riferimento ad un modello oggettivistico mirava ad evitare. Posto, infatti, che la caratteristica qualificante dell'intelligenza artificiale, rispetto ai più tradizionali algoritmi, è proprio l'autonomo e imprevedibile *decision making*, lo spazio per l'individuazione di una responsabilità per il produttore sembra restringersi. La prevedibilità, in concreto, dell'evento algoritmico lesivo è infatti ostacolata non solo dalla capacità dei sistemi di i.a. di prendere decisioni in situazioni di incertezza (vd. *supra*, cap. I, § 5.1), ma anche dall'eventuale interazione con l'ambiente circostante e con altri agenti intelligenti (vd. *supra*, cap. I, § 5.2), nonché, infine, dall'opacità – talvolta insuperabile – dei precisi meccanismi di funzionamento degli algoritmi di *machine learning* (specie se utilizzano reti neurali; vd. *supra*, cap. I, § 5.3).

Le caratteristiche dei sistemi di i.a., d'altra parte, hanno un impatto sulla stessa conformazione strutturale dei difetti di *design*, che cessano di essere vizi necessariamente "seriali", potendo coinvolgere anche soltanto alcuni dei prototipi della serie produttiva¹⁷⁶. Il comportamento dei sistemi di i.a., infatti, non dipende solo dal *codice* installato in fase di progettazione, ma anche dall'acquisizione di nuovi dati attraverso l'interazione con l'ambiente e l'eventuale connessione con altri dispositivi: lo *smart product* è un prodotto le cui attività possono mutare a seconda del contesto in cui si trova e a seconda dell'utilizzo che ne viene fatto dall'utente. Ne consegue che, nell'ambito dell'accertamento di presunti *design defects* relativi a sistemi di i.a., sembra che non potrà applicarsi quella agevolazione probatoria a cui la Corte di Giustizia ha fatto talvolta ricorso, in base alla quale «l'accertamento di un potenziale difetto di prodotti appartenenti

¹⁷⁶ A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit., p. 292; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1264, che giunge a chiedersi se questa asimmetria tra tradizionali difetti di design e difetti derivanti dal *self learning* renda difficilmente ascrivibile la seconda categoria alla prima.

al medesimo gruppo o alla medesima serie di produzione consente di qualificare come difettosi tutti i prodotti di tale gruppo o di tale serie, senza che occorra dimostrare il difetto del prodotto interessato»¹⁷⁷.

3.3.5. Nuovi criteri per la valutazione del difetto: le modifiche proposte dalla proposta di revisione della *Product liability directive*.

Nel tentativo di definire con maggiore precisione la nozione di “difetto” e di rispondere alle difficoltà di applicazione della disciplina con riferimento ai nuovi prodotti intelligenti, la proposta di revisione della direttiva aggiunge nuovi criteri di valutazione della difettosità del prodotto (lett. c, d, f, g, h), riformulando altresì quelli già esistenti all’art. 6 della Direttiva (lett. a, b, e, vd. *supra*, § 3.3.)¹⁷⁸, e sostanzialmente riprodotti all’art. 117, co. 1, cod. cons. Nessuna modifica, invece, è stata proposta con riferimento alle esimenti, nonostante i già citati auspici della dottrina dominante.

L’art. 6 della proposta di revisione prevede:

«Un prodotto è considerato difettoso quando non offre la sicurezza che il grande pubblico può legittimamente attendersi tenuto conto di tutte le circostanze, tra cui:

- a) la presentazione del prodotto, comprese le istruzioni per l'uso, l'installazione e la manutenzione;
- b) l'uso e abuso ragionevolmente prevedibili del prodotto;
- c) gli effetti sul prodotto dell'eventuale capacità di continuare ad imparare dopo la sua diffusione;
- d) gli effetti sul prodotto di altri prodotti che ci si può ragionevolmente attendere siano utilizzati insieme al prodotto;
- e) il momento in cui il prodotto è stato immesso sul mercato o messo in servizio oppure, qualora il fabbricante mantenga il controllo sul prodotto dopo tale momento, il momento in cui il prodotto è uscito dal controllo del fabbricante;

¹⁷⁷ Corte di Giustizia UE, 5 marzo 2015, *Boston Scientific Medizintechnik GmbH*, cit.; si trattava, in questo caso, di difetto inerente a *pacemaker*; vd. in questo senso AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1724.

¹⁷⁸ Nell’attuale formulazione, le circostanze elencate sono le seguenti: «a) la presentazione del prodotto, b) l'uso al quale il prodotto può essere ragionevolmente destinato, c) il momento della messa in circolazione del prodotto».

- f) i requisiti di sicurezza del prodotto, compresi i requisiti di cibersicurezza rilevanti per la sicurezza;
- g) qualunque intervento di un'autorità di regolamentazione o di un operatore economico di cui all'articolo 7 in relazione alla sicurezza del prodotto;
- h) le specifiche aspettative degli utenti finali cui è destinato il prodotto».

Il tentativo di individuare circostanze valutative del difetto tarate sugli *smart products* è da salutare con favore. La proposta prende infatti lodevolmente atto di alcune caratteristiche tipiche dei prodotti intelligenti, che possono incidere sulla formazione delle aspettative da parte dei consumatori e sulla conseguente conformazione dei difetti (ad es. il fatto che spesso i sistemi tecnologici più sofisticati vengono utilizzati *in rete* con altri sistemi, vd. lett. d), o, il fatto che essi sono talvolta rivolti a specifiche categorie di consumatori, vd. lett. h), o, ancora, il fatto che in alcuni casi il produttore mantiene il controllo sul prodotto – attraverso la possibilità di effettuare aggiornamenti del *software* – anche dopo la messa in commercio, vd. lett. e).

Ciò che il testo non chiarisce, tuttavia, è come debba essere valutata la caratteristica più dirompente – nell’ottica dell’attribuzione della responsabilità – dei sistemi di i.a., ossia la capacità di apprendimento. La lett. c) la inserisce tra i criteri che devono essere apprezzati dal giudice; non si capisce, tuttavia, *in che modo* essa debba influire nel giudizio di difettosità. La capacità di *machine learning* deve essere considerata come circostanza che induce a ritenere che il prodotto sia difettoso, o, invece, l’auto-apprendimento deve considerarsi come elemento che contribuisce ad escludere la difettosità del prodotto¹⁷⁹? È ovvio che il *machine learning* non può essere considerato un difetto *ex se*¹⁸⁰; tuttavia, non si può non considerare che è la capacità di auto-apprendimento dei sistemi di i.a. potrebbe sia aumentare le aspettative di sicurezza del

¹⁷⁹ Lo nota anche O. DHEU E AL., *The European Commission's Approach To Extra-Contractual Liability and AI – A First Analysis and Evaluation of the Two Proposals*, CiTiP Working Paper Series, 6 ottobre 2022, in SSRN, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4239792, p. 35-36; vd. anche G. SPINDLER, *Different Approaches for Liability of Artificial Intelligence – Pros and Cons – the New Proposal of the EU Commission on Liability for Defective Products and AI Systems*, 13 febbraio 2023, in SSRN, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354468, p. 11-12.

¹⁸⁰ J. DE BRUYNE – J. TANGHE, *Liability for Damage Caused by Autonomous Vehicles: A Belgian Perspective*, in 8 *Journal of European Tort Law*, iss. 3, p. 370.

prodotto – ai sensi dell’art. 117, co. 1, cod. cons. – sia diminuirle. Sarebbe forse stata necessaria una presa di posizione più trasparente e radicale su quale debba essere lo *standard* di valutazione della difettosità, in caso di comportamento emergente ed imprevedibile di un sistema di i.a.

3.3.6. Verso una nuova nozione di *design defect*? Le proposte della dottrina.

Alla luce delle problematiche che l’autonomia dei sistemi di i.a. pone in relazione alla valutazione normativa della difettosità del prodotto, in dottrina sono state proposte diverse interpretazioni evolutive del *consumer expectations test*.

Ci concentreremo qui sulla specifica questione della configurabilità del difetto nell’ipotesi di comportamento emergente ed imprevedibile del prodotto “soggettivizzato”, che costituisce il *proprium* dei sistemi di intelligenza artificiale. Non ci dilungheremo, invece, sui profili di difettosità che possono derivare dalla mancata predisposizione di meccanismi volti ad ostacolare aggressioni perpetrate da cybercriminali, *hacker* o terroristi – aspetto che, benché di estrema rilevanza nell’attuale contesto di interconnessione tra dispositivi e interazione con l’ambiente circostante, non pone problemi ontologicamente diversi rispetto a quelli posti dai tradizionali prodotti¹⁸¹.

3.3.6.1. Sostanziale eliminazione del requisito del difetto.

Secondo una parte della dottrina, il modo migliore per tutelare il consumatore che abbia subito un danno a causa di un comportamento emergente di un sistema di i.a. consisterebbe nell’eliminazione del requisito del difetto¹⁸²: per ottenere il risarcimento, il

¹⁸¹ Così anche A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit. p. 294.

¹⁸² Evoca questa possibilità – pur scartandola – J.S. BORGHETTI, *Civil Liability for Artificial Intelligence: What Should its Basis Be?*, in *Revue des Juristes de Sciences Po*, 2019, n. 17, p. 99 ss.; sembra invece giungere a tale conclusione G. FINOCCHIARO, *Intelligenza artificiale e responsabilità*, in *Contr. impr.*, 2020, n. 2, p. 730-31, che sottolinea la necessità di «dirigersi verso un modello di responsabilità che sia un sistema puro di allocazione del rischio, prescindendo dalla ricerca dell’errore» (p. 731); nella dottrina statunitense, la tesi è autorevolmente sostenuta da K.S. ABRAHAM – R.L. RABIN, *Automated vehicles and manufacturer responsibility for accidents: a new legal regime for a new era*, in *Va. L. Rev.*, 2019, p. 142 ss.; va ribadito,

danneggiato potrebbe dunque limitarsi alla dimostrazione del nesso di causalità tra prodotto ed evento lesivo, senza che sia necessaria la prova di una “condizione di anormale pericolosità” del prodotto. Si tratta del medesimo orientamento che la giurisprudenza di merito, in varie occasioni, ha cercato di seguire, alleggerendo l’onere probatorio a carico del danneggiato fino a ritenere che fosse sufficiente la dimostrazione di un malfunzionamento del prodotto per dimostrare la sussistenza del difetto, spettando poi al produttore l’onere di fornire un’eventuale prova liberatoria ai sensi dell’art. 118 cod. cons.¹⁸³.

Un tale orientamento faciliterebbe di non poco la posizione del danneggiato, ma renderebbe *puro* un sistema di responsabilità oggettivo che è invece concepito, a livello positivo, come mitigato da valutazioni inerenti, *lato sensu*, la rimproverabilità del produttore – valutazioni che, come abbiamo già rilevato, proprio nel requisito del difetto hanno il loro perno¹⁸⁴ (vd. *supra*, § 3.3. ss.). D’altra parte, la giurisprudenza di legittimità, come già rilevato, afferma costantemente che il difetto del prodotto non può essere dedotto dal verificarsi di un evento dannoso, ma deve invece essere oggetto di un giudizio autonomo¹⁸⁵.

De jure condendo, a tale risultato potrebbe pervenirsi mediante un alleggerimento probatorio, attraverso l’applicazione dell’*inference doctrine* – corollario del principio del *res ipsa loquitur* – in base alla quale il solo fatto che il prodotto, utilizzato secondo le modalità ordinarie, abbia causato un danno fa presumere l’esistenza del difetto¹⁸⁶. È questa la strada che pare essere stata seguita dalla Commissione europea, nella proposta di revisione della direttiva sulla responsabilità per danno da prodotto. Ai sensi dell’art. 9, § 2, lett. c), infatti, la difettosità del prodotto può essere presunta nel caso in cui l’attore provi che «il danno è stato causato da un malfunzionamento evidente del prodotto durante

tuttavia, che negli Stati Uniti la *product liability* è concepita come una vera e propria responsabilità per colpa, a seguito dell’entrata in vigore del già citato *Restatement (Third) of Torts*.

¹⁸³ Vd. ad es. le già citate Trib. Monza, 10 febbraio 2015, in *Danno resp.*, 2015, p. 950; Trib. Pontedera, 16 marzo 2011, in *Danno resp.*, 2012, p. 67; Trib. Roma, 4 dicembre 2003 e Trib. Roma, 3 novembre 2003, in *Danno resp.*, 2004, p. 527; Trib. Benevento, 24 febbraio 2006, in *Danno resp.*, 2006, p. 1254.

¹⁸⁴ G. PONZANELLI, *Causa probabile del danno e responsabilità del produttore*, cit., p. 532.

¹⁸⁵ Cass. civ., sez. III, 29 luglio 2015, n. 15851; Cass. civ., sez. III, 26 giugno 2015, n. 13225; Cass. civ., sez. III, 29 maggio 2013, n. 13458; Cass. civ., ord. 23 maggio 2013, n. 12665; Cass. civ., sez. III, 15 marzo 2007, n. 6007.

¹⁸⁶ D.C. VLADECK, *Machines Without Principals: Liability Rules and Artificial Intelligence*, in 89 *Wash. L. Rev.*, 2014, p. 128; S. DORMONT, *Quel régime de responsabilité pour l’intelligence artificielle*, in *Contrats, concurrence, consommation*, 2018, études 19, § 4.

l'utilizzo normale o in circostanze ordinarie»; ancora, il § 4, lett. a), dell'art. 9 stabilisce che, qualora l'autorità giudiziaria ritenga che l'attore stia incontrando difficoltà eccessive nel provare il carattere difettoso del prodotto, a causa della sua complessità tecnica o scientifica, la difettosità si può presumere nel caso in cui «il prodotto ha contribuito a cagionare il danno».

3.3.6.2. L'agente modello umano.

Secondo alcuni autori, viceversa, la difettosità del prodotto dovrebbe essere valutata in rapporto all'ipotetico comportamento di un agente modello umano¹⁸⁷. Si tratterebbe di verificare – adottando un metodo affine a quello abitualmente utilizzato per valutare la colpa dell'agente – se una persona ordinariamente diligente, nello svolgere le funzioni delegate alla macchina, avrebbe potuto evitare il verificarsi dell'evento lesivo.

La *ratio* di tale orientamento consisterebbe nel fatto che i sistemi di i.a. – essendo progettati, venduti e socialmente accettati come sistemi più funzionali rispetto a quelli tradizionali – creano una legittima aspettativa di sicurezza nei consumatori e, per questo motivo, dovrebbero essere considerati privi della «sicurezza che ci si può legittimamente attendere» quando causano danni che una persona fisica sarebbe stata in grado di evitare.

Un orientamento di questo tipo presta tuttavia il fianco ad alcune critiche. Gli errori che può commettere un sistema di i.a., infatti, sono spesso diversi da quelli che commettono ordinariamente le persone fisiche; difficilmente, dunque, un paragone di questo tipo potrà essere ragionevole¹⁸⁸.

Prendiamo, ancora una volta, l'esempio dell'*automotive*. Le cause degli incidenti automobilistici sono oggi sostanzialmente riconducibili ad errori di distrazione dei conducenti, così come a sonnolenza o alterazione delle percezioni dovuta all'assunzione di alcol o stupefacenti. Queste, di certo, non sono circostanze che possono coinvolgere

¹⁸⁷ A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit. p. 1019; D.C. VLADECK, *Machines Without Principals*, cit., p. 128; R. ABBOTT, *The Reasonable Robot*, cit., p. 59; J.K. GURNEY, *Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles*, in P. Lin e al. (eds.), *Robot Ethics 2.0*, cit., p. 55 ss.

¹⁸⁸ G. WAGNER, *Robot, Inc.*, cit., p. 605; M.A. GEISTFELD, *A Roadmap for Autonomous Vehicles*, cit., p. 1644–47; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1261; A. FUSARO, *Quale modello di responsabilità per la robotica avanzata?*, cit., p. 1350.

sistemi di guida autonoma; di contro, questi ultimi possono causare incidenti a causa di difficoltà nel riconoscimento di immagini o suoni. Come dimostra il fenomeno dell'*adversarial machine learning* (vd. *supra*, cap. I, § 2.2), infatti, può bastare un adesivo o un graffito su un segnale di “stop” per confondere il sistema di guida autonoma¹⁸⁹.

È ovvio che, qualora si intendesse paragonare l'errore umano a quello algoritmico, in un caso come quello appena esemplificato si giungerebbe ad identificare la presenza di un difetto: è evidente, infatti, che qualsiasi conducente umano avrebbe riconosciuto il segnale di “stop” e, in un'ottica di media diligenza, si sarebbe fermato.

Se si aderisse all'orientamento che valuta la difettosità dell'algoritmo sulla base di uno *standard* “umano”, si perverrebbe dunque alla soppressione *de facto* del requisito del difetto, poiché, nei singoli casi concreti, quando l'i.a. sbaglia lo fa in maniera piuttosto *stupida*¹⁹⁰, se comparato con il corrispondente comportamento umano mediamente diligente.

3.3.6.3. L'agente modello algoritmico.

Secondo un'altra parte della dottrina¹⁹¹, la circostanza che un singolo evento lesivo avrebbe potuto essere evitato da un conducente umano diligente è da ritenersi del tutto irrilevante; lo *standard* di valutazione della difettosità del prodotto dovrebbe invece essere costituito dalla comparazione con un modello di algoritmo *mediamente accurato*. È questa la soluzione che, a nostro parere, è maggiormente coerente con lo spirito e con la lettera della normativa: è infatti sulla base del funzionamento e sulle condizioni di sicurezza *medie* dei prodotti che si formano le legittime aspettative dei consumatori.

Qualora si aderisse a tale prospettiva, il rischio che andrebbe in ogni caso scongiurato è che il sistema intelligente che ha causato il danno venga ritenuto difettoso

¹⁸⁹ K. EYKHOLT E AL., *Robust Physical-World Attacks on Deep Learning Visual Classification*, cit.

¹⁹⁰ J.M BISHOP, *Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It*, in *Front. Psychol.*, 5 gennaio 2021, disponibile a questo link: <https://doi.org/10.3389/fpsyg.2020.513474>. Vd., a tal proposito, gli esempi di errori dei sistemi di *machine learning* che abbiamo riportato *supra*, cap. I, § 2.2.

¹⁹¹ G. WAGNER, *Robot, Inc.*, cit., p. 605; ID., *Robot Liability*, cit., p. 44-45; J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, p. 63 ss.; concordi anche G. GUERRA, *Il concetto di difettosità nella realtà che cambia*, cit., p. 278; J. DE BRUYNE E AL., *Tort Law and Damage Caused by AI Systems*, in J., De Bruyne (ed.), *Artificial Intelligence and the Law*, Intersentia, 2021, p. 383.

ogniquale volta si dimostri che sul mercato esiste un algoritmo che, nel caso concreto, avrebbe evitato l'incidente¹⁹²; da un lato, infatti, una simile impostazione sarebbe incompatibile con il già citato art. 117, co. 2, cod. cons., che esclude che un prodotto venga considerato difettoso per il solo fatto che un prodotto più perfezionato sia stato in qualunque tempo messo in commercio (vd. anche art. 6, co. 2, dir. 85/374); dall'altro, non è detto che un algoritmo che, nel caso specifico, avrebbe evitato il danno sia, solo per questo motivo, complessivamente più sicuro¹⁹³. Nella comparazione tra la sicurezza del sistema che ha causato il danno e quella media di prodotti analoghi, insomma, non deve essere valutata la singola *performance* in relazione ad un evento lesivo specifico, ma piuttosto la prestazione complessiva dell'algoritmo, da valutarsi eventualmente sulla base di analisi statistiche¹⁹⁴. A tal fine, potrebbe essere opportuna l'individuazione di una soglia di riferimento – relativa al rapporto tra eventi lesivi verificatisi e utilizzo complessivo del *software* – al di sopra della quale l'algoritmo potrà essere considerato difettoso.

3.4. I rapporti tra *product safety* e *product liability*: verso un'adozione della *preemption doctrine*?

La disciplina della responsabilità per danno da prodotto difettoso trova il suo complemento nella normativa sulla sicurezza dei prodotti, che definisce *ex ante* quali sono i requisiti che devono essere soddisfatti affinché la merce possa essere messa in circolazione.

Nonostante l'obiettivo comune (quello, *lato sensu*, di fornire un'adeguata tutela alla salute dei consumatori), i due *corpus* normativi hanno senz'altro caratteri, funzioni e ambiti di applicazione non coincidenti. La disciplina sulla *product safety* – nella sua veste

¹⁹² G. WAGNER, *Robot, Inc.*, cit., p. 608; J. DE BRUYNE E AL., *Tort Law and Damage Caused by AI Systems*, cit., p. 383.

¹⁹³ J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, cit. p. 70; J. DE BRUYNE E AL., *Tort Law and Damage Caused by AI Systems*, cit., p. 383; A. FUSARO, *Quale modello di responsabilità per la robotica avanzata?*, cit., p. 1351.

¹⁹⁴ J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, cit., p. 63 ss.; J. DE BRUYNE E AL., *Tort Law and Damage Caused by AI Systems*, cit., p. 383; concorde anche G. GUERRA, *Il concetto di difettosità nella realtà che cambia*, cit., p. 278.

generale e in quella settoriale – mira a «*garantire che i prodotti immessi sul mercato ovvero in libera pratica siano sicuri*» (art. 102, co. 1, cod. cons.): essa interviene *ex ante*, delineando le condizioni di accesso al mercato. La normativa sulla *product liability*, al contrario, si applica, per sua natura, *dopo* che si è verificato l'evento lesivo e ha la funzione di assicurare un ristoro economico al soggetto danneggiato dal prodotto; interviene, insomma, laddove la disciplina sulla sicurezza non sia stata capace di impedire la commercializzazione di prodotti difettosi.

Malgrado le due prospettive siano dunque evidentemente diverse, la disciplina sulla *product safety* finisce necessariamente per influenzare l'interpretazione della disciplina sulla responsabilità per danno da prodotto difettoso, fornendo criteri ermeneutici stabili per la valutazione della difettosità del prodotto¹⁹⁵ – valutazione che, come già ribadito più volte, costituisce uno degli aspetti più problematici dell'intera regolazione della *product liability*.

Prendendo atto della crescente complementarietà tra le due discipline, alcuni studiosi – ispirandosi al modello statunitense della c.d. *preemption doctrine* – sono giunti a prospettare una totale integrazione tra la normativa in materia di *product liability* e quella in materia di *product safety*, auspicando un'assimilazione tra la nozione di “prodotto conforme” agli standard in materia di sicurezza e quella di “prodotto non difettoso”¹⁹⁶. Quello del regime applicabile al c.d. “danno da prodotto conforme” è, in

¹⁹⁵ Sui complessi rapporti tra disciplina in tema di sicurezza e disciplina sulla responsabilità per danno da prodotto difettoso la letteratura è ormai vastissima. Ci limitiamo qui a rinviare ad alcuni scritti di particolare rilievo: F. CAFAGGI, *A Coordinated Approach to Regulation and Civil Liability in European Law: Rethinking Institutional Complementarities*, in F. Cafaggi (a cura di), *The Institutional Framework of European Private Law*, Oxford University Press, 2006, p. 191 ss.; G. HOWELLS, *The Relationship Between Product Liability and Product Safety - Understanding a Necessary Element in European Product Liability Through a Comparison with the U.S. Position*, in Wasburn L. J., 2000, vol. 29, p. 305 ss.; M. FAURE, *Product Liability and Product Safety in Europe: Harmonization or Differentiation?*, in *Kyklos*, 2000, vol. 53, p. 467 ss.; A. CORDIANO, *Sicurezza dei prodotti e tutela preventiva dei consumatori*, Cedam, 2005, p. 47 ss. e p. 183 ss.

¹⁹⁶ In particolare, è Enrico Al Mureden che si è fatto esplicito promotore, in Italia, di questo indirizzo; vd. la seconda parte dell'articolo U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1713 ss.; E. AL MUREDEN, *La sicurezza dei prodotti e la responsabilità del produttore*, Torino, 2017; ID., *Il danno da “prodotto conforme”. Le soluzioni europee e statunitensi nella prospettiva del “Transatlantic Trade and Investment Partnership” (T.T.I.P.)*, in *Contr. e impr.*, 2015, p. 388; ID., *Sicurezza “ragionevole” degli autoveicoli e responsabilità del produttore nell'ordinamento giuridico italiano e negli Stati Uniti*, in *Contr. e impr.*, 2012, p. 1506; nello stesso senso, specificamente in materia di sistemi di i.a., C. AMATO, *Product Liability and Product Security: Present and Future*, in S. Lohsse e al. (eds.), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 79 ss.; E. MARCHISIO, *In support of “no-fault” civil liability rules for artificial intelligence*, in *SN Social*

realtà, un interrogativo che, come vedremo a breve, da tempo è dibattuto in dottrina e giurisprudenza. Fin dalle prime applicazioni della disciplina della responsabilità per danno da prodotto, infatti, si è posto il problema di valutare se il rispetto degli *standard* tecnici di sicurezza del prodotto – che regolano le condizioni e i requisiti di accesso al mercato delle merci – basti a far considerare un prodotto “non difettoso”, e, dunque, ad escludere la responsabilità oggettiva del produttore, in caso di eventi lesivi, o se invece la conformità agli stessi debba essere considerata condizione necessaria ma non sufficiente a tale scopo.

La prima opzione, applicata alla materia dei danni causati da sistemi di i.a., avrebbe il vantaggio di superare il già citato *impasse* nell’individuazione del difetto algoritmico – a scapito, tuttavia, della posizione del danneggiato, che si troverebbe di fatto privo di tutela risarcitoria, dal momento che un prodotto immesso sul mercato, attraverso autorizzazione o autocertificazione, sarà presumibilmente conforme agli *standard* stabiliti dalla normativa sulla *product safety*.

3.4.1. Product Safety e tecniche di armonizzazione. Cenni.

In via preliminare, può essere opportuno qualche breve cenno alla normativa sulla sicurezza dei prodotti.

Si tratta, innanzitutto, di una disciplina strutturata su due piani. A livello generale, vi è la *legislazione orizzontale* – costituita dalla dir. 01/95/CE (*General Product Safety Directive*, c.d. GPSD)¹⁹⁷ – che si riferisce a qualsiasi tipologia di prodotto. Ai sensi dell’art. 1, co. 2, essa si applica «nella misura in cui non esistano, nell’ambito della normativa comunitaria, disposizioni specifiche aventi lo stesso obiettivo che disciplinano la sicurezza dei prodotti in questione». Tali “disposizioni specifiche” costituiscono la c.d. *legislazione verticale*, il secondo livello di regolazione della *product safety*, composta da una serie di direttive che disciplinano gli *standard* di sicurezza di specifiche categorie di prodotti: tra le materie regolate, vi è quella dei farmaci (dir. 2001/83/CE); dei presidi

Sciences, 2021, 1: 54, p. 10 ss.; in chiave problematica vd. G. GUERRA, *Il concetto di difettosità nella realtà che cambia*, cit., p. 273 e 276.

¹⁹⁷ Direttiva 2001/95/ce del Parlamento Europeo e del Consiglio del 3 dicembre 2001 relativa alla sicurezza generale dei prodotti.

medico-chirurgici (dir. 2007/47/CE); dei cosmetici (dir. 76/768/CEE); del materiale elettrico (dir. 2006/95/CE); degli autoveicoli (dir. 2007/46/CE); dei motoveicoli a due e tre ruote (dir. 2002/24/CE); degli pneumatici (dir. 1992/23/CE; dir. 2001/43/CE; dir. 2005/11/CE); dei giocattoli (dir. 2009/48/CE), delle macchine (dir. 2006/42/CE)¹⁹⁸. La dir. 01/95/CE resta in ogni caso applicabile anche in questi settori, unicamente per quanto concerne «*gli aspetti e i rischi o le categorie di rischi non soggetti a tali requisiti* [quelli della legislazione verticale]» (art. 1, co. 2, GPSD).

La normativa sulla *product safety* non pretende – com’è inevitabile, in una società come quella odierna, innervata di rischi di ogni tipo – che il prodotto sia immune da rischi. Come per la definizione di “prodotto difettoso”, anche l’art. 2, lett. b), dir. 01/95/CE si accontenta di un concetto “relazionale” di sicurezza, espresso non solo in termini di “assenza di rischi” ma anche di «rischi minimi, compatibili con l’impiego del prodotto e considerati accettabili nell’osservanza di un livello elevato di tutela della salute e della sicurezza delle persone», in funzione di vari elementi, tra i quali la norma considera, in via meramente esemplificativa, le sue «caratteristiche» (n. 1), il suo eventuale «effetto» su altri prodotti (n. 2), la sua «presentazione» (n. 3) e le «categorie di consumatori» cui esso è destinato (n. 4), con particolare riguardo a quelle maggiormente vulnerabili (bambini e anziani)¹⁹⁹. Inoltre, parallelamente a quanto stabilito dall’art. 6, co. 2, della direttiva sul danno da prodotto, l’art. 2 lett. b) dir. 01/95/CE stabilisce che «la possibilità di raggiungere un livello di sicurezza superiore o di procurarsi altri prodotti che presentano un rischio minore non costituisce un motivo sufficiente per considerare un prodotto come non sicuro o pericoloso» (vd. anche art. 103, co. 2, cod. cons.). Il “prodotto pericoloso” è invece concepito come “il negativo” del prodotto sicuro: ai sensi dell’art. 2, lett. c), dir. 01/95/CE è pericoloso «qualsiasi prodotto che non risponda alla definizione di prodotto sicuro»²⁰⁰.

La definizione generale di sicurezza fornita dalla dir. 01/95/CE viene poi concretizzata nelle direttive verticali, che stabiliscono i *requisiti essenziali* che i prodotti

¹⁹⁸ La c.d. “direttive macchine” – particolarmente rilevante nel settore che qui ci interessa, poiché destinata ad applicarsi quantomeno nel settore della robotica – è attualmente soggetta ad un procedimento di revisione, vd. Proposta di Regolamento del Parlamento Europeo e del Consiglio sui prodotti macchina; Bruxelles, COM (2021) 202 final, 21 aprile 2021, che ha proprio l’obiettivo di adeguare la normativa sulla sicurezza dei macchinari ai nuovi prototipi di “robotica intelligente”.

¹⁹⁹ La definizione di “prodotto sicuro” è trasposta invariata nell’art. 103, co. 1, lett. a), cod. cons.

²⁰⁰ La definizione di “prodotto pericoloso” è trasposta invariata nell’art. 103, lett. b), cod. cons.,

devono soddisfare per circolare liberamente nel mercato europeo. Esse, tuttavia, non illustrano le modalità tecniche attraverso le quali tali requisiti devono essere soddisfatti: indicano il traguardo, insomma, ma non il percorso per arrivarci²⁰¹. I dettagli sono delineati attraverso il rinvio mobile a *norme tecniche emanate da organismi privati* (cd. “enti di normalizzazione”), che presentano il vantaggio di essere aggiornate di continuo secondo l’evolversi delle conoscenze tecniche e scientifiche e di rappresentare lo stato dell’arte di una determinata materia²⁰².

Tra le norme tecniche, un rilievo particolare è stato assunto, a partire dalla Risoluzione del Consiglio delle Comunità Europee del 1985²⁰³ sul c.d. “Nuovo Approccio” (*New Legal Framework*, NLF), dalle c.d. “norme armonizzate”, ossia *standard* elaborati dagli organismi europei di normalizzazione (OEN; *European Standardisation Organisations*, ESOs), che hanno l’obiettivo di favorire una maggiore armonizzazione tra le normative interne e, di conseguenza, di agevolare la circolazione dei beni²⁰⁴. Le norme armonizzate sono adottate in base ad una procedura particolare, che comincia con un mandato da parte della Commissione e termina con la pubblicazione dei riferimenti della norma adottata nella Gazzetta ufficiale dell’Unione europea. In particolare, gli organismi europei di normalizzazione ai quali la Commissione può richiedere l’elaborazione degli *standard* tecnici sono il Comitato europeo di

²⁰¹ In questi termini U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, in *Resp. civ. prev.*, fasc. 1, 2005, contenuto in E. Al Mureden (a cura di), *La sicurezza dei prodotti e la responsabilità del produttore. Casi e materiali*, Giappichelli, 2020, p. 70.

²⁰² U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 63.

²⁰³ Risoluzione del Consiglio 85/C,136/01, del 7 maggio 1985, relativa ad una nuova strategia in materia di armonizzazione tecnica e standardizzazione. Il testo, nel corso degli anni, ha subito varie modifiche, che tuttavia ne hanno sempre confermato l’impianto.

²⁰⁴ Per un approfondimento sul funzionamento del processo europeo di armonizzazione e standardizzazione, rinviamo a COMMISSIONE EUROPEA, *La guida blu all’attuazione della normativa UE sui prodotti 2022*, 29 giugno 2022, 2022/C 247/01, nonché ai lavori di M.P. EGAN, *Constructing a European Market: Standards, Regulation, and Governance*, Oxford University Press, 2001 e B. VAN LEEUWEN, *European Standardisation of Services and Its Impact on Private Law: Paradoxes of Convergence*, Hart Publishing, 2017. Sullo specifico problema della delega di poteri regolatori ad enti privati e dello scarso controllo democratico sull’esercizio di tali poteri vd. G. SPINDLER, *Market processes, standardisation and tort law*, in *European Law Journal*, 2002, vol. 4, iss. 3, p. 316 ss. R. VAN GESTEL – H.W. MICKLITZ, *European Integration through Standardization: How judicial review is breaking down the club house of private standardization bodies*, in *50 Common Market Law Review*, 2013, p. 145 ss.; con riferimento all’AI Act vd. M. EBERS, *Standardizing AI. The Case of the European Commission’s Proposal for an ‘Artificial Intelligence Act’*, in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, cit., p. 340 ss.

normalizzazione (CEN), il Comitato europeo di normalizzazione elettrotecnica (CENELEC) e l'Istituto Europeo per le norme di Telecomunicazioni (ETSI, *European Telecommunications Standards Institute*).

Una volta elaborati, gli *standard* vengono inviati alla Commissione, che effettua un controllo sulla rispondenza di questi rispetto alla richiesta di armonizzazione, nonché rispetto ai principi ispiratori della disciplina generale in materia di standardizzazione. Va comunque rilevato che la Commissione normalmente non esamina nel merito gli *standard* tecnici emanati dagli organi di standardizzazione, ma si limita ad un'operazione di mera comparazione formale. Un esame completo del contenuto delle norme armonizzate, compresi gli aspetti tecnici, non solo sovraccaricherebbe la Commissione dal punto di vista tecnico e delle risorse umane, ma sarebbe anche diametralmente opposto alla natura e allo scopo del c.d. "Nuovo approccio", che – tra i suoi obiettivi fondativi – ha proprio quello di delegare la definizione tecnica dei requisiti di sicurezza ad enti dotati di meccanismi decisionali più agili²⁰⁵.

Le norme armonizzate sono poi trasposte in norme tecniche nazionali a cura degli enti nazionali di normalizzazione²⁰⁶ e sostituiscono le norme tecniche di sicurezza vigenti nei diversi paesi europei, che spesso contrastano tra di loro.

La peculiarità delle norme armonizzate (e delle equivalenti versioni nazionali) rispetto agli altri *standard* tecnici elaborati a livello europeo consiste nel fatto che l'adesione alle prime, da parte delle imprese, assicura una presunzione di sicurezza dei prodotti (vd. art. 3, co. 2, secondo periodo, dir. 01/95/CE e art. 105, co. 2, cod. cons.)²⁰⁷, mentre la conformità alle seconde costituisce soltanto uno dei vari elementi che deve essere preso in considerazione dalle autorità preposte al controllo per valutare l'eventuale (in)sicurezza del prodotto (vd. art. 3, co. 3, dir. 01/95/CE e art. 105, co. 3, cod. cons.)²⁰⁸.

²⁰⁵ M. EBERS, *Standardizing AI*, cit. p. 340.

²⁰⁶ U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 63. Tra gli enti nazionali di normalizzazione vi sono l'Ente Nazionale Italiano di Unificazione (UNI) e il Comitato Elettrotecnico Italiano (CEI) in Italia; il *Deutsches Institut für Normung* (DIN) in Germania; la *British Standardization Institution* (BSI) nel Regno Unito; l'*Association Française de Normalisation* (AFNOR) in Francia.

²⁰⁷ L'art. 105, co. 2, cod. cons. si riferisce alle norme armonizzate in questo modo «norme nazionali non cogenti che recepiscono le norme europee i cui riferimenti sono stati pubblicati dalla Commissione europea nella Gazzetta Ufficiale delle Comunità europee a norma dell'articolo 4 della direttiva 2001/95/CE, del Parlamento europeo e del Consiglio, del 3 dicembre 2001».

²⁰⁸ Oltre agli standard europei non armonizzati, le autorità di controllo devono prendere in considerazione anche: le norme in vigore nello Stato membro in cui il prodotto è commercializzato; le raccomandazioni

La suddetta presunzione è da considerarsi *iuris tantum*, determinando dunque un'inversione dell'onere della prova nella valutazione della sicurezza del prodotto: in caso di osservanza delle norme armonizzate, spetterà all'autorità preposta al controllo dimostrare l'eventuale non conformità del prodotto ai requisiti di sicurezza stabiliti dalla legislazione verticale o orizzontale.

Pur essendo liberi di aderire o meno alle norme armonizzate, i produttori sono dunque incentivati a farlo, poiché la conformità consente loro di essere esonerati dalla difficile opera interpretativa dei requisiti stabiliti in via generale dalle direttive e, successivamente, dall'onere della prova circa la sicurezza del prodotto²⁰⁹.

Quanto all'attestazione della conformità del prodotto alle norme tecniche, essa può avvenire – a seconda della tipologia di prodotto – o per mezzo della certificazione di organi specializzati o attraverso una dichiarazione di conformità redatta e documentata dal fabbricante stesso²¹⁰. Per la maggior parte dei prodotti, il controllo *pre-marketing* è effettuato direttamente dal produttore, che – apponendo il marchio CE – certifica la conformità del prodotto agli standard tecnici²¹¹. La *ratio* – tipica del c.d. “Nuovo Approccio” – si ispira all'idea per cui «Il produttore, possedendo le conoscenze dettagliate relative al processo di progettazione e produzione, si trova nella posizione migliore per eseguire la procedura completa di valutazione della conformità. La procedura di valutazione della conformità dovrebbe quindi rimanere obbligo del solo fabbricante»²¹².

Per un numero ridotto di prodotti invece – si pensi, in particolare, ai farmaci – è un'autorità pubblica (ad es., l'Agenzia Europea per i Medicinali) ad effettuare una

della Commissione relative ad orientamenti sulla valutazione della sicurezza dei prodotti; i codici di buona condotta in materia di sicurezza dei prodotti vigenti nel settore interessato; gli ultimi ritrovati della tecnica; la sicurezza che i consumatori possono ragionevolmente attendere (vd. art. 3, co. 3, dir. 01/95/CE e art. 105, co. 3, cod. cons.).

²⁰⁹ M. VEALE, *Demystifying the Draft EU Artificial Intelligence Act*, in *Computer Law Review International*, 2021, n. 4, p. 105; va comunque sottolineato che le norme armonizzate possono acquisire carattere vincolante, se richiamate in leggi o regolamenti che stabiliscono i requisiti essenziali di sicurezza.

²¹⁰ U. CARNEVALI, *La norma tecnica da regola di esperienza a norma giuridicamente rilevante. Riconoscimento storico e sistemazione teorica. Ruolo dell'UNI e del CEI*, in *Resp. civ. e prev.*, 1997, contenuto in E. Al Mureden (a cura di) *La sicurezza dei prodotti e la responsabilità del produttore. Casi e materiali*, cit., p. 52.

²¹¹ Ovviamente, l'auto-attestazione non preclude, in caso di giudizio intentato dal danneggiato, l'indagine sull'effettiva conformità del prodotto ai requisiti di sicurezza.

²¹² Decisione n. 768/2008/CE del Parlamento europeo e del Consiglio del 9 luglio 2008 relativa a un quadro comune per la commercializzazione dei prodotti e che abroga la decisione 93/465/CEE, considerando 21.

valutazione di conformità alle norme tecniche e a concedere l'autorizzazione all'immissione in commercio.

3.4.2. Sistemi di i.a. e norme di armonizzazione: l'AI Act.

La già citata Proposta di Regolamento presentata il 21 aprile 2021 dalla Commissione Europea (c.d. *AI Act*) detta i requisiti di sicurezza che ciascun sistema di i.a. – definito, lo ricordiamo (vd. *supra*, cap. I, § 6.1), come «un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono» (art. 3, lett. a) – dovrà soddisfare per poter essere introdotto sul mercato²¹³.

La proposta introduce una disciplina speciale rispetto al quadro normativo sulla sicurezza generale dei prodotti (dir. 01/95/CE); quest'ultima continuerà ad applicarsi come «rete di sicurezza» al fine di garantire che i «sistemi di IA collegati a prodotti che non sono ad alto rischio in conformità al presente regolamento e che pertanto non sono

²¹³ Sulla proposta di regolamento si vedano i commenti di L. FLORIDI, *The European Legislation on AI: a Brief Analysis of its Philosophical Approach*, in *Philosophy & Technology*, 2021, n. 34, p. 215 ss.; M. EBERS, *Standardizing AI*, cit.; W.G. VOSS, *AI Act: The European Union's Proposed Framework Regulation for Artificial Intelligence Governance*, in 25 *Journal of Internet Law*, 2021, p. 7 ss.; M. VEALE, *Demystifying the Draft EU Artificial Intelligence Act*, cit.; J. SCHUETT, *Risk management in the Artificial Intelligence Act*, arXiv preprint, 2022, disponibile a questo link: <https://arxiv.org/pdf/2212.03109.pdf>; F. SOVRANO E AL., *Metrics, Explainability and the European AI Act Proposal*, in *J*, 2022, n. 5, p. 126 ss.; nella letteratura italiana vd. G. ALPA, *Quale modello normativo europeo per l'intelligenza artificiale?*, cit., p. 1009 ss.; U. RUFFOLO – A. AMIDEI, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design, strumenti di certificazione preventive e «autodisciplina» di settore*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1, cit., p. 489 ss.; C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 2021, 3, p. 415 ss.; G. FINOCCHIARO, *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Dir. inf.*, 2022, n. 2, p. 303 ss.; F. PIZZETTI, *La proposta di Regolamento sull'IA della Commissione Europea presentata il 21.4.2021 (COM (2021) 206 final) tra Mercato Unico e competizione digitale globale*, in *Dir. internet*, 2021, n. 4, p. 591 ss.; per i profili di interesse penalistico vd. F.C. LA VATTIATA, *Brevi note "a caldo" sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Dir. pen. uomo*, 30 giugno 2021; A. LAVORGNA – G. SUFFIA, *La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell'ambito della giustizia penale*, in *Dir. pen. cont. – Riv. Trim.*, 2021, n. 2, p. 88 ss.

tenuti a rispettare i requisiti ivi stabiliti siano comunque sicuri al momento dell'immissione sul mercato o della messa in servizio» (considerando 82)²¹⁴.

In questa sede ci accingiamo ad illustrare le principali disposizioni contenute nella bozza di regolamento presentata dalla Commissione, con l'avvertenza che il testo è ancora oggetto di dibattito parlamentare e sarà poi oggetto di negoziazioni tra rappresentanti del Parlamento, del Consiglio e della Commissione (c.d. *triloghi*)²¹⁵. Il testo definitivamente approvato potrebbe dunque essere parzialmente diverso da quello in commento²¹⁶.

Innanzitutto, prima ancora di delineare i requisiti di sicurezza che i sistemi di i.a. devono soddisfare per essere introdotti nel mercato europeo, la bozza di regolamento propone di vietare l'immissione in commercio di una serie di “pratiche di intelligenza artificiale” (art. 5):

a) tecniche subliminali che agiscono senza che una persona ne sia consapevole al fine di distorcerne materialmente il comportamento in modo tale da poter cagionare – ad essa o ad altra persona – un danno fisico o psicologico;

b) sistemi di i.a. che, al fine di cui alla lett. a), sfruttano le vulnerabilità di uno specifico gruppo di persone;

c) tecniche di valutazione dell'affidabilità delle persone fisiche sulla base del loro comportamento sociale o di caratteristiche personali, il cui risultato può comportare,

²¹⁴ Oltre ai requisiti di sicurezza stabiliti dall'AI Act, i sistemi di i.a. sono inoltre sottoposti a diversi regimi settoriali, a seconda della tipologia di prodotto coinvolto; ad esempio, gli *embodied AI systems* sono già oggi soggetti alla c.d. Direttiva Macchine (Direttiva 2006/42/CE del Parlamento europeo e del Consiglio, del 17 maggio 2006, relativa alle macchine e che modifica la direttiva 95/16/CE), anch'essa oggetto di una proposta di riforma presentata dalla Commissione Europea (Proposta di Regolamento del Parlamento Europeo e del Consiglio sui prodotti macchina; Bruxelles, COM (2021) 202 final, 21 aprile 2021). A tal proposito, l'art. 9 della bozza di revisione stabilisce che «Laddove i prodotti macchina contengono un sistema di intelligenza artificiale, al quale si applicano i requisiti essenziali di sicurezza e di tutela della salute di cui al regolamento (UE) .../..., in relazione a tale sistema di intelligenza artificiale il presente regolamento si applica soltanto per quanto concerne la sua integrazione sicura nella macchina complessiva, in maniera da non compromettere la sicurezza del prodotto macchina nel suo complesso».

²¹⁵ Sulla (criticata) prassi delle negoziazioni informali interistituzionali vd. per tutti F. MARTINES, *I triloghi: analisi di una prassi decisionale alla luce dei principi di democrazia, equilibrio istituzionale, efficienza e trasparenza*, in *Dir. pubbl. comp. eur.*, 2018, fasc. 2, p. 311 ss.; S. NINATTI – A. TANCA, *La natura dei triloghi all'esame del giudice dell'Ue: efficacia "versus" trasparenza del processo legislativo nel caso De Capitani*, in *Quaderni cost.*, 2018, fasc. 2, p. 512 ss.

²¹⁶ Gli ultimi aggiornamenti sul dibattito in sede europea possono essere seguiti sul sito *Legislative train*, a questo link: <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence>, oppure sul sito *The Artificial Intelligence Act* (<https://artificialintelligenceact.eu>).

alternativamente o cumulativamente: (i) un trattamento pregiudizievole di determinate persone o di gruppi di persone che sia ingiustificato o sproporzionato rispetto alla gravità del comportamento adottato; (ii) un trattamento pregiudizievole di determinate persone o di gruppi di persone in contesti che non sono collegati ai contesti in cui i dati sono stati originariamente raccolti²¹⁷;

d) l'uso di sistemi di identificazione biometrica remota “in tempo reale” in spazi accessibili al pubblico e a fini di attività di contrasto, a meno che tale uso sia strettamente necessario per uno dei seguenti obiettivi: i) la ricerca mirata di potenziali vittime di reato, compresi i minori scomparsi; ii) la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone o di un attacco terroristico; iii) il rilevamento, la localizzazione, l'identificazione o l'azione penale nei confronti di un autore o un sospettato di un reato di cui all'articolo 2, 62 paragrafo 2, della decisione quadro 2002/584/GAI del Consiglio²¹⁸, punibile nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà della durata massima di almeno tre anni, come stabilito dalla legge di tale Stato membro.

I sistemi di i.a. che non rientrano nelle quattro categorie vietate possono invece essere commercializzati, purché rispettino i requisiti stabiliti dalla bozza di regolamento, che variano a seconda del “rischio” che essi pongono per la sicurezza degli utenti ed il rispetto dei diritti fondamentali dei cittadini. In particolare, si possono distinguere due diversi regimi: quello per i sistemi di i.a. ad alto rischio (titolo III, artt. 6 ss.) e quello per i sistemi a rischio limitato (art. 52).

Riservando al prossimo paragrafo l'approfondimento dei requisiti per l'immissione sul mercato dei sistemi ad alto rischio – che sono gli unici per i quali è prevista una regolamentazione organica – riportiamo qui brevemente la disciplina

²¹⁷ Si tratta dei c.d. sistemi di *social scoring*. I prototipi di tali algoritmi sono stati sviluppati da alcune amministrazioni locali in Cina e sono finalizzati ad attribuire un punteggio ai cittadini sulla base di una serie di criteri di “affidabilità sociale” (es. puntuale pagamento delle bollette, adempimento delle obbligazioni contrattuali stipulate, acquisti personali, cronologia dei motori di ricerca, conversazioni intrattenute su *chat* e *social network* etc.). Il punteggio così ottenuto può determinare un trattamento favorevole (es. accesso facilitato ai finanziamenti, snellimento della documentazione necessaria per viaggi e spostamenti) o sfavorevole (es. rallentamento della connessione internet, esclusione dalle scuole private, divieti di volo); vd. R. Botsman, *Big data meets Big Brother as China moves to rate its citizens*, in *Wired UK*, 21 ottobre 2017, disponibile a questo link: <https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>.

²¹⁸ Decisione quadro del Consiglio 2002/584/GAI, del 13 giugno 2002, relativa al mandato d'arresto europeo e alle procedure di consegna tra Stati membri.

stabilita per i sistemi a rischio limitato. L'art. 52 dell'*AI Act*, in particolare, impone ai fornitori alcuni (generici) obblighi di trasparenza e *labeling*, al fine di consentire alle persone che si interfacciano con essi di essere a conoscenza della natura "artificiale" di tali sistemi. Tra i sistemi a rischio limitato la proposta annovera i *chatbots* (art. 52, § 1), gli *emotion recognition systems* (art. 52, § 2) e i *deepfakes* (art. 52, § 3). Per quanto riguarda i *chatbots*, sono i fornitori che devono progettare il sistema in modo tale che gli utenti siano informati che stanno interagendo con una macchina piuttosto che con un essere umano; non sono invece previsti obblighi di trasparenza in capo agli utenti. Così, sarà solo la persona (fisica o giuridica) che ha progettato il *chatbot* a dover garantire il rispetto degli obblighi di trasparenza, non anche l'utente che utilizza concretamente il *software*. Al contrario, in relazione agli *emotion recognition systems* e ai *deepfakes*, gli obblighi di trasparenza ricadono soltanto sugli utenti.

Per completezza, può essere infine opportuno sottolineare che vi sono sistemi di i.a. (si pensi, ad es., ad alcuni algoritmi contenuti negli *spam filters* o nei videogiochi) che non rientrano né tra le pratiche vietate, né nelle categorie di intelligenza artificiale a rischio alto o limitato, per le quali la proposta di regolamento individua obblighi di *product safety* specifici. Essi potranno essere sviluppati e utilizzati sulla base della legislazione già esistente, senza alcun obbligo ulteriore; i fornitori potranno comunque scegliere di applicare volontariamente i requisiti di sicurezza stabiliti per i sistemi di i.a. a rischio alto o limitato, così come di aderire ai codici di condotta disciplinati dall'art. 69 *AI Act* ²¹⁹.

3.4.3. (segue) I sistemi di intelligenza artificiale ad alto rischio.

La proposta di regolamento sull'i.a. prevede una disciplina elaborata in materia di *product safety* soltanto in relazione ai sistemi di i.a. ad alto rischio; come abbiamo già rilevato, infatti, per quanto concerne i sistemi a rischio moderato, l'*AI Act* stabilisce soltanto obblighi di trasparenza, mentre nulla dispone con riferimento a tutti gli altri sistemi di i.a. non contemplati dalla normativa.

²¹⁹ Relazione di accompagnamento all'*AI Act*, § 5.2.7.

I sistemi di i.a. ad alto rischio sono individuati dall'art. 6 della proposta attraverso una duplice tecnica normativa: (i) attraverso menzione espressa; (ii) attraverso rinvio al campo di applicazione di normative europee settoriali.

(i) Quanto alla prima categoria, l'art. 6, § 2 rinvia ai settori elencati all'allegato III²²⁰, che comprendono, da un lato, sistemi che possono mettere in pericolo l'incolumità pubblica o la sicurezza fisica delle persone (es. n. 2 “*Gestione e funzionamento delle infrastrutture critiche*”, si pensi alle componenti di sicurezza nella fornitura di acqua o di gas) e, dall'altro, sistemi che, se usati in maniera scorretta, possono causare gravi violazioni dei diritti fondamentali (es. n. 1 “*Identificazione e categorizzazione biometrica delle persone fisiche*”, si pensi ai sistemi di riconoscimento facciale)²²¹. Al fine di assicurare che il regolamento sia aggiornato con le nuove applicazioni di i.a., la Proposta prevede inoltre che la Commissione possa ampliare l'elenco previsto all'allegato III, in base alle procedure specifiche stabilite all'art. 7.

(ii) Ai sensi dell'art. 6, § 1, sono altresì ad alto rischio i sistemi che soddisfano entrambe le seguenti condizioni:

1) sono prodotti – o sono destinati ad essere utilizzati come componenti di sicurezza di un prodotto – disciplinati dalla normativa di armonizzazione dell'Unione elencata nell'allegato II (si tratta, ad esempio, della Direttiva c.d. macchine CE/2006/42, del Regolamento c.d. dispositivi medici UE/2017/745, della Direttiva c.d. giocattoli CE/2009/48, della Direttiva c.d. imbarcazioni di diporto CE/1994/25);

2) sono prodotti soggetti a una valutazione di conformità da parte di terzi, ai fini dell'immissione sul mercato o della messa in servizio, ai sensi della normativa di armonizzazione dell'Unione elencata nell'allegato II.

²²⁰ Questi i settori elencati all'allegato III dell'*AI Act*: 1. Identificazione e categorizzazione biometrica delle persone fisiche (es. sistemi di riconoscimento facciale); 2. Gestione e funzionamento delle infrastrutture critiche; 3. Istruzione e formazione professionale (es. sistemi di valutazione per l'accesso a scuole o università); 4. Occupazione, gestione dei lavoratori e accesso al lavoro autonomo (es. sistemi di valutazione dei curriculum vitae); 5. Accesso a prestazioni e servizi pubblici e a servizi privati essenziali e fruizione degli stessi (es. credit scoring); 6. Attività di *law enforcement* (es. sistemi di valutazione dell'affidabilità della prova nel processo); 7. Gestione della migrazione, dell'asilo e del controllo delle frontiere (es. sistemi di verifica dell'autenticità dei documenti di viaggio); 8. Amministrazione della giustizia e processi democratici (es. sistemi di supporto all'attività dell'autorità giudiziaria).

²²¹ Vd. L. FLORIDI, *The European Legislation on AI*, cit., p. 219, critico sulla scelta di assimilare le due tipologie di pericoli.

Senza inoltrarci nell'ampio dibattito tecnico-legislativo che è seguito alla pubblicazione della proposta, ci limitiamo qui a sottolineare come diversi studiosi abbiano criticato la definizione di “sistemi ad alto rischio”, ritenendola eccessivamente vaga e di difficile delimitazione pratica, dunque inidonea a consentire agli operatori del settore un sufficiente livello di certezza²²²; altri autori hanno espresso scetticismo sulla stessa possibilità di determinare *a priori* quali siano i sistemi ad alto rischio e quali quelli a basso o medio rischio²²³.

Il titolo III, capo II, dell'*AI Act* delinea i requisiti per la messa in commercio dei sistemi ad alto rischio, che possono essere così riassunti:

(i) deve essere predisposto un “sistema di gestione dei rischi”, finalizzato all'individuazione dei rischi prevedibili associati a ciascun sistema di i.a. ad alto rischio e all'adozione di adeguate misure di gestione dei rischi (art. 9)²²⁴;

(ii) i *dataset* impiegati per l'addestramento e la validazione dei sistemi devono rispondere a parametri di alta qualità affinché il loro rendimento sia sicuro e in linea con la destinazione d'uso, dovendo a questo riguardo consistere in dati, relativamente alla destinazione anzidetta, pertinenti, sufficientemente rappresentativi, immuni da errori e completi (art. 10, § 3) – una disposizione, come non abbiamo mancato di rilevare precedentemente (vd. *supra*, cap. I, § 3), che si espone a notevoli perplessità, stante, da un lato, l'assoluta vaghezza precettiva del riferimento ai parametri della “pertinenza” e della “rappresentatività”, e, dall'altro, l'impossibilità pratica di garantire codici algoritmici privi di errori;

(iii) prima di immettere un sistema nel circuito economico, è necessario predisporre la relativa documentazione tecnica, che dimostri il rispetto dei requisiti di cui al Regolamento (art. 11);

(iv) i sistemi devono essere programmati e sviluppati in modo tale da consentire la registrazione automatica degli eventi che si verificano durante il loro impiego (c.d. *logs*), così da assicurare un livello di tracciabilità del funzionamento del dispositivo

²²² L. FLORIDI, *The European Legislation on AI*, cit., p. 219; S. WHITTAM, *Mind the compensation gap: towards a new European regime addressing civil liability in the age of AI*, in *International Journal of Law and Information Technology*, 2022, XX, p. 9; U. RUFFOLO – A. AMIDEI, *La regolazione ex ante dell'intelligenza artificiale*, cit, p. 497.

²²³ S. WHITTAM, *Mind the compensation gap*, cit., p. 9-10.

²²⁴ In argomento vd. diffusamente J. SCHUETT, *Risk management in the Artificial Intelligence Act*, arXiv preprint, 2022, disponibile a questo link: <https://aps.arxiv.org/pdf/2212.03109.pdf>.

intelligente durante tutto il suo ciclo di vita adeguato alla finalità prevista del sistema (art. 12);

(v) i sistemi devono essere sufficientemente trasparenti da consentire agli utenti di interpretare l'*output* del sistema e di utilizzarlo adeguatamente; inoltre, ogni sistema deve essere accompagnato da istruzioni adeguate e comprensibili, riguardanti: l'identità del fornitore ed, eventualmente, del suo rappresentante; le caratteristiche, le capacità e i limiti del sistema; le eventuali modifiche – predeterminate dal fornitore al momento della valutazione di conformità – che possono riguardare il sistema; le misure di controllo umano; una stima della durata di funzionamento del sistema e ogni misura di manutenzione e cura, inclusi gli aggiornamenti (art. 13);

(vi) i sistemi devono essere progettati e sviluppati in modo tale da garantire, secondo modalità specificamente prescritte²²⁵, il loro effettivo controllo da parte di una persona fisica all'atto dell'utilizzo, al fine di prevenire o minimizzare i rischi per la salute, la sicurezza o i diritti fondamentali, che possono emergere in relazione sia alla destinazione d'uso, sia ai prevedibili usi impropri del *device* – principio del c.d. *human-in-the-loop* (art. 14);

(vii) i sistemi di i.a. devono garantire un appropriato livello di precisione e c.d. robustezza (*i.e.* resilienza rispetto tanto alle limitazioni proprie del sistema – quali errori, fallimenti, situazioni impreviste o incongruenze – quanto ad eventuali azioni di compromissione della sua sicurezza); peraltro, con particolare riferimento ai sistemi dotati di *machine learning*, occorrerà assicurare la possibilità di intervento su eventuali *output* affetti da *bias*, attraverso puntuali misure correttive (art. 15);

²²⁵ Ai sensi dell'art. 14, § 4, AI Act, il sistema di i.a. deve essere progettato in modo tale da consentire all'utente: «a) [di] comprendere appieno le capacità e i limiti del sistema di IA ad alto rischio ed essere in grado di monitorarne debitamente il funzionamento, in modo che i segnali di anomalie, disfunzioni e prestazioni inattese possano essere individuati e affrontati quanto prima; b) [di] restare consapevole della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("distorsione dell'automazione"), in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche; c) [di] essere in grado di interpretare correttamente l'output del sistema di IA ad alto rischio, tenendo conto in particolare delle caratteristiche del sistema e degli strumenti e dei metodi di interpretazione disponibili; d) [di] essere in grado di decidere, in qualsiasi situazione particolare, di non usare il sistema di IA ad alto rischio o altrimenti di ignorare, annullare o ribaltare l'output del sistema di IA ad alto rischio; e) [di] essere in grado di intervenire sul funzionamento del sistema di IA ad alto rischio o di interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga».

(viii) i sistemi devono essere dotati di soluzioni tecniche volte a garantirne la sicurezza informatica (art. 15, § 4).

Conformemente all'indirizzo inaugurato con il c.d. "Nuovo Approccio", l'art. 40 dell'*AI Act* rinvia – per la definizione dei requisiti tecnici – alle norme armonizzate, prevedendo una *presunzione di conformità* ai requisiti stabiliti dal titolo III, capo 2, per il caso in cui il produttore si attenga a tali *standard*. Prima ancora che l'*AI Act* delegasse agli organismi europei di normalizzazione la predisposizione di *standard* tecnici in materia di sistemi di i.a., tra l'altro, CEN e CENELEC avevano già creato, nel 2019, un *focus group* sull'i.a., finalizzato ad elaborare standard per la messa in commercio di *smart products*²²⁶.

I sistemi ad alto rischio sono sottoposti al regime di auto-certificazione CE²²⁷; solo le applicazioni ad alto rischio elencate nell'ambito dell'*identificazione biometrica e categorizzazione delle persone fisiche* devono utilizzare organismi notificati specifici per l'*AI Act*. Quest'ultima è tuttavia una previsione destinata ad avere un'applicazione transitoria: una volta che saranno introdotti *standard* armonizzati o specifiche comuni anche in quest'ambito, infatti, sarà sufficiente l'autocertificazione di conformità da parte del produttore (vd. art. 43, co. 1, *AI Act*)²²⁸. È ovvio, d'altra parte, che l'autorizzazione non potrà essere omessa qualora i sistemi di i.a. siano soggetti – oltre che all'*AI Act* – anche a direttive che richiedono l'autorizzazione da parte di un organismo notificato.

²²⁶ Vd. la notizia sul sito di CEN-CENELEC, *CEN and CENELEC launched a new Joint TC on Artificial Intelligence*, 3 marzo 2021, disponibile a questo link: <https://www.cenelec.eu/news-and-events/news/2021/briefnews/2021-03-03-new-joint-tc-on-artificial-intelligence/>. Sul processo di standardizzazione dei sistemi di i.a., a livello europeo ed internazionale, vd. M. EBERS, *Standardizing AI*, cit., p. 8 ss.; G. MOBILIO, *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *Biolaw Journal*, 2020, n. 2, p. 413 ss.

²²⁷ Si noti tuttavia che la legge sull'AI presenta un meccanismo misto, poiché per i sistemi biometrici destinati a essere messi in servizio dalle autorità di polizia, di immigrazione o di asilo o dalle istituzioni dell'UE, la valutazione di conformità è effettuata da un apposito ente pubblico, vd. *AI Act* artt. 43(1), 63(5); lo notano M. VEALE, *Demystifying the Draft EU Artificial Intelligence Act*, cit., p. 106; C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, cit., p. 430 ss.

²²⁸ Dal momento, tuttavia, che la Commissione auspica che gli standard armonizzati siano emanati prima dell'entrata in vigore del Regolamento, gli organismi notificati potrebbero non essere mai necessari, nemmeno per i sistemi biometrici.

3.4.4. La tesi della non difettosità del prodotto conforme agli *standard* tecnici.

Secondo una parte della dottrina, il concetto di “prodotto sicuro” – ai sensi della legislazione, sia verticale che orizzontale, in materia di *product safety* – dovrebbe essere considerato generalmente coincidente con quello di “prodotto non difettoso”: un prodotto che soddisfi tutti i requisiti per la messa in commercio non potrebbe, secondo questa prospettiva, essere considerato “difettoso” e dunque dare luogo a responsabilità ai sensi della disciplina di cui alla dir. n. 85/374/CEE.

Benché tale orientamento non sia nuovo nell’ambito del dibattito sulla responsabilità per danno da prodotto²²⁹, esso è stato oggetto di un rinnovato interesse in relazione alle potenzialità lesive degli *smart products*, poiché pare fornire un criterio stabile per l’individuazione delle ipotesi di difettosità del prodotto – offrendo, in questo modo, una rassicurazione a quanti immetteranno sul mercato un sistema di i.a. conforme agli *standard* tecnici delineati, in generale, dall’*AI Act* e, nello specifico, dalle norme armonizzate che saranno in seguito approvate²³⁰. Stando a tale orientamento, infatti, i sistemi di i.a. che saranno immessi sul mercato in conformità ai requisiti dell’*AI Act* non potranno essere considerati difettosi; di conseguenza, il danneggiato non potrà chiedere il risarcimento del danno subito sulla base del regime oggettivo previsto dalla dir. n. 85/374/CEE.

In particolare, la conformità agli *standard* tecnici concorrerebbe ad integrare la nozione di “sicurezza che ci si può legittimamente attendere”, che rappresenta uno degli elementi costitutivi (negativi) della fattispecie di “prodotto difettoso”, ai sensi degli artt.

²²⁹ Vd., oltre agli autori precedentemente citati, U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 75; ID., *La norma tecnica da regola di esperienza a norma giuridicamente rilevante*, cit., p. 50 ss.; P. TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, cit., p. 419; M. TRONTI, *La direttiva CEE relativa al riavvicinamento delle disposizioni legislative regolamentari ed amministrative degli stati membri in materia di responsabilità per danno da prodotti difettosi*, in *Giur. merito*, 1988, n. 4, p. 695.

²³⁰ U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1713 ss.; E. AL MUREDEN, *La sicurezza dei prodotti e la responsabilità del produttore*, Torino, 2017; ID., *Il danno da “prodotto conforme”. Le soluzioni europee e statunitensi nella prospettiva del “Transatlantic Trade and Investment Partnership” (T.T.I.P.)*, in *Contr. e impr.*, 2015, p. 388; ID., *Sicurezza “ragionevole” degli autoveicoli e responsabilità del produttore nell’ordinamento giuridico italiano e negli Stati Uniti*, in *Contr. e impr.*, 2012, p. 1506; nello stesso senso, specificamente in materia di sistemi di i.a., C. AMATO, *Product Liability and Product Security: Present and Future*, in S. Lohsse e al. (eds.), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 79 ss.; E. MARCHISIO, *In support of “no-fault” civil liability rules for artificial intelligence*, in *SN Social Sciences*, 2021, 1: 54, p. 10 ss.

6 dir. e 117 cod. cons.²³¹. In base a tale impostazione, infatti, le aspettative ragionevoli dei consumatori non potrebbero essere tali da pretendere un prodotto più sicuro rispetto a quello rispettoso delle norme tecniche armonizzate – norme che, contemperando le esigenze di certezza e prevedibilità della normativa da parte dell’industria e il rischio considerato “accettabile” da parte dei consumatori, costituiscono lo “stato dell’arte” in materia di progettazione e fabbricazione dei prodotti²³².

Già Ugo Carnevali, nel 2006, sosteneva che nella generalità di casi – senza precludere, dunque, un giudizio caso per caso, anche alla luce della possibile peculiarità dei casi concreti – il rispetto degli *standard* stabiliti dalle norme armonizzate dovesse tradursi in una valutazione di “non difettosità” dei prodotti, giudicando come «illogico e contraddittorio» che un prodotto valutato come “sicuro” in fase *pre-marketing* potesse al contempo essere considerato “difettoso” nell’ambito di un giudizio di responsabilità promosso dal danneggiato²³³. Secondo tale prospettiva, soltanto in casi eccezionali (es., sopravvenuta conoscenza di nuovi rischi non previsti dagli *standard* tecnici o mancato aggiornamento della normativa), un prodotto conforme alle norme tecniche armonizzate potrebbe essere considerato “difettoso”.

La *ratio* di tale orientamento – agli occhi di chi la propone – è ovviamente quella di trasmettere un messaggio “tranquillizzante” al produttore: attenendosi alla normativa, questi è infatti certo che non potrà essere chiamato a rispondere per gli eventuali danni cagionati da un difetto di *design*. Il danneggiato potrebbe tutt’al più rivolgersi all’utilizzatore, che in alcune circostanze potrebbe essere ritenuto responsabile (si pensi, in materia di circolazione stradale, alla già citata responsabilità oggettiva del conducente e del proprietario, ai sensi dell’art. 2054 c.c.). Qualora, invece, sia lo stesso utilizzatore ad essere danneggiato dal prodotto conforme, i costi non potranno essere trasferiti su terzi, ma rimarranno in capo al consumatore (si pensi – sempre restando sulla circolazione stradale – ai c.d. *one car collisions or accidents*, in cui l’unico soggetto a subire danni è lo stesso conducente del veicolo)²³⁴. Resterebbe ferma, in ogni caso, la possibilità che il

²³¹ U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 75; C. AMATO, *Product Liability and Product Security: Present and Future*, cit., p. 79.

²³² U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 73.

²³³ U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 73.

²³⁴ G. CALABRESI – E. AL MUREN, *Driverless Cars*, cit., p. 158.

produttore sia chiamato a risarcire i danni derivanti da difetti di fabbricazione o da difetti di informazione.

Come abbiamo già accennato, il suddetto orientamento si ispira alla c.d. *preemption doctrine* statunitense, che trova il suo fondamento nella *Supremacy Clause* (art. 6, clausola 2, Cost. Stati Uniti) e che sancisce la prevalenza delle norme federali su quelle statali, in caso di contrasto tra esse. Nel contesto della disciplina del danno da prodotto, il riferimento a tale clausola è ampiamente utilizzato per escludere la responsabilità del produttore nel caso in cui il danno sia stato cagionato da un prodotto conforme agli *standard* tecnici federali ma difforme – per difetto – rispetto alle norme tecniche previste dalle singole giurisdizioni²³⁵. In questo caso, l'applicazione della *preemption doctrine* impone di esimere da responsabilità il produttore, sempre che i suddetti *standard* abbiano la valenza di “livello massimo” di sicurezza (c.d. *ceiling*) e non, invece, di “livello minimo”.

Tale teoria ha avuto larga applicazione con riferimento ai prodotti medici e agli autoveicoli: se un *device* medico risponde ai requisiti di sicurezza stabiliti dalla *Food and Drug Administration*, il produttore di tale *device* non potrà essere chiamato a rispondere nel caso in cui il prodotto abbia cagionato un evento lesivo; lo stesso vale per i veicoli che si attengano agli standard della NHTSA²³⁶. Non sempre, tuttavia, la *regulatory compliance* è stata assunta al rango di *complete defense*; la *preemption doctrine* non gode, infatti, di uniforme applicazione sul suolo statunitense, tanto che in diversi casi si è giunti all'adozione di decisioni *contra preemption*²³⁷.

²³⁵ Sulla *preemption doctrine* nel contesto della *product liability* statunitense vd. T.O. MCGARITY, *The Preemption War: When Federal Bureaucracies Trump Local Juries*, Yale University Press, 2008; K.N. HYLTON, *Preemption and Products Liability: A Positive Theory*, in 16 *Supreme Court Economic Review*, 2008, p. 205 ss.; C.M. SHARKEY, *Products Liability Preemption: An Institutional Approach*, in 76 *Geo. Wash. L. Rev.*, 2008, p. 449 ss.

²³⁶ M.A. GEISTFELD, *The Regulatory Sweet Spot for Autonomous Vehicles*, in 53 *Wake Forest L. Rev.* 2018, 120; in materia di autoveicoli vd. il *leading case* *Geier v. American Honda Company*, 529 U.S.861 (2000); il primo caso in cui, a quanto ci risulta, si è fatto ricorso alla *preemption doctrine* in materia di *product liability* è *Cipollone v. Liggett Group Inc.*, 5050 U.S. 504, 521 (1992), per i danni da tabacco.

²³⁷ Così G. GUERRA, *Il concetto di difettosità nella realtà che cambia*, cit., p. 267, che riporta un caso deciso dalla Corte Suprema degli Stati Uniti nel 2005 (n. 03-388.27 del 27 aprile 2005, in *Resp. civ. prev.*, 2006, p. 392) in cui i giudici stabiliscono – con riferimento ai danni cagionati da un pesticida – che la conformità ai requisiti federali non è idonea ad escludere *ex se* la risarcibilità del danno.

3.4.5. (segue) Separazione concettuale tra disciplina della *product liability* e disciplina della *product safety*.

La maggior parte degli studiosi ritiene che la prospettiva della *preemption doctrine* sia contraria allo spirito della Direttiva sul danno da prodotto e al complessivo rapporto – complementare, ma pur sempre di separazione – tra *product liability* e *product safety*²³⁸.

S'impone, innanzitutto, una considerazione di carattere logico-sistematico. Se la «sicurezza che ci si può legittimamente attendere» (art. 117 cod. cons.) non fosse altro che quella concretizzata dagli *standard* di settore, non si comprenderebbero gli sforzi compiuti dal legislatore per circoscrivere la nozione di difetto, quando sarebbe bastato un mero rinvio alla disciplina sulla sicurezza del prodotto²³⁹. D'altra parte, un indizio in tale senso sembra provenire dalla disciplina codicistica in materia di responsabilità per la circolazione di veicoli: il comma 4 dell'art. 2054 c.c. – nel fare riferimento alla responsabilità del proprietario e del conducente per “i vizi di costruzione” del veicolo – implica infatti logicamente il fatto che i veicoli, *benché omologati*, possano essere ritenuti difettosi e dare luogo a responsabilità (del proprietario e del conducente, così come del produttore)²⁴⁰.

Da più parti si rileva inoltre come i due *corpus* normativi hanno una funzione diversa. La disciplina sulla *product safety* – nella sua veste generale e in quella settoriale

²³⁸ E. BELLISARIO, *Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, cit., *passim*; EAD., *La rilevanza del criterio presuntivo della conformità alle norme armonizzate*, in *Pers. merc.*, 2012, p. 156 s.; C. CASTRONOVO, *La responsabilità civile*, Giuffrè, 2018, p. 802; A. FUSARO, *I prodotti difettosi e pericolosi: le responsabilità*, cit., p. 390 ss.; C.M. VERARDI, *Art. 6 – Esclusione della responsabilità*, in G. Alpa e al. (a cura di), *La responsabilità per danno da prodotti difettosi*, Giuffrè, 1990, p. 79 ss.; M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 352 ss.; vd. per una prospettiva gius-economica G. AFFERNI, *L'analisi economica del diritto*, in G. Alpa (a cura di), *La responsabilità del produttore*, Giuffrè, 2019, p. 349-350; Specificamente con riferimento alla conformità dei sistemi di i.a., vd. A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1721; B.A. KOCH, *Product Liability 2.0*, cit., p. 108; A. PROCIDA MIRABELLI DI LAURO, *Le intelligenze artificiali tra responsabilità civile e sicurezza sociale*, cit., p. 313 ss.; implicitamente, con riferimento allo specifico settore dell'automotive, vd. A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, p. 1020 ss. e F. BARTOLINI, *Auto a guida autonoma e problemi di responsabilità civile*, cit., p. 311; A. FUSARO, *Quale modello di responsabilità per la robotica avanzata?*, cit., p. 1348-1349.

²³⁹ M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 352.

²⁴⁰ U. RUFFOLO, *Intelligenza artificiale e automotive: le responsabilità da veicoli self-driving e driverless*, in Id. (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 164.

– mira a «garantire che i prodotti immessi sul mercato ovvero in libera pratica siano sicuri» (art. 102, co. 1, cod. cons.) e ad assicurare «un'applicazione efficace e coerente del requisito generale di sicurezza» (considerando 14 dir. 2001/95). A tal fine, vengono imposti *standard*, controlli e requisiti che devono essere soddisfatti affinché il prodotto ottenga l'omologazione e, dunque, l'autorizzazione alla messa in commercio. Si tratta di criteri “statici”, che riguardano il prodotto “astratto” e che devono essere valutati *ex ante*, a differenza di ciò che accade in materia di responsabilità per danno da prodotto; ma, soprattutto, si tratta di *condizioni minime* – anche per quanto riguarda le norme armonizzate, che altro non sono se non la *specificazione* dei requisiti di commercializzazione previsti dalla legislazione verticale – il cui rispetto consente la messa in commercio, ma non esenta il produttore dal risarcimento del danno cagionato da un prodotto difettoso²⁴¹.

Complessivamente, d'altra parte, la definizione degli *standard* tecnici comporta l'adozione di scelte politiche, di natura pubblicistica, frutto di un bilanciamento tra rischi industriali e benefici delle innovazioni tecnologiche – nulla a che vedere, dunque, con presunte finalità di razionalizzazione dei giudizi e riduzione della discrezionalità dei giudici nazionali, che, come abbiamo visto, sono spesso addotte a giustificazione della *preemption doctrine*. La normativa sulla *product liability*, al contrario, ha a che fare con il profilo *privatistico* della responsabilità civile, nell'ambito della quale «non interessa tanto se un'attività poteva o non poteva essere esercitata, quanto se si è verificata in concreto una seria minaccia di danno a terzi»²⁴². D'altra parte, a differenza dell'indagine sulla sicurezza, la difettosità di un prodotto, per come è declinata dalla Dir. 85/374/CEE, è per sua natura da condursi *ex post*, in base a valutazioni concernenti non solo l'astratto *design* dell'artefatto, ma anche le modalità di “reazione” rispetto all'utilizzo concreto da parte del consumatore²⁴³.

²⁴¹ Proprio in relazione alla commercializzazione dei sistemi di i.a., è la stessa relazione di accompagnamento alla bozza di regolamento del 2021 che definisce i requisiti ivi stabiliti come “minimi” (p. 3).

²⁴² Così M. COMPORTI, *Esposizione al pericolo e responsabilità civile*, cit., p. 175-176, con riferimento al rapporto tra disciplina autorizzativa allo svolgimento di attività pericolose e applicazione degli artt. 2049-2054 c.c.

²⁴³ A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1722; U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1709.

Che le discipline europee di *product safety* e *product liability*, per quanto connesse, debbano rimanere tra loro separate è previsto dallo stesso diritto europeo e, in particolare, dalla Dir. 2001/95/CE, che, al sedicesimo considerando, precisa che la Direttiva medesima non deve “interferire” con i diritti previsti dalla Dir. 85/374/CEE in materia di responsabilità per danno da prodotti difettosi²⁴⁴. Allo stesso modo, nel codice del consumo, la “sicurezza dei prodotti” (art. 102 ss.) è trattata separatamente rispetto alla “responsabilità da prodotto difettoso” (art. 114 ss.) e l’art. 111 c. cons. sancisce, in tema di *product safety*, che «sono fatte salve le disposizioni...in materia di responsabilità per danno da prodotti difettosi».

Per come concepita in Europa, infine, la disciplina sulla responsabilità per danno da prodotto ha proprio l’obiettivo di porre il costo dei danni derivanti dal prodotto sul produttore, anche qualora questi sia incolpevole: un’esonazione da responsabilità per il fabbricante di prodotto conforme abbasserebbe l’asticella della tutela del consumatore in maniera non indifferente²⁴⁵.

Se, dunque, come già accennato, il mancato rispetto della legislazione verticale o orizzontale sulla *product safety* determina pacificamente la difettosità del prodotto, non è vera l’affermazione inversa: un prodotto conforme potrà, infatti, essere considerato difettoso, qualora presenti un’anomala condizione di pericolosità. Un superamento di tale orientamento in relazione ai sistemi di i.a. non pare auspicabile, sia perché difettano ragioni specifiche per un discostamento dal regime ordinario, sia perché non sembra opportuno che – proprio in relazione a prodotti per i quali il rischio di eventi dannosi non è del tutto preventivabile – sia sufficiente il rispetto di *requisiti minimi* di commercializzazione per beneficiare di un’immunità.

²⁴⁴ A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell’Unione Europea*, cit., p. 1722; U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1709. La stessa Risoluzione del Parlamento del 20 ottobre 2020, cit., al considerando 8, sottolinea che «qualsiasi aggiornamento del quadro della responsabilità per danno da prodotti difettosi dovrebbe andare di pari passo con l’aggiornamento della direttiva 2001/95/CE del Parlamento europeo e del Consiglio, del 3 dicembre 2001, relativa alla sicurezza generale dei prodotti, al fine di garantire che i sistemi di IA integrino la sicurezza e la protezione fin dalla progettazione», considerando, dunque, implicitamente, ancora valida la distinzione tra le due discipline.

²⁴⁵ U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1709.

3.5. L'accertamento del nesso di causalità. Profili sostanziali.

L'attribuzione al produttore della responsabilità per i danni derivanti da sistemi di i.a. incontra ulteriori problematiche con riferimento all'accertamento del nesso di causalità, sia sotto il profilo *sostanziale*, sia sotto il profilo più strettamente *processuale*, relativo alla ripartizione dell'onere probatorio.

L'art. 1 della dir. 85/374/CEE – recepito dall'art. 114 cod. cons. – stabilisce che «Il produttore è responsabile del danno causato da un difetto del suo prodotto»; l'onere di fornire la prova del nesso di causalità, così come di tutti gli altri elementi costitutivi della fattispecie, è posta a carico del danneggiato (vd. art. 120 cod. cons., che recepisce il disposto di cui all'art. 4 della Dir. 85/374/CEE). Com'è noto, il nesso causale, anche ai fini della responsabilità civile aquiliana, è regolato dagli artt. 40 e 41 c.p. La dimensione logica e sostanziale della causalità è identica nel diritto civile e in quello penale²⁴⁶ – anticipiamo dunque in questa sede alcune riflessioni che riprenderemo anche nel prossimo capitolo (vd. *infra*, cap. III, § 4.3.), con riferimento all'accertamento del nesso causale in sede penale. Diversi restano invece gli *standard* probatori richiesti l'accertamento: in un caso, quello del “più probabile che non”, ispirato alla regola della normalità causale e della “preponderanza dell'evidenza”; nell'altro caso, quello dell’“oltre il ragionevole dubbio”²⁴⁷.

Restando per ora sui profili sostanziali, ci si chiede innanzitutto se il carattere adattivo ed imprevedibile del sistema di i.a. possa determinare un'interruzione del nesso di causalità tra difetto e danno, ai sensi dell'art. 41, co. 2, c.p.²⁴⁸ (nel mondo anglosassone,

²⁴⁶ Cass. civ., sez. un., 11 gennaio 2008, n. 576; Cass. civ., sez. III, 11 maggio 2009, n. 10741; Cass. civ., sez. III, 8 luglio 2010, n. 16123.

²⁴⁷ Il principio è stato affermato dalle note dieci sentenze “gemelle” emanate dalle Sezioni unite civili l'11 gennaio 2008 (numeri dal 576 al 585), in tema di responsabilità aquiliana del Ministero della salute per danni da trasfusione di sangue infetto; vd. la nota di R. BLAIOTTA, *Casualità e colpa: diritto civile e diritto penale si confrontano*, in *Cass. pen.*, 2009, fasc. 1, p. 78 ss.; più di recente, il principio è stato ribadito in giurisprudenza da Cass., sez. L, 3 gennaio 2017, n. 47; Cass., sez. I, 30 giugno 2021, n. 18584; sull'impatto che la differenza tra i due *standard* probatori ha sugli esiti dei processi vd. F. STELLA, *Giustizia e modernità*, cit., p. 97-98.

²⁴⁸ In argomento, nella letteratura straniera, vd. M. BASHAYREH – F.N. SIBAI – A. TABBARA, *Artificial Intelligence and legal liability: towards an international approach of proportional liability based on risk sharing*, in *Information and Communications Technology Law*, 2020, 30 (2), p. 173–174; Z.Y. LEE, *Deep learning artificial intelligence and the law of causation: application, challenges and solutions*, in *Information and Communications Technology Law*, 2021, vol. 30, iss. 3, p. 255 ss.; in materia penale vd. S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame?*, cit., p. 431-432; C.

l'istituto è conosciuto come *novus actus interveniens*²⁴⁹). La questione è di rilievo sia che si aderisca all'impostazione *scientifico-naturalistica*, basata sul criterio della *condicio sine qua non*, sia che si ritengano preferibili le teorie della causalità umana e della causalità adeguata.

Nella prima ipotesi, si sarebbe di fronte ad un caso di *causalità sorpassante*, in cui il fattore causale sopravvenuto (la decisione autonoma adottata dal dispositivo intelligente) «*soppianta, azzerandolo, il rischio iniziale* ascrivibile al programmatore»²⁵⁰, introducendone uno nuovo, dipendente dalla decisione algoritmica “intelligente”. Anche sulla base delle impostazioni c.d. normative, il comportamento emergente del sistema di i.a. – che ha causato un danno – potrebbe essere considerato quale fattore *eccezionale* (per quanto riguarda la causalità umana) o *non ordinario* (per quanto concerne la causalità adeguata) e, dunque, interruttivo del nesso causale tra condotta del produttore e danno.

In realtà, tali ricostruzioni – seppur suggestive – non paiono soddisfacenti²⁵¹. In sede di programmazione, infatti, i più moderni sistemi di i.a. sono *appositamente* dotati di capacità di apprendimento. L'abilità di assumere decisioni autonome altro non è, dunque, che la “prevedibile” estrinsecazione dei *features* attribuiti nella fase, *lato sensu*, di produzione, con il risultato che l'evento lesivo derivante dal *decision making* artificiale non potrà essere considerato né frutto di una serie causale autonoma, né conseguenza di fattori anormali o eccezionali. Difficilmente, dunque, il comportamento emergente del sistema di i.a. potrà essere considerato come «un rischio nuovo e incommensurabile, del tutto incongruo rispetto al rischio originario attivato dalla prima condotta»²⁵², soprattutto

PIERGALLINI, *Intelligenza artificiale*, cit., p. 1760 ss.; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1211; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 716 ss.

²⁴⁹ A. DU BOIS-PEDAIN, *Novus actus and beyond: attributing causal responsibility in the criminal courts*, in *The Cambridge Law Journal*, 2021, Vol. 80, suppl. S1, p. S61 ss.; S. CHOPRA – L.F. WHITE, *A Legal Theory for Autonomous Artificial Agents*, cit., p. 122.

²⁵⁰ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1761.

²⁵¹ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1761-1762; S. CHOPRA – L.F. WHITE, *A Legal Theory for Autonomous Artificial Agents*, cit. p. 137; C. MINELLI, *La responsabilità “penale” tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, in *Dir. pen. cont. – Riv. trim.*, 2022, n. 2, p. 56; S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame?*, cit., p. 432-433; *contra* S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 716 ss., in cui l'autore implicitamente pare sostenere che “l'interposizione” algoritmica interrompa il nesso di causalità.

²⁵² Cass. pen., 3 marzo 2022, n. 19882; Cass. pen., sez. IV, 13 maggio 2019, n. 20270; Cass. pen., sez. IV, 3 maggio 2016, n. 25689.

stante la tradizionale riluttanza, in giurisprudenza, a riconoscere cause sopravvenute interruttrive²⁵³.

Abbandonata l'opzione ermeneutica fondata sull'art. 41, co. 2, c.p., resta da capire su quali basi sia possibile accertare il nesso di causalità tra difetto e danno, dal momento che, ad oggi, il settore dell'i.a. è privo di un consolidato corredo nomologico di leggi scientifiche che consentano di collegare – ancorché statisticamente – l'immissione di un prodotto sul mercato e l'evento lesivo.

Va innanzitutto sottolineato che la dimostrazione del rapporto causale nell'ambito della responsabilità per danno da prodotto difettoso è sempre stata considerata uno degli elementi più problematici della disciplina²⁵⁴. Il già citato art. 1 della dir. 85/374/CEE non precisa se sia sufficiente la dimostrazione che il difetto sia stato *condicio sine qua non* dell'evento lesivo o se, al contrario, sia necessario che l'evento corrisponda ad una conseguenza normale, prevedibile o non eccezionale del fatto. Nel silenzio normativo, la giurisprudenza dei diversi Stati membri ha mostrato un'ampia eterogeneità – frutto di tradizioni giuridiche non sempre coincidenti – nell'interpretazione degli elementi necessari a soddisfare l'onere della prova relativa al nesso di causalità tra il difetto di un prodotto e l'evento dannoso²⁵⁵.

In generale, nell'ambito del diritto civile, si è assistito ad un progressivo slittamento che, dal criterio della *condicio sine qua non*, di derivazione “newtoniana”, ha portato ad adottare in via generalizzata il criterio dell'aumento del rischio²⁵⁶. A ciò ha fatto seguito un ulteriore slittamento, che ha portato all'affermazione del criterio del *più*

²⁵³ R. BLAIOTTA, *Causalità giuridica*, Giappichelli, 2010, p. 209 ss., che ripercorre le (pochissime) sentenze che hanno riconosciuto un'interruzione del nesso causale; in argomento vd. anche l'approfondita indagine di A. VALLINI, “Cause sopravvenute da sole sufficienti” e nessi tra condotte. Per una collocazione dell'art.41, comma 2, c.p. nel quadro teorico della causalità “scientifica”, in *Studi in onore di Franco Coppi*, vol. I, Torino, Giappichelli, 2011, p. 341 ss.

²⁵⁴ Vd., nella dottrina statunitense, D.G. GIFFORD, *The Challenge to the Individual Causation Requirement in Mass Products Torts*, in 62 *Wash. & Lee L. Rev.* 873, 900–15 (2005); J. STAPLETON, *Product Liability*, Cambridge University Press, 1994, p. 280 ss.

²⁵⁵ S. WHITTAKER, *Liability for Products. English Law, French Law, and European Harmonisation*, cit., p. 512 ss.; D. WUYTS, *The product liability directive – More than two decades of defective products in Europe*, cit., p. 25 ss.

²⁵⁶ Per una concezione probabilistica della causalità vd., nella dottrina statunitense, D. ROSENBERG, *The causal connection in mass exposure cases: a «public law» vision of the tort system*, in 97 *Harvard Law Review*, 1984, p. 851; G.O. ROBINSON, *Probabilistic causation and Compensation for tortious risk*, in 14 *Journal of Legal Studies*, 1985, p. 779; E.D. ELLIOTT, *Why Courts? Comment on Robinson*, in 14 *Journal of Legal Studies*, 1985, p. 799.

probabile che non – in cui l’attore deve dimostrare che è stata la condotta del convenuto, con un tasso di probabilità superiore al 50%, a causare il danno – ad un livello di probabilità causale molto diluito: oggi si giudica sufficiente che la condotta del convenuto sia stata tale da aumentare il rischio del danno; sono invece escluse le conseguenze dannose che, in assenza di un fatto illecito, si sarebbero verificate con uguale probabilità²⁵⁷. Anche in questo settore, il diritto civile è riuscito ad interiorizzare – con la flessibilità strutturale che gli è propria e senza scossoni dogmatici rispetto ai capisaldi della responsabilità aquiliana classica – quel “modello del rischio” tipico della società globalizzata e post-moderna²⁵⁸, che invece tante preoccupazioni scatena nell’ordinamento penale, per sua natura rigidamente determinato dal dettato costituzionale (in particolare, dai principi di colpevolezza, legalità, offensività) e dalla regola processuale dell’*oltre ragionevole dubbio*.

È proprio nell’ambito degli studi sulla responsabilità per danno da prodotto, d’altra parte, che si è assistito all’affermarsi del concetto di *black box*²⁵⁹, che descrive quei sistemi tecnologici dei quali si conoscono *input* e *output*, senza che sia possibile decrittare scientificamente i procedimenti interni che trasformano i primi nei secondi – con la conseguenza che non è verificabile con certezza *quale input* abbia cagionato *quale output* e in che modo. Alla tipica opacità degli artefatti moderni – spesso di per sé dotati di notevole complessità tecnica, si pensi ad un autoveicolo, ad un elicottero, ad uno *smartphone* – l’intelligenza artificiale aggiunge la capacità di elaborazione statistica di quantità immense di dati e, soprattutto quando si ha a che fare con sistemi che utilizzano reti neurali, una tecnica di computazione (e, dunque, di decisione) che non sempre è comprensibile secondo i tradizionali schemi di ragionamento umani (sul punto, vd. *supra*, cap. I, § 5.3).

Una nota vicenda riguardante alcuni autoveicoli Toyota è emblematica di come, negli artefatti a tecnologia molto avanzata, sia complesso isolare il malfunzionamento all’origine dell’evento lesivo e, di conseguenza, individuare la società (e, nel processo penale, la persona) responsabile del fatto illecito. In quel caso, i ricorsi presentati dalle persone danneggiate allegavano che gli incidenti si erano verificati a causa di

²⁵⁷ G. PONZANELLI, *La responsabilità civile: profili di diritto comparato*, cit., p. 90.

²⁵⁸ C.E. PALIERO, *L’autunno del patriarca*, cit., p. 1240.

²⁵⁹ C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., *passim*.

un'improvvisa accelerazione dei veicoli, alla quale i conducenti non potevano rimediare nemmeno azionando il freno. Nonostante l'intensa attività investigativa portata avanti dagli organi inquirenti e dalla già citata *National Highway Traffic Safety Administration* (NHTSA), le indagini non furono in grado di individuare la precisa causa del malfunzionamento²⁶⁰.

Problemi di accertamento ancora più intensi potrebbero porsi con riferimento agli incidenti derivanti da macchine che incorporano sistemi di i.a.: in questi casi, infatti, oltre che con la complessità tecnologica degli artefatti, sarà necessario fare i conti con la capacità di apprendimento e di autonomo *decision-making* degli algoritmi, nonché con la loro permeabilità ai dati provenienti dall'esterno.

Secondo alcuni studiosi, a tale *gap* conoscitivo potrebbe sopperire – sull'esempio di quanto già avviene nel settore degli aeromobili – l'installazione di *event data recorders* (c.d. EDR, comunemente conosciuti come “scatole nere” o “logs”), funzionali ad illuminare, *ex post*, la ricostruzione della dinamica del fatto lesivo²⁶¹. A tal proposito, il

²⁶⁰ Va comunque rilevato che, nonostante l'impossibilità di accertare con precisione la causa degli incidenti, la Corte distrettuale dell'Oklahoma, nel 2013, ha condannato Toyota al risarcimento del danno, nonché al pagamento di *punitive damages*, ritenendo che l'incidente fosse più probabilmente causato da un difetto del veicolo che da una condotta del conducente – di fatto, applicando dunque la teoria del *res ipsa loquitur* (United States District Court, Oklahoma County, *Jean Bookout and Charles Schwarz v. Toyota Motor Corporation*, Case no. CJ-2008-7969, 24 ottobre 2013), a tal proposito vd. M. Cronin Fisk, *Toyota Settles Oklahoma Acceleration Case After Jury Verdict*, in *Bloomberg*, 25 ottobre 2013, disponibile a questo link: <https://www.bloomberg.com/news/articles/2013-10-25/toyota-settles-oklahoma-acceleration-case-after-jury-verdict>. Successivamente a tale condanna, Toyota ha proceduto a chiuso con accordi transattivi (c.d. *settlement*) tutti i procedimenti che erano aperti a suo carico per gli incidenti derivanti dalle suddette accelerazioni improvvise. Sulla complessa vicenda giudiziaria che ha coinvolto la casa automobilistica vd. T. WINKLE, *Development and Approval of Automated Vehicles: Considerations of Technical, Legal, and Economic Risks*, in M. Maurer e al. (eds.), *Autonomous Driving*, Springer, 2016, p. 598; R.I. STEINZOR, (Still) “Unsafe at Any Speed”: Why Not Jail for Auto Executives?, in 9 *Harvard Law & Policy Review*, 2015, p. 443 ss.

²⁶¹ Così C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1760; K. BÖHM E AL., *New developments on EDR (Event Data Recorder) for automated vehicles*, in *Open Engineering*, 2020, vol. 10, iss. 1, p. 140 ss.; G. SPINDLER, *User liability and strict liability in the Internet of Things and for Robots*, cit., p. 139; G. WAGNER, *Robot, Inc.*, cit., p. 612; R. BORGOGNO, *La responsabilità penale nei processi ad elevata automazione*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 736-737; vd. anche M.F. LOHMANN, *Liability Issues Concerning Self-Driving Vehicles*, in 7 *European Journal of Risk Regulation*, 2016, n. 2, p. 339, che non manca di notare come l'utilizzo degli *event data recorder* potrebbe porre alcuni problemi di tutela della *privacy* – dal momento che, nel caso di registrazione video, potrebbero venire coinvolti anche passanti o veicoli non coinvolti nell'incidente – e, nel processo penale, potrebbe porsi in contrasto con le garanzie del diritto di difesa dell'imputato; sul tema vd. anche S. GLESS., *AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials*, in 51 *Georgetown Journal*

Reg. (UE) 2019/2144²⁶² ha disposto che, a partire dal 6 luglio 2022, tutte le automobili di nuova immatricolazione debbano essere equipaggiate con EDR²⁶³. Una misura di questo tipo potrebbe essere opportunamente estesa anche ad altri settori; va in questa direzione il citato art. 12 dell'*AI Act*, che – tra i vari requisiti inerenti alla *product safety* nel campo dell'i.a. – prevede l'installazione obbligatoria, sui sistemi ad alto rischio, di strumenti di registrazione automatica degli eventi che si verificano durante il loro impiego²⁶⁴. Tale disposizione, tra l'altro, è da leggersi in combinato disposto con l'art. 8 della proposta di revisione della dir. 85/374/CEE, che, come vedremo al prossimo paragrafo, prevede che il giudice possa ordinare al convenuto la divulgazione degli elementi di prova a sua disposizione, qualora la domanda di risarcimento presentata dall'attore appaia fondata: tra tali elementi di prova si ritiene che possano rientrare anche i dati raccolti dai *logs* durante l'incidente (e, eventualmente, nelle fasi immediatamente precedenti e successive).

In ogni caso, se è evidente che l'installazione di EDR potrebbe dimostrarsi utile per escludere, nel caso concreto, una responsabilità – concorrente o alternativa – dell'utilizzatore del sistema di i.a., potrebbe restare comunque complesso, in alcune circostanze, risalire con certezza all'azione concretamente causativa del danno. Questa potrebbe infatti consistere in un difetto (di fabbricazione o di costruzione), ma potrebbe altresì dipendere da fattori esterni e ambientali, che potrebbero persino saldarsi tra loro in una “*web of causation*”²⁶⁵, destinata a rilasciare «effetti di tipo *interattivo* (dovuti alla *sommatoria* di alcune cause) o *sinergico* (capaci di esitare in un *fenomeno qualitativamente diverso* rispetto alla mera *sommatoria*)»²⁶⁶.

Le semplificazioni probatorie basate sul meccanismo del *res ipsa loquitur* potrebbero, in ogni caso, rendere più probabile l'accoglimento del ricorso, per il

of International Law, 2020, p. 195 ss.; K. QUEZADA-TAVÁREZ, *Legal challenges in bringing AI evidence to the criminal courtroom*, in *New Journal of European Criminal Law*, Vol. 12(4), p. 531 ss.

²⁶² Regolamento (UE) 2019/2144 del Parlamento europeo e del Consiglio del 27 novembre 2019 relativo ai requisiti di omologazione dei veicoli a motore e dei loro rimorchi.

²⁶³ Sul tema vd. E. AL MUREDEN, *Event Data Recorder e Advanced Driver Assistance Systems: la “spinta gentile” verso la mobilità del futuro*, in *Contr. Impr.*, 2022, n. 2, p. 390 ss.

²⁶⁴ Tale misura era già stata proposta dall' EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 47;

²⁶⁵ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1762.

²⁶⁶ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1762; vd. anche A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 9;

danneggiato: già oggi esse sono *de facto* utilizzate dai giudici e, se la proposta di revisione della direttiva sul danno da prodotto dovesse essere approvata, esse alleggeriranno notevolmente l'onere probatorio per il ricorrente (vd. *infra*, § 3.5.1., in questo cap.).

3.5.1. (segue) Profili processuali.

Sotto il profilo dell'*onus probandi*, l'accertamento del nesso di causalità dell'evento lesivo algoritmico incontra problematiche già osservate nell'ambito della tradizionale responsabilità per danno da prodotto. L'art. 120 cod. cons., recependo il disposto dell'art. 4 della Dir. 85/374/CEE, pone a carico del danneggiato l'onere di fornire la prova, oltre che del danno subito, dell'esistenza del difetto del prodotto e del nesso di causalità sussistente non semplicemente tra utilizzo del prodotto e danno, bensì tra *difetto del prodotto* e danno. Tale riparto dell'onere probatorio – che, di fatto, corrisponde a quello dell'ordinaria responsabilità aquiliana, ai sensi degli artt. 2043 e 2697 c.c. – è sempre stato segnalato, dalla dottrina, come un grave ostacolo all'ottenimento del risarcimento da parte del consumatore²⁶⁷.

Il contesto della produzione industriale ha sempre fisiologicamente determinato un'asimmetria informativa tra produttore e consumatore, in relazione alle specifiche caratteristiche del prodotto. Tale carenza di informazioni tecniche pare accentuarsi con riferimento al funzionamento dei sistemi di i.a., dal momento che la capacità di auto-apprendimento ed interazione aggiunge un livello di complessità all'individuazione del fatto causativo del danno²⁶⁸; *persino il produttore* – stante la già citata *opacità* del prodotto (vd. *supra*, cap. I, § 5.3) – potrebbe avere difficoltà ad individuarlo.

Sebbene abbia sempre confermato la vigenza del suddetto riparto dell'onere probatorio, la Corte di Giustizia ha ritenuto compatibili con la Direttiva le regole nazionali che consentono al giudice di ritenere provato il nesso di causalità in presenza di indizi

²⁶⁷ Vd., quando ancora non esisteva una disciplina speciale in materia di responsabilità per danno da prodotto difettoso, le riflessioni di U. CARNEVALI, *La responsabilità del produttore. Problemi generali*, cit., p. 154-157 e G. ALPA, *Responsabilità dell'impresa e tutela del consumatore*, Giuffrè, 1975, p. 430 ss.

²⁶⁸ P. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 175.; S. CHOPRA – L.F. WHITE, *A Legal Theory for Autonomous Artificial Agents*, cit., p. 144; S. WHITTAM, *Mind the compensation gap*, cit., p. 3 e p. 7; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1724.

gravi, precisi e concordanti della connessione eziologica, anche in assenza di dati scientifici univoci al riguardo. In particolare, i giudici lussemburghesi riconoscono che – specialmente quando oggetto di contenzioso sono prodotti ad alta complessità tecnica e/o scientifica – precludere tale possibilità significherebbe imporre uno *standard* probatorio talmente alto da escludere ogni modalità di prova diversa dalla prova certa a livello scientifico, rendendo così «eccessivamente difficile [...] o impossibile l'affermazione della responsabilità del produttore, in tal modo compromettendo l'effetto utile della direttiva 85/374»²⁶⁹.

Il carattere eccessivamente gravoso dell'onere della prova a carico del danneggiato è stato sottolineato anche dalle istituzioni europee, che, nella quarta relazione sull'applicazione della Direttiva, hanno prospettato, per un verso, l'ipotesi di introdurre un rovesciamento dell'onere della prova a favore del danneggiato e, per altro verso, l'ipotesi di prevedere dei meccanismi di *disclosure* obbligatoria in capo al produttore. Tali posizioni sono poi confluite nella proposta di revisione della direttiva sulla *product liability*.

L'art. 9 della proposta di revisione della Direttiva sul danno da prodotto cerca di rispondere alle problematiche emerse nell'applicazione concreta della disciplina, prevedendo una presunzione di difettosità del prodotto (§ 2) e una presunzione di causalità tra difetto e danno (§ 3), entrambe *iuris tantum* (vd. ult. §: «*Il convenuto ha il diritto di confutare le presunzioni di cui ai paragrafi 2, 3 e 4*»).

Quanto al rapporto di causalità tra difetto e danno, la proposta prevede che «*si presume l'esistenza del nesso di causalità tra il carattere difettoso del prodotto e il danno nel caso in cui sia stato provato che il prodotto è difettoso e che la natura del danno cagionato è generalmente coerente con il difetto in questione*» (art. 9, § 3). Anche al di fuori di tale ipotesi, il § 4 dell'art. 9 stabilisce che, qualora la complessità tecnologica del prodotto renda eccessivamente difficile, per il ricorrente, la prova del nesso di causalità – così come la prova del difetto – il giudice può presumere la loro sussistenza nel caso in

²⁶⁹ Corte di Giustizia UE, 21 giugno 2017, N. W, L. W, C. W/Sanofi Pasteur MSD SNC, Caisse primaire d'assurance maladie des Hauts-de-Seine, Carpimko, causa C-621/15, citata da A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1724. In quel caso, la Corte di Giustizia aveva ritenuto difettoso un vaccino, ritenendo «pacifico che la ricerca medica non ha permesso né di stabilire né di escludere l'esistenza di un nesso di causalità» tra il danno subito e l'iniezione del vaccino.

cui: «(a) il prodotto ha contribuito a cagionare il danno; e b) è probabile che il prodotto fosse difettoso, oppure che il carattere difettoso dello stesso è una causa probabile del danno, o entrambi tali elementi».

Non possiamo non rilevare, in relazione a quest'ultima disposizione, come il riferimento alla "probabilità" di causazione del danno è talmente vago da apparire privo di portata precettiva, dal momento che l'indagine sul nesso di causalità – nella sua duplice veste di *causalità generale* e *causalità individuale*²⁷⁰ – è necessariamente basata su stime probabilistiche.

Nell'ottica di porre rimedio, quantomeno *ex post*, all'asimmetria informativa tipica del rapporto consumatore-produttore, l'art. 8 della proposta prevede inoltre che l'autorità giudiziaria – qualora la domanda di risarcimento presentata dall'attore appaia plausibile – possa ordinare al convenuto la divulgazione di «*pertinenti elementi di prova a sua disposizione*» (§ 1), nei limiti di quanto sia «*necessario e proporzionato per sostenere la domanda*» (§ 2) e tenendo conto «*dei legittimi interessi di tutte le parti, compresi i terzi interessati, specialmente per quanto riguarda la protezione delle informazioni riservate e dei segreti commerciali*» (§ 3). Qualora al convenuto venga ordinato di divulgare informazioni che costituiscono un segreto commerciale (o presunto tale), gli organi giurisdizionali devono adottare «*le misure specifiche necessarie per tutelarne la riservatezza se tali informazioni vengono utilizzate o se ad esse è fatto riferimento nel corso di un procedimento giudiziario*» (§ 5)²⁷¹.

Tali disposizioni vanno necessariamente messe in relazione con l'art. 9, § 2, lett. c, della proposta, ai sensi del quale vi è una presunzione di difettosità del prodotto nel caso in cui il convenuto non abbia *rispettato l'obbligo di divulgare i pertinenti elementi di prova a sua disposizione a norma dell'articolo 8, paragrafo 1*».

²⁷⁰ Il tema è troppo vasto e ricco di implicazioni dogmatiche per essere trattato in questa sede. Rinviamo a G. VISINTINI, *Causalità e danno*, in *Contr. impr.*, 2015, n. 2, p. 262 ss., e F. PIRAINO, *Il nesso di causalità*, in *Eur. dir. priv.*, 2018, n. 2, p. 399, per un'efficace ricostruzione – in chiave problematizzante – dei due differenti aspetti che devono essere accertati in ordine al nesso di causalità (e sulle relative, diverse, tecniche di indagine probabilistica che devono essere utilizzate). Nella dottrina penalistica vd. F. STELLA, *Giustizia e modernità*, cit., p. 291 ss.

²⁷¹ Tale meccanismo è assimilabile a quanto stabilito dalla Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, cit., art. 3, vd. *supra*, § 1.2.

4. Responsabilità per esercizio di attività pericolose (art. 2050 c.c.).

Accanto alla responsabilità per danno da prodotto, ai sensi della dir. 85/374/CE (così come eventualmente revisionata a seguito dell'entrata in vigore della proposta presentata nel settembre 2022 dalla Commissione europea), una parte della dottrina ritiene configurabile, in capo al produttore di sistemi di i.a., il regime di responsabilità oggettiva previsto dall'art. 2050 c.c. per l'esercizio di attività pericolose²⁷². Se le ipotesi codicistiche di responsabilità oggettiva viste precedentemente (vd. *supra*, § 2 ss., in questo cap.) parevano attagliarsi in maniera più adeguata alla posizione dell'utilizzatore o fruitore di un sistema di i.a., l'art. 2050 c.c. delinea invece un'ipotesi di responsabilità oggettiva che potrebbe essere applicata nei confronti del produttore, in quanto esercente di un'attività pericolosa²⁷³.

In particolare, secondo tale prospettiva, il ricorso all'intelligenza artificiale può essere talvolta suscettibile di rendere pericolose attività che la giurisprudenza normalmente non considera come tali: si pensi, ad esempio, ai sistemi di i.a. in campo medico o ai veicoli a guida autonoma, che – sebbene promettano un complessivo beneficio in termini di riduzione degli incidenti – possono essere fonte di nuovi tipi di rischi²⁷⁴, generando “pericolosità settoriali anomale”²⁷⁵.

Il ricorso alla disciplina oggettiva di cui all'art. 2050 c.c. potrebbe rivelarsi particolarmente utile, per il danneggiato, laddove emerga un rischio da sviluppo, che, come abbiamo già evidenziato, è suscettibile di escludere la responsabilità da prodotto

²⁷² In questo senso, U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 108 ss.; ID., *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 248 ss.; U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, in *Giur. it.*, luglio 2019, p. 1712.

²⁷³ Sulla cumulabilità tra la tutela di cui alla dir. 85/374/CEE e quella prevista dall'art. 2050 c.c. vd. diffusamente U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 251 ss.

²⁷⁴ U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 249-250; ID., *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 108-109; *contra* A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 625 e A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit., nota 89, che, con particolare riferimento alle macchine a guida autonoma, escludono l'applicabilità dell'art. 2050 c.c., rilevando che, aderendo a questa lettura, «si finirebbe per qualificare come pericolosa un'attività (la guida automatizzata), la quale, se conforme alle regole della tecnica, è paradossalmente più sicura della normale attività di guida 'umana'».

²⁷⁵ U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 248.

difettoso (vd. art. 7, lett. e, dir. 85/374/CEE). Dal momento che, ad oggi, non paiono ancora chiaramente delineati i rapporti tra esimente *de qua* e capacità di autoapprendimento dei sistemi di i.a. (vd. *supra*, § 3.3.1., in questo cap.), la disciplina di cui all'art. 2050 c.c. potrebbe costituire un'adeguata soluzione per ristorare la persona che abbia subito un danno cagionato da un prodotto che, al momento della messa in commercio, non poteva essere considerato come difettoso. L'esistenza di un rischio di sviluppo, anzi, potrebbe costituire un indice di specifica pericolosità del prodotto²⁷⁶.

La responsabilità ai sensi dell'art. 2050 c.c. si perfeziona quando l'esercizio dell'attività pericolosa non venga accompagnata dall'adozione di misure idonee ad evitare il danno. In base alla lettura tipica dell'analisi economica del diritto, l'art. 2050 c.c. impone all'esercente i *costi* delle misure preventive, addossandogli, in assenza di queste, i *costi* da risarcimento dei successivi eventi dannosi²⁷⁷. La questione della "prevedibilità", dunque, non si pone sul piano del *difetto* – atteso che, come sottolineato, l'esistenza di un rischio da sviluppo è irrilevante in questa sede –, quanto, piuttosto, in relazione alle misure preventive, la cui idoneità deve essere valutata *ex ante*, alla luce delle tecnologie e delle conoscenze esistenti al momento della produzione²⁷⁸.

5. La personalità elettronica: una proposta europea, presto abbandonata.

Una delle proposte più radicali nel campo dell'allocazione dei costi derivanti da danni da i.a. è quella di introdurre una forma di responsabilità diretta patrimoniale dell'agente algoritmico, attraverso la creazione di una personalità elettronica²⁷⁹. Al sistema di i.a. sarebbe attribuito un *patrimonio separato* (sulla cui formazione e

²⁷⁶ U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 250-251.

²⁷⁷ U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 109.

²⁷⁸ U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 110.

²⁷⁹ Tra i maggiori sostenitori vd. G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, cit., *passim*; S. BECK, *The problem of ascribing legal responsibility in the case of robotics*, cit., p. 479 ss., nonostante quest'ultima riconosca come la personalità elettronica possa costituire soltanto una tra le soluzioni che il nostro ordinamento necessita per adattarsi al nuovo fenomeno dell'i.a.; per un'approfondita panoramica sulle diverse posizioni, da quelle più risalenti a quelle più recenti vd. M. HILDEBRANDT, *Criminal Liability and 'Smart' Environments*, in R.A. Duff (ed.), *Philosophical Foundations of Criminal Law*, Oxford University Press, 2011, p. 507 ss.; EAD, *Law for Computer Scientists and Other Folk*, Oxford University Press, 2020, p. 237 ss.

composizione si prospettano varie opzioni), destinato in via esclusiva al soddisfacimento degli obblighi risarcitori, senza che sia necessario l'accertamento di una colpa. La proposta richiama da vicino l'istituto del *peculium* di derivazione romanistica, che consisteva in un patrimonio che il *dominus* assegnava al *servus* (o al *filius*) e che era destinato a sostenere i costi derivanti dal risarcimento degli eventuali eventi dannosi cagionati da essi²⁸⁰.

La motivazione principale alla base di quest'idea consiste nella presunta inadeguatezza degli attuali regimi di responsabilità – tutti fondati, in maniera più o meno accentuata, sulla capacità di governo di una persona fisica su una *cosa* o su un'*attività* – a regolare la casistica dei danni provocati dai sistemi di i.a. autonomi, che sfuggono al controllo di produttori, utilizzatori e, più in generale, *operatori*. È ovvio, dunque, che tale proposta è condivisa soltanto da coloro che ritengono non applicabile la disciplina tradizionale in materia di responsabilità extracontrattuale: ove, infatti, quest'ultima si giudicasse adeguata, non vi sarebbe motivo per ricorrere ad una costruzione giuridica così radicale²⁸¹.

L'*electronic personhood* consentirebbe di contemperare le pretese risarcitorie del danneggiato – che avrebbe accesso ad una tutela indennitaria indipendente dall'accertamento di un difetto del prodotto o di una colpa dell'utilizzatore – con le esigenze di innovazione e di prevedibilità delle decisioni giudiziarie che fanno capo alle imprese. La previsione di una responsabilità civile diretta di un sistema di i.a. apporterebbe benefici principalmente di natura probatoria per il ricorrente²⁸². Inoltre, all'istituzione di una *e-personality* potrebbe accompagnarsi l'assegnazione di un numero identificativo a ciascun sistema di i.a., con corrispondente creazione di un "registro degli agenti artificiali", sul modello dei più tradizionali registri delle imprese: in questo modo, le persone che dovessero interagire con gli agenti intelligenti potrebbero informarsi sul

²⁸⁰ U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 264.

²⁸¹ Vd. ad es. E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 467, che considera la soluzione "ridondante", poiché ritiene che molti dei problemi che il nuovo status giuridico si propone di affrontare possono essere risolti attraverso i classici schemi della responsabilità per danno da prodotto o attraverso l'introduzione di un obbligo assicurativo.

²⁸² G. WAGNER, *Robot, Inc.*, cit., p. 611-2.

loro patrimonio, sui loro proprietari e, eventualmente, sugli eventi lesivi a loro riconducibili, al fine di valutarne l'affidabilità²⁸³.

L'ipotesi di introdurre una responsabilità patrimoniale diretta per i sistemi di i.a., d'altra parte, non deve far pensare al riconoscimento di una capacità morale (c.d. *moral standing*) in capo agli agenti artificiali, con conseguente attribuzione di uno *status* legale equiparabile a quello umano²⁸⁴. La proposta della c.d. *electronic personhood*, infatti, è dettata da *esigenze pratiche* di allocazione dei costi economici dei danni algoritmici, non dalla presa d'atto di una coscienza o autoconsapevolezza degli agenti elettronici, tale da spingere il legislatore a prendere in considerazione la loro dignità, attraverso l'ampliamento della nozione di soggettività giuridica umana²⁸⁵.

La *ratio* della personalità elettronica sarebbe dunque assimilabile, *mutatis mutandis*, a quella che ha condotto al riconoscimento di un'*autonomia patrimoniale* alle società e alle persone giuridiche in generale (associazioni, fondazioni, enti pubblici etc.), ovvero la separazione (perfetta o imperfetta) di un centro d'imputazione della responsabilità patrimoniale rispetto a quelli delle singole persone fisiche che fanno parte dell'ente.

Dalla personalità elettronica così intesa deve dunque essere distinto il dibattito sull'eventuale creazione di uno *status* legale in capo al sistema di i.a. – paragonabile a quello umano o a quello di altri esseri senzienti, quali gli animali – al fine di riconoscerli diritti e tutele. Suggerimenti di questo tipo – che spesso risalgono addirittura alla fine del secolo scorso o all'inizio degli anni Duemila – si fermano in realtà sul piano della speculazione, e sono da ricollegare alla prospettiva, lontana e puramente eventuale,

²⁸³ S. BECK, *The problem of ascribing legal responsibility in the case of robotics*, cit., p. 475. La proposta era già stata avanzata da C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, cit., p. 147 ss.

²⁸⁴ Distinguono accuratamente i due piani A. BERTOLINI – F. EPISCOPO, *Robots and AI as Legal Subjects? Disentangling the Ontological and Functional Perspective*, in *Front. Robot. AI*, 2022, vol. 9, p. 3; G. WAGNER, *Robot, Inc.*, cit., p. 608; U. PAGALLO, *Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots*, in *Information*, 2018, n. 9, p. 231 ss.; U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 264; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 445 ss.; S. BECK, *The problem of ascribing legal responsibility in the case of robotics*, cit., p. 479.

²⁸⁵ U. PAGALLO, *Vital, Sophia, and Co.*, cit., p. 231 ss.; U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 264.

dell'introduzione di sistemi di *artificial general intelligence*, in grado di svolgere attività intellettive ad ampio spettro e non limitate all'esecuzione di specifici *tasks*²⁸⁶.

L'ipotesi di introdurre una personalità elettronica in capo ai sistemi di i.a. (*rectius*, ai *robot*) sembrava aver trovato una sponda istituzionale nella Risoluzione del Parlamento europeo sulla regolazione della robotica (2017)²⁸⁷, che, al § 59, lett. f), proponeva – seppur in via residuale rispetto ad altre soluzioni giuridiche citate, quali, ad esempio, l'istituzione di un regime assicurativo obbligatorio o di un fondo di risarcimento – la creazione di «uno status giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi».

Da questa sintetica descrizione, in realtà, non è del tutto agevole stabilire se la Risoluzione prefigurasse il riconoscimento di una piena *capacità d'agire* in capo ai sistemi di i.a. o se invece fosse propensa all'introduzione di una responsabilità patrimoniale diretta dell'agente algoritmico²⁸⁸. In ogni caso, quale che fosse la prospettiva, tale soluzione pare essere stata abbandonata, dal momento che tutte le più recenti iniziative europee non accennano più alla proposta²⁸⁹ – complice anche, probabilmente, la massiccia dose di critiche che essa aveva ricevuto²⁹⁰.

²⁸⁶ L.B. SOLUM, *Legal Personhood for Artificial Intelligences*, in N.C. L. Rev., 1992, n. 70, p. 1231; G. TADDEI ELMI, *I diritti dell'intelligenza artificiale, tra soggettività e valore: fantadiritto o jus condendum?*, in L. Lombardi Vallauri (a cura di), *Il meritevole di tutela. Studi per una ricerca coordinata da L. Lombardi Vallauri*, Milano, 1990, p. 685 ss.; G. SARTOR, *L'intenzionalità dei sistemi informatici e il diritto*, cit., p. 23 ss.; ID., *Gli agenti software: nuovi soggetti del ciberdiritto?*, in *Contr. Impr.*, 2002, p. 465 ss.; D.J. CALVERLEY, *Imagining a Nonbiological Machine as a Legal Person*, in *Artificial Intelligence & Society*, 2008, p. 22; G. TEUBNER, *Rights of Non-Humans? Electronic Agents and Animals as New Actors*, in 33 *Journal of Law and Society*, 2006, p. 497 ss.

²⁸⁷ Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, cit., vd. § 59, lett. f).

²⁸⁸ Su quest'ambiguità vd. U. PAGALLO, *Vital, Sophia, and Co.*, cit., p. 239.

²⁸⁹ Vd. Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale, cit; Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL)); Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *L'intelligenza artificiale per l'Europa*, COM/2018/237 final.

²⁹⁰ Vd. COMITATO ECONOMICO E SOCIALE EUROPEO, *L'intelligenza artificiale – Le ricadute dell'intelligenza artificiale sul mercato unico (digitale), sulla produzione, sul consumo, sull'occupazione e sulla società*, INT/806, 2017, § 1.12 e 3.33; vd. anche la lettera firmata da numerosi esperti nel settore dell'intelligenza

In particolare, la dottrina maggioritaria non ha mancato di sottolineare gli aspetti critici della proposta di *electronic personhood*²⁹¹. Sebbene l'idea sia tutt'altro che peregrina, la sua realizzazione pratica potrebbe dar luogo a fraintendimenti e avere effetti sistemici e di lungo periodo imprevedibili²⁹².

Tralasciando gli ostacoli di ordine tecnico-normativo, primo tra tutti la non agevole delimitazione dei requisiti per l'attribuzione della personalità²⁹³, molti autori notano come l'introduzione di una *electronic personhood* – benché limitata al mero soddisfacimento di obbligazioni patrimoniali – potrebbe aggravare quella tendenza, già in corso, alla antropomorfizzazione dei sistemi di i.a. e alla creazione di aspettative sociali nei confronti del loro funzionamento, nonché, sul lungo periodo, all'estensione di diritti e garanzie un tempo riconosciuti soltanto agli esseri umani²⁹⁴.

La responsabilità diretta del sistema di i.a. potrebbe inoltre limitare in maniera irragionevole la tutela risarcitoria del danneggiato, dal momento che circoscriverebbe la responsabilità di produttori e/o utenti all'ammontare del patrimonio dell'ente robotico, così assunto a *limitato capitale di rischio*²⁹⁵. Come accade in relazione alle società e, in generale, agli enti collettivi, d'altra parte, la limitata esposizione economica dei finanziatori-azionisti della *persona elettronica* probabilmente sarebbe un incentivo ad assumere più rischi a costi inferiori, con conseguente prevedibile aumento degli eventi dannosi²⁹⁶.

artificiale, *Open Letter to the European Commission Artificial Intelligence And Robotics*, 2018, disponibile a questo link: www.robotics-openletter.eu.

²⁹¹ Ex multis U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 2, il Mulino, 2022, p. 234 ss.; G. COMANDÉ, *Intelligenza artificiale e responsabilità tra liability e accountability*, cit., p. 180; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 467; A. BERTOLINI – F. EPISCOPO, *Robots and AI as Legal Subjects?*, cit., passim.

²⁹² U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 264; R. ABBOTT – A. SARCH, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, in 53 *U. Cal. Davis L.R.*, 2019, p. 377

²⁹³ G. COMANDÉ, *Intelligenza artificiale e responsabilità tra liability e accountability*, cit., p. 180

²⁹⁴ R. ABBOTT – A. SARCH, *Punishing Artificial Intelligence*, cit., p. 377. Sul possibile riconoscimento di diritti in capo ai sistemi di i.a. vd., per tutti, D.J. GUNKEL, *Robot Rights*, MIT Press, 2018; J.C. GELLERS, *Rights for Robots. Artificial Intelligence, Animal and Environmental Law*, Routledge, 2021.

²⁹⁵ U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 263-264; B.A. KOCH, *Product Liability 2.0*, cit., p. 115; vd. anche EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 38; G. WAGNER, *Robot, Inc.*, cit., p. 611-2.

²⁹⁶ G. WAGNER, *Robot, Inc.*, cit., p. 612.

6. *Manufacturer Enterprise Liability* (MER) e assicurazione obbligatoria.

Prendendo atto delle difficoltà probatorie che il danneggiato potrebbe incontrare nell'ottenere il risarcimento del danno da intelligenza artificiale ai sensi della disciplina sul danno da prodotto, alcuni autori hanno proposto l'introduzione di un fondo indennitario, finanziato dalle società produttrici, al quale il ricorrente potrebbe accedere qualora dimostri di aver subito il danno²⁹⁷. La proposta è stata avanzata soprattutto nel settore delle macchine a guida autonoma, con l'appellativo di *Manufacturer Enterprise Responsibility* (MER)²⁹⁸, ma non vi è motivo per escludere che tale soluzione possa trovare applicazione anche con riferimento ad altri settori.

Si tratterebbe di un regime di responsabilità oggettiva *puro*, nell'accezione *supra* ricordata (vd. § 3.3., in questo cap.), dal momento che il danneggiato sarebbe esonerato dalla necessità di provare il difetto del prodotto – prova che, come abbiamo sottolineato (vd. *supra*, § 3.3. ss., in questo cap.), costituisce uno degli aspetti più controversi dell'intera disciplina contenuta nella dir. n. 85/374/CEE. D'altra parte, l'istituzione di un fondo indennitario avrebbe altresì la funzione di coprire i danni anonimi, attraverso una parziale socializzazione dei rischi, e avrebbe il vantaggio di consentire un risparmio delle spese legali per entrambe le parti coinvolte²⁹⁹. Ne risulterebbe, insomma, un incentivo, per i consumatori, a sostituire i prodotti tradizionali con i sistemi di i.a., con tutti i benefici, per la società nel suo complesso, che potrebbero derivarne (si pensi, per restare al campo degli *autonomous vehicle*, alla prevedibile riduzione degli incidenti).

²⁹⁷ Nella dottrina statunitense, la proposta è stata avanzata, con accenti diversi, da D.C. VLADECK, *Machines without principals: liability rules and artificial intelligence*, cit., p. 146 ss.; K.S. ABRAHAM – R.L. RABIN, *Automated vehicles and manufacturer responsibility for accidents*, cit., p. 105. In Italia, sono favorevoli A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 628; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1274; A. DAVOLA, *A Model for Tort Liability in a World of Driverless Cars: Establishing a Framework for the Upcoming Technology*, in 54 *Idaho L. Rev.*, 2018, p. 591 ss.; G. FINOCCHIARO, *Intelligenza artificiale e responsabilità*, cit.; sui vantaggi e gli svantaggi dell'introduzione di *compensation funds* vd. diffusamente G. BORGES, *New Liability Concepts: the Potential of Insurance and Compensation Funds*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 153 ss.

²⁹⁸ K.S. ABRAHAM – R.L. RABIN, *Automated vehicles and manufacturer responsibility for accidents*, cit., p. 146 ss.

²⁹⁹ U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1274.

In dottrina, per la verità, sono state proposte diverse “versioni” dell’istituto, che differiscono soprattutto per l’individuazione delle fonti di finanziamento e dei criteri in base ai quali dovrebbero ripartirsi gli oneri tra le società produttrici. Secondo alcuni, il fondo dovrebbe essere alimentato *pro quota* da tutti gli operatori che progettino, programmino e impieghino sistemi di i.a., sulla base delle quote di mercato occupate: in questo modo, il contributo di ciascun produttore al fondo sarebbe proporzionato all’entità del rischio introdotto nella società, quantificato sulla base del numero dei dispositivi venduti³⁰⁰. Altri autori obiettano che una simile socializzazione dei costi incentiverebbe il problema dei c.d. *free rider*³⁰¹ – dal momento che il fatto che il costo dei danni venga ripartito tra *tutte* le aziende disincentiva l’adozione di misure di sicurezza più performanti – e propongono di utilizzare, come parametro per la suddivisione dei costi, il criterio della frequenza e della gravità degli incidenti in cui il prodotto sia coinvolto³⁰².

Va comunque rilevato che c’è una parte della dottrina che si oppone al modello puro di responsabilità oggettiva delineato dalla *manufacturer enterprise liability*, obiettando che una tale soluzione potrebbe ritardare l’ingresso sul mercato di sistemi di i.a. che sono già sicuri (c.d. *chilling effect*), poiché i produttori potrebbero preferire attendere a commercializzare i prodotti fino a quando non siano in grado di poter calcolare con precisione i danni derivanti da essi, così da internalizzare il costo degli incidenti, trasferendolo sui consumatori³⁰³. L’attuale modello di responsabilità oggettiva, basato

³⁰⁰ D.C. VLADECK, *Machines without principals: liability rules and artificial intelligence*, cit., p. 146 ss.

³⁰¹ Parla di perdita del “*steering effect*” tipico del sistema di responsabilità (per colpa o limitatamente oggettivo) G. BORGES, *New Liability Concepts: the Potential of Insurance and Compensation Funds*, cit., p. 162. Simili problemi di *free riding* erano già emersi in relazione alla *Market Share Liability*, ossia a quell’orientamento giurisprudenziale che, negli Stati Uniti, ha consentito ai ricorrenti di ottenere il ristoro per i danni subiti, nonostante questi non fossero in grado di identificare la società produttrice responsabile della componente del prodotto che aveva causato il danno, vd. A.M. MARINO, *Market Share Liability and Economic Efficiency*, in 57 *Southern Economic Journal*, 1991, p. 667 ss. In generale, sulla *Market Share Liability* vd. *Sindell v. Abbott Labs.*, 607 P.2d 924 (Cal. 1980), il primo caso in cui una corte ha adottato tale orientamento giurisprudenziale; per una ricostruzione delle criticità di tale orientamento – che, di fatto, costituisce una deroga alla necessità di provare il nesso di causalità – vd. M.A. GEISTFELD, *The Doctrinal Unity of Alternative Liability and Market-Share Liability*, in 155 *University of Pennsylvania Law Review*, 2006, p. 447 ss.

³⁰² K.S. ABRAHAM – R.L. RABIN, *Automated vehicles and manufacturer responsibility for accidents*, cit., p. 169; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1274.

³⁰³ A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 628; A. DAVOLA, *A Model for Tort Liability in a World of Driverless Cars*, cit., p. 606-607; una soluzione analoga è proposta da C. SCHROLL, *Splitting the bill: creating a national car insurance fund to pay for accidents in autonomous vehicles*, in 109 *Northwestern University L. Rev.*, 2015, p. 803 ss.

sulla prova di una difettosità del prodotto, potrebbe tuttavia essere accompagnato dalla costituzione di un fondo di garanzia pubblico, destinato a ristorare il danneggiato “in via surrogata”, nel caso in cui il prodotto sia privo di difetti e, dunque, il patrimonio del produttore non sia aggredibile: in questo modo, si raggiungerà comunque l’obiettivo di evitare che il costo del danno resti in capo alla vittima³⁰⁴.

In alternativa alla creazione di un fondo indennitario *ad hoc*, il Parlamento europeo, nella Risoluzione sul diritto civile e la robotica del 16 febbraio 2017³⁰⁵, ha invece proposto un regime di assicurazione obbligatorio per tutti i dispositivi intelligenti, sul modello di quanto avviene per i veicoli a motore (ma che, a differenza di quest’ultimo, non coinvolgerebbe solo il proprietario o l’utilizzatore, ma l’intera pletora di soggetti che fanno parte della catena produttiva)³⁰⁶, con eventualmente l’integrazione di un fondo pubblico finalizzato a garantire la possibilità di ristoro nel caso in cui i sistemi di i.a. siano privi di copertura assicurativa³⁰⁷.

7. Intelligenza artificiale e responsabilità civile: indicazioni e spunti per il prosieguo dell’indagine.

Nel ripercorrere i processi che hanno portato al superamento del paradigma colpacentrico nella responsabilità aquiliana, abbiamo evidenziato alcune analogie tra il contesto industriale in cui si sono sviluppate le prime applicazioni dell’imputazione oggettiva del rischio di impresa e l’attuale dibattito sull’allocazione dei costi derivanti dai danni algoritmici. Come allora, anche oggi emerge l’esigenza di temperare l’innovazione industriale e sicurezza dei consumatori, e, come allora, il rischio da scongiurare è che si creino delle sacche di immunità giustificate dall’obiettivo di incentivare il progresso tecnologico e di evitare *chilling effects*.

³⁰⁴ A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 628-629.

³⁰⁵ Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, cit.

³⁰⁶ Risoluzione del Parlamento europeo del 16 febbraio 2017, cit., § 57 e 59, lett. a).

³⁰⁷ Risoluzione del Parlamento europeo del 16 febbraio 2017, cit., § 58 e 59, lett. b); in argomento vd. G. BORGES, *New Liability Concepts: the Potential of Insurance and Compensation Funds*, cit., p. 153 ss.

La “scoperta” dei modelli di imputazione oggettiva, a metà del secolo scorso, ha coinciso con la decisa recisione del cordone ombelicale che ancora univa responsabilità civile e responsabilità penale. Compito della responsabilità civile, infatti, non è la “ricerca” del “vero” autore del fatto dannoso, meritevole di biasimo da parte dell’ordinamento, quanto, piuttosto, l’individuazione di *criteri giuridici* per l’attribuzione del danno, che determinino *artificialmente* le condizioni in presenza delle quali il costo di un danno subito da un soggetto debba essere trasferito su un altro soggetto.

Nonostante, negli ultimi decenni, si sia assistito ad un ritorno a forme di responsabilità miste, non puramente oggettive (vd. § 1.2.), il criterio di allocazione del danno resta, nel diritto civile, intrinsecamente politico, non necessariamente legato alla ricerca del soggetto personalmente rimproverabile per aver cagionato l’evento lesivo. Accanto alla clausola generale prevista dall’art. 2043 c.c., infatti, l’ordinamento civile prevede già dei regimi (codicistici ed extracodicistici) in grado di imputare il danno da intelligenza artificiale all’utente e al produttore di un sistema di i.a., senza che sia necessario provarne la *colpa*.

La tipica *artificialità* del diritto civile moderno – unita alla pluralità di soluzioni, cumulative o alternative, finalizzate a ristorare il danneggiato – ne rende infatti plasmabile la conformazione rispetto alle scelte di gestione del rischio assunte dal decisore pubblico. Così, se la normativa esistente – eventualmente, con alcune lievi modifiche o interpretazioni estensive – potrebbe essere in grado di fornire un ristoro effettivo al soggetto che subisca un danno da intelligenza artificiale, non è escluso che il legislatore – come rappresentante degli interessi generali – possa intervenire per chiarire qual è l’allocazione dei costi che ritiene ottimale, a seconda degli scopi che si prefigge (es. sperimentazione di sistemi innovativi o maggiore cautela).

Nella nostra trattazione, in particolare, ci siamo concentrati sull’applicabilità dei regimi di responsabilità oggettivi codicistici (per quanto riguarda la posizione dell’utente, vd. § 2. ss.) e del regime contenuto nella direttiva sul danno da prodotto (per la posizione del produttore vd. § 3. ss.), tralasciando invece la disciplina generale della responsabilità per colpa (art. 2043 c.c.): sono questi i regimi, infatti, come d’altronde pronosticato dalla dottrina unanime, ad essere destinati all’applicazione quasi totalitaria nel caso di danni da intelligenza artificiale.

Quanto all'utilizzatore, innanzitutto, molti studiosi si sono già espressi sulla possibilità di applicare i regimi speciali di cui agli artt. 2049-2054 c.c. (con l'esclusione, tuttavia, dell'art. 2050 c.c. sulla responsabilità per l'esercizio di attività pericolose – sul quale vd. § 4 – che pare attagliarsi meglio alla posizione del produttore). Tali disposizioni si riferiscono a situazioni in cui vi è una persona che *ha creato* un rischio di evento dannoso, oppure che *beneficia* della creazione di un rischio o, ancora, che *si trova nella posizione migliore* per controllare un rischio. Benché l'individuazione di un criterio unificatore sia tutt'altro che agevole, ci pare di poter affermare che, quantomeno in senso lato, la *ratio* delle disposizioni citate possa essere riassunta nel principio del *cuius commoda, eius et incommoda esse debent*. Così, in questa prospettiva, si potrebbe ritenere ragionevole che chi decide di beneficiare dei più sofisticati algoritmi di intelligenza artificiale debba anche rispondere civilmente degli eventuali effetti collaterali che ne derivino.

Il riferimento ad una fattispecie in luogo dell'altra dipende qui dalla *concezione* che si ha dei sistemi di i.a., nonché dalla *ratio* delle norme codicistiche. Così, se si considera il sistema di i.a. come una cosa inanimata, si potrà ipotizzare di fare ricorso all'art. 2051 c.c. (a cui si aggiunge, nello specifico campo dell'*automotive*, la responsabilità del proprietario per i danni cagionati dall'autoveicolo ai sensi dell'art. 2054 c.c.). Se si ritiene che il sistema di i.a. possa essere considerato come una forma di intelligenza "minorata", si potrà fare riferimento all'art. 2052 c.c. Infine, qualora si consideri l'agente algoritmico alla stregua di una forma di intelligenza paragonabile a quella umana – quantomeno limitatamente alle specifiche funzioni che gli sono attribuite – la disciplina di riferimento sarà quella della responsabilità vicaria (art. 2049 c.c.).

Se la *ratio* di ritenere responsabile oggettivamente l'operatore del sistema di i.a. parrebbe in linea con le disposizioni codicistiche in materia di imputazione oggettiva, in assenza di un intervento legislativo – che definisca il bilanciamento, tutto *politico*, tra istanze di tutela del danneggiato ed esigenze di produzione e diffusione dei sistemi tecnologici più evoluti – l'applicazione estensiva delle norme codicistiche dovrebbe comunque ispirarsi a criteri di prudenza.

Nel lungo periodo, tuttavia, la progressiva attribuzione di autonomia ai dispositivi intelligenti – e la corrispondente perdita di controllo da parte degli utilizzatori – aprirà probabilmente la strada ad una maggiore responsabilizzazione delle società produttrici,

che si trovano nella posizione più adeguata per incidere sulla sicurezza dei dispositivi intelligenti.

A tal proposito, abbiamo preso in considerazione i principali elementi costitutivi della disciplina del danno da prodotto difettoso (nozioni di “prodotto”, “produttore”, “difetto”, “nesso causale”), cercando di dare rilievo ai problemi specifici che essi possono porre quando il prodotto sia costituito da un agente intelligente; in quest’ottica, abbiamo altresì dato atto delle modifiche che, sulla scorta del mutato contesto tecnologico nella fruizione dei beni di consumo, sono state proposte a livello europeo. In particolare, la proposta di revisione della direttiva sul danno da prodotto³⁰⁸ promette di aggiornare i concetti di “prodotto” e “produttore” alla comparsa sul mercato di “*prodotti fluidi*”, ossia di prodotti che subiscono delle modifiche, nel loro ciclo vitale, per effetto di aggiornamenti e modifiche del sistema operativo. A tal proposito, la Commissione europea propone di estendere la nozione di “prodotto” agli algoritmi che possono essere venduti anche separatamente rispetto all’*hardware* e la nozione di “produttore” ai programmatori e sviluppatori di *software* e codici algoritmici.

Se tali istanze di adeguamento riguardano, in generale, le più moderne modalità di produzione di beni di consumo – e possono essere soddisfatte piuttosto agevolmente, come mostrano le modifiche proposte dalla Commissione Europea –, la presunta inadeguatezza della nozione di “difettosità” è invece legata specificamente alla capacità di autonomo *decision making* dei sistemi di i.a.

Il concetto di “difettosità” – da sempre considerato il perno, ma anche l’elemento più controverso, della disciplina sulla responsabilità per danno da prodotto – introduce infatti una valutazione della condotta soggettiva del produttore, che, nella prassi, rischia di scolorire in un vero e proprio *test* di diligenza professionale. L’approssimarsi della disciplina *de qua* ad un modello di responsabilità per colpa – già oggetto di incertezze e perplessità da parte della dottrina in relazione ai prodotti tradizionali – incontra problemi peculiari in relazione all’imputazione dei danni cagionati da sistemi di i.a.

A meno di non voler considerare la stessa capacità di apprendimento alla stregua di un difetto, sarà infatti difficile stabilire quando i sistemi di i.a. siano difettosi, dal momento che questi sono *progettati* per discostarsi dalle informazioni ricevute in sede di

³⁰⁸ Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi, COM (2022) 495 final, 28 settembre 2022.

sviluppo. D'altra parte, anche qualora si ritenesse che la deviazione dell'algoritmo rispetto al percorso decisionale ordinario integri la nozione di "difetto", la responsabilità del produttore potrebbe essere esclusa sulla base della *later defect defense* (art. 118, lett. b, cod. cons.) o della *risk development defense* (vd. art. 118, lett. e, cod. cons.).

Ecco che, allora, se la tendenza in atto è quella di approssimare sempre di più la nozione di *difetto* ad una valutazione circa la *prevedibilità* dell'esito lesivo, si finisce per cascare nuovamente in quel *responsibility gap* che il modello oggettivistico mirava ad eludere.

Benché alcune delle modifiche contenute nella proposta di revisione della dir. 85/374/CEE, relative alla definizione di difetto, rispondano senza dubbio ad esigenze di aggiornamento del testo normativo, resta dunque il problema di fondo, non affrontato direttamente dalle istituzioni europee, di definire come incida il *machine learning* sulla configurabilità di un "difetto" del prodotto. La proposta di introdurre, tra i criteri per la valutazione di difettosità, quello relativo agli «*effetti sul prodotto dell'eventuale capacità di continuare ad imparare dopo la sua diffusione*» (art. 6, lett. c, proposta di revisione) non pare di grande aiuto, dal momento che non spiega *in che modo* l'autoapprendimento possa influire sul giudizio di difettosità. La capacità di *machine learning*, insomma, deve essere considerata come indice di difettosità del prodotto, o, al contrario, deve indurre ad escludere la suddetta difettosità?

Alla luce di tali problematiche, ci siamo interrogati su alcune interpretazioni evolutive del *consumer expectations test* che sono state proposte in dottrina: la sostanziale eliminazione del requisito del difetto (§ 3.3.6.1.) – che trasformerebbe, di fatto, la disciplina in esame in un regime di responsabilità oggettiva pura –, il riferimento all'ipotetica condotta di un agente modello *umano* (§ 3.3.6.2.) o *algoritmico* (§ 3.3.6.3.) e, infine, l'adozione della *preemption doctrine* (§ 3.4. ss.), che consentirebbe di ritenere sempre "non difettoso" il prodotto che rispetti gli *standard* di settore. Se quest'ultimo orientamento pare senz'altro da rigettare, stante la differenza strutturale e funzionale tra disciplina di *product safety* e quella di *product liability* (§ 3.4.5), alcune delle predette soluzioni sembrano delineare prospettive di rilievo, che potrebbero essere valorizzate, in via giurisprudenziale, nella risoluzione dei primi casi relativi all'imputazione dei danni da prodotto intelligente.

Particolare interesse suscita, nello specifico, la proposta di valutare la difettosità del prodotto sulla base delle sue *performance* complessive, paragonate alle prestazioni di un modello di algoritmo mediamente accurato. In questo modo, si eviterebbe l'*impasse* di mettere a confronto due modelli decisionali (quello algoritmico e quello umano) che sono intrinsecamente *incomparabili*, valorizzando altresì le prestazioni complessive dell'algoritmo – da valutarsi, eventualmente, sulla base di analisi statistiche.

Per concludere sulla responsabilità del produttore, un ultimo sguardo è stato dedicato al tema dell'accertamento del nesso di causalità, nei suoi profili sostanziali (§ 3.5.) e processuali (§ 3.6.). Posto che è da escludersi che l'autonomo *decision making* del sistema di i.a. possa determinare un'interruzione del nesso di causalità, ai sensi dell'art. 41, co. 2, c.p., l'opacità algoritmica potrebbe rendere particolarmente complessa la valutazione del decorso eziologico interno alla *black box*. A nostro parere, tuttavia, il criterio dell'aumento del rischio, ormai entrato a pieno titolo nell'armamentario logico-argomentativo delle corti civili, dovrebbe rendere più agevole la prova del nesso; lo stesso può dirsi della presunzione del *res ipsa loquitur*, utilizzata, nella prassi, da una parte della giurisprudenza e destinata a diventare criterio cogente qualora venisse approvata la proposta di revisione della direttiva sul danno da prodotto³⁰⁹.

Infine, la versatilità delle categorie della responsabilità aquiliana si apprezza anche nella docilità con cui queste si prestano ad accogliere l'eventuale introduzione di nuovi istituti di allocazione economica del rischio, da affiancarsi a quelli tradizionali. Così, le proposte volte ad istituire una personalità elettronica (§ 5) o un fondo di compensazione per i danni da i.a. (*Manufacturer Enterprise Liability*, § 6) possono, tuttalpiù, essere criticate per la *ridondanza* rispetto agli scopi che si prefiggono. Per il resto, esse potrebbero inserirsi senza frizioni nel contesto della responsabilità aquiliana, che – negli ampi confini dei principi costituzionali di solidarietà (art. 2), uguaglianza (art. 3), e libera iniziativa economica privata (art. 41) – assume, *liquidamente*, le forme della struttura economica di riferimento e delle scelte politiche di allocazione dei costi.

Nel diritto civile, insomma, ci pare che siano prospettabili diverse soluzioni per offrire ristoro economico alla persona danneggiata dai sistemi di i.a., fondate, da un lato,

³⁰⁹ Vd. art. 9, § 3 della proposta: «*si presume l'esistenza del nesso di causalità tra il carattere difettoso del prodotto e il danno nel caso in cui sia stato provato che il prodotto è difettoso e che la natura del danno cagionato è generalmente coerente con il difetto in questione*».

su semplificazioni probatorie in materia di accertamento del nesso di causalità, e, dall'altro lato, su criteri di imputazione *lato sensu* oggettivi. Anche laddove il modello oggettivistico rischia di scolorire in valutazioni sulla diligenza dell'agente – come in relazione al requisito del “difetto” nella responsabilità per danno da prodotto –, il tentativo portato avanti dalla dottrina è quello di limitare il più possibile tale giudizio, individuando i precisi parametri soggettivi ai quali ancorare tale valutazione giudiziale.

Il modello di ristoro offerto dal diritto civile per i danni da intelligenza artificiale si pone, dunque, in antitesi con il paradigma penalistico, che non ammette né semplificazioni probatorie – in forza dello *standard* dell'*oltre ogni ragionevole dubbio* (art. 533 c.p.p.), portato processuale della presunzione di innocenza (art. 27, co. 2, Cost.) – né criteri di imputazione oggettivistici, stante il principio di personalità della responsabilità penale (art. 27, co. 1, Cost.).

CAPITOLO III

INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ PENALE

SOMMARIO: 1. Diritto penale e intelligenza artificiale. Alcune premesse. – 2. La responsabilità penale diretta dei sistemi di i.a.: la tesi di Gabriel Hallevy. – 2.1. Le critiche: il principio di personalità della responsabilità penale. – 2.2. (*segue*) Le funzioni della pena e il paragone con la responsabilità (para) penale delle persone giuridiche. – 3. La responsabilità dell'utilizzatore di un sistema di i.a. – 3.1. Utilizzatore di sistema di i.a. totalmente autonomo: responsabilità colposa. – 3.2. Utilizzatore di sistema di i.a. parzialmente autonomo: responsabilità colposa. – 3.2.1. Il fatto tipico. – 3.2.2. La colpa. – 3.2.2.1. Il principio di affidamento. – 3.2.2.2. Un limite alla rimproverabilità dell'utilizzatore? Prospettive *de jure condendo*. – 3.3. La divergenza tra voluto e realizzato: *aberratio ictus* e *aberratio delicti*. – 3.4. Quale rilievo per il *dolus generalis*? Il caso degli *high frequency trading*. – 4. Responsabilità del produttore per l'evento lesivo derivante dal sistema di i.a. – 4.1. Il diritto penale di fronte all'incertezza: principio di precauzione e tecniche di incriminazione. – 4.2. La struttura del fatto tipico, tra commissione e omissione. – 4.3. L'accertamento del nesso di causalità. – 4.3.1. L'individuazione delle leggi scientifiche di copertura: un'incompatibilità tra ragionamento causale e approccio stocastico del *decision making* algoritmico? – 4.4. L'individuazione del soggetto personalmente responsabile: il c.d. *many hands problem*. – 4.4.1. La tipizzazione delle "aree di competenza" nella gestione del rischio ad opera dell'*AI Act*. – 4.5. La colpa del produttore. – 4.5.1. Le regole cautelari scritte: la colpa specifica. – 4.5.2. Il rapporto tra regole cautelari scritte e regole cautelari non scritte: quale spazio per il rischio consentito? – 4.5.2.1. Individuazione delle aree di rischio consentito e limiti all'operatività dell'istituto. – 4.5.2.2. Gli obblighi di monitoraggio e il "fallimento" delle regole cautelari positive. – 4.5.2.3. Intelligenza artificiale e rischio consentito: verso un approccio "quantitativo"? – 4.5.3. Le regole cautelari non scritte: la colpa generica. – 4.5.3.1. La prevedibilità dell'evento: è rimproverabile la *prevedibilità* dell'*imprevedibilità* algoritmica? – 4.5.3.2. Principio di precauzione e giudizio di prevedibilità dell'evento. – 4.5.4. Intelligenza artificiale, rischio consentito ed individuazione delle cautele doverose. – 5. Una tutela anticipata nei confronti dei sistemi di i.a. pericolosi? Prospettive *de jure condito*. – 5.1. (*segue*) Prospettive *de jure condendo*. – 6. Riepilogo: intelligenza artificiale e responsabilità penale.

1. Diritto penale e intelligenza artificiale. Alcune premesse.

Nel capitolo precedente, abbiamo cercato di mostrare come la responsabilità aquiliana può costituire lo strumento cardine per la tutela delle persone danneggiate dai sistemi di i.a.: da un lato, attraverso l'adeguamento della direttiva sul danno da prodotto alla nuova realtà tecnologica; dall'altro, attraverso la predisposizione di meccanismi di agevolazione probatoria per l'attore. Sebbene non manchino le questioni problematiche, la responsabilità aquiliana si conferma un istituto duttile, capace di adattarsi ai nuovi fenomeni tecnologici e di allinearsi alle decisioni sulla gestione del rischio assunte a livello politico.

Di contro, le garanzie che circondano il sistema costituzionale del diritto penale (prime tra tutte: legalità, offensività, colpevolezza, presunzione di innocenza) lo rendono un diritto *rigido*, inidoneo a contenere quei fenomeni sistemici, globalizzati ed incerti che caratterizzano la *post modernità*. Il danno da intelligenza artificiale, in particolare, sembrerebbe acuire quella crisi che il diritto penale vive di fronte alla società del rischio.

Di qui l'interrogativo, che si staglierà sullo sfondo di tutte le nostre riflessioni (e che è ben noto ai penalisti che da tempo si occupano del diritto penale nell'era della complessità): dobbiamo rassegnarci ad un certo grado di irrilevanza del diritto penale – inadatto ad attagliarsi ai danni derivanti da fenomeni tecnologici diffusivi e ubiquitari – oppure è auspicabile una rinuncia al diritto penale così come lo conosciamo, sempre più piegato ad istanze solidaristiche e *lato sensu* simboliche? Posto che, a nostro parere, non è auspicabile un superamento delle garanzie costituzionali, sulla scorta di pretese esigenze di “giustizia sostanziale”, in questa sede ci guardiamo bene dal voler fornire risposte ad interrogativi di carattere generale e di politica criminale, che vanno ben oltre il tema da noi indagato. Piuttosto, qui ci limiteremo ad illustrare le principali problematiche che i sistemi di i.a. pongono, dal punto di vista dell'imputazione penale degli eventi lesivi da essi derivanti: prenderemo in considerazione, in particolare, la posizione degli *utilizzatori* di dispositivi intelligenti e quella dei *produttori* – riconducendo, all'interno di questo termine, tutte le persone “dietro” allo sviluppo e alla commercializzazione dei sistemi di i.a. (fabbricanti, sviluppatori, programmatori, fornitori di dati, etc.), pur consci che, ai fini del riconoscimento di una responsabilità penale, sarà sempre necessario individuare il soggetto *personalmente rimproverabile* all'interno degli organigrammi aziendali.

Abbiamo già rilevato, nel primo capitolo, che i caratteri dell'intelligenza artificiale che rendono più problematica l'imputazione penale del danno cagionato dai dispositivi intelligenti sono l'autonomia, l'interattività e l'opacità tecnologica (vd. *supra*, cap. I, par. § 5). Tutti questi tre elementi contribuiscono, in vario modo, ad *allontanare* il fattore algoritmico lesivo dal controllo umano, determinando un'interposizione dell'imponderabile ed imperscrutabile *decision making* algoritmico tra produttore/utilizzatore ed evento lesivo.

In questo contesto, com'è evidente, un problema si pone, innanzitutto, sotto il profilo dell'accertamento del *nesso di causalità* tra condotta umana ed evento lesivo. Se

la futuribile introduzione di presunzioni di causalità, nel processo civile, potrà di certo agevolare la posizione del ricorrente (vd. *supra*, cap. II, § 3.5.1.), una simile strada non è ovviamente percorribile nella prova della causalità penale, in cui la presunzione di innocenza si riflette, in termini processuali, nella regola dell'*oltre ogni ragionevole dubbio*. Se, insomma, l'ordinamento civile – che fonda l'accertamento processuale proprio sul criterio del *più probabile che no* – può sopportare l'ipotesi di condanna al risarcimento di *una persona che non ha causato il danno*, lo stesso non può dirsi per l'applicazione della pena, che può essere inflitta soltanto qualora non residuino ragionevoli dubbi circa la colpevolezza dell'imputato.

La crisi del modello nomologico, determinata da una tendenziale inesplicabilità dei rapporti causa-effetto che governano l'agire algoritmico, si ripercuote, poi, a cascata, sull'intera struttura del reato e, in particolare, sull'accertamento della *colpa* in capo al produttore. L'assenza di generalizzazioni causali sul funzionamento del sistema di i.a. determina la perdita di controllo dell'utilizzatore e del produttore sull'agire algoritmico, con la conseguenza che potrebbe venir meno la possibilità di muovere un rimprovero personale all'agente¹. Il necessario rispetto del principio di colpevolezza (art. 27, co. 1 e 3, Cost.) impone infatti che sussista una connessione psicologica – quantomeno nella forma della colpa – tra azione umana ed evento lesivo: tale connessione rischia di saltare, se consideriamo che non esiste “agente modello” che possa pienamente controllare i rischi connessi al funzionamento dei sistemi di i.a.

Rinviando ai prossimi paragrafi un'analisi più approfondita di questi temi, ci limitiamo qui ad avvertire il lettore che la varietà e l'ampiezza delle possibili applicazioni dell'intelligenza artificiale – e di conseguenza, delle possibili offese ai beni giuridici – richiederebbero un'indagine settoriale di ben altra portata rispetto a quella contenuta nel presente lavoro. Il taglio che si è scelto di seguire in questo capitolo – come, del resto, nei capitoli precedenti – è tuttavia ispirato al tentativo di fornire una panoramica generale sulla tenuta degli elementi costitutivi del reato e dei principi del diritto penale, in relazione al nuovo contesto tecnologico. Ci occuperemo, insomma, della *parte generale* del “diritto

¹ Vd. per tutti L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 816; C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, 2020, n. 4, p. 1769 ss.

penale dell'i.a.", tralasciando invece quella che potremmo definire come *parte speciale* del diritto penale algoritmico.

In particolare, nella trattazione che segue ci soffermeremo sulla configurabilità di *reati causalmente orientati* e sulla possibile imputazione di essi all'utilizzatore (vd. *infra*, § 4. ss.) e al produttore del sistema di i.a. (§ 5. ss.); soltanto in maniera cursoria ci occuperemo della possibile configurabilità di reati dolosi a forma vincolata (§ 4.2.4) e sulle tecniche di anticipazione della tutela penale (§ 6).

Preliminare alla trattazione ci pare, tuttavia, la convinta confutazione di quelle tesi che – prendendo atto della difficile riconduzione del “reato da intelligenza artificiale” ai tradizionali modelli di imputazione penale – ipotizzano il riconoscimento di una responsabilità penale diretta in capo ai sistemi di i.a. È proprio con l'affermazione dell'indissolubile legame tra *intelligenza umana* e *diritto penale* che iniziano le nostre riflessioni sull'impatto dell'*intelligenza artificiale* sui modelli ascrittivi della responsabilità penale.

2. La responsabilità penale diretta per i sistemi di i.a.: la tesi di Gabriel Hallevy.

Partendo dalla constatazione circa la difficoltà nell'accertare una responsabilità in capo alle persone “intorno all'algoritmo”, alcuni autori hanno ipotizzato il riconoscimento di una responsabilità penale diretta dei sistemi di i.a.².

² Il principale teorizzatore della possibile configurazione di una responsabilità penale diretta in capo ai sistemi di i.a. è il penalista israeliano Gabriel Hallevy, vd. G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015; ID., “*I, Robot – I, Criminal*”—When Science Fiction Becomes Reality: Legal Liability of AI Robots committing Criminal Offenses, in 22 *Syracuse Sci. & Tech. L. Rep.* 1, 2010, p. 1 ss.; ID., *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in 4 *Akron Intellectual Property Journal*, 2010, p. 171 ss.; ID., *Dangerous Robots – Artificial Intelligence vs. Human Intelligence*, 11 febbraio 2018, in SSRN, disponibile a questo link: <https://ssrn.com/abstract=3121905>; nello stesso senso, ma con accezioni spesso molto diverse tra loro, vd. Y. HU, *Robot Criminals*, in 52 *U. Mich. J. L. Reform*, 2019, p. 487 ss.; C. MULLIGAN, *Revenge Against Robots*, in *South Carolina Law Review*, 2018, vol. 69, p. 579; M. SIMMLER – N. MARKWALDER, *Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in *Crim. Law Forum*, 2019, n. 30, p. 1 ss.; F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 433 ss. Possibilisti, in una prospettiva di lungo periodo, in cui elemento oggettivo e soggettivo del reato siano configurabili anche in capo ai sistemi di i.a., sono F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2019, fasc. 10, p. 31 ss.; ID., *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche*

Va premesso, anzitutto, che la tesi qui in esame non ha nulla a che vedere con le riflessioni svolte nel precedente capitolo (vd. *supra*, cap. II, § 5.) circa la possibilità di attribuire ai sistemi di i.a. una personalità elettronica. La proposta di introdurre una responsabilità patrimoniale diretta per gli agenti intelligenti, infatti, mira a frapporre uno schermo tra patrimonio personale del produttore o utilizzatore e patrimonio devoluto alla soddisfazione delle eventuali obbligazioni da fatto illecito algoritmico. Si tratta di una soluzione proposta per rispondere ad esigenze pragmatiche di *allocazione economica* dei costi derivanti dai danni da algoritmo, non dalla presa d'atto di una coscienza o autoconsapevolezza degli agenti elettronici. Come abbiamo già sottolineato, la *ratio* della personalità elettronica sarebbe assimilabile, *mutatis mutandis*, a quella che ha condotto al riconoscimento di un'*autonomia patrimoniale* alle società e alle persone giuridiche in generale (associazioni, fondazioni, enti pubblici etc.), ovverosia la separazione (perfetta o imperfetta) di un centro d'imputazione della responsabilità patrimoniale rispetto a quelli delle singole persone fisiche che fanno parte dell'ente.

L'introduzione di una *electronic personhood*, salvi i rilievi già svolti, non avrebbe l'impatto dirompente di scardinare la tradizionale dicotomia tra soggetto e oggetto, riconoscendo in capo ad artefatti frutto di ingegno umano una *capacità morale*. Sarebbe questo, invece, l'esito di un eventuale (ma, lo anticipiamo, improbabile) accoglimento delle teorie che sostengono la possibilità di riconoscere in capo ad un sistema di intelligenza artificiale una responsabilità penale diretta.

In particolare, il più noto sostenitore di questa tesi è il penalista israeliano Gabriel Hallevy, che in diverse occasioni ha cercato di dimostrare come la categorica negazione di una *capacità penale* in capo agli agenti artificiali non sia che il frutto avvelenato di un pregiudizio antropocentrico.

A parere di Hallevy, già negli attuali fatti dannosi commessi dai sistemi di i.a.³ sarebbero infatti ben identificabili, in capo all'agente artificiale, l'*actus reus* e la *mens rea*, ovverosia i due requisiti che, nei paesi di *common law*, costituiscono il fondamento

nuova riflessione, cit., p. 16-17; M. HILDEBRANDT, *Criminal Liability and 'Smart' Environments*, in R.A. Duff (ed.), *Philosophical Foundations of Criminal Law*, Oxford University Press, 2011, p. 526.

³ Parla di sistemi già esistenti negli articoli (per tutti G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities*, cit., 201); più prudente appare invece nella monografia (G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, cit., p. 229).

della responsabilità penale⁴ – e che sono sostanzialmente analoghi all’elemento oggettivo e soggettivo nella teoria bipartita. L’Autore si riferisce in particolare ai fatti dannosi derivanti dall’agire di quelli che abbiamo definito come *embodied AI systems* (vd. *supra*, cap. I, § 7.1.1.), ossia i *robot* dotati di intelligenza artificiale; non ci sono motivi, tuttavia, per non ritenere che tali argomentazioni non siano estendibili anche agli atti lesivi commessi dai *bots*.

Innanzitutto, lo studioso israeliano, partendo da una concezione schiettamente materialistica dell’elemento oggettivo del reato⁵ («*any material performance through factual–external presentation is considered an act*»⁶), ritiene che i robot siano capaci di compiere azioni penalmente rilevanti. Il caso portato ad esempio è quello di un evento lesivo che può essere cagionato dal braccio idraulico di un macchinario dotato di intelligenza artificiale («*when a machine, (e.g., robot equipped with artificial intelligence technology) moves its hydraulic arms or other devices of its, it is considered act*»⁷). Ancora, dal momento che *la macchina è capace di non agire di fronte ad un legittimo dovere di agire*, si ritiene che il *robot* potrà commettere altresì *fatti tipici omissivi*⁸.

Ci limitiamo soltanto a notare che, in quest’ottica, come d’altra parte rilevato dallo stesso Hallevy, anche il movimento di un braccio meccanico riferito ad un sistema *privo di intelligenza artificiale* può essere considerato un *actus reus*, stante la non necessarietà della *suitas* nella nozione di elemento oggettivo così delineata (e lo stesso dicasi con riferimento ai fatti omissivi)⁹.

Più ardita è la dimostrazione circa la possibile sussistenza dell’elemento soggettivo nell’agire robotico, sia sotto il profilo del dolo (c.d. *general intent*, che negli ordinamenti di *common law* comprende *intention*, *knowledge*, e *recklessness*) sia sotto il profilo della colpa (*negligence*).

⁴ Sulla distinzione tra *actus reus* e *mens rea* vd., per tutti, P.H. ROBINSON, *Structure and Function in Criminal Law*, Oxford University Press, 1997, p. 16 ss.

⁵ In questo senso A. CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018, p. 12 e 14.

⁶ G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, cit., 62.

⁷ G. HALLEVY, *ult. op. cit.*, p. 61.

⁸ G. HALLEVY, *ult. op. cit.*, p. 63.

⁹ Concordano con Hallevy – pur ritenendo necessario il carattere volontaristico del fatto tipico – F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 440-441.

Quanto al dolo, Hallevy sostiene che il comportamento emergente del sistema intelligente può essere caratterizzato sia da *cognizione* che da *volizione*: da un lato, infatti, l'agente di i.a. è in grado di acquisire, attraverso sensori, dati dall'ambiente esterno e di rielaborarli, apprezzando altresì la probabilità di verificazione di un evento lesivo¹⁰; dall'altro, il sistema è in grado di conformare il proprio comportamento agli obiettivi specifici che gli sono attribuiti, sulla base di un autonomo *decision making*¹¹.

Per quanto invece concerne la sussistenza di una “colpa algoritmica”, Hallevy ritiene che il sistema, talvolta, possa pervenire ad una percezione o ad una rappresentazione errata delle circostanze di fatto corrispondenti all'astratta fattispecie incriminatrice, eventualmente a causa di un processo di auto-apprendimento fallace: Hallevy definisce tale circostanza come “*unawareness*”, che dunque parrebbe corrispondere ad una *cognition* erronea. In questi casi, sarebbe possibile pervenire ad una valutazione di *negligenza* del sistema di i.a., qualora emerga che l’“agente modello algoritmico” avrebbe potuto riconoscere l'errore percettivo ed evitare l'esito lesivo¹².

Nel tentativo di definire una sorta di “sistematica del reato algoritmico”, l'Autore si premura altresì di verificare la compatibilità del modello così delineato con il paradigma classico della responsabilità penale, giungendo in alcuni casi ad una risposta positiva (ad esempio, per quanto concerne la configurabilità della scriminante della legittima difesa¹³) e rilevando invece, in altri casi, la necessità di adattamenti interpretativi (è il caso, ad esempio, della *intoxication defense*, che potrebbe trovare applicazione nel caso di malfunzionamento del sistema dovuto all'esposizione a materiali tossici¹⁴).

Come argomento a sostegno delle proprie tesi, l'Autore riporta un parallelismo suggestivo con la responsabilità “da reato” degli enti¹⁵. Hallevy osserva che l'esclusione degli agenti intelligenti dai soggetti del diritto penale ricorda quel pregiudizio antropocentrico che per secoli ha impedito di ipotizzare una responsabilità “penale” degli enti. Il superamento del brocardo *societas delinquere non potest* ha seguito un percorso lungo e irto di ostacoli, ma infine il pregiudizio è stato vinto da ragioni di ordine

¹⁰ G. HALLEVY, *ult. op. cit.*, p. 86 ss.

¹¹ G. HALLEVY, *ult. op. cit.*, p. 93 ss.

¹² G. HALLEVY, *ult. op. cit.*, p. 124 ss.

¹³ G. HALLEVY, *ult. op. cit.*, p. 169 ss.

¹⁴ G. HALLEVY, *ult. op. cit.*, p. 159 ss.

¹⁵ G. HALLEVY, *ult. op. cit.*, p. 199 ss.

pragmatico e dalla necessità di individuare un centro di imputazione penale che non si risolvesse nella ricerca di un capro espiatorio all'interno delle società.

D'altra parte, l'individuazione di una sanzione penale per i *robot* potrebbe rivelarsi operazione più immediata rispetto al processo di elaborazione delle pene nei confronti delle *corporations*. Se, infatti, una delle pregiudiziali alla configurazione di una responsabilità penale nei confronti degli enti è stata, per lungo tempo, quella sintetizzata nell'espressione «*No Soul to Damn: No Body to Kick*»¹⁶, con riferimento ai *robot* sarebbe quantomeno individuabile, se non un'*anima*, quantomeno un *corpo* da punire¹⁷.

Le “pene robotiche” alle quali andrebbero incontro i sistemi di i.a. ritenuti responsabili di fatti penalmente rilevanti potrebbero essere, nell'ottica di Hallevy, il *training* robotico, la privazione della “libertà personale” (o presunta tale) e la disattivazione¹⁸. Le sanzioni penali nei confronti degli agenti intelligenti verrebbero così ad assumere una finalità principalmente *special-preventiva*: in prima istanza, di *riabilitazione*, attraverso un programma di addestramento finalizzato a re-indirizzare i processi di *machine learning*, eventualmente accompagnato da una “detenzione”¹⁹; in seconda istanza, qualora il percorso di *riabilitazione* si riveli inadeguato, la pena dovrebbe consistere nell'*incapacitazione* del sistema di i.a., attraverso la disattivazione del sistema – accumulata alla *pena capitale*²⁰.

Nella prospettiva proposta da Hallevy, dunque, le pene robotiche non potrebbero svolgere alcuna funzione di tipo general-preventivo o retributivo, atteso che lo stesso autore riconosce che i sistemi di i.a. non sono capaci di *autoconsapevolezza* o *sofferenza*,

¹⁶ L'espressione è notissima nel dibattito angloamericano in tema di *corporate liability*, vd. J.C. COFFEE, “*No Soul to Damn: No Body to Kick*”: *An Unscandalized Inquiry into the Problem of Corporate Punishment*, in *Michigan Law Review*, 1981, p. 386 ss. L'espressione è stata poi ripresa da vari autori, proprio con riferimento al dibattito sulla responsabilità diretta dei sistemi di i.a.; vd. P.M. ASARO, *A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, in Lin e al. (eds.), *Robot Ethics*, cit.

¹⁷ Questa considerazione, ovviamente, si presta a trovare applicazione soltanto con riferimento a agli *embodied AI systems*, non ai *bots*.

¹⁸ G. HALLEVY, *ult. op. cit.*, p. 212 ss. Altre pene sono immaginate da Y. HU, *Robot Criminal Liability Revisited*, cit., p. 529; vd. anche M.A. LEMSLEY – B. CASEY, *Remedies for Robots*, in 86 *U. Chi. L. Rev.*, 2019, p. 1311 ss., p. 1389 ss.; F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 460.

¹⁹ G. HALLEVY, *ult. op. cit.*, p. 211 e 220.

²⁰ G. HALLEVY, *ult. op. cit.*, p. 217.

né sono in grado di conformare il proprio comportamento nel rispetto di norme imperative²¹.

La teoria di Gabriel Hallevy – così come degli autori che, con accezioni anche diversificate tra loro, condividono la possibilità di muovere un rimprovero ad un agente artificiale – riecheggia la teoria funzionalista di Günther Jakobs, in base alla quale il concetto di libero arbitrio non avrebbe un fondamento biofisico, ontologico, ma dovrebbe invece essere considerato alla stregua di un costrutto sociale²².

Strettamente legata al concetto di libero arbitrio, anche la *personalità giuridica* non dipenderebbe, così, da caratteristiche intrinseche e ontologiche di un *agente*, ma non sarebbe altro che un artificio socialmente indotto, capace di essere proiettato in direzione di qualsiasi entità capace di deludere aspettative sociali di comportamento²³. Così, qualora si riconoscesse ai sistemi di i.a. tale capacità, sarebbe legittima, di conseguenza, una risposta penale da parte dell'ordinamento.

2.1. Le critiche: il principio di personalità della responsabilità penale.

La dottrina largamente maggioritaria – salvo alcune isolate voci di cui abbiamo dato conto – giudica la tesi di Gabriel Hallevy incompatibile con i principi cardine del diritto penale.

Innanzitutto, alcuni autori contestano la ricostruzione dogmatica già sul piano della presunta sussistenza di un *actus reus* robotico²⁴. Sia negli ordinamenti di *common*

²¹ In realtà, nei suoi primi lavori, Hallevy giunge ad individuare uno spazio anche per la funzione general-preventiva, poi abbandonata nella sua monografia. Riconoscono invece una possibile efficacia general-preventiva alle pene robotiche F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 460. Secondo altri, invece, la pena nei confronti delle macchine dovrebbe avere finalità di “vendetta”, rispondendo così, catarticamente, alle aspettative retributiviste dei consociati vd. C. MULLIGAN, *Revenge Against Robots*, cit., *passim*; in argomento vd. le acute riflessioni di A. CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, cit., p. 33.

²² Vd. G. JAKOBS, *Sistema dell'imputazione penale*, Napoli, 2017, espressamente citato da M. SIMMLER – N. MARKWALDER, *Guilty Robots?*, cit., p. 14.

²³ M. SIMMLER – N. MARKWALDER, *Guilty Robots?*, cit., p. 14.

²⁴ C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., p. 1769; D. LIMA, *Could AI Agents Be Held Criminally Liable?*, cit., p. 680; M. FREITAS – F. ANDRADE – P. NOVAIS, *Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible*, in P. Casanovas – U. Pagallo – M. Palmirani – G. Sartor (eds.) *AI Approaches to the Complexity of Legal Systems*, Springer, 2014, p. 152; U. PAGALLO, *AI and bad robots. The criminology of automation*, in M. R. McGuire – T. Holt (eds.), *The*

law²⁵, sia in quelli di *civil law*, infatti, il fatto tipico deve essere *imputabile psicologicamente all'autore*. Nel nostro ordinamento, com'è noto, il riferimento normativo è l'art. 42, co. 1, c.p., che esclude che l'agente possa essere punito per un fatto che non ha commesso con “coscienza e volontà”²⁶.

Senza addentrarci nel dibattito tra coloro che ritengono la *suïtas* come afferente al fatto tipico²⁷ e coloro che, invece, la collocano nell'ambito della colpevolezza²⁸, basterà qui rilevare che di certo l'azione robotica non pare essere espressione di *coscienza e volontà*.

Come abbiamo illustrato nel primo capitolo (vd. *supra*, cap. I, § 5.1. ss.), il sistema di i.a. rielabora le informazioni esterne attraverso un algoritmo, che è tuttavia privo della capacità di autodeterminazione che – seppur soltanto come postulato epistemologico – caratterizza l'agire umano. I comportamenti emergenti dei sistemi di i.a. sono pur sempre derivanti dal codice e dall'interazione di questo con l'ambiente esterno, sebbene la molteplicità e la complessità dei fattori interagenti sia talmente enorme da renderne ostica la ricostruzione *ex post* e praticamente impossibile la prevedibilità *ex ante*. Nell'agire algoritmico non esiste, dunque, uno spazio di *scelta* e di *governo* di un corpo –

Routledge Handbook of Technology, Crime and Justice, Routledge, 2017, p. 645; B. PANATTONI, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, in *Dir. inf.*, 2021, fasc. 1, p. 345, nt. 112; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law. Do we need a “new” criminal law for autonomous robots?*, in WWU Münster, disponibile a questo link: <https://miami.uni-muenster.de/Record/ff3df8a9-0daf-4dd8-a015-af91cb499f1a>, p. 12.

²⁵ Si vd. ad es., nel contesto statunitense il *voluntary act principle*, espresso dal § 2.01 del Model Penal Code: «2.01 Requirement of Voluntary Act; Omission as Basis of Liability; Possession as an Act. (1) A person is not guilty of an offense unless his liability is based on conduct which includes a voluntary act or the omission to perform an act of which he is physically capable. (2) The following are not voluntary acts within the meaning of this Section: (a) a reflex or convulsion; (b) a bodily movement during unconsciousness or sleep; (c) conduct during hypnosis or resulting from hypnotic suggestion; (d) a bodily movement that otherwise is not a product of the effort or determination of the actor, either conscious or habitual». Nella dottrina statunitense vd. anche M.S. MOORE, *Act and Crime: The Philosophy of Action and its Implications for Criminal Law*, Oxford University Press, 1993; per una ricostruzione delle teorie tradizionali sul *voluntary act requirement* vd. M. CORRADO, *Is There an Act Requirement in the Criminal Law?*, in 142 *University of Pennsylvania Law Review*, 1994, p. 1529 ss.

²⁶ F. ANTOLISEI, *L'azione e l'evento nel reato*, Istituto Editoriale Scientifico, 1928, *passim* e p. 59.

²⁷ G. FIANDACA – E. MUSCO, *Diritto penale. Parte Generale*, Zanichelli, VIII ed., 2019, p. 232; T. PADOVANI, *Diritto penale*, Giuffrè, XII ed., 2019, p. 146 ss.; F. PALAZZO, *Corso di diritto penale. Parte generale*, Giappichelli, VI ed., 2016, p. 316-317; F. MANTOVANI, *Diritto penale. Parte generale*, Cedam, XI ed., 2020, p. 327; M. DONINI, *Teoria del reato. Una Introduzione*, Cedam, 1996, p. 114.

²⁸ G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale. Parte generale*, Giuffrè, IX ed., 2022, p. 465.

caratteristiche che devono ritenersi imprescindibili affinché il fatto possa dirsi cosciente e volontario.

In ogni caso, anche a voler riconoscere in capo al sistema di i.a. una capacità di azione penalmente rilevante, il problema insormontabile alla configurabilità di una responsabilità penale algoritmica parrebbe essere rappresentato dall'impossibilità di muovere un rimprovero di colpevolezza all'agente intelligente, così come è invece imposto dal principio di *personalità della responsabilità penale* (art. 27, co. 1, Cost.)²⁹.

Il sistema di i.a., come abbiamo già ribadito in più occasioni, non è capace di *discernimento semantico* o di ragionamento, ma si limita a riconoscere *pattern* ricorrenti in un *dataset*. Tale attività *simula* l'intelligenza e la comprensione umana, ma è priva di un corrispettivo psicologico e cognitivo³⁰.

Concentriamo la nostra attenzione, in particolare, sull'ipotetica configurabilità di un dolo algoritmico, e, più precisamente, di un dolo intenzionale. È vero che il sistema di i.a. può *rappresentarsi* la commissione di un fatto; è altresì vero che la rappresentazione può abbracciare anche la contrarietà di tale fatto ad una norma penale; è, infine, innegabile

²⁹ In Italia, C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1769 ss.; A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 14; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in Riv. it. dir. proc. pen., 2021, n. 1, p. 96; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 71; M.B. MAGRO, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in Leg. pen., 10 maggio 2020, p. 9 ss.; EAD., *Robot, cyborg e intelligenze artificiali*, in A. Cadoppi – S. Canestrari – A. Manna – M. Papa (a cura di), *Cybercrime*, Utet, 2019, p. 1202 ss.; A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 747; F. MAZZACUVA, *The impact of AI on corporate criminal liability: algorithmic misconduct in the prism of derivative and holistic theories*, in 92 *International Review of Penal Law*, 2021, cit., p. 153-154; R. BORSARI, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, cit., p. 262 ss.; L. PICOTTI, *Cybercrimes and Criminal Relevance of Artificial Intelligence*, in F.J. Castro Toledo (coord.), *La transformación algorítmica del sistema de justicia penal*, Thomson Reuters, 2022, p. 70; nella dottrina straniera, vd. U. PAGALLO, *The Laws of Robots*, Springer, 2015 p. 50-51; ID., *Vital, Sophia, and Co. – The Quest for the Legal Personhood of Robotics*, in 9 *Information*, 2018, p. 7; S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame?*, cit., p. 417; M. FREITAS – F. ANDRADE – P. NOVAIS, *Criminal Liability of Autonomous Agents*, p. 153-154; P. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 176 ss.; T.C. KING E AL., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, cit., p. 108; M. COECKELBERGH, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, in 26 *Science and Engineering Ethics*, 2020, p. 2051 ss.; A. MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, in 92 *International Review of Penal Law*, 2021, p. 118; S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 77; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 12 ss.

³⁰ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 14; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 12; EAD., *Robot, cyborg e intelligenze artificiali*, cit., p. 1195 ss.

che l'agente artificiale possa agire, valutando la convenienza – per gli scopi che gli sono attribuiti – della commissione di quel fatto. Il presunto dolo così delineato, tuttavia, non è altro che un *simulacro*, una scatola vuota, che si limita a replicare esteriormente processi psicologici inesistenti. Come è stato osservato in dottrina, insomma, «i soggetti artificiali intelligenti di tal tipo possano agire con tipicità dolosa. Ma questa nulla ancora ci dice riguardo alla colpevolezza dolosa»³¹. Ordinariamente, la dimostrazione della prima è sufficiente a provare la seconda: non potendo indagare direttamente il contegno psicologico dell'agente, la prova dell'intenzionalità è affidata necessariamente a meccanismi presuntivi basati sull'osservazione dell'atteggiamento esteriore del soggetto. Soltanto eccezionalmente il dolo così individuato non sarà sufficiente a muovere un rimprovero di colpevolezza all'agente: ciò avverrà quando siano intervenute delle circostanze anormali che abbiano influito in maniera irresistibile sulla volontà o sulle capacità psicofisiche del soggetto (scusanti), quando l'agente sia stato incapace di intendere o volere o, infine, quando, al momento di agire, egli non conoscesse, senza colpa, la legge penale violata.

Lo stesso meccanismo presuntivo non potrà risultare valido, com'è evidente, in relazione agli algoritmi intelligenti³². Non esiste infatti, ad oggi, una regola di esperienza che consenta di ricollegare l'esteriorizzazione di *cognizione* e *volizione* da parte dell'agente di i.a. ad una capacità interiore di libera autodeterminazione³³.

Anche volendo aderire alla tesi funzionalista implicitamente fatta propria da Gabriel Hallevy, va comunque rilevato che, ad oggi, la percezione sociale dei sistemi di i.a. è ben lontana dal nutrire aspettative di autonoma conformazione alle regole della convivenza sociale³⁴. Se anche si considerassero i concetti di autodeterminazione e libero arbitrio – fondamentali per il riconoscimento di una colpevolezza in capo all'agente – come un'illusione, una *fictio* giuridica, in ogni caso non pare che tale illusione coinvolga anche, nella percezione sociale, i sistemi di i.a. Gli agenti artificiali sono infatti concepiti come *oggetti*, che, ancorché in una certa misura imprevedibili, rispondono deterministicamente agli *input* e agli stimoli derivanti dall'ambiente. L'attribuzione di

³¹ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 14.

³² A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 14.

³³ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 14-15.

³⁴ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 14-15; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 13 ss.

fatti penalmente rilevanti ai sistemi di i.a., piuttosto che agli esseri umani che stanno dietro allo sviluppo e all'utilizzo di tali sistemi, sembrerebbe, d'altra parte, creare un'ingiustificata area di impunità a favore di questi ultimi – nei cui confronti, al contrario, saranno presumibilmente indirizzati i sentimenti di riprovazione e biasimo da parte dei consociati³⁵. La responsabilità penale diretta dei sistemi di i.a., dunque, non avrebbe in ogni caso l'effetto di colmare il *retribution gap*³⁶ determinato dall'impossibilità di identificare una persona personalmente responsabile per la commissione dell'illecito algoritmico.

2.2. (segue) Le funzioni della pena e il paragone con la responsabilità (para) penale delle persone giuridiche.

Infine, le ultime due critiche che possono essere mosse alla ricostruzione di Gabriel Hallevy riguardano la compatibilità della teoria con le funzioni tradizionalmente attribuite alla pena e la pertinenza del paragone con la responsabilità da reato degli enti.

Quanto al primo profilo, la (consapevole) soppressione dagli orizzonti della pena della funzione retributiva e, soprattutto, di quella general-preventiva non pare coerente con le tradizionali teorie della pena né, per quanto concerne il nostro ordinamento, parrebbe sostenibile dal punto di vista costituzionale³⁷. Di retribuzione, a rigore, non si potrebbe parlare oggi, perché le attuali macchine intelligenti continuano a essere incapaci di percepire sensi di colpa, riprovazione sociale, vergogna o sofferenza³⁸. Allo stesso

³⁵ S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 79-80; F. CONSULICH, *Flash offenders. Le prospettive di accountability penale nel contrasto alle intelligenze artificiali devianti*, in corso di pubblicazione sulla Riv. it. dir. proc. pen., 2022, fasc. 4, p. 1048.

³⁶ J. DANAHER, *Robots, Law and the Retribution Gap*, in 18 *Ethics and Information Technology*, 2016, p. 299 ss..

³⁷ L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 832; ID., *Cybercrimes and Criminal Relevance of Artificial Intelligence*, cit., p. 70; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 9; A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 15; nella letteratura straniera vd. P.M. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 181; U. PAGALLO, *Killers, fridges, and slaves: a legal journey in robotics*, in *AI & Society*, vol. 26, p. 350; D. LIMA, *Could AI Agents Be Held Criminally Liable?*, cit., p. 688-689; M. HILDEBRANDT, *Ambient Intelligence, Criminal Liability and Democracy*, cit., p. 167.

³⁸ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 15; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1770; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 10; F. CONSULICH, *Flash offenders*,

modo, gli algoritmi non possono provare nemmeno il *timore della sanzione*, dal momento che né la previsione incriminatrice in sé, né l'afflizione della pena nei confronti di un altro agente robotico, potrebbero veicolare un messaggio di *dissuasione* rispetto ad un determinato comportamento. La condanna all'esecuzione di una pena, dunque, non avrebbe alcun significato comunicativo nei confronti degli altri "consociati algoritmici", ma sarebbe destinata a restare un episodio isolato, con effetti limitati al solo sistema sottoposto a pena³⁹.

E quali effetti, sul piano special-preventivo? Gabriel Hallevy, come accennato, individua un possibile spazio applicativo per la funzione riabilitativa, immaginando un procedimento di *training* che ripristini il normale funzionamento dell'algoritmo. Anche in relazione a quest'ipotesi, tuttavia, le similitudini con l'ordinaria attività rieducativa svolta in sede di esecuzione penale paiono ridotte⁴⁰: il cammino di risocializzazione, infatti, è possibile soltanto attraverso un impegno personale del condannato, che decide di intraprendere un percorso interiore di analisi e di messa in discussione del proprio passato – un percorso che, per le motivazioni anzidette, non pare replicabile con riferimento all'agire algoritmico. Il *training* dell'algoritmo e la sua eventuale riprogrammazione non costituiscono uno strumento di *pressione indiretta* assimilabile alla pena, ma – proprio come la disattivazione, la "pena di morte algoritmica" nella teoria di Hallevy – corrispondono piuttosto ad un intervento manipolativo sull'algoritmo, volto a modificarne in maniera "coercitiva" le modalità di agire⁴¹.

cit., p. 1026; A. MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, cit., p. 119; R. SPARROW, *Killer robots*, cit., p. 72; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 16.

³⁹ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 16; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1770; F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 32; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 99; P.M. ASARO, *A body to Kick, but Still No Soul to Damn*, cit., 181; F. CONSULICH, *Flash offenders*, cit., p. 1022; A. MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, cit., p. 118.

⁴⁰ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1770; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 9; A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 16; P.M. ASARO, *A body to Kick, but Still No Soul to Damn*, cit., 181; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 98-99.

⁴¹ F. CONSULICH, *Flash offenders*, cit., p. 1025.

Le “pene robotiche” prese in considerazione dai teorici della responsabilità penale diretta dei sistemi di i.a., insomma, hanno un fondamento puramente simbolico⁴² e non hanno nulla a che vedere con la pena tradizionalmente intesa – concetto che, fin dal suo etimo, riporta ad una *sofferenza*, sentimento che, quantomeno ad oggi, pare inconcepibile in assenza di substrato (*hardware*) biologico⁴³. Al di là dei profili di responsabilità penale delle “persone intorno alla macchina” – che tratteremo più diffusamente più avanti (vd. *infra*, § 4. ss.) – ci limitiamo qui a rilevare che l’ordinamento penale dispone già di strumenti idonei a neutralizzare la pericolosità intrinseca delle cose: ci riferiamo, in particolare, alla confisca (art. 240 c.p.) e all’obbligo di richiamo e ritiro dei prodotti pericolosi (previsto, in generale, dall’art. 104, co. 3 ss., cod. cons., e, nello specifico, dalle normative settoriali).

Per concludere, anche il parallelismo con la responsabilità da reato degli enti non appare del tutto pertinente. Il percorso secolare che ha portato al riconoscimento di una capacità penale in capo alle persone giuridiche (o para-penale, com’è il caso dell’Italia) è stato dettato dalla necessità di attribuire autonomo rilievo ad una colpevolezza *propria* dell’ente, nella convinzione che la *colpa di organizzazione* non si risolve nell’insieme delle “colpe” riconducibili a quelle persone che di tale organizzazione, a vario titolo, fanno parte. Gli enti restano, tuttavia, delle costruzioni “naturalisticamente fittizie”⁴⁴, nel senso che non possono agire *direttamente* sul mondo, ma operano attraverso le condotte di persone fisiche, tanto che la responsabilità amministrativa degli enti presuppone la commissione di un reato proprio da parte dei soggetti apicali (art. 5, lett. a), d.lgs. 231/2001) o subordinati (art. 5, lett. b), d.lgs. 231/2001).

La società costituisce dunque soltanto il destinatario *apparente* della funzione dissuasiva esercitata dalla previsione di una pena per la società; il comando sotteso alla norma è invece inevitabilmente rivolto al *management* della società, affinché adotti misure idonee a minimizzare il rischio-reato⁴⁵.

⁴² Così S. AIRES DE SOUSA, “*Não fui eu, foi a máquina*”: teoria do crime, responsabilidade e inteligência artificial, cit., p. 78; implicitamente anche J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 18.

⁴³ M.B. MAGRO, *Decisione umana e decisione robotica*, cit., *passim*; F. CONSULICH, *Flash offenders*, cit., p. 1025 ss.

⁴⁴ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 17; vd. anche C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1768.

⁴⁵ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1768; P. SEVERINO, *Intelligenza artificiale e diritto penale*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 535; M.B.

I sistemi di i.a., al contrario, possono *incidere* direttamente sul mondo reale, fisico o virtuale che sia (vd. la classificazione che abbiamo proposto *supra*, cap. I, § 7), senza che sia necessario un intervento di mediazione da parte delle persone fisiche. La previsione di una sanzione penale nei confronti di un sistema di i.a., insomma, non potrà avere alcuna funzione di pressione sulle persone che hanno prodotto o messo in commercio la macchina⁴⁶ – se non assolutamente indiretta, come ripercussione negativa derivante dalla disattivazione di alcuni prodotti intelligenti o da un eventuale calo delle vendite dovuto ad una crisi reputazionale. Tuttalpiù, potrà essere il proprietario o l'utilizzatore della macchina che potrà subire un pregiudizio economico in caso di “afflizione della pena” all'algoritmo⁴⁷ – niente, tuttavia, che sia paragonabile all'efficacia dissuasiva che la previsione di una sanzione penale per l'ente può avere sulla gestione del rischio societario.

Per concludere, non si può non rilevare come il rischio che la tesi di Hallevy porta con sé è quello di una *deresponsabilizzazione* delle persone fisiche che producono o utilizzano i sistemi di i.a., che potrebbero beneficiare di uno “schermo penale” algoritmico e aggirare le responsabilità derivanti dalle proprie condotte colpose (o, eventualmente, dolose).

Va in ogni caso rilevato che le stesse istituzioni europee non paiono orientate nel senso di configurare una responsabilità penale diretta dei sistemi di i.a. La Risoluzione del Parlamento europeo del 6 ottobre 2021 su *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale* è chiara sul punto, quando al punto 13 stabilisce che «la responsabilità giuridica e l'imputabilità devono sempre ricadere su una persona fisica o giuridica, che deve sempre essere identificata per le decisioni assunte con il sostegno dell'IA»⁴⁸.

MAGRO, *Decisione umana e decisione robotica*, cit., p. 9; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 70; A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 17; P.M. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 181; S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame?*, cit., p. 423-4; D. LIMA, *Could AI Agents Be Held Criminally Liable?*, cit., p. 688; S. AIRES DE SOUSA, *“Não fui eu, foi a máquina”: teoria do crime, responsabilidade e inteligência artificial*, cit., p. 78.

⁴⁶ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 17; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 9.

⁴⁷ A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 17; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1205.

⁴⁸ Risoluzione del Parlamento europeo, 6 ottobre 2021, *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*; vd. anche il considerando (J):

3. La responsabilità dell'utilizzatore di un sistema di i.a.

Escluso che i sistemi di i.a. possano essere chiamati a rispondere direttamente per gli eventi lesivi da essi derivanti, ci accingiamo ora a valutare se una responsabilità penale possa essere configurata in capo all'utilizzatore del dispositivo intelligente. Come abbiamo più volte ribadito, la progressiva perdita di controllo dell'operatore sul sistema di i.a., unita all'opacità tipica degli algoritmi di *machine learning*, potrebbe costituire un limite alla rimproverabilità dell'utilizzatore, impossibilitato ad esercitare un governo sul dispositivo. A nostro parere, tuttavia, alcuni profili di responsabilità penale per l'utilizzatore di un sistema di i.a. possono emergere sia in relazione alla futura ed eventuale messa in commercio di dispositivi totalmente autonomi (si pensi ad un incidente mortale cagionato dalla carente manutenzione di una *self driving car*), sia in relazione agli attuali dispositivi parzialmente autonomi, che prevedono un ruolo dell'utente nel governo dell'agire algoritmico (si pensi, per restare all'esempio dell'*automotive*, al conducente di un *autonomous vehicle* di livello 2 che si distrae alla guida e che non intervenga in una situazione di emergenza, cagionando così un incidente mortale). Ci occuperemo principalmente dei profili di responsabilità colposa (§§ 3.1. e 3.2.), riservando soltanto brevi cenni alle ipotesi di divergenza tra voluto e realizzato (§ 3.3) e alle ipotesi in cui viene in rilievo un *dolus generalis* dell'operatore (§ 3.4.).

3.1. Utilizzatore di sistema di i.a. totalmente autonomo: responsabilità colposa.

Consideriamo brevemente il caso dell'utilizzatore di un sistema di i.a. totalmente autonomo, riservando poi un esame più approfondito alla posizione dell'utilizzatore di un sistema di i.a. parzialmente autonomo.

«considerando che è necessario un modello chiaro per attribuire la responsabilità per i potenziali effetti nocivi dei sistemi di IA nel settore del diritto penale; che le norme regolamentari in questo ambito dovrebbero sempre sostenere la responsabilità umana e che il loro primo e principale scopo deve innanzi tutto essere la prevenzione di qualunque effetto negativo».

Innanzitutto, in relazione ai sistemi di i.a. totalmente autonomi, a rigore non potrebbe nemmeno parlarsi di “utilizzatori”, ma piuttosto di “utenti” o “fruttori” del servizio intelligente. All’assoluta perdita di governo sull’agire tecnologico corrisponderà, com’è ovvio, l’impossibilità di muovere all’utente un rimprovero penale per la verifica di eventi lesivi sui quali questi non ha alcun dominio. In capo agli utenti, la legge potrebbe comunque configurare, a nostro parere, una limitata *posizione di controllo*⁴⁹, nei limiti ovviamente del principio dell’*ad impossibilia nemo tenetur*⁵⁰. Così, per restare all’esempio dell’*automotive*, il proprietario di un veicolo di livello 4 o 5 sarà comunque soggetto all’obbligo di controllo previsto all’art. 79 del Codice della Strada, che stabilisce che «I veicoli a motore ed i loro rimorchi durante la circolazione devono essere tenuti in condizioni di massima efficienza, comunque tale da garantire la sicurezza» – o, in ogni caso, dovrà essere introdotta una norma analoga che assicuri un idoneo controllo sulle condizioni del veicolo. L’utilizzatore di un veicolo di livello 4 o 5 dovrà dunque essere tenuto a curarne la manutenzione, verificare che non ci siano segnali di malfunzionamento, curare l’aggiornamento del sistema operativo, ed eventualmente evitare di mettere in moto la vettura quando le circostanze del caso concreto lo sconsiglino (si pensi a condizioni di scarsa visibilità stradale): nel caso in cui, a causa di tali imprudenze, si verifichi un incidente mortale, egli potrà essere chiamato a rispondere di omicidio stradale per omissione ai sensi degli artt. 589-bis e 40, co. 2, c.p.

Ovviamente, una responsabilità potrà emergere anche sul piano commissivo – ma paiono più casi di scuola che reali ipotesi applicative: si pensi ad un incidente mortale che si verifichi perché il proprietario di una *self driving car* appone un adesivo su un sensore del veicolo.

Va comunque rilevato che, qualora venisse approvato l’*AI Act* senza modifiche sostanziali al suo impianto valoriale, l’introduzione di sistemi di i.a. “ad alto rischio” (vd. *supra*, cap. II, § 3.4.3.) totalmente autonomi in assenza di supervisione umana sarebbe da considerarsi vietata. L’art. 14 della bozza di regolamento, rubricato “Sorveglianza

⁴⁹ Sulla distinzione tra posizioni di controllo e di protezione vd. G. FIANDACA, *Il reato commissivo mediante omissione*, Giuffrè, 1979, p. 189; G. GRASSO, *Il reato omissivo improprio. La struttura della fattispecie*, Giuffrè, 1983, p. 320; F. MANTOVANI, *Causalità, obbligo di garanzia e dolo nei reati omissivi*, in *Riv. it. dir. proc. pen.*, fasc. 4, 2004, p. 1007.

⁵⁰ A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit, p. 338, con riferimento agli utenti di *self driving cars*.

umana”, prevede infatti che i dispositivi intelligenti debbano essere sviluppati in modo tale da garantire che il loro funzionamento sia sempre soggetto al controllo di una persona fisica. In particolare, la persona deputata alla vigilanza deve essere messa nelle condizioni di «decidere, in qualsiasi situazione particolare, di non usare il sistema di IA ad alto rischio o altrimenti di ignorare, annullare o ribaltare l'output del sistema di IA ad alto rischio» (art. 14, § 4, lett. d) e di «intervenire sul funzionamento del sistema di IA ad alto rischio o di interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga» (art. 14, § 4, lett. e). Ancora, il considerando 48 stabilisce che «I sistemi di IA ad alto rischio dovrebbero essere progettati e sviluppati in modo da consentire alle persone fisiche di sorvegliarne il funzionamento. Il fornitore del sistema dovrebbe a tal fine individuare misure di sorveglianza umana adeguate prima dell'immissione del sistema sul mercato o della sua messa in servizio. Tali misure dovrebbero in particolare garantire, ove opportuno, che il sistema sia soggetto a vincoli operativi intrinseci che il sistema stesso non può annullare e che risponda all'operatore umano, e che le persone fisiche alle quali è stata affidata la sorveglianza umana dispongano delle competenze, della formazione e dell'autorità necessarie per svolgere tale ruolo».

In realtà, resta in ogni caso aperto uno spiraglio per l'introduzione di sistemi di i.a. completamente autonomi, purché *costantemente sorvegliati*. È quanto è accaduto in Germania, con l'approvazione della legge sulla circolazione di autoveicoli di livello 4 di autonomia (vd. *supra*, cap. I, § 7.1.1.1.2.), che non prevede alcun obbligo di monitoraggio per gli utilizzatori, ma stabilisce che sia sempre necessaria la vigilanza da parte di un “supervisore tecnico” – anche *esterno* rispetto all'abitacolo del veicolo – con l'obbligo di intervenire quando vi sia una richiesta in questo senso da parte del sistema, riportando il veicolo in una situazione di rischio minimo. In questo caso,

3.2. Utilizzatore di sistema di i.a. parzialmente autonomo: responsabilità colposa.

Il principio del c.d. *human-in-the-loop* determina una rinuncia, per ora, ai benefici – in termini di tempo, risorse e, talvolta, efficienza – che potrebbero derivare dall'attribuzione ai sistemi di i.a. di una completa autonomia di azione, senza alcuna sorveglianza da parte degli esseri umani: un traguardo già tecnologicamente possibile in

molti settori. Tale rinuncia si giustifica, d'altra parte, con l'obiettivo di impedire che gli errori algoritmici cagionino offese che potrebbero invece essere evitate attraverso un intervento umano (si pensi, ad esempio, ai banali errori di riconoscimento che emergono dall'*adversarial machine learning* o ai vari episodi già verificatisi di discriminazioni algoritmiche, vd. *supra*, cap. I, § 2.2 e 3). È quanto esplicitato dallo stesso *AI Act*, che all'art. 14, § 2 chiarisce che «La sorveglianza umana mira a prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali».

La necessaria individuazione di un garante per l'attività algoritmica sconta, tuttavia, il rischio che l'utilizzatore si trasformi in un "capro espiatorio" per l'attribuzione della responsabilità penale – in quanto immediato centro di accollo degli eventi dannosi – anche nei casi in cui questi non aveva concretamente la possibilità di evitare l'evento lesivo⁵¹. Nonostante il costante incremento di autonomia dei sistemi di i.a. sembrerebbe preludere ad uno spostamento del *focus* della responsabilità penale, da quella dell'utilizzatore a quella del produttore, le note difficoltà in ordine all'accertamento della responsabilità del produttore – difficoltà che si acuiscono quando il prodotto è "intelligente" (vd. *supra*, cap. II, § 3. ss., e *infra*, § 4. ss.) – rischiano di scoraggiare le indagini in quella direzione, favorendo invece una ricerca di responsabilità nei confronti del c.d. *human-in-the-loop*⁵², che verrebbe così a colmare il presunto *retribution gap*⁵³ e a soddisfare il "bisogno di pena" serpeggiante nella società.

3.2.1. Il fatto tipico.

Procediamo con ordine e incominciamo dalla disamina del fatto tipico che potrebbe essere imputato all'utilizzatore di un sistema di i.a. parzialmente autonomo.

⁵¹ Sul possibile sviluppo di un meccanismo di "capro espiatorio" nei confronti dell'utilizzatore di sistemi di i.a. parzialmente autonomi vd. M.C. ELISH, *Moral crumple zones*, cit., *passim*; A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 337; B. PANATTONI, *Intelligenza artificiale*, cit., p. 340; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale: prove d'impatto a livello europeo*, cit., p. 16; C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 57.

⁵² Per considerazioni simili, basate sull'analisi della giurisprudenza statunitense in materia di incidenti aerei, vd. M.C. ELISH, *Moral crumple zones*, cit., p. 12 ss.; vd. anche W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit., p. 390 ss.

⁵³ J. DANAHER, *Robots, Law and the Retribution Gap*, cit., p. 299 ss.

Innanzitutto, una responsabilità dell'utilizzatore potrà rilevare sia sul piano commissivo sia sul piano omissivo, a seconda che il sistema di i.a. sia un *intelligent decision support system* (c.d. IDSS) – e necessiti, dunque, della mediazione di una persona fisica per influire sulla realtà – o, invece, il dispositivo possa agire direttamente sul mondo, fisico o virtuale (vd. la classificazione da noi proposta *supra*, cap. I, § 7 ss.)⁵⁴.

Quanto alla prima ipotesi, ci riferiamo ai sistemi che si limitano a fornire *raccomandazioni* e necessitano di un filtro umano per incidere concretamente sui beni giuridici. Nel primo capitolo (vd. § 7.2.1.) ci siamo concentrati soprattutto sui sistemi di supporto alla diagnosi e al trattamento in ambito medico, che sono attualmente la forma più diffusa di IDSS. In questi casi, com'è evidente, la condotta penalmente rilevante sarà commissiva: se il trattamento sanitario prescritto dal medico utilizzatore di un *Clinical Decision Support System* determina l'insorgere di una patologia, al medico potrà essere imputata la commissione del reato di *lesioni colpose* o di *omicidio colposo* (sempre, ovviamente, che sussistano tutti gli elementi della fattispecie).

In relazione all'utilizzo di sistemi di i.a. che possono recare offese immediate ai beni giuridici, con l'incremento del livello di autonomia si assisterà invece, probabilmente, ad un sostanziale cambio di paradigma nell'imputazione penale, da commissiva ad omissiva⁵⁵. L'utilizzatore, infatti, più che una persona che adopera direttamente il dispositivo intelligente, è il soggetto che beneficia dell'attività algoritmica e che mantiene l'obbligo di supervisionarla e di intervenire qualora emergano eventuali malfunzionamenti.

Prendiamo ancora una volta l'esempio dei veicoli a guida parzialmente autonoma. Come abbiamo illustrato nel primo capitolo (vd. *supra*, cap. I, § 7.1.1.1.1.), i veicoli di livello 2 o 3 possono svolgere interamente, in determinate circostanze (es. tempo soleggiato, specifica area georeferenziata) e per un certo periodo di tempo, la funzione di

⁵⁴ In ogni caso, valgono anche con riferimento ai sistemi parzialmente autonomi le considerazioni già svolte nel precedente paragrafo: l'utilizzatore potrà dunque cagionare un evento lesivo anche manomettendo colposamente il dispositivo intelligente (si ricordi l'esempio dello sticker) o non effettuando la doverosa attività manutentiva.

⁵⁵ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1753; B. PANATTONI, *Intelligenza artificiale*, cit., p. 340; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale*, cit., p. 167; con specifico riferimento agli *autonomous vehicles* vd. L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 826; A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 334 ss.; ID., *Reati colposi e tecnologie dell'intelligenza artificiale*, cit., p. 28-29; M. GIUCA, *Disciplinare l'intelligenza artificiale*, cit., p. 24.

guida. Il conducente è tuttavia obbligato a supervisionare l'attività di guida e ad intervenire ogniqualvolta sia necessario: la differenza tra i due modelli consiste nel fatto che i sistemi di livello 3 sono in grado di riconoscere – quando si presentano – le situazioni nelle quali non sono in grado di operare, avvertendo con un certo preavviso il conducente affinché acquisisca il controllo del veicolo; un simile meccanismo di allarme non è invece previsto nei veicoli di livello inferiore, con la conseguenza che il conducente deve monitorare costantemente il sistema⁵⁶. In entrambi i casi, l'utilizzatore è chiamato a supervisionare la funzione di guida e ad intervenire in caso di necessità: in un caso (veicoli di livello 2), la posizione di garanzia potrà dirsi “attivata” nel momento della messa in funzione del *software* di guida autonoma; nell'altro (livello 3), la posizione di garanzia sarà “attivata” soltanto eventualmente, nel caso in cui l'allarme acustico del veicolo segnali un problema.

Perché sia configurabile una posizione di garanzia in capo all'utilizzatore sarà necessaria, ovviamente, l'introduzione di specifiche norme. È questa la strada intrapresa dal legislatore francese, che, al neo-introdotta art. L123-1, co. 2, *code de la route*, stabilisce che «*Le conducteur doit se tenir constamment en état et en position de répondre à une demande de reprise en main du système de conduite automatisé*» (vd. *supra*, cap. I, 8.1.1.1.2.)⁵⁷. Ancora, la Sezione 1b del codice della strada tedesco stabilisce che il conducente di un veicolo di livello 3 deve sempre rimanere consapevole delle condizioni di guida dell'*autonomous vehicle*, in modo tale da poter riprendere immediatamente il controllo del veicolo quando richiesto dal sistema o, comunque, quando motivi di sicurezza lo impongano⁵⁸. Lo stesso Decreto c.d. Smart Road⁵⁹ – che disciplina la circolazione sperimentale di veicoli a guida autonoma nel nostro Paese – parrebbe già istituire una posizione di garanzia per il “supervisore” del veicolo, laddove stabilisce che «Il supervisore deve essere in grado di commutare tempestivamente tra operatività del veicolo in modo automatico e operatività dello stesso in modo manuale e viceversa. Il

⁵⁶ *Contra* C. PIERGALLINI, *Intelligenza artificiale*, cit., 1753 e A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 335, che ritengono che il paradigma commissivo resti valido in relazione ai veicoli di livello 2.

⁵⁷ In argomento vd. più diffusamente M. GIUCA, *Disciplinare l'intelligenza artificiale*, cit., p. 24 ss.

⁵⁸ M. EBERS, *Civil liability for autonomous vehicles in Germany*, cit., p. 8.

⁵⁹ Decreto del Ministero delle Infrastrutture e dei Trasporti del 28 febbraio 2018 recante *Modalità attuative e strumenti operativi della sperimentazione su strada delle soluzioni di Smart Road e di guida connessa e automatica* (vd. *supra*, cap. I, *supra*, § 7.1.1.1.2.).

supervisore ha la responsabilità del veicolo in entrambe le modalità operative» (art. 10, co. 2)⁶⁰.

Affronteremo successivamente le complesse questioni attinenti all'accertamento del nesso di causalità che collega messa in commercio del prodotto intelligente e verificarsi dell'evento lesivo (c.d. *causalità reale*, vd. *infra*, § 4.3.). Qui ci occuperemo invece brevemente della *seconda fase* del giudizio sulla causalità omissiva, che concerne il nesso tra omissione dell'azione doverosa da parte dell'utilizzatore ed evento (c.d. *causalità ipotetica*). Come noto, infatti, ai fini dell'accertamento di un reato omissivo improprio, alla verifica circa il decorso eziologico reale all'origine dell'evento lesivo deve necessariamente seguire il giudizio "predittivo" relativo all'efficacia impeditiva dell'azione doverosa⁶¹. Tale accertamento si basa su un giudizio controfattuale volto a verificare se, aggiungendo mentalmente l'azione doverosa omessa, individuata alla luce della regola cautelare che imponeva quell'azione, si sarebbe reciso il decorso causale sfociato nell'evento (*rectius*, con quante probabilità tale evento non si sarebbe verificato, secondo un giudizio prognostico).

Così, dopo aver accertato il nesso reale (es. l'algoritmo di *image recognition* non ha riconosciuto un segnale di stop e ha impegnato l'incrocio stradale in cui stava passando un veicolo, scontrandosi con questo e causando la morte del conducente), il giudice dovrà accertare se l'azione richiesta dalla norma giuridica avrebbe potuto impedire l'evento. Il nesso tra omissione ed evento andrà dunque escluso se, ad esempio, la richiesta di prendere il controllo del veicolo di livello 3 sia giunta troppo tardi al conducente, così che la pressione sul pedale del freno non avrebbe in ogni caso impedito lo scontro.

Sempre restando sul tema delle macchine a guida autonoma, una responsabilità commissiva potrebbe comunque residuare nei casi in cui l'utilizzatore, dopo aver ripreso correttamente il controllo della vettura, commetta un errore nell'effettuare la manovra necessaria: in questi casi, infatti, si ristabilisce il tradizionale rapporto di strumentalità tra

⁶⁰ In questo senso, L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 826-827.

⁶¹ Sulla struttura della causalità nei reati omissivi impropri vd. per tutti G. GRASSO, *Il reato omissivo improprio*, cit., p. 385 ss.; F. STELLA, *La nozione penalmente rilevante di causa: la condizione necessaria*, in *Riv. it. dir. proc. pen.*, 1988, p. 1249 ss.; ID., *Causalità omissiva, probabilità, giudizi controfattuali. L'attività medico-chirurgica*, in *Cass. pen.*, 2005, 422 ss.; C.E. PALIERO, *La causalità dell'omissione: formule concettuali e paradigmi prasseologici*, in *Riv. it. med. leg.*, 1992, p. 821 ss.; G. MARINUCCI, *Causalità reale e causalità ipotetica nell'omissione impropria*, in *Riv. it. dir. proc. pen.*, 2009, 523 ss.

conducente e veicolo⁶². Lo stesso può dirsi, ovviamente, anche con riferimento agli *autopilot* utilizzati sui velivoli: una volta ripreso il controllo dell'aereo, il pilota che cagioni un incidente sarà imputabile a titolo commissivo – benché, ovviamente, nella sua condotta possano persistere dei momenti *omissivi*, tipici di qualsiasi contestazione colposa⁶³.

L'evento lesivo potrà poi ovviamente essere il risultato di una convergenza causale di più condotte, poste in essere da diverse persone. Si pensi, ad esempio, ad un *healthcare robot* di livello L2 (vd. *supra*, cap. I, § 7.1.1.4) gravemente difettoso, che, durante l'esecuzione di un'operazione chirurgica, recida un'arteria, causando la morte del paziente, mentre il supervisore medico – nonostante l'obbligo di sorveglianza – si trovi al di fuori della sala operatoria. In questo caso, l'evento lesivo parrebbe il risultato di due concause: il funzionamento difettoso del prodotto e, sul lato omissivo, il mancato intervento doveroso da parte del medico – sempre che venga dimostrata, in chiave ipotetica, l'efficacia impeditiva dell'azione doverosa. In un caso di questo tipo, potrebbe configurarsi un *concorso di condizioni* (colpose) *indipendenti*, ai sensi dell'art. 41, co. 3, c.p.⁶⁴. Si tratta di un'ipotesi che potrebbe rivelarsi molto frequente nella prassi, atteso che – man mano che i sistemi di i.a. diventeranno più sofisticati, acquisendo una maggiore autonomia – la necessità di un intervento impeditivo da parte del supervisore potrebbe denotare un qualche malfunzionamento dell'algoritmo intelligente. Non si può, d'altra parte, non osservare come la dimostrazione di una responsabilità in capo al produttore incontra problemi talmente rilevanti – sia per quanto riguarda l'accertamento dell'elemento oggettivo, sia per quanto concerne l'elemento soggettivo (vd. *infra*, § 4 ss.)

⁶² A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 336.

⁶³ F. VIGANÒ, *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, in *Dir. pen. cont. – Riv. Trim.*, 2013, n. 3, p. 391; F. GIUNTA, *Illiceità e colpevolezza nella responsabilità colposa*, Cedam, 1993, p. 90; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie" nella prospettiva delle fattispecie colpose causalmente orientate*, Cedam, 2003, p. 44.

⁶⁴ In questo senso I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 103. Con riferimento ai reati colposi causalmente orientati, di cui ci stiamo occupando, non potrà ovviamente configurarsi una cooperazione colposa ai sensi dell'art. 113 c.p. Sulla differenza tra i due istituti vd. L. RISICATO, *Il concorso colposo tra vecchie e nuove incertezze*, in *Riv. dir. proc. pen.*, 1998, p. 137 ss.; L. CORNACCHIA, *Concorso di colpe e principio di responsabilità per fatto proprio*, Giappichelli, 2004, p. 142 ss. Recentemente, sulla funzione incriminatrice svolta dall'art. 113 c.p. vd. F. CONSULICH, *Errare commune est. Il concorrente colposo, il nuovo protagonista del diritto penale d'impresa (e non solo)*, in *Leg. pen.*, 28 marzo 2022 e M. DI FLORIO, *La cooperazione nel delitto colposo: una fattispecie con una (problematica) funzione incriminatrice*, in *Arch pen.*, 2021, n. 1.

– che il rischio è che l’ipotesi delle concause venga in concreto abbandonata, con conseguente addebito al solo utilizzatore dell’evento lesivo.

3.2.2. La colpa.

Come abbiamo già accennato, il principio dell’*human-on-the-loop* mira ad evitare, in chiave precauzionale, che l’attribuzione di una totale autonomia ai sistemi di i.a. possa determinare rischi eccessivi per la tutela di beni giuridici di fondamentale importanza. L’altra faccia della medaglia è costituita, tuttavia, dal rischio che all’utilizzatore vengano accollati tutti gli eventi lesivi derivanti dalla messa in funzione del sistema di i.a., indipendentemente dalla specifica individuazione della regola cautelare violata e dalla concreta possibilità di governo sull’attività algoritmica. L’utilizzatore si trasformerebbe così in un vero e proprio “capro espiatorio”, a cui sarebbe attribuita una sorta di *responsabilità (oggettiva) da posizione*.

Per quanto riguarda la valutazione della colpa specifica, le norme cautelari positive potranno essere rintracciate nell’ordinaria regolazione del settore, ove applicabile (si pensi, ad esempio, alle regole cautelari in ambito medico-sanitario, che dovranno ritenersi tendenzialmente valide anche in relazione all’utilizzo di CDSS), ma potranno altresì essere oggetto di specifica normazione.

In ogni caso, sia nell’elaborazione di norme cautelari specifiche, sia nella valutazione circa la sussistenza della colpa generica, particolare attenzione dovrà essere posta in relazione alla valorizzazione della *misura soggettiva* della colpa, relativa al concreto “potere” dell’utilizzatore nella situazione in cui ha operato⁶⁵. L’impatto dell’automazione sulle capacità di azione dell’utilizzatore è ormai oggetto di ampi studi

⁶⁵ Sulla misura soggettiva della colpa vd. il saggio fondamentale di V. DE FRANCESCO, *Sulla misura soggettiva della colpa*, in *Studi Urbinati*, 1977, p. 273 ss.; più recentemente, D. CASTRONOVO, *La colpa penale*, Giuffrè, 2009; ID., *La colpa “penale”. Misura soggettiva e colpa grave*, in *Riv. it. dir. proc. pen.*, 2013, n. 4, p. 1723 ss.; A. CANEPA, *L’imputazione soggettiva della colpa. Il reato colposo come punto cruciale nel rapporto tra illecito e colpevolezza*, Giappichelli, 2011; M. GROTTI, *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica*, Giappichelli, 2012; A.R. DI LANDRO, *Dalle linee guida e dai protocolli all’individualizzazione della colpa penale nel settore sanitario. Misura oggettiva e soggettiva della malpractice*, Giappichelli, 2012.

nell'ambito della psicologia comportamentale⁶⁶. Il costante aumento delle capacità di svolgere attività autonome, in capo ai sistemi di i.a., può, da una parte, determinare una contrazione del rischio, in relazione a tutte quelle attività che normalmente richiederebbero una costante attenzione da parte dell'essere umano; d'altra parte, l'automazione può essere causa di *nuovi rischi*, determinati proprio dalla necessaria condivisione del *decision-making* tra persona fisica e algoritmo⁶⁷.

Quanto all'accertamento della colpa in capo all'utilizzatore, si pone dunque il problema di stabilire i confini della diligenza esigibile, tenendo in considerazione il fatto che la perdita di controllo sull'attività tecnologica ha necessariamente un impatto sulla ricostruzione dell'agente modello.

Senza alcuna pretesa di individuare criteri univoci per l'identificazione della colpa – compito che, in concreto, spetterà all'autorità giudiziaria attraverso una valutazione casistica⁶⁸ – ci limitiamo in questa sede a rilevare che lo *standard* di diligenza dovrà necessariamente tenere conto degli effetti dell'autonomia algoritmica sulle capacità di azione dell'operatore di un sistema di i.a.⁶⁹. Così, ad esempio, l'agente modello di conducente di veicolo parzialmente autonomo non potrà essere ricostruito utilizzando i medesimi criteri – come noto, oltretutto, particolarmente rigorosi⁷⁰ – che sono impiegati per l'individuazione dello *standard of care* nella guida tradizionale: non si potrà infatti pretendere, dal sorvegliante, la medesima capacità reattiva della persona che, nei veicoli ordinari, si occupa costantemente della funzione di guida. Nella costruzione dell'*homo*

⁶⁶ D.B. KABER – M.R. ENDSLEY, *The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task*, in 5 *Theoretical Issues in Ergonomics Science*, 2004, p. 113 ss.; R. PARASURAMAN E AL., *A Model for Types and Levels of Human Interaction with Automation*, in 30 *IEEE Transactions on Systems Man and Cybernetics*, p. 286 ss.

⁶⁷ Su questi aspetti vd. diffusamente W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit.; M.C. ELISH, *Moral crumple zones*, cit.; M.C. ELISH – T. HWANG, *Praise the machine! Punish the human!*, cit.

⁶⁸ Vd. sul punto E. HILGENDORF, *Automated Driving and the Law*, cit., p. 181 e 188.

⁶⁹ In questo senso, nella dottrina statunitense, P. HUBBARD, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation*, cit., p. 1861 ss.; A.D. SELBST, *Negligence and AI's Human Users*, cit., p. 1327 ss.

⁷⁰ È nota la riluttanza della giurisprudenza a dare spazio al principio di affidamento nell'ambito della circolazione stradale. L'agente viene spesso ritenuto responsabile del fatto cagionato anche quando non era concretamente prevedibile l'altrui comportamento. Vd. *ex multis* Cass., sez. IV, 6 maggio 2021, n. 24414, Busdraghi; Cass., sez. IV, 27 aprile 2017, n. 25552, Luciano; Cass., sez. IV, 19 febbraio 2019, n. 20823, Farimbella; Cass. Pen., Sez. 4, 22/6/2017, n. 45795, Passero, e Cass. Pen., Sez. 4, 25/6/2014, n. 46818, Nuzzolese.

eiusdem professionis et condicionis dovrà dunque prendersi a modello un agente dotato di medie conoscenze e abilità, con l'aggiunta delle eventuali speciali e maggiori conoscenze maturate dal soggetto individuale⁷¹ – con l'ovvia precisazione che non esisterà un *modello unico* di “utilizzatore di sistema di i.a.”, ma *tanti modelli* quanti sono i distinti ambiti di attività nei quali il dispositivo intelligente potrà trovare applicazione, in relazione altresì al livello di autonomia dei sistemi stessi. Al giudice, soprattutto nelle prime fasi di commercializzazione dei sistemi di i.a., si chiederà uno sforzo di astrazione notevole, nell'identificazione dell'agente modello, stante la scarsità, ad oggi, di una conoscenza diffusa circa il preciso impatto dell'intelligenza artificiale sulle capacità psicofisiche dell'agente. Dall'utilizzatore, in ogni caso, non potrà pretendersi l'esercizio di una sorveglianza *generica*, in grado di minimizzare o neutralizzare *qualsiasi rischio* derivante dall'impiego del prodotto intelligente.

Nel prosieguo ci accingiamo ad analizzare tre problemi che, a nostro modesto parere, possono ostacolare la capacità di governo dell'utilizzatore sul sistema di i.a., e che, dunque, dovranno essere presi in considerazione nella valutazione della *diligenza* dell'utilizzatore – *sub specie* di *esigibilità* della conformazione alle norme cautelari. Questi i principali problemi:

- a) l'opacità del sistema di i.a.;
- b) la c.d. “distorsione dell'automazione” (*technological complacency*).
- c) la ridotta capacità di reazione durante le situazioni di emergenza.

Le prime due questioni possono riguardare sia i c.d. *intelligent decision support systems* sia i dispositivi che possa agire direttamente sulla realtà (fisica o virtuale); il problema tematizzato *sub c)* riguarda invece esclusivamente questi ultimi.

a) Opacità del sistema di i.a.

Come abbiamo cercato di far emergere nel primo capitolo (vd. *supra*, cap. I, § 5.3.), il funzionamento dei sistemi di i.a. è caratterizzato da *opacità*, dal momento che – pur essendo individuabili i dati di *input* e *output* dell'algoritmo – non sempre è invece possibile ricostruire la catena causale che ha determinato il passaggio dagli *input* agli

⁷¹ G. MARINUCCI, *La colpa per inosservanza*, Giuffrè, 1965, p. 195.

output (c.d. *black box*). Tale caratteristica ha evidenti ripercussioni sull'accertamento della colpa in capo all'utilizzatore del dispositivo intelligente⁷².

La ricostruzione della dinamica degli incidenti aerei che hanno coinvolto i Boeing 737 MAX – e che abbiamo riportato nel primo capitolo (vd. *supra*, cap. I, § 7.1.1.2.) – è in tal senso emblematica. In quei casi, i sensori avevano percepito erroneamente che l'aereo si trovava in situazione di stallo e avevano dunque trasmesso il segnale di attivazione al sistema di i.a. anti-stallo, nonostante non ve ne fosse necessità. Il tentativo dei piloti di riportare l'aereo nella corretta posizione si scontrava così con il tentativo opposto da parte del dispositivo di intelligenza artificiale, che aveva l'obiettivo di risolvere una situazione di stallo in realtà inesistente. La mancata comprensione, da parte dei piloti, di cosa stesse accadendo era determinata, tra gli altri motivi, dall'*opacità* del sistema di i.a., che non consentiva ai piloti di identificare l'errore in cui era incorso il dispositivo e di decidere, eventualmente, di disattivarlo⁷³.

Lo stesso problema si verifica, *mutatis mutandis*, in relazione alla conduzione di auto a guida parzialmente autonoma. Così, se il conducente di un veicolo autonomo di livello 2 non ha modo di accorgersi che il sistema di *image recognition* ha scambiato un semaforo rosso per un lampeggiante dell'ambulanza, sarà difficile pretendere che egli riprenda il controllo del veicolo e freni in luogo del sistema.

Infine, l'*opacità* caratterizza anche i sistemi di i.a. che forniscono raccomandazioni agli utenti – anche se in questi casi *non* si pone, viceversa, un problema di *tempestività* nella reazione da parte dell'utilizzatore. Se il medico che si serve di un sistema di supporto alla diagnostica e al trattamento ha accesso alla sola raccomandazione finale fornita dal dispositivo – non potendo verificare *su quali basi* il sistema sia pervenuto a tale esito – potrebbe risultare complesso, per il medico, riconoscere l'errore algoritmico e discostarsi dal suggerimento (sempre che, ovviamente, questo sia quantomeno plausibile). Come tutte le attività scientifiche, infatti, quella diagnostica e quella finalizzata all'individuazione del trattamento sanitario più appropriato si fondano

⁷² In questo senso vd. D. SELBST, *Negligence and AI's Human Users*, cit. p. 1331 ss.; A. CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, cit., p. 30.

⁷³ Sugli specifici problemi della *black box* e della *technological complacency* nella gestione dell'emergenza in ambito di aviazione vd. M.C. ELISH, *Moral crumple zones*, cit.; M.C. ELISH – T. HWANG, *Praise the machine! Punish the human!*, cit.; W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit.

su ipotesi e valutazioni la cui fondatezza deve essere argomentata, affinché possa essere, eventualmente, contestata e messa in discussione. Questo problema è strettamente correlato con il tema della *technological complacency*, che tratteremo poco oltre.

Recependo le raccomandazioni derivanti da *report* e documenti di *policy*, la bozza di regolamento dell'UE sull'intelligenza artificiale cerca di affrontare il problema del *black box*, prevedendo che i sistemi di IA ad alto rischio siano «*progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e utilizzarlo adeguatamente*» (art. 13, § 1). L'art. 14, § 4, *AI Act* prescrive inoltre che l'interfaccia del dispositivo debba assicurare un'agevole interpretazione dell'*output* del sistema, così che il garante sia posto nelle condizioni ottimali per verificare l'eventuale presenza di anomalie, disfunzioni e prestazioni inattese, e per intervenire di conseguenza (art. 14, § 4, lett. a) e c)⁷⁴.

Se l'intento del legislatore europeo è certamente condivisibile, il rischio è che tali disposizioni si traducano in mere dichiarazioni di principio⁷⁵, specie in relazione a quei sistemi che utilizzano le reti neurali artificiali, la cui opacità pare caratteristica intrinseca e *trade off* dell'accuratezza nelle previsioni (vd. *supra*, cap. I, § 5.3.). Anche qualora i sistemi di i.a. venissero dotati di interfacce *user-friendly*, non è detto, infatti, che l'utente sia messo nelle condizioni di *ribaltare* la decisione dell'intelligenza artificiale: o perché il tempo di reazione è minimo (si pensi ancora al conducente che, in prossimità di un semaforo, si renda conto che il sistema di *image recognition* lo ha scambiato per un lampeggiante) oppure perché la spiegazione fornita dal *machine learning* è descrittiva e probabilistica, non causale (si pensi ad un CDSS che si limiti a fornire l'elenco, immenso, di precedenti dai quali ha dedotto la concreta ipotesi diagnostica e trattamentale).

⁷⁴ Vd. anche considerando 48 dell'*AI Act*: «I sistemi di IA ad alto rischio dovrebbero essere progettati e sviluppati in modo da consentire alle persone fisiche di sorvegliarne il funzionamento. Il fornitore del sistema dovrebbe a tal fine individuare misure di sorveglianza umana adeguate prima dell'immissione del sistema sul mercato o della sua messa in servizio. Tali misure dovrebbero in particolare garantire, ove opportuno, che il sistema sia soggetto a vincoli operativi intrinseci che il sistema stesso non può annullare e che risponda all'operatore umano, e che le persone fisiche alle quali è stata affidata la sorveglianza umana dispongano delle competenze, della formazione e dell'autorità necessarie per svolgere tale ruolo».

⁷⁵ A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 2, cit., p. 290; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale*, cit., p. 16; U. RUFFOLO – A. AMIDEI, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design*, cit., p. 492; E.M. COTUGNO, *Intelligenza artificiale e tutela del consumatore. I nuovi obblighi informativi*, cit., p. 319.

b) *Distorsione dell'automazione (Technological Complacency)*.

Gli effetti collaterali derivanti dalla c.d. *black box* dei sistemi di i.a. si accentuano se consideriamo che uno dei principali problemi attinenti al fenomeno dell'automazione consiste nella tendenza delle persone ad affidarsi eccessivamente alla tecnologia e a non metterne in discussione i risultati (c.d. *technological complacency*)⁷⁶. Il fenomeno della c.d. *distorsione dell'automazione* si verifica soprattutto in relazione a quei sistemi di i.a. che non incidono direttamente sulla realtà, ma che si limitano a fornire suggerimenti al decisore umano.

Si pensi, ad esempio, alle raccomandazioni fornite da un sistema di supporto alla diagnosi e al trattamento sanitario, che sono basate sull'analisi di milioni e milioni di dati su patologie e cure precedenti e, in certi settori, raggiungono già un livello di accuratezza superiore a quello umano (vd. *supra*, cap. I, § 7.2.1.). Il medico che utilizzi uno di questi *device* – eventualmente dopo averne testato in prima persona e in più occasioni l'efficacia – sarà prevedibilmente portato a fidarsi di tali raccomandazioni, giungendo, in ipotesi estreme, a mettere in discussione la sua stessa capacità di giudizio clinico. È il c.d. *effetto moutonnier*⁷⁷ (traducibile come “effetto pecorone” o “effetto gregge”), che può

⁷⁶ La dottrina definisce tale fenomeno come “*attitudinal problem*”, vd. B. PANATTONI, *Intelligenza artificiale*, cit., p. 337, che a sua volta cita J. ZERILLI E AL., *Algorithmic decision-Making and the Control Problem*, in 29 *Minds and Machines*, 2019, p. 559 ss. Sul tema vd. anche S. MASSI, *Affidamento sull'intelligenza artificiale e “disimpegno morale” nella definizione dei presupposti della responsabilità penale*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 665 ss.; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale*, cit., p. 23; in ambito filosofico vd. C. BAGNOLI, *Teoria della responsabilità*, Il Mulino, 2019, p. 77 ss.

⁷⁷ C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 2021, 3, p. 429; A. SANTOSUOSSO – G. SARTOR, *La giustizia predittiva: una visione realistica*, in *Giur. it.*, 2022, n. 7, p. 1779 ss.; che riprendono l'espressione di A. GARAPON – J. LASSÈGUE, *Justice digitale. Révolution graphique et rupture anthropologique*, PUF, 2018, p. 279. In argomento vd. anche O. DI GIOVINE, *Il judge-Bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Cass. pen.*, 2020, p. 960; D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi*, cit., p. 117, con specifico riferimento agli algoritmi di supporto alla decisione giudiziaria; C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, cit., § 4.1., che parla di *anchoring bias* o *effect*.

trasformare la supervisione dell'essere umano in una mera formalità, che nasconde una sostanziale adesione acritica alle decisioni algoritmiche.

A tal proposito, l'art. 14, § 4, lett. b) dell'*AI Act* prevede che il produttore del sistema intelligente debba fare in modo che l'utente sia sempre consapevole della «possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("distorsione dell'automazione"), in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche».

Come notazione di *policy*, ci limitiamo qui a rilevare che, sul lungo periodo, il tendenziale affidamento di professionisti ed utenti sul *decision making* algoritmico potrà comportare un *calo delle competenze* in diversi settori: è quanto, tra l'altro, è stato notato con riferimento alle prestazioni dei piloti d'aereo, sempre più abituati a gestire i voli attraverso l'ausilio di sistemi di i.a.⁷⁸

c) Ridotta capacità di reazione durante le situazioni di emergenza.

Infine, strettamente legato al problema della *technological complacency* – ma avente rilievo, a nostro avviso, autonomo – è quello del prevedibile incremento degli episodi di distrazione, da parte dei soggetti tenuti alla sorveglianza dei sistemi di i.a. parzialmente autonomi. Pare infatti evidente – ed è stato altresì dimostrato da vari studi scientifici – che il livello di attenzione di un soggetto che si limiti a *monitorare l'attività altrui* è più basso rispetto a quello di chi *agisce* in prima persona⁷⁹. Si tratta di un problema, come anticipato, che concerne esclusivamente l'utilizzo di sistemi di i.a. che possono ledere direttamente beni giuridici, nel mondo reale o in quello virtuale: soltanto in relazione ad essi, infatti, l'utente potrà essere chiamato a reagire tempestivamente ed

⁷⁸ M.C. ELISH, *Moral crumple zones*, cit., p. 12-13; W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit., *passim*.

⁷⁹ Il problema è stato definito anche "*attentional problem*", vd. B. PANATTONI, *Intelligenza artificiale*, cit., p. 337, che a sua volta cita J. ZERILLI E AL., *Algorithmic decision-making and the Control Problem*, cit., p. 559 ss.; vd. anche S. BECK, *Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood*, in *86 Robotics and Autonomous Systems*, 2016, p. 141; E. HILGENDORF, *Automated Driving and the Law*, cit., p. 187; A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 337; P.S. GÜNSBERG, *Automated vehicles – is a dilution of human responsibility the answer?*, in *New Journal of European Criminal Law*, 2022, Vol. 13(4), p. 443.

in situazioni di emergenza (si pensi all'utilizzatore di un veicolo di livello 3, che sia richiamato alla guida mentre sta leggendo un libro o guardando un film). Lo stesso non potrà dirsi in relazione all'utilizzatore di un *intelligent decision support system*, che avrà tempo per valutare e ponderare la raccomandazione fornita dal dispositivo intelligente (pur con i limiti sottolineati precedentemente).

Nell'accertamento della colpa del "sorvegliante" per non aver impedito il verificarsi di un evento lesivo, sarà allora necessario prendere in considerazione il preavviso con cui viene fornito l'allarme da parte del dispositivo intelligente: se questo è inferiore rispetto al tempo di reazione medio, è ovvio che all'utilizzatore non potrà essere imputata la commissione (attraverso omissione, secondo lo schema dell'art. 40, cpv., c.p.) dell'evento lesivo.

Come abbiamo visto nel precedente capitolo (vd. *supra*, cap. II, § 3.3.2), la mancata predisposizione di adeguati meccanismi di allerta potrà risolversi, di converso, in un difetto di progettazione del prodotto, che, qualora ne ricorrano i presupposti, potrà essere imputato al produttore, sotto il profilo della responsabilità civile per danno da prodotto difettoso e altresì sotto il profilo della responsabilità penale (vd. *infra*, § 4 ss.).

Sul lungo periodo, qualora i sistemi di i.a. parzialmente autonomi diventino largamente diffusi, potrebbe inoltre porsi – anche in relazione alla reattività degli operatori – un problema di disabitudine e di erosione delle competenze: così, per restare sull'esempio degli *autonomous vehicles*, si arriverebbe al paradosso di esigere interventi emergenziali da persone *non abituate a guidare*⁸⁰.

3.2.2.1. Il principio di affidamento.

Il rapporto tra sistemi artificiali e operatori umani è ormai ad un tale livello di compenetrazione che alcuni studiosi hanno parlato, a tal proposito, di "soggetti ibridi", generati dallo stretto legame tra uomo ed agente *software*⁸¹. Una tale ibridazione non

⁸⁰ Per considerazioni simili vd. F. DOUMA – S.A. PALODICHUK, *Criminal Liability Issues Created by Autonomous Vehicles*, in 52 *Santa Clara Law Review*, p. 1164.

⁸¹ Questa la tesi di Gunther Teubner. Vd. G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, a cura di Pasquale Femia, Edizioni Scientifiche Italiane, 2019; A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, Bloomsbury, 2022.

trova tuttavia riconoscimento, ad oggi, a quanto ci consta, nella disciplina positiva, in alcuna branca del diritto – tantomeno potrà trovare accoglimento nel diritto penale, un *corpus* normativo, come abbiamo già rilevato, pensato per influenzare il comportamento delle sole *persone fisiche*, anche quando sottopone a sanzione gli enti giuridici collettivi.

Per quanto riguarda i sistemi di i.a. parzialmente autonomi, la condivisione di un ruolo decisionale tra sistema di i.a. e utilizzatore rischia di complicare la delimitazione delle aree di responsabilità, dal momento che si pone la necessità di separare con precisione le mansioni e le aree di controllo di entrambe le forme di intelligenza, per evitare che il principio dell’*human in the loop* diventi un facile espediente per accollare all’essere umano anche gli errori derivanti dal dispositivo algoritmico – e, dunque, le “negligenze” imputabili al produttore.

Alcuni autori hanno notato come il rapporto di “collaborazione” tra essere umano e sistema di i.a. evochi scenari di colpa relazionale già noti, relativi all’attività svolta in *équipe*, in cui i processi di *decision making* sono frazionati e distribuiti in capo ad una pluralità di persone – con l’ovvia differenza che, nei contesti tradizionali, il processo decisionale ed esecutivo è parcellizzato esclusivamente tra *intelligenze umane*⁸². Come nell’attività di *équipe*, dunque, il problema che si profila è quello di stabilire se e in quale misura l’utilizzatore possa fare affidamento sul corretto funzionamento del sistema di intelligenza artificiale, e se l’impiego di quest’ultimo possa avere una qualche impatto sulla configurabilità di una responsabilità penale a suo carico.

Com’è noto, sulla scorta del principio di affidamento, si ritiene che l’agente – nello svolgimento di attività pericolose lecite svolte in collaborazione con altri soggetti – possa legittimamente fare affidamento sul fatto che le altre persone si comportino in maniera diligente, prudente e perita, nel rispetto, dunque, delle regole cautelari, scritte o non scritte, proprie del settore di riferimento⁸³.

Nell’ambito di nostro interesse, in particolare, ci si chiede se, a fronte di una lesione cagionata da un dispositivo intelligente rivelatosi “difettoso” – con tutte le

⁸² F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 27.

⁸³ F. MANTOVANI, *Il principio di affidamento nel diritto penale*, in *Riv. it. dir. proc. pen.*, 2009, p. 536; M. MANTOVANI, (voce) *Affidamento (principio di)*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, 2022, p. 2; G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 433; D. PULITANÒ, *Diritto penale*, cit., p. 310; F. MANTOVANI, *Diritto penale. Parte generale*, cit., p. 380; G. FIANDACA – E. MUSCO, *Diritto penale. Parte generale*, cit., p. 589 ss.

difficoltà che tale dimostrazione comporta (vd. *supra*, cap. II, § 3.3.) –, possa configurarsi un limite all'imputazione colposa dell'evento lesivo in capo all'utilizzatore, per avere questi confidato nella *sicurezza* di un prodotto immesso sul mercato conformemente agli *standard* fissati dall'*AI Act* (così come eventualmente concretizzati dalle norme armonizzate)⁸⁴.

La risposta, a nostro parere, può variare, a seconda che in capo all'agente sia configurabile una posizione di controllo, ai sensi dell'art. 40, co. 2, c.p., il cui rispetto comporti *la vigilanza dell'operato altrui*⁸⁵ (in questo caso, dell'operato del sistema di i.a.). In particolare, può essere utile tornare all'esempio delle macchine a guida autonoma, distinguendo tra veicoli di livello 2 e veicoli di livello 3. Con riferimento ai primi, si ritiene che non possa operare il principio di affidamento: il conducente, infatti, è titolare di una posizione di garanzia che si “attiva” nel momento stesso in cui entra in funzione il *software* di guida autonoma (vd. *supra*, cap. I, § 7.1.1.1.1.). Egli non potrà fare affidamento sul corretto comportamento del sistema di i.a., dal momento che la diligenza gli imponeva proprio di controllare che quel comportamento non fosse pericoloso. Tutt'al più, un'esclusione della colpevolezza in capo al conducente potrà fondarsi sull'*inesigibilità* della condotta alternativa lecita, nel caso in cui si dimostri che l'agente si sia trovato in circostanze tali per le quali non potesse pretendersi un comportamento conforme al precetto penale⁸⁶.

Diverso il discorso per i veicoli di livello 3: in questo caso, la posizione di garanzia si “attiva” soltanto a seguito di una richiesta di intervento da parte del veicolo. Nelle fasi precedenti, il conducente potrà fare ragionevole affidamento sul funzionamento del veicolo.

La via seguita dal legislatore francese nella regolazione delle macchine a guida parzialmente autonoma merita, a tal proposito, qualche attenzione. La già citata *Loi*

⁸⁴ S. BECK, *Intelligent agents and criminal law*, cit., p. 141; C. MINELLI, *La responsabilità “penale” tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 58; D'AMICO, *Intelligenza artificiale e auto a guida autonoma. tra prevenzione primaria, colpa penale e rischio consentito*, in *Riv. it. med. leg.*, 2022, fasc. 3, p. 608.

⁸⁵ È questo uno dei limiti all'operare del principio di affidamento, vd. F. MANTOVANI, *Il principio di affidamento nel diritto penale*, cit., p. 546; C. PIERGALLINI, (voce) *Colpa (diritto penale)*, in *Enc. dir.*, Annali, X, 2017, p. 238; G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 435.

⁸⁶ Sul concetto di *inesigibilità* vd per tutti, G. FORNASARI, *Il principio di inesigibilità nel diritto penale*, Cedam, 1990.

PACTE del 2019⁸⁷ ha infatti distinto con chiarezza le diverse sfere di responsabilità che coinvolgono produttore e conducente, utilizzando, come criterio di demarcazione, la circostanza dell’attivazione o meno del *software* di guida autonoma, al momento della verifica dell’evento lesivo. In particolare, come già rilevato (vd. *supra*, cap. I, § 7.1.1.1.2.), l’art. L123-2, co. 1, *code de la route* stabilisce che, per il lasso di tempo in cui il sistema esercita il “controllo dinamico” del veicolo, sarà il produttore a rispondere degli eventuali *délits* di omicidio o lesioni stradali – sempre che, ovviamente, ne ricorrano tutti i presupposti in punto di elemento oggettivo e soggettivo; d’altro canto, per il medesimo lasso di tempo, l’art. L123-1, co. 1, esclude la vigenza degli obblighi di vigilanza ordinariamente previsti per il conducente di veicolo tradizionale.

L’art. 123-1, co. 2, delinea inoltre – seppur con una certa vaghezza – gli obblighi di sorveglianza in capo all’utilizzatore di *autonomous vehicle*; in particolare, nel periodo di tempo in cui è attivo il sistema di guida autonoma, il conducente deve mantenersi sempre in stato di allerta, così da poter riprendere tempestivamente la funzione di guida, qualora vi sia una richiesta in questo senso da parte del sistema. Al comma 3, l’art. L123-1 stabilisce infine che laddove il conducente non risponda alla richiesta di riprendere il controllo avanzata dal sistema, tornerà ad applicarsi la disciplina generale sulla responsabilità penale del conducente (art. L121-1), e, dunque, le ordinarie norme cautelari ivi previste.

Sebbene, all’apparenza, la novella potrebbe apparire come priva di una reale portata innovatrice, ribadendo una disciplina già desumibile dai principi generali del diritto penale⁸⁸, in realtà la normativa individua una chiara divisione delle sfere di competenza tra macchina e utilizzatore, stabilendo che quest’ultimo non è tenuto a monitorare costantemente l’attività del *software*, quando questo è attivo, essendo sufficiente che questi sia sufficientemente “allerta” da poter intervenire in caso di emergenza, quando questa sia debitamente comunicata dal sistema.

In ogni caso, come noto, uno dei limiti all’affidamento dell’agente è da individuarsi nella *riconoscibilità del rischio*⁸⁹ – da valutarsi, ovviamente, al momento

⁸⁷ Legge n. 486 del 22 maggio 2019, cit. Sulla disciplina francese vd. più approfonditamente M. GIUCA, *Disciplinare l’intelligenza artificiale*, cit.

⁸⁸ M. GIUCA, *Disciplinare l’intelligenza artificiale*, cit., p. 22 ss.

⁸⁹ M. MANTOVANI, *Il principio di affidamento nel diritto penale*, cit., p. 545; C. PIERGALLINI, (voce) *Colpa (diritto penale)*, cit., p. 238; G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p.

della condotta. Così, nel caso in cui il prodotto abbia già dimostrato malfunzionamenti (ad es., il sistema di allarme di un *autonomous vehicle* di livello 3 ha già dato segnali di difettosità), l'utilizzatore non potrà sottrarsi alla responsabilità allegando di avere fatto affidamento sul corretto funzionamento del sistema di i.a. Il modello «a precauzione bilaterale» tipico della responsabilità per danno da prodotto⁹⁰ – nel quale entrambe le parti (produttore e consumatore) sono chiamate a contribuire alla prevenzione del danno – è infatti destinato a caratterizzare, in maniera ancora più pregnante, l'interpretazione del principio di affidamento nel rapporto tra produttore di dispositivi intelligenti ed utilizzatore. In questi casi, come abbiamo già sottolineato, potrebbe rinvenirsi un'ipotesi di concorso di cause colpose, ai sensi dell'art. 41, co. 3, c.p.

Merita, infine, un cenno il fatto che il principio di affidamento – in quanto principio *biunivoco* e *reciproco* – opera non soltanto a favore dell'utilizzatore, *ma anche* a favore del produttore: quest'ultimo, infatti, non potendo garantire una *sicurezza assoluta*, ma soltanto *relativa*, del prodotto, fa ragionevole affidamento sul fatto che il consumatore utilizzi il prodotto conformemente alle istruzioni fornite⁹¹.

3.2.2.2. Un limite alla rimproverabilità dell'utilizzatore? Prospettive *de jure condendo*.

L'impatto dell'automazione sulle capacità cognitive e psico-fisiche degli utilizzatori ha indotto alcuni autori ad ipotizzare una limitazione della responsabilità in capo a questi ultimi, attraverso la valorizzazione delle norme cautelari scritte che – si auspica – saranno presto introdotte.

Il modello di riferimento è l'art. 590-*sexies* c.p., che, come noto, per le lesioni colpose e l'omicidio colposo commessi nell'esercizio della professione sanitaria, prevede una causa di esclusione della responsabilità, nel caso in cui l'evento si è verificato a causa

434-435; D. CASTRONUOVO, *Responsabilità da prodotto*, cit., p. 334; G. FIANDACA – E. MUSCO, *Diritto penale. Parte generale*, cit., p. 590.

⁹⁰ D. CASTRONUOVO, *Responsabilità da prodotto*, cit., p. 334.

⁹¹ D. CASTRONUOVO, *Responsabilità da prodotto*, cit., p. 335. Va comunque ribadito quanto sottolineato nel capitolo precedente (vd. *supra*, cap. II, § 3.3.2), ossia che soltanto eccezionalmente le istruzioni possono costituire mezzi di prevenzione del rischio adatti a sopperire ad un difetto di progettazione del prodotto.

di imperizia e «sono rispettate le raccomandazioni previste dalle linee guida come definite e pubblicate ai sensi di legge ovvero, in mancanza di queste, le buone pratiche clinico-assistenziali, sempre che le raccomandazioni previste dalle predette linee guida risultino adeguate alle specificità del caso concreto»⁹².

Nate originariamente per altri scopi (*i.e.* innalzare lo *standard* di accuratezza nelle attività di diagnosi e terapia, garantendo ai pazienti un servizio migliore), le linee guida cliniche si sono rivelate un preziosissimo strumento nell'accertamento della colpa del personale sanitario, per eventi lesivi legati all'esercizio dell'attività professionale. Le linee guida, in particolare, presentano l'indiscutibile vantaggio di fornire sia all'operatore medico, *ex ante*, sia al giudice penale, *ex post*, un parametro comune, sebbene astratto e generale, per individuare la condotta doverosa al momento del fatto⁹³.

Come abbiamo già sottolineato, parte della dottrina suggerisce che il meccanismo previsto dall'art. 590-sexies c.p. possa essere replicato – attraverso l'introduzione di una specifica disposizione normativa – in relazione alle raccomandazioni provenienti dai sistemi di supporto alla diagnosi e al trattamento clinico (vd. *supra*, cap. I, § 7.2.1.). Prendendo atto che le raccomandazioni fornite dai CDSS acquisiranno, nel corso del tempo, sempre maggiore accuratezza e che esse potranno condizionare l'autonomia professionale del medico, potrebbe dunque essere opportuna l'introduzione di una tutela dell'affidamento del medico nei confronti della correttezza dei risultati algoritmici, che verrebbe meno soltanto nelle ipotesi di colpa grave⁹⁴.

Uno studioso che è recentemente intervenuto sul tema, Antonio Fiorella, ritiene che la causa di esclusione della responsabilità di cui all'art. 590-sexies c.p. possa fungere da modello per una norma che escluda *tout court* la responsabilità penale dell'utilizzatore

⁹² La disposizione deve essere comunque interpretata come se contenesse il limite della colpa lieve: se la colpa da imperizia è grave, la responsabilità penale non potrà essere esclusa, anche nel caso in cui siano state individuate correttamente dal medico le linee-guida adeguate; vd. in questo senso la nota sentenza Mariotti delle Sezioni Unite (Cass., sez. un., 21 dicembre 2017, n. 8770, Mariotti) – una soluzione ermeneutica, com'è stato osservato, difficilmente compatibile con la lettera della norma; vd., per tutti, C. CUPELLI, *L'art. 590-sexies c.p. nelle motivazioni delle Sezioni Unite: un'interpretazione 'costituzionalmente conforme' dell'imperizia medica (ancora) punibile*, in *Dir. pen. cont.*, 2018, n. 3; R. BARTOLI, *Riforma Gelli-Bianco e Sezioni Unite non placano il tormento: una proposta per limitare la colpa medica*, in *Dir. pen. cont.*, 2018, n. 5, p. 245.

⁹³ Così F. CONSORTE, *Colpa e linee guida*, nota a Cass. pen., sez. IV, 23 novembre 2010, n. 8254, in *Dir. pen. proc.*, 2011, n. 10, p. 1228.

⁹⁴ A. PERIN, *Standardizzazione, automazione e responsabilità medica*, cit., p. 231-232; sostanzialmente nello stesso senso J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 122 ss..

del sistema di i.a., nel caso in cui siano rispettate «le raccomandazioni previste dai protocolli sull'IA come definiti e pubblicati ai sensi di legge ovvero, in mancanza di questi, secondo le buone pratiche operative, sempre che le raccomandazioni previste dai predetti protocolli o le buone pratiche risultino adeguate alle specificità del caso concreto»⁹⁵.

La previsione di una norma di siffatto genere mira evidentemente a schermare l'operatore di i.a. da quelle istanze di colpevolizzazione che – prevedibilmente – sarà destinato a subire, in quanto centro di accollo della responsabilità “più vicino” rispetto all'evento algoritmico lesivo. Una simile norma, tuttavia, presuppone l'introduzione di protocolli *chiari e rigidi*, che definiscano in maniera compiuta gli obblighi di sorveglianza dell'utilizzatore⁹⁶. Non sempre, tuttavia, questo sarà possibile: come dettagliare, ad esempio, tutti gli obblighi di diligenza che possono emergere nel co-governo di un veicolo parzialmente autonomo?

D'interesse appare inoltre la seconda disposizione proposta dal medesimo Autore, che, riprendendo sostanzialmente il contenuto dell'art. 2236 c.c., recita: «La responsabilità è esclusa se, tenendo conto della speciale difficoltà dell'attività di controllo dell'I.A., l'errore, in cui il garante sia incorso, non sia stato determinato da colpa grave», laddove «la colpa è senz'altro grave se il garante trascuri di mantenere costantemente monitorati i possibili margini di autonomia non dominabile dell'IA»⁹⁷. Posto che quest'ultima definizione è ai limiti della contraddittorietà o, quantomeno, della vaghezza (come si può *monitorare* un'*autonomia non dominabile*?), la parte precettiva della disposizione, pur non rinunciando ad un accertamento circa la *colpa generica* dell'operatore di i.a., sembrerebbe tenere debitamente in considerazione le già citate difficoltà tecniche che possono emergere nel governo di un sistema di intelligenza artificiale. In particolare, tra i parametri cumulativi che, ai fini della valutazione circa la sussistenza di una colpa *grave*, devono essere presi in considerazione⁹⁸, specifico rilievo

⁹⁵ Così A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 662-663.

⁹⁶ È lo stesso autore a richiederlo, vd. A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 658-659.

⁹⁷ Le disposizioni sarebbero collocate, in ipotesi, all'art. 55-bis c.p., eventualmente rubricato “*Limiti all'autonomia ammissibile dell'IA. La colpa punibile del garante*”; vd. A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 662-663.

⁹⁸ L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, in *Arch. pen.*, 2022, n. 2, p. 6-7. Per una ricostruzione sistematica del concetto di colpa grave si rinvia a P.F. POLI, (voce) *Colpa*

dovrebbe essere attribuito al grado di violazione della regola cautelare, ovvero sia al livello di divergenza tra condotta astrattamente doverosa (da individuarsi sulla base del parametro dell'agente modello) e comportamento concretamente tenuto dall'agente.

In dottrina, tra l'altro, è ormai all'ordine del giorno il dibattito sull'opportunità di delimitare la colpa penale alla sola forma *grave* e *cosciente*, lasciando alla responsabilità aquiliana – nell'ottica dell'attuazione del principio di *extrema ratio* – la sanzione delle forme di colpa residuali⁹⁹. Che il cambiamento culturale possa partire proprio dalla regolazione del rapporto dell'agente con l'intelligenza artificiale?

3.3. La divergenza tra voluto e realizzato: *aberratio ictus* e *aberratio delicti*.

Nel primo capitolo abbiamo osservato che l'intelligenza artificiale può essere utilizzata come *strumento* di commissione di illeciti penali. In questi casi, la dottrina maggioritaria ritiene che a rispondere del delitto commesso, a titolo di dolo, sarà colui che si è consapevolmente servito del dispositivo intelligente (vd. *supra*, cap. I, § 4.)¹⁰⁰. Alla casistica *standard* – in cui il sistema di i.a. cagiona l'evento lesivo, così come pianificato dal suo utilizzatore – si possono tuttavia affiancare ipotesi più complesse, in cui i) l'agente intelligente *devia* rispetto al percorso consumativo ipotizzato dall'utilizzatore; oppure in cui ii) il programmatore si limita a fornire istruzioni *generali*,

grave, in Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 103. ss.; ID. *La colpa grave. I gradi della colpa tra esigenze di extrema ratio ed effettività della tutela penale*, Giuffrè, 2021.

⁹⁹ Così M. DONINI, *Teoria del reato. Una introduzione*, Cedam, 1996, p. 179; M. RONCO, *Scritti patavini*, Giappichelli, I, 2017, p. 406; recentemente vd. L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., *passim*, al quale si rimanda per un'ampia ricostruzione delle ragioni che militano per questa tesi, non nascondendosi, d'altra parte, il rischio che il criterio della gravità si possa risolvere in un incremento di indeterminatezza della fattispecie colposa. Vd. anche D. CASTRONUOVO, *La colpa "penale". Misura soggettiva e colpa grave*, cit., p. 1723 ss. che intravede nella "colpa grave" un grimaldello per introdurre un giudizio "individualizzante" sulla colpa.

¹⁰⁰ S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame?*, cit., p. 425; A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 656-657; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, *ivi*, p. 713-714; S. MASSI, *Affidamento sull'intelligenza artificiale e "disimpegno morale" nella definizione dei presupposti della responsabilità penale*, *ivi*, p. 665 ss.; A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 7; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 100; S. RIONDATO, *Robotica e diritto penale (robots, ibridi, chimere e "animali tecnologici"*, in D. Provolo e al. (a cura di), *Genetics, Robotics, Law, Punishment*, Padova University Press, p. 601-602.

lasciando al *decision making* algoritmico la “scelta” di quando e come commettere il fatto illecito.

Partiamo dal caso *sub i*). Posto che, come abbiamo già sottolineato, un’eventuale ipotesi di *aberratio causae* non ostacolerebbe la configurazione del dolo (vd. *supra*, cap. I, § 4.), i casi problematici concernono la verifica di eventi *diversi* da quelli programmati dall’utente del sistema di i.a. L’autonomia del dispositivo intelligente, infatti, non si presta ad essere imbrigliata nella pianificazione criminosa dell’agente¹⁰¹.

Si pensi alla seguente ipotesi: Tizio programma un drone militare, di cui sia illecitamente entrato in possesso, per uccidere Caio. Il drone, tuttavia, sbaglia mira e, invece di colpire Caio, colpisce sua moglie Sempronio che camminava al suo fianco. Ancora, si immagini che Tizio abbia programmato il drone per gambizzare Caio, ma quest’ultimo, per un errore nell’esecuzione, colpisca Caio al petto, causandone la morte.

Nel primo caso, parrebbe configurarsi un’ipotesi di *aberratio ictus* monolesiva¹⁰², nel secondo di omicidio preterintenzionale. Per entrambi i casi, la disciplina codicistica accolla oggettivamente all’agente il verificarsi dell’evento lesivo *diverso* da quello voluto dall’agente (vd. artt. 82 e 584 c.p.). Sono ipotesi di responsabilità oggettiva: si attribuiscono all’agente, a titolo di dolo, fatti che *solo parzialmente* sono coperti da *rappresentazione e volontà*.

Per armonizzare la disciplina dell’*aberratio ictus* e della preterintenzione (così come dell’*aberratio delicti*) con il principio costituzionale di colpevolezza (art. 27 Cost., co. 1 e 3), si dovrà aggiungere il limite della colpa. Come noto, il principio di colpevolezza impone infatti che «tutti e ciascuno degli elementi che concorrono a contrassegnare il disvalore della fattispecie siano soggettivamente collegati all’agente (siano, cioè, investiti dal dolo o dalla colpa)»¹⁰³

Quanto all’*aberratio ictus*, l’agente risponderà soltanto quando l’offesa a persona diversa sia dovuta a colpa, in quanto, cioè, una persona ragionevole si sarebbe accorta, in

¹⁰¹ Ci riferiamo, qui, a tutte le ipotesi in cui si verifichi un evento diverso da quello voluto dall’agente: oltre alle ipotesi citate nel paragrafo, si pensi altresì all’*aberratio ictus* plurilesiva (art. 82, cpv., c.p.) e alle varie figure di delitti aggravati dall’evento.

¹⁰² I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 110.

¹⁰³ C. Cost., sent. 1085 del 30 novembre 1988. Per una completa ricostruzione delle ipotesi di colpa in attività illecite si rinvia, per tutti, a F. BASILE, *La colpa in attività illecite: un’indagine di diritto comparato sul superamento della responsabilità oggettiva*, Giuffrè, 2005.

quelle circostanze concrete, che l'offesa da lui progettata si sarebbe potuta verificare nei confronti di una persona diversa¹⁰⁴. All'esempio che abbiamo riportato – il drone, invece che colpire Caio, colpisce Sempronia – va assimilato il caso in cui il sistema di *image recognition* confonda Caio con il suo gemello, colpendo quest'ultimo. L'errore nell'identificazione della persona, infatti, rileva, da parte dell'utilizzatore, come *errore nell'uso dei mezzi di esecuzione del reato*, con la conseguenza che, per l'appunto, all'agente si applicherà il disposto di cui all'art. 82, co.1, c.p., piuttosto che la più rigorosa disciplina dell'*error in persona* (art. 60 c.p.).

Nel caso di preterintenzione/*aberratio delicti*, l'agente risponderà di omicidio preterintenzionale o del *diverso delitto* cagionato soltanto qualora la verifica dell'evento diverso avrebbe ragionevolmente potuto essere prevista¹⁰⁵.

Il medesimo meccanismo imputativo, *mutatis mutandis*, è tra l'altro indicato dal codice penale per il concorso di persone anomalo: l'art. 116 c.p. stabilisce infatti che risponde a titolo di concorso doloso chi, volendo concorrere in un determinato reato, abbia fornito invece un contributo causale alla commissione di un altro reato che, per decisione di un altro complice, è risultato diverso da quello da lui voluto. Anche in questo caso, un'interpretazione costituzionalmente conforme impone di imputare al concorrente il reato diverso soltanto se una persona ragionevole poteva prevedere che sarebbe stato commesso quel reato diverso¹⁰⁶.

Sia, dunque, che il sistema di i.a. sia considerato quale *mezzo di esecuzione* del reato – come pare auspicabile –, sia che esso sia considerato una sorta di *tertium genus* tra *mezzo* e *concorrente*, in ogni caso, affinché il *fatto diverso* sia imputabile all'utilizzatore, sarà necessario provare che l'agente versi in una “colpa in attività

¹⁰⁴ G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 454; D. PULITANÒ, *Diritto penale*, cit., p. 332-333; G. FIANDACA – E. MUSCO, *Diritto penale. Parte generale*, cit., p. 409. Per una recente e approfondita ricostruzione ermeneutica dell'istituto vd. T. TRINCHERA, *Questioni in tema di aberratio ictus*, in *Riv. it. dir. proc. pen.*, 2015, fasc. 3, p. 1412 ss.

¹⁰⁵ D. LIMA, *Could AI Agents Be Held Criminally Liable?*, cit., p. 691. Sull'interpretazione costituzionalmente orientata dell'*aberratio delicti* e della preterintenzione vd. G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 458-459; D. PULITANÒ, *Diritto penale*, cit., p. 331-332; FIANDACA – MUSCO, *Diritto penale. Parte generale*, cit., p. 413.

¹⁰⁶ C. Cost., sent. 42 del 13 maggio 1965.

illecita”¹⁰⁷, avendo potuto ragionevolmente prevedere, sulla base delle circostanze concrete, una deviazione rispetto al piano criminoso.

Ci scontriamo, qui, con la difficoltà di delineare un concetto di *prevedibilità in concreto*, in relazione al comportamento di sistemi di i.a. che sono intrinsecamente imprevedibili. Si pone, qui, il medesimo problema che affronteremo quando discuteremo della colpa generica del produttore (vd. *infra*, § 4.5.3). O si ritiene che il requisito colposo della prevedibilità dell’evento/del fatto diverso sia integrato di *default*, quando entra in gioco l’agire di dispositivi intelligenti, oppure si dovrà discernere caso per caso – e con estrema difficoltà – se il sistema intelligente fosse un esecutore “affidabile” o meno. Quest’ultima verifica pare particolarmente complessa ove si consideri che i sistemi di i.a. – specialmente se acquistati sul mercato sommerso e, dunque, non sottoposti a controlli o autorizzazioni, com’è il caso, ad esempio, dei droni *killer* – rischiano di essere *ancora meno prevedibili di quanto lo siano gli esseri umani*.

In questa sede, ci limitiamo a rilevare che la giurisprudenza maggioritaria, in relazione all’applicazione dell’art. 116 c.p., è solita accontentarsi della mera prevedibilità *in astratto*, da accertarsi attraverso un accostamento delle due fattispecie legali – quella del reato voluto e quella del reato diverso che è stato realizzato – senza tener conto delle concrete modalità di realizzazione del fatto¹⁰⁸. Così, come la giurisprudenza considera che sia sempre prevedibile che la rapina possa sfociare in omicidio – anche se, in concreto, *quella rapina* non lasciava presagire un esito letale – allo stesso modo, vi è da ritenere che le corti tenderanno ad utilizzare il medesimo criterio ove l’agente intelligente commetta un reato *diverso* da quello richiesto.

Tornando ai nostri esempi, se il drone uccide Caio, invece di gambizzarlo, a Tizio sarà presumibilmente imputabile l’evento morte, ai sensi dell’art. 584 c.p. Diverso potrebbe essere il caso in cui, ad un ipotetico androide intelligente, si richieda di commettere un furto in un supermercato e questi, invece, commetta una strage: in quest’ipotesi, infatti, i due reati presentano modalità esecutive talmente diverse che potrebbe mettersi in discussione la prevedibilità del fatto da parte dell’utilizzatore. Va rilevato, comunque, che un’ipotesi di questo tipo è, ad oggi, pura fantascienza, non

¹⁰⁷ Così la celebre sentenza Ronci (Cass, sez. un., 22 gennaio 2009, n. 22676), che ha interpretato l’art. 586 c.p. come se già contenesse il limite della colpa.

¹⁰⁸ *Ex multis* Cass., sez. II, 9 dicembre 1985, n. 3633.

vedendosi all'orizzonte una prossima introduzione di androidi in grado di poter eseguire fatti qualificabili come “furto”.

Infine, nel caso in cui il sistema di i.a. leda una persona diversa da quella che intendeva ledere l'operatore, dovrà presumibilmente procedersi ad una valutazione caso per caso, tenendo altresì conto del livello di accuratezza del dispositivo.

3.4. Quale rilievo per il *dolus generalis*? Il caso degli *high frequency trading*.

La seconda ipotesi (*sub ii*) in cui la concreta integrazione di fattispecie dolose incontrerà un ostacolo si ha nel caso in cui l'operatore di un sistema algoritmico si limiti a fornire istruzioni *generali* per la commissione del delitto, lasciando al *decision making* algoritmico la “scelta” di come e quando commettere il fatto illecito. A differenza delle ipotesi esaminate nel paragrafo precedente – in cui l'utilizzatore pianifica in concreto la commissione del fatto illecito, ma l'algoritmo si discosta da tale progetto criminoso – in questo caso il programmatore si limita a definire il *genus* del reato, senza indicare una preferenza in termini di *tempus* e *modus*.

Com'è noto, l'addebito a titolo doloso richiede che l'agente voglia consapevolmente realizzare un fatto concreto che corrisponde al fatto descritto dalla fattispecie incriminatrice¹⁰⁹. L'agente deve rappresentarsi *quello specifico fatto*: non basta, invece, a differenza di quanto accade in relazione al tipo colposo, che questi si rappresenti una *classe di avvenimenti lesivi*¹¹⁰. Nel perseguire i suoi propositi illeciti attraverso i.a., tuttavia, per l'operatore potrebbe essere più proficuo fornire soltanto alcune direttrici di intervento, lasciando all'algoritmo il compito di selezionare il momento, il luogo e le modalità migliori per agire. L'elevatissima potenza computazionale dei sistemi di *machine learning*, idonea a delineare modelli probabilistici previsionali di stupefacente accuratezza, nonché la loro capacità di adattamento a contesti in continuo e rapido cambiamento, potrebbero infatti identificare le condizioni più opportune per la commissione del reato, specie per quanto concerne la criminalità da profitto.

¹⁰⁹ G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 404 ss.

¹¹⁰ F. CONSULICH, *Flash offenders*, cit., p. 1035.

Quella che è una precisa scelta sistematica di politica criminale – l'irrelevanza del *dolus generalis*¹¹¹ – potrebbe rivelarsi, con lo sviluppo e la commercializzazione di sofisticati sistemi di i.a., una lacuna normativa di non poco conto, favorendo comportamenti opportunistici da parte degli utilizzatori¹¹². L'agente potrebbe infatti sfruttare la già sottolineata caratteristica di *tertium genus* dei dispositivi intelligenti per commettere impunemente fatti illeciti. Se l'agente si comportasse con le medesime modalità nei confronti di una persona fisica (es. la istigasse a commettere un reato, senza conoscere le precise modalità con cui tale reato sarà realizzato), egli sarebbe punito ai sensi dell'art. 110 c.p. Dal momento che – per le motivazioni già più volte evidenziate – i sistemi di i.a. non possono essere considerati alla stregua di persone fisiche, la disciplina del concorso non potrà trovare applicazione – a meno di non volerne estendere analogicamente il meccanismo imputativo: una soluzione, a nostro parere, che entrerebbe diametralmente in contrasto con il principio di legalità, *sub specie* di tassatività. Tuttavia, pur non essendo *persone fisiche*, i sistemi di i.a. sono comunque dotati di autonomo *decision making*: la qualificazione come “strumenti” o “oggetti” nelle mani dell'utilizzatore – pur doverosa, nel nostro sistema – non riesce a cogliere a pieno la particolare capacità algoritmica di *attualizzare* e *concretizzare* le istruzioni generiche fornite dal programmatore.

A rigore, dunque, il programmatore che addestrì *social bots* fornendogli un *dataset* di migliaia di *tweet* ingiuriosi e diffamatori, non potrà essere chiamato a rispondere dei *post* diffamatori pubblicati dal *bot* sui *social network*; ancora, al *trader* che fornisca ai sistemi di *high frequency trading* (HFT) parametri manipolativi del mercato non sarà addebitabile il reato di *market abuse*, difettando la rappresentazione del *concreto fatto di reato*¹¹³. Né potrà individuarsi, in capo all'utilizzatore, una responsabilità

¹¹¹ Sul tema vd. G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 406.

¹¹² In dottrina c'è addirittura chi propone di estendere l'oggetto del dolo, fino a comprendervi il *dolus generalis* (identificato come “*opaque recklessness*”): «*Under this approach, the use of an AI systems with the awareness that it might engage in criminal activities of a certain kind (e.g. unlawful commercial transactions) suffices for the user's criminal liability for the specific criminal activity performed by the system (e.g. purchase of Ecstasy), even when the user did not foresee that the system would engage in the specific activity*» (F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 459). Una soluzione che ci pare sproporzionata e in potenziale contrasto con i principi di legalità e colpevolezza.

¹¹³ Così F. CONSULICH, *Il nastro di Möbius*, cit., p. 218-219; A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 749-750.

omissiva “da posizione”, in assenza di un obbligo giuridico di impedire l’evento, ai sensi dell’art. 40, cpv., c.p.¹¹⁴.

Com’è stato osservato, questo fenomeno si traduce nel paradosso per cui si viene a creare «*a great incentive for human agents to avoid finding out what precisely the ML system is doing, since the less the human agents know, the more they will be able to deny liability*»¹¹⁵. Una soluzione – quella dell’impunità – del tutto insoddisfacente, in relazione a condotte per le quali la meritevolezza della risposta penale appare intensa ed immediatamente percepibile¹¹⁶.

Il suddetto *responsibility gap* è stato evidenziato soprattutto con riferimento al fenomeno degli *high frequency trading*, sempre più diffusi nel mondo *Fintech*. Oltre alle ipotesi – su cui siamo già soffermati precedentemente (vd. *supra*, cap. I, § 7.1.2.2.) – in cui è l’HFT che, autonomamente, reagisce a situazioni di contesto (es. pubblicazione di notizia falsa), vendendo o acquistando titoli in rapida successione, vi è per l’appunto la possibilità che le pratiche manipolative siano frutto di programmazione da parte del *trader*, al quale sarebbe dunque riconducibile un *dolo antecedente e generale*, irrilevante ai fini penalistici.

Federico Consulich – tra i primi, in Italia, ad occuparsi della rilevanza penale delle condotte manipolative poste in essere da sistemi di *algorithmic trading*¹¹⁷ – ha intravisto, in questa casistica, un’assonanza con l’istituto dell’*actio libera in causa*, meccanismo in base al quale si colpisce una condotta priva dei necessari connotati soggettivi, a condizione che essa sia conseguenza di una precedente libera scelta di agire *in quella maniera*. Nonostante la suggestività della ricostruzione, è lo stesso Autore a riconoscerne i limiti: nell’ordinamento italiano, infatti, un simile meccanismo ascrittivo è circoscritto allo *stato preordinato d’incapacità di intendere o di volere* (art. 87 c.p.; vd. anche artt. 92, co. 2, e 93 c.p. per le ipotesi specifiche di ubriachezza e intossicazione da stupefacente) e non potrebbe, dunque, essere applicato *in malam partem* ad ipotesi diverse; d’altra parte, anche in relazione all’art. 87 c.p., per applicare il modello dell’*actio*

¹¹⁴ D. PIVA, *Machina discere, (deinde) delinquere et puniri potest*, cit., p. 686.

¹¹⁵ R. WILLIAMS, *Lords select committee, artificial intelligence committee, written evidence (AIC0206)*, 2017, disponibile a questo link: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/70496.html%23_ftn13#_ftnref7.

¹¹⁶ F. CONSULICH, *Il nastro di Möbius*, cit., p. 219.

¹¹⁷ F. CONSULICH, *Il nastro di Möbius*, cit., *passim*; vd. anche F. CONSULICH, *Flash offenders*, cit., *passim*.

libera in causa è necessario che vi sia piena corrispondenza tra reato commesso e reato programmato – una corrispondenza difficile da registrare nel caso di cui stiamo trattando¹¹⁸.

In questi casi – quantomeno *de jure condito* – la possibilità di muovere un rimprovero al *trader*-persona fisica pare dunque preclusa¹¹⁹; sorge dunque la necessità di concentrarsi su quello che, in una buona maggioranza di casi, è il reale fulcro degli interessi che stanno dietro le condotte di *market abuse*: l'ente¹²⁰. In particolare, nel nostro ordinamento¹²¹, il riferimento è alla possibile valorizzazione dell'art. 8 d.lgs. 231/2001, nella parte in cui stabilisce (lett. a) che la responsabilità dell'ente sussiste anche quando «l'autore del reato non è stato identificato»: la soluzione, proposta con riferimento alle fattispecie di manipolazione del mercato, parrebbe in realtà trasponibile anche in altri settori interessati dall'attività algoritmica. In particolare, secondo tale prospettiva, l'art. 8 delineerebbe un terzo modello imputativo (ulteriore rispetto a quello previsto agli artt. 6 e 7 del decreto, rispettivamente per i reati commessi da apicali e per i reati commessi da sottoposti), fondato proprio sulla mancata identificazione dell'autore del reato, fatto di per sé prova di una disorganizzazione¹²². L'uso abusivo di HFT, secondo questa tesi,

¹¹⁸ Vd. F. CONSULICH, *Il nastro di Möbius*, cit., p. 222-223.

¹¹⁹ F. CONSULICH, *Il nastro di Möbius*, cit., p. 224 ss.; F. MAZZACUVA, *The impact of AI on corporate criminal liability: algorithmic misconduct in the prism of derivative and holistic theories*, cit., p. 144; M. DIAMANTIS, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, cit., *passim*, spec. p. 898; S. AIRES DE SOUSA, "Não fui eu, foi a máquina": *teoria do crime, responsabilidade e inteligência artificial*, cit., p. 69; vd. anche A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, cit., p. 750-751, che sottolinea come, nel nostro ordinamento, una sanzione per il *trader* sarà comunque sempre applicabile a titolo di colpa, attraverso il ricorso alla disciplina amministrativa in materia di *market abuse* (vd. art. 187-ter T.u.f.).

¹²⁰ F. CONSULICH, *Il nastro di Möbius*, cit., p. 224 ss.; M. DIAMANTIS, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, cit., *passim*, spec. p. 898; S. AIRES DE SOUSA, "Não fui eu, foi a máquina": *teoria do crime, responsabilidade e inteligência artificial*, cit., p. 73-74; F. MAZZACUVA, *The impact of AI on corporate criminal liability: algorithmic misconduct in the prism of derivative and holistic theories*, cit., p. 144 ss., al quale si rinvia per una panoramica sui regimi applicabili negli ordinamenti di *common law* (in particolare, *respondeat superior* e *identification principle*), basati su regimi di imputazione *lato sensu* oggettivi; A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, cit., p. 750-751.

¹²¹ Negli ordinamenti di *common law*, la soluzione proposta è quella di applicare estensivamente meccanismi di *vicarious liability*, vd. M. DIAMANTIS, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, cit., *passim*, spec. p. 926; F. MAZZACUVA, *The impact of AI on corporate criminal liability: algorithmic misconduct in the prism of derivative and holistic theories*, cit., p. 144 ss.

¹²² F. CONSULICH, *Il nastro di Möbius*, cit., p. 225-226. Sul meccanismo imputativo previsto dall'art. 8, d.lgs. 231/2001, vd. diffusamente F. CONSULICH, *Il principio di autonomia della responsabilità dell'ente. Prospettive di riforma dell'art. 8*, in *Rivista 231 – La resp. amm. delle soc. e degli enti*, 2018, n. 4, p. 197

potrebbe rientrare a pieno titolo nella casistica: «vi possono essere casi di accettazione aziendale o vero e proprio doloso sfruttamento della loro propensione distorsiva del mercato, senza che si possa ricondurre ad una precisa persona fisica la scelta di quella specifica operazione manipolativa»¹²³.

In realtà, come una parte della dottrina non ha mancato di osservare¹²⁴, il problema della difficile attribuzione della responsabilità penale per un'attività criminosa svolta da un algoritmo di i.a. non è del tutto coincidente con quello della mancata identificazione dell'autore del reato. È vero che, in alcuni casi, potrà risultare ostico risalire al singolo *trader* che abbia attivato l'HFT; il problema, tuttavia, è *a monte*: l'operatore dell'HFT potrà anche essere identificato, ma non è rinvenibile il dolo di *quello specifico fatto tipico* di *market abuse*¹²⁵. A difettare, dunque, non è tanto l'*identificazione* dell'autore del reato, quanto la stessa *integrazione* del reato-presupposto.

In un'ottica precauzionale, si potrebbe riflettere, *de lege ferenda*, sull'opportunità di *vietare* alcune tecniche algoritmiche o di impedirne l'utilizzo in determinati ambiti¹²⁶. Qui il discorso necessiterebbe di una ben più approfondita analisi, che prenda in considerazione, settore per settore, i beni giuridici che vengono in rilievo, nonché i rischi e i benefici di ciascuna particolare applicazione dell'attività algoritmica. Restando al settore dei HFT, ci limitiamo a constatare che, ad oggi, un'esclusione degli algoritmi di *trading* ad alta frequenza dai mercati finanziari pare del tutto improbabile, dal momento che essi sono alla base della grandissima parte delle negoziazioni che avvengono sui mercati: una loro scomparsa potrebbe dunque determinare distorsioni e oscillazioni sull'andamento dei titoli, la cui portata è imprevedibile.

ss.; M.A. BARTOLUCCI, *L'art. 8 d.lgs. 231/2001 nel triangolo di Penrose. Tra minimizzazione del rischio-reato d'impresa e 'nuove forme' di colpevolezza*, in *Dir. pen. cont.*, 9 gennaio 2017.

¹²³ F. CONSULICH, *Il nastro di Möbius*, cit., p. 232.

¹²⁴ A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, cit., p. 752-753; S. PREZIOSI, *Responsabilità da reato degli enti e intelligenza artificiale*, in *Rivista 231 – La resp. amm. delle soc. e degli enti*, 2020, n. 4, p. 176 ss.

¹²⁵ A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, cit., p. 753.

¹²⁶ F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 459.

4. Responsabilità del produttore per l'evento lesivo derivante dal sistema di i.a.

Come abbiamo già rilevato nel precedente capitolo (vd. *supra*, cap. II, § 3. ss.), l'impresa produttrice di sistemi di i.a. costituisce il soggetto più idoneo a valutare i rischi della commercializzazione dei prodotti intelligenti e ad incidere sulla loro sicurezza¹²⁷. Con la progressiva attribuzione di autonomia ai dispositivi intelligenti, sarà possibile esercitare, da parte degli utilizzatori, un controllo sempre minore sulle fonti di rischio: anche qualora il principio dell'*human in the loop* venisse confermato sul lungo periodo, una responsabilità penale in capo all'operatore del sistema sarà configurabile soltanto qualora questi sia posto nelle condizioni di poter intervenire concretamente sull'attività algoritmica. Una soluzione opposta, come già sottolineato, equivarrebbe ad attribuire all'utilizzatore una responsabilità (oggettiva) da posizione, in assoluto contrasto con il principio di colpevolezza sancito dalla Costituzione, all'art. 27, co. 1 e 3.

D'altra parte, la perdita di controllo sul comportamento degli algoritmi di *machine learning* caratterizza anche le società produttrici, che introducono sul mercato artefatti pericolosi, nella consapevolezza di non conoscerne a pieno tutti i potenziali effetti collaterali.

Nei paragrafi che seguono, ci ricollegheremo a molti dei problemi già emersi precedentemente, quando ci siamo occupati della responsabilità civile del produttore per il danno da prodotto difettoso (vd. *supra*, cap. II, § 3. ss.); in questa sede, tuttavia, li affronteremo dalla prospettiva penalistica, cercando di verificare se e come possa configurarsi una responsabilità penale del produttore per i "reati da intelligenza artificiale". Sottolineiamo fin da subito che la problematica riconducibilità del fatto algoritmico alla "persona" del produttore si innesta in un contesto, già esistente, di grave messa in discussione (fino, talvolta, alla forzatura) delle categorie dogmatiche tradizionali, incapaci di contenere – se non attraverso notevoli flessibilizzazioni – la complessità dei fenomeni (post) moderni¹²⁸.

¹²⁷ G. WAGNER, *Robot liability*, cit., p. 38 ss.

¹²⁸ Sulla responsabilità penale per danno da prodotto si vd. l'accurata ricostruzione teorica e giurisprudenziale di C. PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, Giuffrè, 2004, cit.; si vd. inoltre ID., *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, in *Riv. it. dir. proc. pen.*, 1996, p. 354 ss.; ID., *La responsabilità del produttore: una nuova frontiera del diritto penale?*, in *Dir. pen. proc.*, 2007, n. 9, p. 1125 ss; F. STELLA, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Giuffrè, III ed., 2003; A.

Prima di entrare nel merito delle questioni, ci pare opportuna un'avvertenza. Per comodità espositiva, nei paragrafi che seguono faremo spesso riferimento alla figura del “produttore” dei sistemi di i.a. come presunto autore del reato. Con l'espressione “produttore” intendiamo riferirci a tutte quelle persone che, a vario titolo, contribuiscono ai processi di *sviluppo* e *produzione* dei dispositivi intelligenti (progettisti, *data trainer*, programmatori, fabbricanti, installatori, distributori etc.) – ben consapevoli, tuttavia, delle asperità relative all'accertamento: (i) della specifica *causa* dell'evento lesivo (es. difetto di programmazione o difetto di installazione); (ii) della persona responsabile all'interno delle organizzazioni complesse (es. professionista addetto alla programmazione dell'algoritmo o amministratore delegato della società stessa). Dedicheremo a questo aspetto un paragrafo specifico (vd. *infra*, § 4.4.).

4.1. Il diritto penale di fronte all'incertezza: principio di precauzione e tecniche di incriminazione.

Prima di valutare se sia configurabile una responsabilità del produttore per gli eventi lesivi derivanti dai sistemi di i.a., può essere utile una sintetica panoramica sul ruolo che il principio di precauzione può giocare nell'incriminazione delle condotte di sviluppo e immissione sul mercato di dispositivi intelligenti¹²⁹. L'incertezza scientifica

BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, in *Riv. trim. dir. pen. ec.*, 2003, p. 1 ss.; A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, in R. Bartoli (a cura di), *Responsabilità penale e rischio nelle attività mediche e d'impresa: (un dialogo con la giurisprudenza)*, Firenze University press, 2010, p. 437 ss.; C.E. PALIERO, *L'autunno del patriarca, Rinnovo o trasmutazione del diritto penale dei codici*, in *Riv. it. dir. proc. pen.*, 1994, p. 1220 ss.; C. PERINI, *Il concetto di rischio nel diritto penale moderno*, Giuffrè, 2010; D. CASTRONUOVO, *Responsabilità da prodotto e struttura del fatto colposo*, in *Riv. it. dir. proc. pen.*, 2005, p. 301 ss.; J.M. SILVA SANCHEZ, *L'espansione del diritto penale*, cit.; G. AZZALI, *La responsabilità penale del produttore per danni alla salute*, in *Riv. trim. dir. pen. ec.*, 1991, p. 849 ss.

¹²⁹ Sul rapporto tra principio di precauzione e diritto penale si vedano gli studi monografici di D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, Aracne, 2012; E. CORN, *Il principio di precauzione nel diritto penale. Studio sui limiti all'anticipazione della tutela penale*, Giappichelli, 2013; F. CONSORTE, *Tutela penale e principio di precauzione*, Giappichelli, 2013; vd. anche C. RUGA RIVA, *Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*, in E. Dolcini – C.E. Paliero (a cura di) *Studi in onore di Giorgio Marinucci*, II, Giuffrè, 2006, p. 1743 ss.; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, in *Criminalia*, 2006, p. 227 ss.; A. MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole?*, in *Dir. pen. cont.*, 2011; G. FORTI, *La “chiara luce della*

sul funzionamento dei sistemi di *machine learning* potrebbe infatti indurre il legislatore a vietare alcune forme di intelligenza artificiale che possano comportare un rischio per la salute e l'incolumità fisica dei consumatori.

Com'è noto, il principio di precauzione – nelle diverse varianti nelle quali si può esplicitare – impone l'astensione dall'esercizio di un'attività (o il *limite* al suo esercizio, al di sotto di una soglia “di sicurezza”), nel caso in cui vi sia *incertezza scientifica* circa la possibilità che l'impiego di una determinata tecnologia o lo svolgimento di una data attività possano comportare danni alla salute e all'integrità fisica delle persone¹³⁰. A differenza del concetto di prevenzione – che mira a ridurre l'incidenza di rischi noti e scientificamente dimostrati – la precauzione concerne i *rischi ignoti* che, allo stato delle conoscenze nomologiche, *non* possono essere *interamente dimostrati*, né possono essere *ragionevolmente esclusi*¹³¹.

Il principio ha origine nel campo della tutela ambientale e, in questo settore, trova espresso riconoscimento nell'art. 191 TFUE (*ex art.* 174 TCE), ove, al § 2, si stabilisce che «La politica dell'Unione in materia ambientale mira a un elevato livello di tutela, tenendo conto della diversità delle situazioni nelle varie regioni dell'Unione. Essa è fondata sui principi della precauzione e dell'azione preventiva, sul principio della correzione, in via prioritaria alla fonte, dei danni causati all'ambiente, nonché sul

verità” e l’“ignoranza del pericolo”. *Riflessioni penalistiche sul principio di precauzione*, in M. Bertolino – G. Forti (a cura di), *Scritti per Federico Stella*, Jovene, 2007, p. 573 ss.; D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, in *Dir. pen. e proc.*, 2008, p. 647 ss.; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, Giuffrè, 2018, p. 324 ss.; C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto “a distanza” tra disvalore d’azione e disvalore d’evento*, in *Riv. trim. dir. pen. econ.*, 2010, p. 225 ss.; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, in G. De Francesco – G. Morgante (a cura di), *Il diritto penale di fronte alle sfide della «società del rischio»*, Giappichelli, 2017, p. 293; da ultimo, vd. F. D’ALESSANDRO, (voce) *Precauzione (principio di)*, in Donini (diretto da), *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 1001 ss.

¹³⁰ C. PIERGALLINI, *Il paradigma della colpa nell’età del rischio: prove di resistenza del tipo*, cit., in *Riv. it. dir. proc. pen.*, 2005, p. 1695. In generale, sul principio di precauzione come criterio di gestione (spesso irrazionale) del rischio vd. C. SUNSTEIN, *Laws of Fear. Beyond the precautionary principle*, Cambridge University Press, 2005. M.A. GEISTFELD, *Reconciling Cost-Benefit Analysis with the Principle That Safety Matters More Than Money*, in 76 *New York University Law Review*, April 2001, p. 173 ss.; con riferimento al tema del riscaldamento globale vd. D.A. FARBER, *Coping with Uncertainty: Cost-Benefit Analysis, the Precautionary Principle, and Climate Change*, in 90 *Washington Law Review*, 2015, p. 1659 ss.

¹³¹ D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell’incertezza nella struttura del reato*, cit., pp. 26-27; A. MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole?*, cit., p. 4 ss.; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 237-238.

principio "chi inquina paga"¹³². Il campo di applicazione del principio è stato poi progressivamente esteso anche ad altri settori e, oggi, trova un generale riconoscimento nella Comunicazione della Commissione sul principio di precauzione del 2000¹³³, che stabilisce che il principio «trova applicazione in tutti i casi in cui una preliminare valutazione scientifica obiettiva indica che vi sono ragionevoli motivi di temere che i possibili effetti nocivi sull'ambiente e sulla salute degli esseri umani, degli animali e delle piante possano essere incompatibili con l'elevato livello di protezione prescelto dalla Comunità» (§ 4).

Il principio di precauzione, in diversi casi, ha ispirato il legislatore penale nella definizione del bilanciamento di interessi tra esercizio di attività pericolose utili e protezione della salute e dall'integrità personale dei cittadini – si pensi, ad esempio, al settore della tutela del consumatore (art. 114 ss., cod. cons.), alla disciplina degli organismi geneticamente modificati (d.lgs. 12 aprile 2001, n. 206 e d.lgs. 8 luglio 2003, n. 224), alla tutela dell'ambiente (d.lgs. 3 aprile 2006, n. 152, c.d. t.u. ambiente) e, da ultimo, alla prevenzione del contagio da Covid-19¹³⁴. Ancora, proprio nel settore dell'intelligenza artificiale, si può affermare che l'*AI Act* debba molto del suo complessivo impianto valoriale al principio di precauzione – quantomeno nella sua accezione moderata¹³⁵. Il divieto di commercializzare una serie di “pratiche di intelligenza artificiale” (art. 5, *AI Act*, vd. *supra*, cap. II, § 3.4.2.) – tra le altre, l'uso di sistemi di identificazione biometrica remota “in tempo reale” e l'uso di tecniche di *social scoring* – è evidentemente un'applicazione del principio di precauzione. Lo stesso si può dire del trattamento riservato ai sistemi di i.a. ad alto rischio, che possono essere introdotti sul mercato soltanto qualora vengano rispettate una serie di cautele volte a minimizzare i rischi per i diritti fondamentali e per la salute e l'integrità fisica delle persone – con

¹³² Per un panorama completo dei richiami al principio di precauzione contenuti nella normativa internazionale e comunitaria v. G. FORTI, “Accesso” alle informazioni sul rischio e responsabilità, cit., 160 ss., e, più recentemente, F. D'ALESSANDRO, (voce) *Precauzione (principio di)*, cit., p. 977 ss.

¹³³ Commissione europea, Comunicazione della Commissione sul principio di precauzione, 2 febbraio 2000, COM (2000) 1 final.

¹³⁴ Per una ricostruzione della normativa italiana ispirata al principio di precauzione vd. F. D'ALESSANDRO, (voce) *Precauzione (principio di)*, cit., p. 980 ss.

¹³⁵ B. PANATTONI, *Intelligenza artificiale*, cit., p. 356-357; C. MINELLI, *La responsabilità “penale” tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 52.

l'obbligo ulteriore di una verifica periodica circa gli eventuali nuovi rischi che emergano dopo la commercializzazione.

In generale, la dottrina maggioritaria tende ad escludere che il principio di precauzione possa giustificare l'introduzione di reati di pericolo concreto¹³⁶; maggiore apertura vi è invece verso la possibilità che si possa ricorrere alla formulazione di fattispecie di pericolo astratto anche in assenza di una base scientifica corroborata circa l'effettiva pericolosità di determinate condotte (si tratterebbe, comunque, di un "pericolo" del tutto ipotetico)¹³⁷. Tale posizione, pur nella sua problematicità, sarebbe giustificata dalla necessità di tutelare beni giuridici di rango primario – quali, per l'appunto, l'integrità fisica e la salute delle persone – nei casi in cui la rischiosità di una determinata condotta, anche se non certa, pare seriamente possibile, sulla base dei dati scientifici disponibili. Va comunque rilevato che l'orientamento dominante è nel senso di ritenere che il principio di offensività imponga che la previsione di reati di pericolo astratto possa essere giustificata soltanto sulla base di dati eziologici nomologicamente consolidati¹³⁸.

Ancora, l'approccio precauzionale può essere utilizzato in sede di definizione delle norme cautelari¹³⁹. È quanto accade con la definizione di limiti-soglia quantitativi all'esercizio di un'attività, che, in contesti di incertezza scientifica, spesso si ispira a logiche iperprudenziali¹⁴⁰. Con riferimento alle cautele positive *elastiche*, invece, è

¹³⁶ D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 652; C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1748; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 46; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 240.

¹³⁷ D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 652; G. FORTI, "Accesso alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione", in *Criminalia*, 2006, p. 155 ss.; p. 204 ss.; R. BARTOLI, (voce) *Fonti della colpa*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, 2022, p. 526.

¹³⁸ M. DONINI, *Il volto attuale dell'illecito penale*, Giuffrè, 2004, p. 120; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 504 ss.; C. RUGA RIVA, *Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*, cit., p. 1749; A. MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole?*, cit., p. 5; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 47 ss.; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 291-292.

¹³⁹ D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 652; R. BARTOLI, (voce) *Fonti della colpa*, cit., p. 526.

¹⁴⁰ F. CONSULICH, (voce) *Rischio consentito*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 1114; F. D'ALESSANDRO, *Pericolo astratto e limiti soglia. Le promesse non mantenute del diritto penale*, Giuffrè, 2012, p. 266. Sul rapporto tra logiche precauzionali e tecnica dei limiti-soglia vd. diffusamente E. PENCO, *Limiti-soglia e responsabilità colposa. Il ruolo incerto delle soglie quantitative, dalla colpa specifica al rischio consentito*, in *Riv. it. dir. proc. pen.*, 2019, fasc. 1, p. 195 ss.

quanto accade, con riferimento alla materia di nostro interesse, nella definizione dei requisiti per la messa in commercio dei sistemi di i.a., ai sensi dell'*AI Act*.

La cautela consistente nella *minimizzazione* di attività *potenzialmente pericolose* potrebbe infatti rientrare legittimamente in quel bilanciamento tra istanze contrastanti che costituisce il *proprium* della definizione positiva delle cautele: così il legislatore, o, comunque, gli enti pubblici o di rappresentanza di categoria, nel rappresentare gli interessi dei consociati, potranno stabilire che la prudenza imponga di *non agire* in determinate circostanze, anche qualora questo significhi la rinuncia allo svolgimento di attività vantaggiose per la collettività. Inoltre, la definizione positiva di regole cautelari, pur se ispirata al principio di precauzione, non perde la sua funzione di garanzia nei confronti dell'agente. D'altra parte, va ricordato che la violazione di regole cautelari rigide ispirate alla logica precauzionale potrà dare luogo ad un addebito basato sulla *colpa specifica* soltanto qualora, nel corso del giudizio, venga accertata l'effettiva idoneità preventiva tipica della regola precauzionale, nonché l'efficacia della condotta alternativa lecita¹⁴¹.

In tutti questi casi – sia che giustifichi l'introduzione di reati di pericolo concreto o astratto, sia che ispiri la codificazione di norme cautelari – il principio di precauzione è assunto dal legislatore/regolatore come direttrice di intervento politico. Sulla base di un bilanciamento tra costi (incerti) e benefici, il decisore pubblico preferisce rinunciare ai vantaggi che potrebbero derivare dall'esercizio di un'attività (o ad alcuni di essi), piuttosto che rischiare che si verifichino eventi lesivi di cui non è prevedibile la portata¹⁴².

Più complessi sono invece i rapporti tra *principio di precauzione* e *colpa generica* – e lo stesso può dirsi dei rapporti tra logica precauzionale e regole cautelari elastiche. La questione, in questo caso, è quella di valutare se, in un contesto di *incertezza scientifica*, la diligenza imponga di astenersi dall'esercitare l'attività potenzialmente pericolosa, o comunque di ridurne l'impatto (ma allora sorge la domanda: *di quanto* e *come* ridurne l'impatto?). Promettendoci di ritornare sul principio di precauzione – e, in particolare, sul suo rapporto con la colpa generica (vd. *infra*, § 4.5.3.2.) – ci limitiamo qui a rilevare che, nonostante i molteplici rilievi critici avanzati in dottrina, il principio di precauzione gode

¹⁴¹ D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 652.

¹⁴² R. BARTOLI, (voce) *Fonti della colpa*, cit., p. 526

di un certo successo applicativo nella *law in action*, sia in funzione estensiva della tipicità delle fattispecie, sia in chiave di integrazione extrapenale dei precetti¹⁴³.

4.2. La struttura del fatto tipico, tra commissione e omissione.

Stando al modello di intelligenza artificiale che sembra emergere dal quadro normativo fornito dall'*AI Act*, una responsabilità del produttore potrà configurarsi – riprendendo la celebre distinzione operata da Franco Bricola¹⁴⁴ – sia per il “tipo” di produzione, sia per il “modo” di produzione. La responsabilità per il “tipo” di produzione si configura in relazione a quei settori della produzione che presentano livelli di pericolosità altissimi, difficilmente fronteggiabili mediante cautele relative al processo produttivo: è il caso, nel settore che ci interessa, delle pratiche di i.a. vietate, ai sensi dell’art. 5 dell’*AI Act*. Una volta autorizzata la produzione di dispositivi intelligenti, tuttavia, l’attività produttiva, ancorché rischiosa, dovrebbe considerarsi “consentita”: ecco che, allora, una responsabilità del produttore sarà eventualmente da individuarsi in relazione alle “modalità” di produzione, con riferimento, cioè, alle concrete cautele apprestate dal produttore nell’*iter* di sviluppo e commercializzazione del prodotto, al fine di eliminare o minimizzare i rischi¹⁴⁵.

Da sempre discussa in dottrina è la qualificazione della condotta del produttore come commissiva o omissiva: se è innegabile, infatti, che il produttore, immettendo il prodotto in commercio, compie un’azione naturalisticamente qualificabile come *commissiva*, è anche vero che, nella fase successiva, il suo contegno è quasi unicamente *omissivo* – un’omissione che si discute se possa essere qualificata come mero *momento omissivo della colpa* o come *omissione penalmente rilevante*.

¹⁴³ Per una ricostruzione delle principali applicazioni giurisprudenziali del principio di precauzione vd. D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell’incertezza nella struttura del reato*, cit., p. 87 ss.; F. D’ALESSANDRO, (voce) *Precauzione (principio di)*, cit., p. 985 ss.

¹⁴⁴ F. BRICOLA, *Responsabilità penale per il tipo e per il modo di produzione* in AA.VV., *La responsabilità dell’impresa per i danni all’ambiente e ai consumatori*, Milano, 1978, p. 75 ss.

¹⁴⁵ Riprendono la classificazione di Bricola, nel settore della produzione di sistemi di i.a., C. PIERGALLINI, *Intelligenza artificiale*, cit., 1753-1754; C. MINELLI, *La responsabilità “penale” tra persona fisica e corporation alla luce della Proposta di Regolamento sull’Intelligenza Artificiale*, cit., p. 61-62; F.C. LA VATTIATA, *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, in R. Giordano e al. (a cura di), *Il diritto nell’era digitale*, cit., p. 672.

Quello dell'individuazione di criteri stabili per la distinzione tra condotte commissive ed omissive è in realtà un problema classico del diritto penale¹⁴⁶, aggravato dal contesto *fluid* della società post-moderna, in cui gli eventi lesivi sono spesso frutto di trame causali complesse e frazionate, con elementi naturalisticamente commissivi e omissivi. Non è certo questa la sede per una compiuta tematizzazione del problema. Possiamo limitarci ad assumere, quale punto di riferimento, il criterio proposto da una parte della dottrina, in base al quale l'addebito è da considerarsi *commissivo* qualora si imputi al soggetto di avere *introdotto nella situazione concreta un fattore di rischio in precedenza assente*, poi sfociato nell'effettiva produzione di un evento lesivo; l'addebito dovrà invece dirsi *omissivo* qualora al soggetto si imputi di *non avere contrastato fattori di rischio già presenti nella situazione concreta*, i quali siano effettivamente sfociati nella produzione dell'evento lesivo¹⁴⁷.

Anche così definita nelle sue linee generali, la distinzione tra commissione ed omissione appare fluida nel concreto¹⁴⁸. Proprio con riferimento alla responsabilità penale per danno da prodotto, in particolare, è stata osservata una tendenza alla "interscambiabilità" o "fungibilità", in giurisprudenza, tra qualificazione commissiva e omissiva del fatto¹⁴⁹, che non si afferma attraverso la mediazione di ben individuate posizioni di garanzia, quanto piuttosto in funzione delle contingenze dettate dalle circostanze di fatto del caso concreto¹⁵⁰.

La questione, com'è evidente, non si pone in relazione agli eventi lesivi che siano conseguenza della messa in circolazione di prodotti la cui pericolosità (*rectius*, la cui pericolosità *ulteriore* rispetto ai confini del rischio consentito, così come delimitato dalla

¹⁴⁶ Vd. in questo senso R. BLAIOTTA, *Causalità giuridica*, Giappichelli, 2010, p. 317 ss., che ricostruisce i principali orientamenti dottrinali e giurisprudenziali sul tema; vd. anche G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., p. 103; G. FORTI, *Colpa ed evento nel diritto penale*, Giuffrè, 1990, p. 179 ss.; C.E. PALIERO, *La causalità dell'omissione: formule concettuali e paradigmi prasseologici*, cit., p. 828.

¹⁴⁷ F. VIGANÒ, *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, cit., p. 384, al quale si rinvia per gli opportuni riferimenti.

¹⁴⁸ V. MONGILLO, *Imputazione oggettiva e colpa tra "essere" e normativismo: il disastro di Viareggio*, in *Giur. it.*, aprile 2022, p. 955.

¹⁴⁹ C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1241; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 232 ss.; ID., *La responsabilità del produttore: una nuova frontiera del diritto penale?*, cit., p. 1126; ID., *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 360; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse: problemi e prospettive*, in *Riv. trim. dir. pen. econ.*, 2017, n. 3-4, *passim* e spec. p. 538.

¹⁵⁰ C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1241.

normativa di riferimento sulla sicurezza del prodotto, vd. *infra*, § 4.5.2. ss.) fosse conosciuta o comunque conoscibile al momento dell'immissione nel mercato: in questi casi, la responsabilità del produttore deve essere ovviamente ricostruita sul piano commissivo, dal momento che l'atto di immissione in commercio corrisponde alla *colposa* (o, eventualmente, dolosa) introduzione nella società di un nuovo fattore di pericolo¹⁵¹. Piuttosto, la distinzione tra omissione e commissione rischia di scolorire nei casi in cui il prodotto – dopo essere stato messo in commercio in conformità a tutte le cautele doverose – riveli successivamente di comportare rischi *maggiori* o *diversi* rispetto a quelli che le norme cautelari miravano a minimizzare o neutralizzare.

Con riferimento ai prodotti tradizionali, è questa la casistica *standard*. Il produttore immette un prodotto sul mercato in buona fede, all'oscuro cioè della lesività dello stesso¹⁵². La pericolosità del prodotto potrà emergere successivamente, nella concreta interazione con i consumatori e, eventualmente, a causa di un'inadeguatezza sopravvenuta delle istruzioni fornite in sede di commercializzazione.

In questi casi, la *colpa* eventualmente rimproverabile all'agente difficilmente potrà avere ad oggetto il fatto *commissivo* della messa in commercio (a meno che il produttore, dopo l'emersione della pericolosità del prodotto, abbia messo in commercio una nuova *tranche* di prodotti della stessa serie). Il mancato ritiro del prodotto pericoloso, così come la mancata informazione ai consumatori e alle autorità competenti circa la pericolosità del prodotto, costituiscono fatti che, già *ictu oculi*, appaiono *omissivi*.

La casistica giurisprudenziale internazionale documenta, tuttavia, alcuni casi in cui – per eludere le difficoltà strutturali connesse all'accertamento dell'illecito omissivo improprio (in particolare, l'individuazione della posizione di garanzia, dell'azione doverosa e, infine, dell'efficacia impeditiva ipotetica dell'azione doverosa) – sono state qualificate come *commissive* condotte naturalisticamente *omissive*. Si pensi, ad esempio, al noto caso *Lederspray*, uno spray per pellami che conteneva sostanze tossiche per il sistema polmonare. In quel caso, la decisione assunta dai vertici della società di *non ritirare* il prodotto – nonostante si fossero già verificati eventi lesivi correlati alla messa in commercio del prodotto stesso – è stata considerata *attiva*, perché consistente in un

¹⁵¹ Vd. per tutti C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 236.

¹⁵² C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 261.

asserito *facere* (la decisione)¹⁵³. Il ragionamento è fallace: eliminando mentalmente la decisione dei vertici della società, secondo il criterio della *condicio sine qua non*, la situazione di rischio rimane infatti inalterata; la decisione, nella sostanza, costituisce un'omissione. Il problema sarà, piuttosto, quello di verificare se esistano, per il produttore, obblighi di impedimento dell'evento lesivo.

In alcuni ordinamenti (in specie, quello tedesco e quello spagnolo), la giurisprudenza è solita riconoscere la responsabilità del produttore per il fatto di aver omesso quelle cautele doverose (informazioni al consumatore o ritiro del prodotto), la cui fonte deriverebbe *de facto* dall'aver posto in essere una *pregressa azione pericolosa*, ossia dall'aver messo in commercio un prodotto pericoloso¹⁵⁴. Nello stesso caso *Lederspray*, ad esempio, il tribunale ha ritenuto i dirigenti della società responsabili del reato di *lesioni colpose per omissione*, in relazione ai *fatti verificatisi prima della decisione di mantenere i prodotti sul mercato*, per aver omesso di ritirare il prodotto dannoso dopo che già si erano verificati degli eventi lesivi nei confronti dei consumatori.

Il “fare pericoloso precedente” che fonda la posizione di garanzia è stato in questo caso ricollegato all'atto di immissione in commercio del prodotto, da cui sarebbe scaturito, in capo al produttore, l'obbligo di impedire che da tale atto derivassero eventi lesivi (posizione di garanzia c.d. *da ingerenza* o *ex ingerenz*)¹⁵⁵. Secondo tale prospettiva, la posizione di garanzia “si attiverebbe” soltanto a seguito della *percepibilità* del rischio¹⁵⁶.

Non è di certo questa la sede per approfondire il dibattito sulle fonti dell'obbligo di impedimento, che vede contrapposti i sostenitori della teoria formale a quelli della

¹⁵³ In argomento vd. diffusamente C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 232 ss. Un approccio simile è riscontrabile nella giurisprudenza italiana in materia di esposizione dei lavoratori a sostanze tossiche, in cui la condotta del datore di lavoro è considerata commissiva; per alcuni rilievi critici sul punto si vd. M. ROMANO – F. D'ALESSANDRO, *Nesso causale ed esposizione ad amianto. Dall'incertezza scientifica a quella giudiziaria: per un auspicabile chiarimento delle Sezioni Unite*, in *Riv. it. dir. proc. pen.*, 2016, fasc. 3, p. 1147 ss.

¹⁵⁴ C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 261.

¹⁵⁵ C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 360; ID., *Danno da prodotto e responsabilità penale*, cit., p. 233; C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 261; sul tema vd. G. GRASSO, *Il reato omissivo improprio*, cit., p. 277 ss.

¹⁵⁶ Per la ricostruzione del caso giudiziario, con particolare attenzione ai profili attinenti alla qualificazione della condotta materiale, vd. C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 233 ss.; vd. anche F. STELLA, *Giustizia e modernità*, cit., pp. 232-233; C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 261.

teoria sostanzialistico-funzionale, con la posizione intermedia della teoria mista¹⁵⁷. La giurisprudenza, in diverse occasioni, e in relazione alla casistica più disparata¹⁵⁸, ha mostrato di condividere l'assunto che una posizione di garanzia possa sorgere dalla creazione di una preesistente situazione di pericolo.

A nostro avviso, invece, è più persuasiva la tesi sostenuta dalla dottrina maggioritaria¹⁵⁹, che, in ossequio al principio di legalità, non possa configurarsi una responsabilità per reato omissivo improprio in assenza di norme giuridiche che impongano di attivarsi per impedire uno specifico evento. Possono integrare il disposto di cui all'art. 40 co. 2 c.p. norme giuridiche extrapenali (di qualsiasi fonte: leggi, regolamenti, decreti ministeriali, atti emanati da enti locali etc.), così come norme di diritto privato (contratti o atti unilaterali), ma di certo non meri rapporti *de facto*, né precedenti attività pericolose. Piuttosto, a chi abbia creato una situazione di pericolo sarà eventualmente imputabile, a titolo commissivo, l'evento lesivo verificatosi.

Le medesime perplessità in ordine alla configurabilità di un posizione di garanzia *de facto* sono state manifestate da parte della dottrina che si è occupata nello specifico della responsabilità del produttore per il danno derivante da prodotto¹⁶⁰: il suddetto processo di deformalizzazione determina infatti la diluizione della posizione di garanzia in un vago "obbligo generale di prevenzione del danno"¹⁶¹, attraverso il quale tutti coloro

¹⁵⁷ Sul dibattito vd. per tutti G. FIANDACA, *Il reato commissivo mediante omissione*, cit., p. 129 ss.; G. GRASSO, *Il reato omissivo improprio*, cit., p. 184 ss.; F. SGUBBI, *Responsabilità penale per omesso impedimento dell'evento*, Cedam, 1975, p. 165; I. LEONCINI, *Obbligo di attivarsi, obbligo di garanzia e obbligo di sorveglianza*, Giappichelli, 1999, p. 180 ss.

¹⁵⁸ Si è affermata ad esempio la responsabilità del proprietario di un vivaio per il decesso del conducente di un autocarro che, a causa dell'innalzamento del piano stradale realizzato senza rispettare le distanze, rimaneva folgorato per il contatto delle piante trasportate con la linea elettrica sovrastante (Cass., Sez. IV, 10 gennaio 2013, n. 27591); nello stesso senso è stato risolto il caso di un bambino morto dopo essere stato travolto da una traversa per uso ferroviario, mentre si arrampicava per gioco su una catasta non messa in sicurezza (Cass., Sez. IV, 21 maggio 1990, n. 11112).

¹⁵⁹ G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 292; F. MANTOVANI, *Causalità, obbligo di garanzia e dolo nei reati omissivi*, cit., p. 984; G. FIANDACA, *Il reato commissivo mediante omissione*, cit., pp. 204 ss.; G. GRASSO, *Il reato omissivo improprio*, cit., p. 277 ss.; I. LEONCINI, *Obbligo di attivarsi, obbligo di garanzia e obbligo di sorveglianza*, cit., p. 103 ss.

¹⁶⁰ C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 260 ss.; A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, cit., p. 26; A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, cit., p. 460; C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 360-361.

¹⁶¹ C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 361.

che hanno partecipato al processo di produzione e commercializzazione del bene rispondono dell'evento lesivo per il mero fatto di aver *aumentato il generico rischio della produzione*.

In ogni caso, a nostro parere – posto che una responsabilità commissiva sarà sempre rinvenibile in capo al produttore, nel caso in cui la pericolosità del prodotto fosse conosciuta o conoscibile al momento della messa in commercio – uno spazio per la configurabilità di una *tipicità omissiva* sarà comunque individuabile in capo al produttore, nel nostro ordinamento, in forza degli specifici *obblighi di monitoraggio ed intervento post market*, collocati nel codice del consumo e nelle normative settoriali, che hanno proprio l'obiettivo di individuare difetti del prodotto che siano emersi successivamente alla messa in commercio.

In particolare, il codice del consumo delinea una serie di obblighi di controllo, monitoraggio, scambio di informazioni con le autorità amministrative (art. 104, co. 6 ss., cod. cons.) ed intervento (*sub specie* di ritiro, di richiamo e di informazione al consumatore, vd. art. 104, co. 5, cod. cons.), nel caso in cui emergano *rischi ulteriori* rispetto a quelli che la cristallizzazione delle norme cautelari mirava a prevenire. Ecco che, allora, qualora il prodotto, *dopo* essere stato introdotto sul mercato sulla base di tutte le cautele previste dalla legge o dai regolamenti, riveli un'attitudine lesiva *diversa o maggiore* rispetto a quella che le norme cautelari *accettavano* come “rischio residuo” (vd. *infra*, § 4.5.2. ss.), il produttore dovrà provvedere a fornire le necessarie informazioni al consumatore o a richiamare o a ritirare il prodotto. Qualora non lo faccia, l'evento lesivo derivante dal prodotto gli sarà rimproverabile a titolo omissivo – sempre che, ovviamente, si dimostri che l'azione doverosa avrebbe impedito l'evento lesivo (c.d. *causalità ipotetica*)¹⁶².

I medesimi principi saranno tendenzialmente validi anche per quanto riguarda la produzione di sistemi di i.a. – con l'avvertenza che l'abilità di questi ultimi di auto-apprendere e di modificarsi nel tempo potrebbe rendere opportuna la messa a punto, in via legislativa, di obblighi di controllo e monitoraggio rafforzati (è questa la strada

¹⁶² Va tuttavia rilevato che questa è una tipica situazione in cui vi è incertezza sulla qualificazione del fatto. Il mancato rispetto degli obblighi di informazione, di ritiro e di richiamo, così come il giudizio sull'evitabilità dell'evento, potrebbero infatti essere interpretati come elementi attinenti alla *colpa*, e non al nesso condizionalistico omissivo. Sul punto vd. R. BLAIOTTA, *Causalità giuridica*, cit., p. 318 ss.; C.E. PALIERO, *La causalità dell'omissione*, cit.

intrapresa dalla bozza di regolamento europeo, vd. *infra*, § 4.5.2.2.). Sistemi perfettamente sicuri e conformi agli *standard* tecnici potrebbero infatti rivelarsi, a seguito della concreta interazione con l'ambiente, più pericolosi di quanto previsto in sede di elaborazione delle norme cautelari. Come già sottolineato nel precedente capitolo (vd. *supra*, cap. II, § 3.3.4), oltretutto, i difetti di *design* dei sistemi intelligenti potranno essere frutto della *singolare* attività esperienziale del sistema di i.a., con la conseguenza che essi potranno coinvolgere anche soltanto alcuni dei prototipi della serie produttiva: il monitoraggio potrebbe dunque richiedere un'attività di vigilanza capillare da parte del produttore.

4.3. L'accertamento del nesso di causalità.

Posto che la struttura del fatto tipico imputabile al produttore potrà essere, alternativamente, commissiva o omissiva, a seconda del momento in cui sorga il *rischio residuo* derivante dal prodotto, in questa sede ci occuperemo principalmente della valutazione del nesso di *causalità reale*¹⁶³ tra immissione del prodotto intelligente sul mercato ed evento lesivo, convinti che sia con riferimento a questo momento che si porranno *problemi specifici* inerenti ai sistemi di intelligenza artificiale. L'accertamento della *causalità ipotetica* tra omissione della condotta doverosa ed evento lesivo riprodurrà, invece, problematiche che – seppur non trascurabili – sono già note in letteratura¹⁶⁴.

¹⁶³ In generale, sulla causalità, nella dottrina italiana, è doveroso il richiamo ai primi scritti di Federico Stella, che è stato il più illustre sostenitore della teoria nomologico-deduttiva, vd. F. STELLA, *Leggi scientifiche e spiegazione causale nel diritto penale*, Giuffrè, II ed., 2000; ID., *La nozione penalmente rilevante di causa: la condizione necessaria*, in Riv. it. dir. proc. pen., 1988, p. 1217 ss.; ID., (voce) *Rapporto di causalità*, in Enc. Giur. Treccani, XXV, 1991; sul rapporto di causalità vd. anch. *ex multis*, e i lavori di G. FIANDACA, (voce) *Causalità*, in Dig. Disc. Pen., vol. II, 1988, Utet, p. 119 ss.; O. DI GIOVINE, *Lo statuto epistemologico della causalità penale tra cause sufficienti e condizioni necessarie*, in Riv. it. dir. proc. pen., 2002, fasc. 2, p. 634 ss.; M. DONINI, *Il garantismo della condicio sine qua non e il prezzo del suo abbandono. Contributo all'analisi dei rapporti fra causalità e imputazione*, in Riv. it. dir. e proc. pen., 2011, n. 2, p. 494 ss.; R. BLAIOTTA, *Causalità giuridica*, cit.

¹⁶⁴ In materia vd., per tutti, C.E. PALIERO, *La causalità dell'omissione*, cit., p. 821 ss.; M. DONINI, *La causalità omissiva e l'imputazione "per l'aumento del rischio"*, in Riv. it. dir. proc. pen., 1999, fasc. 1, p. 32 ss., con riferimento ad un contesto "pre-Franzese"; successivamente vd. F. STELLA, *Causalità omissiva, probabilità, giudizi controfattuali. L'attività medico-chirurgica*, in Cass. pen., 2005, 1062 ss.; G. MARINUCCI, *Causalità reale e causalità ipotetica nell'omissione impropria*, in Riv. it. dir. proc. pen., 2009,

La valutazione del nesso eziologico tra immissione del prodotto intelligente ed evento lesivo incontra diversi ostacoli, a causa dell'assenza, ad oggi, di consolidati paradigmi nomologico-deduttivi che siano in grado di descrivere (prima) e di spiegare (dopo) il dispiegarsi della catena causale nel funzionamento degli algoritmi di *machine learning*¹⁶⁵. Posto che, come abbiamo già rilevato nel capitolo precedente (vd. *supra*, cap. II, § 3.5.), i danni cagionati dalla capacità di apprendimento del sistema di i.a. non paiono assurgere a fattore interruttivo del nesso causale, ai sensi dell'art. 41, co. 2, c.p., sarà allora necessario individuare, di volta in volta, il fattore causativo del danno.

Come abbiamo già sottolineato (vd. *supra*, cap. II, § 3.5.), i sistemi di i.a. – pur caratterizzati da assoluta opacità quanto alla rappresentazione dei rapporti di causa-effetto interni ai processi decisionali – possono essere dotati, così come tutti gli artefatti a tecnologia più avanzata, di *event data recorder* (c.d. EDR), ossia di meccanismi di supporto alla ricostruzione dinamica degli eventi lesivi. Gli *event data recorder* non sono in grado di rimediare al *deficit* di *comprendibilità* delle modalità decisionali, ma sono invece capaci di fornire dati utili per la ricostruzione della dinamica dell'incidente. Ad esempio, per quanto riguarda gli autoveicoli, l'EDR, a seconda delle versioni disponibili in commercio, può registrare la velocità di crociera, l'accelerazione, gli eventuali *input* da parte del conducente (es. la pressione sul pedale del freno), l'attivazione di spie o segnali da parte del veicolo¹⁶⁶.

523 ss.; F. VIGANÒ, *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, cit., p. 380 ss.

¹⁶⁵ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1760; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 107; F.C. LA VATTIATA, *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, cit., p. 703; D. PIVA, *Machina discere, (deinde) delinquere et puniri potest*, cit., p. 686; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 720, che parla di “anomia causale”; D.F. LLORCA E AL., *Liability regimes in the age of AI: a use-case driven analysis of the burden of proof*, submitted to *Journal of Artificial Intelligence Research*, 3 novembre 2022, disponibile in Arxiv a questo link: <https://arxiv.org/abs/2211.01817>, p. 6-7; E. KARNER – B.A. KOCH – M.A. GEISTFELD, *Comparative Law Study on Civil Liability for Artificial Intelligence* – Study requested by the Directorate-General for Justice and Consumers (EU Commission), novembre 2020, p. 23 ss.; vd. anche C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, cit., p. 191, U. PAGALLO, *The Law of robots*, cit., p. 73 e Y BATHAEE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, cit., p. 922, i quali, tuttavia, nel parlare di “*failure of causation*”, fanno riferimento alla teoria della *proximate cause*, assimilabile alle teorie della causalità adeguata, in cui vi è dunque una valutazione della *reasonable foreseeability* dell'evento.

¹⁶⁶ T. LEHOULLIER E AL., *The Use of Event Data Recorders in the Analysis of Unintended Acceleration Incidents*, testo della relazione tenuta alla 23e *Conférence canadienne multidisciplinaire sur la sécurité*

Prendiamo l'ipotesi di un incidente stradale causato da un veicolo autonomo. L'*event data recorder* potrà ad esempio indicare che il sistema di guida autonoma non ha rallentato di fronte ad un semaforo rosso o che, in una condizione di manto stradale bagnato, il veicolo non ha applicato lo spazio di frenata che l'algoritmo prevedeva per tale situazione. L'EDR, tuttavia, *non dice nulla sulle cause dell'incidente*: si limita a fornire *dati grezzi* da interpretare – attraverso l'ausilio di periti e consulenti tecnici¹⁶⁷ – ma non è di per sé risolutivo ai fini della ricostruzione del nesso causale. Insomma, le questioni restano aperte: perché il veicolo non ha rallentato di fronte al semaforo rosso? Perché non ha applicato lo spazio di frenata previsto?

Come noto, ai fini dell'accertamento processuale della causalità non potranno bastare semplici *intuizioni* del giudice, eventualmente suffragate dai dati raccolti dagli EDR; sarà invece necessario che il rapporto eziologico tra fatto ed evento sia dimostrato nei suoi sotto-eventi – con l'ovvia precisazione che «è assai difficile, se non impossibile, che il giudice riesca a spiegare, sulla base di teorie e leggi della scienza universali, l'intero «meccanismo di produzione» dell'evento lesivo, ricostruendo le catene di subeventi che, partendo dall'evento iniziale, si collegano l'uno all'altro fino a raggiungere l'effetto finale»¹⁶⁸.

I dati che vengono forniti dai *logs* costituiscono soltanto uno degli elementi che possono essere valorizzati dal giudice nell'accertamento del nesso causale – soprattutto nella fase attuale, in cui la tecnologia di registrazione e conservazione dei dati non ha ancora raggiunto apprezzabili livelli di accuratezza ed affidabilità (tanto che può accadere

routière Montréal, Québec, 26-29 mai 2013, disponibile a questo link: https://www.roadsafetyresearch.ca/papers/edr_unintended_acceleration.pdf;

¹⁶⁷ Sull'utilizzo nel processo penale dei dati raccolti dagli *event data recorder* vd. F.C. LA VATTIATA, *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, cit., p. 707, che, in maniera condivisibile, sostiene che i dati raccolti dall'EDR non potranno essere “auditi” direttamente dal giudice, ma saranno introdotti nel processo in contraddittorio, attraverso consulenza tecnica o perizia. Sul tema della prova scientifica, al centro di ampie riflessioni dottrinali negli ultimi anni, si vedano i contributi raccolti in G. Canzio – L. Lupária Donati (a cura di), *Prova scientifica e processo penale*, Cedam, II ed., 2022; vd. inoltre O. DOMINIONI, *La prova penale scientifica: gli strumenti scientifico-tecnici nuovi o controversi e di elevata specializzazione*, Giuffrè, 2005; F. GIUNTA, *Questioni scientifiche e prova scientifica tra categorie sostanziali e regole di giudizio*, in *Criminalia*, 2014, p. 561 ss.; nonché i contributi raccolti in M. Bertolino – G. Ubertis (a cura di), *Prova scientifica, ragionamento probatorio e decisione giudiziale*, Atti del Convegno, Milano, 10-11 ottobre 2014, Jovene, 2015; sulla prova scientifica nel processo civile si vd., per tutti, M. TARUFFO, *La prova scientifica nel processo civile*, in M. Bertolino – G. Forti (a cura di), *Scritti per Federico Stella*, cit., vol. II, p. 1375 ss.

¹⁶⁸ F. STELLA, *Giustizia e modernità*, cit., p. 348.

che i dati grezzi contenuti nell'EDR siano contraddittori rispetto agli altri dati provenienti da *smartphone*, videocamere di sicurezza installate nelle strade urbane, infrastrutture intelligenti etc.)¹⁶⁹.

L'individuazione dei decorsi eziologici lesivi all'interno dei manufatti complessi è questione già ben conosciuta in dottrina¹⁷⁰: si pensi, ad esempio, alla vicenda Toyota, in cui anni e anni di indagini non sono stati in grado di *spiegare* la causa delle improvvise accelerazioni dei veicoli (vd. *supra*, cap. II, § 3.5.). Alla tipica opacità degli artefatti moderni, l'intelligenza artificiale aggiunge la capacità di interazione con ambienti spesso imprevedibili, l'abilità di modificare le proprie *performance* e di adattarsi a seconda del contesto in cui agisce, e, soprattutto, la *modalità di assunzione delle decisioni*, che non si basa su un paradigma *deterministico* ma sulla creazione di *modelli probabilistici*. Come abbiamo già ribadito precedentemente, ciò non significa che i sistemi di *machine learning* decidono *liberamente*, così come fanno gli esseri umani (*rectius*, postuliamo che essi *facciano*); ciò significa, piuttosto, che l'assunzione delle decisioni da parte degli algoritmi di ML avviene sulla base di una quantità talmente enorme di variabili e dati (sia *codificati*, in sede di progettazione, sia *acquisiti* con l'esperienza) che l'esito di tale processo può essere ricostruito soltanto con grande difficoltà *ex post*, in sede processuale.

Posto, dunque, che, nei prodotti più sofisticati, sarà difficile persino distinguere tra potenziali fattori causali che riguardano il sistema di i.a. e fattori che riguardano l'*hardware* nel suo complesso¹⁷¹, ci preme qui concentrare l'attenzione soltanto sul decorso eziologico interno al micro-sistema *black box* costituito dall'algoritmo di *machine learning*. In questo modo, potremo isolare i *problemi specifici* che concernono l'accertamento del nesso causale in relazione ai sistemi intelligenti.

Prendiamo, dunque, un caso semplificato, in cui, in ipotesi, è già stato dimostrato che l'evento lesivo è stato causato dall'algoritmo di i.a.: un sistema di *image recognition* installato su un *autonomous vehicle* non riconosce un semaforo rosso; il veicolo impegna l'incrocio e si scontra con una moto proveniente dal senso di marcia opposto, causando

¹⁶⁹ K. BÖHM E AL., *New developments on EDR (Event Data Recorder) for automated vehicles*, in *Open Engineering*, 2020, vol. 10, iss. 1, p. 142.

¹⁷⁰ Vd., per tutti, C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 140 ss.; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *Leg. pen.*, 2011, n. 2, p. 401 ss.

¹⁷¹ Vd. per tutti E. KARNER – B.A. KOCH – M.A. GEISTFELD, *Comparative Law Study on Civil Liability for Artificial Intelligence*, cit., p. 23-24.

la morte del motociclista. Ecco che, in un caso di questo tipo, il mancato riconoscimento del semaforo rosso potrà essere alternativamente determinato: dall'esecuzione di tecniche di *adversarial machine learning* da parte di un gruppo terroristico (vd. *supra*, cap. I, § 2.2); dal mancato aggiornamento del *software* da parte del proprietario; da un insufficiente addestramento dell'algoritmo; da condizioni metereologiche che non consentivano al sistema di operare efficientemente; da un *bug* nel codice etc... Oltre a tali fattori – capaci di interferire con il *decision making* algoritmico – il sistema di i.a. può anche *semplicemente sbagliare*¹⁷²: ad esempio, l'algoritmo di *image recognition* può scambiare il semaforo rosso per una decorazione natalizia – così come, in alcune occasioni, ha scambiato una tartaruga per un fucile (vd. *supra*, cap. I, § 2.2). È questa, ovviamente, l'ipotesi più problematica.

In passato, con riferimento ai prodotti tradizionali (“non intelligenti”), una parte della giurisprudenza¹⁷³ ha tentato di “flessibilizzare” il modello nomologico-deduttivo attraverso l'elusione dell'accertamento del concreto decorso eziologico. I giudici, prendendo atto dell'opacità sottesa alle vicende eziologiche relative alla difettosità dei prodotti, hanno in un certo senso “adattato” il giudizio di imputazione alla fisionomia del fenomeno della *black box*, ritenendo sufficiente, ai fini dell'accertamento causale: i) che sussistano un rischio iniziale e un evento lesivo finale, che di tale rischio costituisce concretizzazione; ii) che vi sia il ragionevole convincimento che sussista un rapporto di causalità tra rischio ed evento lesivo; iii) che non abbiano concretamente operato fattori esterni ed indipendenti rispetto al prodotto (c.d. “accertamento *black box*”). Si delinea così, com'è stato osservato, un nuovo concetto di *causalità*: si intuisce che un prodotto ha un nesso con determinati danni, si sa che non vi sono fattori estranei che possono aver

¹⁷² E. KARNER – B.A. KOCH – M.A. GEISTFELD, *Comparative Law Study on Civil Liability for Artificial Intelligence*, cit., p. 23: «it may have worked as intended, but nevertheless produced an undesired outcome due to its self-learning abilities and the capability of adapting its behaviour to new challenges».

¹⁷³ I casi principali nella giurisprudenza internazionale sono, in Germania, quelli relativi all'*Holzschutzmittel* (un preparato utilizzato per la protezione dei mobili di legno, contenente un'alta concentrazione di sostanze insetticide) e al *Lederspray* (uno spray per pelli presunta causa di edemi polmonari) e, in Spagna, quello relativo all'"olio di colza". In Italia si ricorda il processo delle macchie blu, vd. Trib. Rovereto, 17 gennaio 1969, in *Riv. it. dir. proc. pen.*, 1971, p. 1021 ss., su cui vd. F. STELLA, *Leggi scientifiche*, cit., p. 47 ss. e 82 ss. Per una precisa ricostruzione della casistica si rinvia a F. STELLA, *Giustizia e modernità*, cit., p. 224-235; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 50 ss.; sulla giurisprudenza italiana vd. A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, cit., p. 455 ss.

cagionato il danno, ma *non si conosce il fattore dannoso all'interno del prodotto*¹⁷⁴. I giudici ricavano induttivamente il nesso sulla base, da un lato, del collegamento temporale tra l'uso del prodotto e gli eventi lesivi occorsi e, dall'altro, sull'interpretazione della *condicio sine qua non* come condizione necessaria *non* dell'evento, ma dell'aumento del rischio¹⁷⁵.

I giudici hanno talvolta individuato il nesso di condizionamento senza neppure individuare la sostanza causalmente lesiva all'interno di un novero di elementi chimici o naturali "sospetti"¹⁷⁶. In questi casi, com'è stato osservato, il giudice non solo cessa di essere "consumatore" di leggi scientifiche di copertura, ma rischia di non assurgere più nemmeno a "produttore" di tali leggi¹⁷⁷, dal momento che rinuncia alla stessa possibilità di conoscere il fattore eziologico determinante nel caso concreto¹⁷⁸.

Il rischio che simili scorciatoie processuali si verifichino anche in relazione ai c.d. prodotti intelligenti non è trascurabile: ad oggi, infatti – e, stando, alle dichiarazioni degli scienziati su un'incipiente crisi di comprensione dei modelli computazionali più avanzati, sempre di più anche in futuro – non esiste un consolidato corredo nomologico che consenta di *spiegare* il comportamento dei sistemi di i.a., tanto che sono gli stessi scienziati, talvolta, a riferirsi alle attività algoritmiche come ad un fenomeno quasi magico (vd. *supra*, cap. I, § 5.3.1).

L'identificazione della causa dell'errore algoritmico potrebbe apparire all'apparenza superflua: in fondo, si potrebbe pensare, l'importante è che si dimostri che la *condicio sine qua non* dell'evento lesivo sia costituita da un generico malfunzionamento del sistema di i.a., potendo in questo modo accollare l'addebito al "produttore", indipendentemente dall'identificazione dello specifico fattore causativo del danno. A differenza di prodotti in cui è la stessa dannosità ad essere incerta (si pensi ad un prodotto contenente sostanze insetticide, di cui sia dubbia l'efficacia lesiva), nel caso

¹⁷⁴ C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 359, che riporta le considerazioni di W. HASSEMER, *Produktverantwortung im modernen Strafrecht*, 2. Auf., Heidelberg, C.F. Müller, 1996, p. 53-54.

¹⁷⁵ F. STELLA, *Giustizia e modernità*, cit., p. 237 e 245; vd. anche A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, cit., p. 23 ss.

¹⁷⁶ Vd. in particolare i casi *Holzschutzmittel* e *Lederspray*; C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 359; M. DONINI, *La causalità omissiva e l'imputazione "per l'aumento del rischio"*, cit., p. 66.

¹⁷⁷ Sul giudice come "consumatore" di leggi causali vd. F. STELLA, *Leggi scientifiche*, cit., p. 80.

¹⁷⁸ M. DONINI, *La causalità omissiva e l'imputazione "per l'aumento del rischio"*, cit., p. 66.

dei sistemi di i.a. spesso la potenzialità lesiva è *evidente*, poiché legata al dinamismo dell'oggetto: si pensi ad un *healthcare robot* che, mentre operi, recida improvvidamente un'arteria, o ad una *self driving car* che corra all'impazzata per le strade cittadine, investendo qualsiasi persona che incontri sulla sua traiettoria.

Una simile prospettiva ricorda quell'approccio al ragionamento causale basato sul convincimento intuitivo del giudice, che ha dominato nel nostro Paese fino alla fine degli anni Ottanta¹⁷⁹. Emblematica, in questo senso, è la sentenza sul disastro del Vajont, in cui si legge: «Se nessuno è in grado di spiegare perché la frana si sia verificata, ciò nondimeno non si può minimamente dubitare che la frana sia dovuta all'opera dell'uomo»¹⁸⁰. O, ancora, si pensi alla sentenza di condanna del Tribunale Supremo spagnolo nel celebre processo dell'olio di colza, in cui si leggeva: «se è stato provato in modo giuridicamente ineccepibile che la composizione del contenuto del prodotto (...) causa i danni, non è richiesto, per la prova della causalità, che si dimostri la ragione per la quale quel prodotto possa considerarsi causale rispetto al danno, che si provi cioè, secondo l'analisi delle conoscenze scientifiche e naturalistiche, il fondamento di questa causalità»¹⁸¹.

L'approccio nomologico-deduttivo – seppur “corretto” dai criteri di probabilità logica, come risultanti dalla sentenza Franzese del 2002 – richiede invece che vengano rintracciate le leggi scientifiche in base alle quali sia possibile spiegare l'evento. Nel caso in cui il giudice, pur sospettando un legame causale tra una data azione ed un dato evento, si trovi nell'impossibilità di corroborare quel sospetto, dovrebbe risolversi per l'assoluzione dell'imputato, nel rispetto della presunzione di innocenza¹⁸², sancita dall'art. 27, co. 2, Cost, nonché dagli artt. 6, § 2, CEDU e 48 CDFUE.

Anche nell'ipotesi, dunque, in cui il nesso di causalità tra immissione del prodotto sul mercato ed evento lesivo sia “evidente” (perché, ad esempio, possono essere esclusi

¹⁷⁹ Decisiva per un cambio di passo è stata senza dubbio la sentenza relativa al disastro di Stava, vd. Cass., sez. IV, 6 dicembre 1990, n. 4793, Bonetti; in questo senso vd. G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 264; M. RONCO, *Aporie scientiste e certezza logico razionale: note in tema di nesso causale*, in *Dir. pen. proc.*, 2007, n. 2, p. 247; per alcuni esempi di ragionamento causale, “intuitivo” nella giurisprudenza internazionale in materia di responsabilità per danno da prodotto si rinvia a C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 192 ss.

¹⁸⁰ Trib. L'Aquila, 17 dicembre 1969, in *Rass. Giur. Enel*, 1970, p. 197. Sul carattere “intuitivo” delle argomentazioni adottate dal Tribunale in quell'occasione vd. F. STELLA, *Leggi scientifiche*, cit., p. 75 ss.

¹⁸¹ L'estratto della sentenza è riportato in F. STELLA, *Giustizia e modernità*, cit., p. 233, al quale si rinvia per gli opportuni riferimenti; vd. anche, per una ricostruzione della vicenda, C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 57 ss. e 193 ss.

¹⁸² C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 254.

decorsi causali alternativi, quali, ad esempio, la distrazione dell'utente e l'attacco informatico), sarà comunque necessario dimostrare il singolo fatto causativo dell'errata decisione algoritmica.

Se nel processo civile, come abbiamo visto nel precedente capitolo (vd. *supra*, cap. II, § 3.5.), il problema dell'inscrutabilità della *black box* algoritmica potrà essere aggirato attraverso il ricorso alla logica del *res ipsa loquitur* – già utilizzata, nella prassi, dai giudici europei e destinata ad essere incardinata nel testo della direttiva sul danno da prodotto difettoso, qualora la proposta di revisione venisse approvata –, nel processo penale sarà invece necessario individuare il *preciso fattore* che ha cagionato l'evento lesivo. Non solo perché lo impongono il principio di legalità e la presunzione d'innocenza (concretizzata, in termini processuali, dallo *standard* dell'«oltre il ragionevole dubbio»), ma anche perché da tale individuazione dipenderà l'identificazione della sfera di responsabilità all'interno della società produttrice, o, eventualmente, in capo al programmatore, al progettista, al fornitore del *dataset* di addestramento etc. Un problema, quest'ultimo, che pare non porsi con tale forza in relazione al processo civile, in cui il produttore (e più spesso, la società produttrice) risponde oggettivamente per i danni cagionati al consumatore da *qualsiasi tipo* di difetto del prodotto – eventualmente solidalmente con progettisti, programmatori, fornitori di dati etc.

4.3.1. L'individuazione delle leggi scientifiche di copertura: un'incompatibilità tra ragionamento causale e approccio stocastico del *decision making* algoritmico?

Come abbiamo già accennato, in relazione ai danni cagionati dai sistemi di i.a., uno degli aspetti più controversi è rappresentato dall'individuazione delle leggi scientifiche di copertura, in grado di spiegare i decorsi eziologici lesivi.

In alcuni casi, l'errore algoritmico potrà essere spiegato individuando fattori di interferenza esterna: si pensi, ad esempio, ad un attacco *hacker* che mandi in *tilt* un sistema di riconoscimento delle immagini; oppure al mancato aggiornamento del *software* da parte dell'utilizzatore del prodotto intelligente – aggiornamento che avrebbe risolto un *bug* nel codice. In questi casi, si potrà dire che l'ordinario funzionamento degli algoritmi di *machine learning* sia stato deviato da cause esterne, commissive o omissive:

nonostante la complessità della materia, periti e consulenti tecnici possono essere in grado di individuare la legge scientifica che collega il fatto A (ad es., inoculazione di un *virus*) al sotto-evento B (ad es., modifica dei parametri decisionali dell'algoritmo) fino all'evento lesivo C (ad es., investimento di un pedone e morte). Nelle suddette ipotesi, si ripropongono sostanzialmente le difficoltà già incontrate nell'accertamento causale in relazione a sistemi complessi dotati di *black box*.

Ciò che, invece, è peculiare dei sistemi di *machine learning* di ultima generazione è il modello di ragionamento *probabilistico*, che potrà indurre l'algoritmo a commettere "errori" (*rectius*, quelli che dal "nostro" punto di vista, umano, sono considerati errori) anche in condizioni di ordinario e corretto funzionamento¹⁸³.

Come abbiamo osservato nel primo capitolo (vd. *supra*, cap. I, § 5.3.1.), il funzionamento delle tecniche più sofisticate di *machine learning* è ancora in parte ignoto, nonostante se ne sfruttino già le potenzialità applicative. Ci si affida ai modelli previsionali algoritmici come nell'antichità ci si affidava agli oracoli: non si sa bene *perché*, ma le previsioni effettuate dagli algoritmi di i.a. sono, nella maggioranza dei casi, corrette, e spesso più accurate di quelle umane. Quando l'oracolo sbaglia e tale errore determina un evento lesivo, quest'ultimo potrà essere ricondotto eziologicamente ed *ex post* alla decisione oracolare. Ma come potranno individuarsi delle regolarità causali nelle decisioni di sistemi che sono *unpredictable by design*¹⁸⁴?

L'unica regolarità che ci pare che possa essere rilevata è quella relativa ad un'ipotetica *statistica degli errori dell'i.a.* Tornando all'esempio del sistema di *image recognition*, si potrà ad esempio rilevare che, in un caso su mille, l'algoritmo non è in grado di riconoscere correttamente le immagini: è quanto, ad esempio, accade quando si valuta che un *Clinical Decision Support System* presenta un'accuratezza del 90% nelle diagnosi di cancro – ciò significa, è evidente, che nel 10% dei casi l'algoritmo *sbaglia* (vd. *supra*, cap. I, § 7.2.1). Tale rilevazione statistica, tuttavia, *nulla* dice sui fattori che possono aver cagionato l'errore: si è trattato dell'esposizione a *dataset* troppo limitati?

¹⁸³ Vd. Z.C. LIPTON, *The Mythos of Model Interpretability*, in 16 *Queue*, May-June 2018, p. 5.

¹⁸⁴ Concorda sull'impossibilità di trovare una legge scientifica di copertura S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 717, che ritiene altresì che la decisione autonoma dell'algoritmo di *machine learning* possa interrompere il nesso causale, ai sensi dell'art. 41, cpv., c.p.

Oppure l'immagine sottoposta a riconoscimento aveva uno spettro di colori troppo ampio per le capacità computazionali dell'algoritmo?

Breve: non si tratta di una legge scientifica, dal momento che la valutazione statistica non è in grado di isolare il motivo scatenante dell'errore (e, di conseguenza, del danno) e non dice nulla su quegli "eventi intermedi" della catena causale la cui prova è invece necessaria per poter verificare compiutamente l'esistenza di un rapporto di causalità¹⁸⁵. L'affermazione "un sistema di *image recognition* sbaglia nel 10% dei casi" è assimilabile ad un'affermazione di questo tipo: "il 10% della popolazione italiana soffre di obesità": lo ripetiamo, non si tratta di leggi causali, ma di mere rilevazioni statistiche, prive di qualsiasi funzione esplicativa.

In concreto, qualora non sussistano interferenze causali esterne, sarà complicatissimo risalire ai singoli fattori causali che abbiano determinato quella precisa decisione algoritmica. Il *decision making* del sistema di i.a. è infatti frutto di una stratificazione di stimoli e di esperienze – del sistema stesso e, talvolta, di tutti gli altri *software* che sono interconnessi sul *cloud* – di difficile analisi retrospettiva, se non attraverso semplificazioni estreme, di difficile utilità nel processo penale¹⁸⁶. Anche qualora una ricostruzione *ex post* fosse tecnicamente possibile – magari proprio attraverso lo sviluppo di sistemi di i.a. di *reverse engineering*, in grado di ricostruire retrospettivamente il percorso eziologico¹⁸⁷ – dalla costante verifica empirica sarà difficile trarre una *regolarità causale*, proprio perché i fattori in base ai quali l'algoritmo di *deep learning* decide sono, di volta in volta, talmente tanti e vari da non poter essere ridotti ad una formula¹⁸⁸.

¹⁸⁵ F. STELLA, *Giustizia e modernità*, cit., p. 416.

¹⁸⁶ P.M. ASARO, *Determinism, machine agency, and responsibility*, in *Pol e soc.*, 2014, n. 2, p. 287; già C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, cit., p. 188 ss., prima ancora che fossero sviluppate le tecniche di *machine learning*, individuava nella complessità dei fattori che potevano incidere su un sistema di intelligenza artificiale un ostacolo all'accertamento processuale del nesso di causalità.

¹⁸⁷ Ovviamente, anche in questo caso, come per quanto concerne gli attuali EDR, l'algoritmo di ricostruzione causale *ex post* dovrebbe essere introdotto nel processo nel contraddittorio tra le parti.

¹⁸⁸ J.A. KROLL E AL., *Accountable Algorithms*, in 165 *U. Pa. L. Rev.*, 2017, p. 661: «*Approaches that attempt to review system failures simply by looking at how the output responds to changes in input are limited by either an inability to attribute a cause to those changes or an inability to interpret whether or why a change is significant*»; vd. anche D. LEHR – P. OHM, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, in 51 *University of California Davis Law Review*, 2017, p. 707, che notano come la logica del *ceteris paribus* sia totalmente estranea ai meccanismi decisionali dei sistemi di *machine learning*. Vd. anche D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1094,

In un certo senso, i procedimenti decisionali degli algoritmi che utilizzano reti neurali – se guardati dall'esterno – si avvicinano effettivamente ai meccanismi cognitivi dell'uomo. Si tratta ovviamente, come abbiamo sottolineato più volte, di una mera apparenza: l'algoritmo non ha coscienza di sé né di quello che sta facendo, non si autodetermina, ma reagisce ad una quantità enorme di stimoli, esterni ed interni al codice. Se guardate *da fuori*, tuttavia, le decisioni algoritmiche appaiono tanto imprevedibili quanto quelle umane. Perché, ad esempio, l'algoritmo di *image recognition* un giorno riconosce un segnale di stop e il giorno dopo non riconosce quel medesimo segnale, nonostante ci siano le stesse condizioni di luce, di visibilità, di traffico? Il sistema non è “impazzito”, ma la sua capacità di riconoscimento ha subito interferenze che – quantomeno ad oggi – non possono essere ricostruite con certezza.

Un esempio concreto può essere tratto dall'utilizzo di *ChatGPT*, un sistema di *natural language processing*, liberamente accessibile *online*¹⁸⁹, che, attraverso algoritmi di apprendimento automatico, genera risposte agli *input* scritti forniti dall'utente. Come esperimento, abbiamo fornito al sistema il medesimo *input* due volte, a distanza di un minuto, ottenendo come risultato due risposte diverse. L'*input* era il seguente: *scrivi un comunicato stampa per pubblicizzare la vendita di veicoli a guida autonoma in Italia, assicurando i potenziali acquirenti sulla sicurezza del prodotto*. Queste le due risposte:

Ecco un possibile comunicato stampa: "Siamo lieti di annunciare la vendita di veicoli a guida autonoma in Italia. La sicurezza dei nostri veicoli è la nostra massima priorità. Abbiamo sviluppato sistemi di guida autonoma altamente avanzati che utilizzano tecnologie di ultima generazione per garantire la massima sicurezza durante il viaggio. I nostri veicoli sono dotati di sensori e telecamere che forniscono una visione a 360 gradi del mondo circostante, permettendo loro di reagire rapidamente a qualsiasi situazione imprevista. Inoltre, abbiamo implementato rigorosi protocolli di manutenzione e ispezione per garantire che i nostri veicoli siano sempre in perfette condizioni. Siamo convinti che i nostri veicoli a guida autonoma rappresentino il futuro della mobilità e siamo entusiasti di offrirli ai nostri clienti in Italia. Siamo convinti che questa tecnologia possa contribuire a rendere le strade più sicure per tutti gli utenti. Non esitate a contattarci per ulteriori informazioni o per prenotare un test drive dei nostri veicoli a guida

che parlano di “*inscrutability*” con riferimento alla imperscrutabilità causale tipica degli algoritmi di *machine learning*.

¹⁸⁹ Vd. <https://openai.com/blog/chatgpt/>. L'accesso è libero previa creazione di un account.

autonoma. Siamo felici di rispondere a qualsiasi domanda o dubbio che possiate avere."

Comunicato Stampa. La nostra azienda è lieta di annunciare la vendita di veicoli a guida autonoma in Italia. Siamo convinti che questa tecnologia rappresenti il futuro del trasporto e siamo entusiasti di offrirla ai nostri clienti. I veicoli a guida autonoma sono stati sottoposti a rigorosi test di sicurezza e hanno dimostrato di essere estremamente affidabili. Grazie all'utilizzo di sensori e algoritmi avanzati, questi veicoli sono in grado di percepire e reagire in modo rapido e preciso a qualsiasi situazione di emergenza che possa verificarsi durante il viaggio. Inoltre, i nostri veicoli a guida autonoma sono dotati di una serie di sistemi di sicurezza aggiuntivi, come airbag e sistemi di frenata di emergenza, che proteggono i passeggeri in caso di incidente. Siamo fieri di offrire ai nostri clienti veicoli a guida autonoma di alta qualità e sicuri, che offrono un'esperienza di guida confortevole e senza stress. Siamo convinti che questi veicoli rappresentino il futuro del trasporto e siamo entusiasti di farne parte. Invitiamo i nostri clienti a visitare il nostro showroom per scoprire di più sui nostri veicoli a guida autonoma e per prenotare un test drive. Siamo sicuri che rimarrete stupiti dalla loro capacità e dalla loro sicurezza.

Il medesimo *input* – fornito due volte, a brevissima distanza l'una dall'altra, e dal medesimo dispositivo – ha generato *output* diversi¹⁹⁰: si tratta di una rappresentazione plastica di come il *decision making* possa reagire in maniera *imprevedibile* rispetto ad *input* analoghi.

Di qui la suggestione – che, allo stato delle conoscenze scientifiche e del dibattito giuridico sul tema, non può che rimanere tale – di un certo grado di analogia con la *causalità psichica*. Benché, infatti, siano reperibili delle generalizzazioni sulle modalità comportamentali umane, gli eventi psicologici mal si prestano ad essere descritti da leggi scientifiche di copertura, poiché sono irripetibili – essendo legati ad esperienze cognitive strettamente personali – e non rispondono a regolarità¹⁹¹. Allo stesso modo, si direbbe,

¹⁹⁰ L'esperimento è stato poi effettuato a più riprese, a distanza di vari intervalli di tempo. Il risultato ottenuto è sempre stato differente.

¹⁹¹ O. DI GIOVINE, *Lo statuto epistemologico della causalità penale*, cit., p. 679; M. RONCO, *Le interazioni psichiche nel diritto penale: in particolare sul concorso psichico*, in *Ind. pen.*, 2004, n. 3, p. 834 ss.; L. CORNACCHIA, *Il problema della c.d. causalità psichica rispetto ai condizionamenti mentali*, in S. Canestrari – G. Fornasari (a cura di), *Nuove esigenze di tutela nell'ambito dei reati contro la persona*, Clueb, 2001, p. 200 ss.; L. RISICATO, *La causalità psichica tra determinazione e partecipazione*, Giappichelli, 2007, p. 7 ss.; F. CINGARI, *Causalità psichica e massime di esperienza: un modello differenziato di causalità?*, in *Dir. pen. proc.*, 2009, fasc. 6, p. 770 ss. In realtà, la dottrina dominante è dell'idea che anche per dimostrare il nesso eziologico tra comportamenti umani ed eventi psichici sia necessario fare ricorso al paradigma

ogni algoritmo vive di un'interazione con il contesto ambientale, nonché, eventualmente, con altri soggetti intelligenti, che non è riducibile esclusivamente al codice, e, dunque, allo sviluppatore – tanto che, come abbiamo notato nel precedente capitolo, il difetto di *design*, con riferimento ai sistemi di i.a., cessa di essere un vizio necessariamente seriale, potendo coinvolgere anche soltanto alcuni esemplari della serie (vd. *supra*, cap. II, § 3.3.4). Se nella causalità psichica, d'altra parte, l'evento non è rappresentato da un accadimento naturalistico, bensì da una decisione altrui¹⁹², lo stesso, *mutatis mutandis*, può dirsi in relazione alla decisione algoritmica, che è frutto dell'esperienza e di un'enorme quantità di stimoli interni ed esterni, tra i quali sarà difficilissimo individuare la *condicio sine qua non*. La teoria che considera la causalità psichica come irriducibile al paradigma nomologico-deduttivo, d'altra parte, non contesta che le azioni umane abbiano una causa, ma esclude, piuttosto, che la causa possa essere descritta attraverso una legge di carattere generale¹⁹³.

Il criticato orientamento giurisprudenziale che, con riferimento ai prodotti tradizionali “non intelligenti”, faceva utilizzo del c.d. accertamento della causalità *black box* è stato oggetto di diverse critiche in dottrina¹⁹⁴, che ne hanno sottolineato la non conformità con il paradigma nomologico-deduttivo. In quei casi, il meccanismo della “scatola nera” impediva alla comunità di esperti di provare al di là di ogni ragionevole dubbio il nesso di causalità tra il fatto A e l'evento B. Sebbene problematico nel contesto dell'accertamento della responsabilità penale, quello della *black box* tradizionale è un fenomeno che, con l'avanzare del tempo e l'acquisizione di nuove conoscenze scientifiche, può essere minimizzato, prodotto dopo prodotto. Tutti i prodotti tecnologicamente avanzati, infatti, quando inizialmente introdotti sul mercato, sono *black box systems*: l'osservazione sistematica degli effetti che essi possono produrre può

nomologico-deduttivo, vd. per tutti F. STELLA, *Leggi scientifiche*, cit., p. 102 ss.; S. SEMINARA, *Riflessioni sulla condotta istigatoria come forma di partecipazione al reato*, in *Riv. it. dir. proc. pen.*, 1983, p. 1123 ss.

¹⁹² F. D'ALESSANDRO, *Spiegazione causale mediante leggi scientifiche, a dieci anni dalla sentenza Franzese*, in *Criminalia*, 2012, p. 353.

¹⁹³ In questi termini, M. RONCO, *Le interazioni psichiche nel diritto penale: in particolare sul concorso psichico*, cit., p. 835.

¹⁹⁴ F. STELLA, *Giustizia e modernità*, cit., p. 233 ss.; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 199 ss.; L. STORTONI, *Angoscia tecnologica ed esorcismo penale*, in *Riv. it. dir. proc. pen.*, 2004, fasc. 1, p. 76-77; C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1240; C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 22.

tuttavia portare a spiegarne le inferenze causali. Così, inizialmente può esserci incertezza circa gli effetti nocivi di una sostanza contenuta in un prodotto. In assenza di una legge scientifica di copertura – che potrebbe avere anche probabilità statistica *bassa*, come insegnano le Sezioni Unite nella sentenza Franzese¹⁹⁵, qualora sussista un alto tasso di *probabilità logica* – il giudice dovrebbe risolversi per l’assoluzione dell’imputato, nel rispetto del principio dell’*in dubio pro reo*¹⁹⁶. Attraverso lo studio sistematico degli effetti della sostanza sui consumatori, gli scienziati potrebbero tuttavia giungere alla formulazione di una legge scientifica (statistica) che ricollegghi l’esalazione della sostanza al verificarsi degli eventi lesivi.

Più ancora di quanto ha fatto, fino ad ora, la casistica in materia di responsabilità per danno da prodotto – così come quella sulla responsabilità per esposizione a sostanze tossiche e per danno ambientale –, gli eventi lesivi derivanti da sistemi di i.a. rischiano di mettere in crisi il modello tradizionale del diritto penale, pensato per rispondere ad episodi dannosi *singoli*, non a fenomeni sistemici e caratterizzati da trame causali imperscrutabili ed imprevedibili. A differenza della fenomenologia che ormai potremmo definire “classica” della società del rischio – che, seppur “classica”, è ben lungi dall’aver trovato una collocazione dogmatica sufficientemente condivisa – il problema dell’accertamento del nesso causale con riferimento ai danni da algoritmo sconta la discrepanza tra *ragionamento causale* tipico del modello di imputazione penalistico (quantomeno sul piano dell’accertamento della causalità *reale*¹⁹⁷) e *decision-making* di tipo probabilistico, che invece attiene alla logica del *deep learning*.

L’alternativa ci pare, allo stato attuale delle conoscenze scientifiche, netta – e ci riporta a quel dilemma, a cui abbiamo accennato all’inizio di questo capitolo (vd. *infra*, § 1), già emerso in relazione a tutte le fenomenologie tipiche del c.d. “diritto penale del rischio”: o si prende atto dell’inconciliabilità tra accertamento causale e approccio probabilistico dei sistemi di *machine learning* – rinunciando, così, alla possibilità di muovere un rimprovero al produttore per la commissione di *reati di evento* – oppure si piega la categoria della causalità alle contingenti esigenze di accertamento processuale,

¹⁹⁵ Cass., sez. un., 11 luglio 2022, n. 30328, Franzese.

¹⁹⁶ C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 254.

¹⁹⁷ Sulla natura stocastica della causalità omissiva vd., per tutti, C.E. PALIERO, *La causalità dell’omissione*, cit., p. 821 ss.

in potenziale collisione, tuttavia, con il principio di legalità¹⁹⁸. A nostro parere, il rigoroso rispetto dei principi costituzionali impone senz'altro di propendere per la prima soluzione.

Ciò non toglie che, in futuro, l'acquisizione di ulteriori dati scientifici sul funzionamento del *decision making* algoritmico – così come il costante incremento delle capacità tecniche e computazionali dei sistemi di i.a. – potrebbe suggerire un ripensamento del paradigma nomologico-deduttivo, in questo settore. Qualora, infatti, si perfezionassero le tecniche di *explainability* e *interpretability* alle quali abbiamo fatto cenno nel primo capitolo (vd. *supra*, cap. I, § 5.3), potrebbe giungersi alla ricostruzione *ex post* del nesso di causalità tra evento lesivo e specifico *input*, senza che, tuttavia, possa addivenirsi ad una generalizzazione scientifica del rapporto tra *input* e *output*, impedita dai processi stocastici alla base delle decisioni algoritmiche.

Ecco che, a nostro parere, in quest'eventualità, senza tornare ad un modello di accertamento della causalità basato sulla mera intuizione del giudice¹⁹⁹ – e senza sfociare in un acritico cedimento ad istanze “general-preventive” e simboliche –, potrebbe emergere la necessità, in questo ristretto ambito, di allontanarsi dal modello della sussunzione sotto leggi scientifiche pur nella cornice del rispetto dei principi costituzionali di legalità, colpevolezza, offensività.

D'altra parte, la determinazione del nesso causale, nel processo penale, non risponde al problema tipico delle scienze della natura, ossia quello di formulare leggi generali che descrivano la realtà e – quanto alle scienze di tipo pratico e applicativo, quali ad esempio la medicina – quello di elaborare modelli prognostici per rispondere adeguatamente a determinati avvenimenti; piuttosto, lo scopo dell'accertamento causale a fini penalistici è quello ricostruire *ex post* un evento singolare e particolare²⁰⁰. Qualora fosse possibile risalire a tale evento *al di là di ogni ragionevole dubbio*, attraverso i citati

¹⁹⁸ L'altra possibilità, ovviamente, resta quella di vietare la commercializzazione dei sistemi di i.a. dei quali non sia possibile ricostruire, in via generale, i nessi causa-effetto. Non sembra, tuttavia, che il legislatore europeo si stia muovendo in questa direzione, dal momento che l'*AI Act* – pur richiedendo il rispetto di alcuni, vaghi, requisiti di trasparenza – consente la messa in commercio di sistemi di *deep learning*, per loro natura caratterizzati da “inscrutabilità”.

¹⁹⁹ Su cui vd. in senso critico C.E. PALIERO, *Le fattispecie causalmente orientate*, cit., p. 1501; G. UBERTIS, *Sistema di procedura penale*, vol. 1, *Principi generali*, Utet, 2004, p. 56 ss.

²⁰⁰ M. RONCO, *Aporie scientiste e certezza logico razionale: note in tema di nesso causale*, cit., p. 248, al quale si rinvia per gli opportuni riferimenti bibliografici sulla distinzione tra accertamento causale generale, tipico delle scienze naturali, e accertamento individuale, tipico del diritto.

strumenti computazionali di ricostruzione *ex post* della causalità individuale, ci pare che potrebbe ritenersi provata l'esistenza di una *condicio sine qua non*, ai sensi dell'art. 40, cpv., c.p.

La sentenza Franzese, nel 2002²⁰¹, ha segnato un primo passo nella direzione della valorizzazione del metodo induttivo, stabilendo che anche le leggi di copertura aventi probabilità statistica medio-bassa possano essere utilizzate in sede di accertamento del nesso di causalità, purché dal compendio probatorio acquisito nel corso del processo emerga una spiegazione causale convincente e «la sicura non incidenza nel caso di specie di altri fattori interagenti in via alternativa».

Alcune voci in dottrina, d'altra parte, da tempo mostrano un'apertura alla valorizzazione di modelli induttivi della spiegazione causale, anche in contesti nei quali siano *inesistenti* delle generalizzazioni scientifiche²⁰²: tale posizione, avanzata, in passato, in relazione all'insorgere di nuovi problemi causali, può rappresentare un'interessante prospettiva, *a fortiori*, nel settore dell'intelligenza artificiale, in cui si potrebbe assistere ad un'ontologica impossibilità di individuare *regolarità causali* nel funzionamento degli algoritmi.

Anche quando non si disponga di preesistenti generalizzazioni *ad hoc*, l'indagine causale potrebbe condurre a risultati di *alta credibilità razionale*, quando i fatti siano rigorosamente ricostruiti e, considerati nell'insieme, possano condurre ad una spiegazione causale convincente, nonché ad escludere la possibilità di decorsi causali alternativi. Nonostante l'assenza di leggi di copertura, la scienza dovrebbe in ogni caso essere protagonista nelle aule di giustizia²⁰³, poiché soltanto la conoscenza del

²⁰¹ Cass., sez. un., 11 luglio 2022, n. 30328, Franzese.

²⁰² M. RONCO, *Aporie scientiste e certezza logico razionale: note in tema di nesso causale*, cit., *passim*; R. BLAIOTTA, *Incidenti da aviolancio, problemi causali nuovi, attività rischiose*, cit., p. 766-767; ID., *Causalità giuridica*, cit., p. 397: «sebbene non accada frequentemente, un'indagine di tipo induttivo, alimentata dalle peculiarità particolarmente caratteristiche e chiare dei dati empirici di una determinata vicenda, può in alcuni casi condurre alla razionale spiegazione dell'evento; anche se non si dispone di informazioni scientifiche *ad hoc*»; *contra* R. BARTOLI, *Il problema della causalità penale. Dai modelli unitari al modello differenziato*, Giappichelli, 2010, p. 44-45, che sostiene che la spiegazione causale di un fatto non possa essere fornita in assenza di un sapere nomologico-deduttivo; critico anche A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, cit., p. 413, che coglie una tendenza "anti-scientista" in queste tesi, così come in alcuni sviluppi giurisprudenziali del metodo Franzese.

²⁰³ R. BLAIOTTA, *Incidenti da aviolancio, problemi causali nuovi, attività rischiose*, cit., p. 767; M. RONCO, *Aporie scientiste e certezza logico razionale: note in tema di nesso causale*, cit., p. 252. Diversa questione, ma non meno rilevante, resta quella relativa alla selezione del sapere scientifico in sede processuale, così

funzionamento dei sistemi di *machine learning* – unita alla capacità tecnica di interpretare i dati informatici – può aiutare a delineare le inferenze causali sussistenti nel caso di specie.

È ovvio, in ogni caso, che in assenza di paradigmi nomologici consolidati, l'indagine sulla c.d. *causalità individuale* dovrà essere ancora più rigorosamente basata su solidi elementi probatori.

4.4. L'individuazione del soggetto personalmente responsabile: il c.d. *many hands problem*.

Quando, nel capitolo precedente, abbiamo fatto riferimento alla responsabilità del “produttore” per il danno derivante da un prodotto difettoso (vd. *supra*, cap. II, § 3. ss.), ci riferivamo, ovviamente, non soltanto al produttore-persona fisica – figura, nella società di massa e globalizzata in cui viviamo, ormai sostanzialmente scomparsa – ma soprattutto al produttore-ente, quale centro di imputazione privilegiato (e autonomo) per il risarcimento dei danni derivanti dal prodotto. È scontato sottolineare, tuttavia, come la Costituzione, all'art. 27, co. 1, imponga, ai fini della responsabilità penale, l'individuazione di una persona fisica che sia personalmente rimproverabile all'interno della struttura societaria.

L'identificazione della persona fisica a cui imputare la verifica dell'evento lesivo, all'interno delle complesse strutture societarie, ha sempre costituito un problema

da discernere tra “scienza attendibile” e “junk science”; in materia vd., per tutti, F. CENTONZE, *Scienza “spazzatura” e scienza “corrotta” nelle attestazioni e valutazioni dei consulenti tecnici nel processo penale*, in *Riv. it. dir. proc. pen.*, 2001, fasc. 4, p. 1232 ss. Sullo specifico tema della valutazione di affidabilità di tesi scientifiche contrastanti vd. la sentenza Cozzini (Cass., sez. IV, 17 settembre 2010, n. 43786, Cozzini e altri, in *Dir. pen. cont.*, 11 ottobre 2012, con nota di S. ZIRULIA), che ha fornito una griglia di criteri, sia di ordine oggettivo (relativi alle caratteristiche dello studio scientifico) sia soggettivo (attinenti all'autorevolezza e imparzialità del perito o consulente tecnico) per selezionare il sapere scientifico rilevante. Sulle questioni ancora aperte nella giurisprudenza *post Cozzini*, nello specifico settore degli eventi lesivi derivanti dall'esposizione all'amianto, vd. S. ZIRULIA, *Contrasti reali e contrasti apparenti nella giurisprudenza post-Cozzini su causalità e amianto. Riflessioni per un rinnovato dibattito sul c.d. effetto acceleratore nei casi di morte per mesotelioma*, in *Riv. it. dir. proc. pen.*, 2019, fasc. 3, p. 1289 ss.; vd. anche M. ROMANO – F. D'ALESSANDRO, *Nesso causale ed esposizione ad amianto*, cit., p. 1133 ss.

nell'ambito della responsabilità penale per danno da prodotto²⁰⁴ e, in generale, nell'ambito della criminalità d'impresa²⁰⁵. La progettazione, la fabbricazione e la commercializzazione di un prodotto sono infatti attività che coinvolgono una pluralità di centri decisionali, aventi spesso carattere collegiale: nella realtà delle imprese, nessun individuo governa da solo il processo decisionale. La ricostruzione dell'evento tipico, inoltre, è spesso resa complessa dal frazionamento delle funzioni decisionali, che sono articolate in diversi centri gestionali ed organizzativi, spesso oltretutto dotati di facoltà di delega e sub-delega²⁰⁶: è quello che la dottrina angloamericana chiama “*many hands problem*”²⁰⁷.

Com'è stato osservato, d'altra parte, pur essendo tipicamente *plurisoggettiva*, la responsabilità del produttore si allontana dalle classiche categorie di *autorità* e *partecipazione* che costituiscono il fulcro del concorso di persone nel reato²⁰⁸, risolvendosi piuttosto in forme di *esecuzione frazionata* – in cui spesso, oltretutto,

²⁰⁴ C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1241-1242; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 303 ss.; ID., *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 370; A. BERNARDI, *La responsabilità da prodotto nel sistema italiano*, cit., p. 25; A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, cit., pp. 464 ss.; F. CENTONZE, *La normalità dei disastri tecnologici*, cit., p. 267 ss.; nello specifico settore della responsabilità per danno da prodotto alimentare vd. L. CORNACCHIA, *Posizioni di garanzia nell'ambito della sicurezza agroalimentare*, in L. Foffani – A. Doval Pais – D. Castronuovo (a cura di), *La sicurezza agroalimentare nella prospettiva europea*, Giuffrè, 2014, p. 457 ss. e, più di recente, V. MONGILLO, (voce) *Industria alimentare e responsabilità colposa*, in M. Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, cit., p. 704 ss.

²⁰⁵ A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), *Manuale di diritto penale dell'impresa*, Monduzzi, II ed., 2000, p. 59; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 515 ss.; V. ATTILI, *L'agente-modello “nell'era della complessità”: tramonto, eclissi o trasfigurazione?*, in *Riv. it. dir. e proc. pen.*, 2006, fasc. 4, p. 1247 ss.

²⁰⁶ G. DE VERO, *Il nesso causale e il diritto penale del rischio*, in *Riv. it. dir. e proc. pen.*, 2016, p. 695 ss. Sull'istituto della delega nel diritto penale vd., per tutti, G. FIANDACA, *Il reato commissivo mediante omissione*, cit., p. 200 ss.; A. FIORELLA, *Il trasferimento di funzioni nel diritto penale dell'impresa*, Nardini, 1985, p. 112 ss.

²⁰⁷ La prima tematizzazione del problema si deve, nella letteratura anglosassone, a D. THOMPSON, *The moral responsibility of public officials: the problem of many hands*, in *74 American Political Science Review*, 1980, p. 905 ss.; il dibattito è stato poi portato avanti soprattutto attraverso i lavori di K.M. MCGRAW, *Avoiding blame: an experimental investigation of political excuses and justifications*, in *20 British Journal of Political Science*, 1990, p. 119 ss.; M. BOVENS, *The quest for responsibility: accountability and citizenship in complex organisations*, Cambridge University Press, 1998; R. MASTOP, *Characterising responsibility in organisational structures: The problem of many hands*, in G. Governatore – G. Sartor (eds.), *Deontic logic in computer science*, Springer, 2010, p. 274 ss.

²⁰⁸ C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1241-1242.

l'azione tipica è realizzata «più per sovrapposizione autonoma di decisioni che per comune concerto»²⁰⁹.

L'agente modello che davvero rileva, nella definizione della diligenza esigibile, è di fatto quello dell'organizzazione, non quello del singolo individuo. È infatti la *persona giuridica* il destinatario naturale delle norme cautelari sulla gestione del rischio, l'unico soggetto che ha la capacità organizzativa ed economica per riconoscere i rischi e per minimizzarli, adeguandosi tempestivamente alle più recenti conoscenze scientifiche²¹⁰. La discrasia tra fenomeno criminale e struttura costituzionale dell'illecito rischia di rendere macchinosa l'attribuzione della responsabilità penale per gli eventi lesivi derivanti da scelte aziendali, con il conseguente pericolo che la ricerca del soggetto rimproverabile all'interno delle società si trasformi nella ricerca di un capro espiatorio²¹¹.

Ci limitiamo a sottolineare, in questa sede, che per un modello di criminalità sostanzialmente *collettiva* – com'è quello della responsabilità per danno da prodotto – avrebbe forse potuto essere opportuna l'introduzione di una responsabilità amministrativa dell'ente, così da poter colpire il centro immediato degli interessi²¹². Nel catalogo dei reati-presupposto previsto dal d.lgs. 231/2001, tuttavia, non si fa cenno alle fattispecie di pericolo o di danno relative alla sicurezza dei prodotti: la legge n. 123 del 2007 (poi modificata dal d.lgs. 81/2008) ha introdotto, tra i reati-presupposto, le fattispecie di omicidio colposo e lesioni colpose commesse con violazione delle norme sulla tutela della salute e della sicurezza del lavoro (art. 25-septies d. lgs. 231/2001), ma non ha richiamato le ipotesi omicidio o lesioni colpose derivanti da difetti di produzione.

Qualora in futuro, come auspicabile, le suddette fattispecie dovessero rientrare nel campo di applicazione del decreto 231, l'eventuale mancata identificazione della persona responsabile all'interno dell'organigramma aziendale non impedirebbe la configurazione di una responsabilità in capo all'ente. In particolare, il già citato art. 8 d.lgs. 231/2001 – nello stabilire che «*la responsabilità dell'ente sussiste anche quando l'autore del reato non è stato identificato o non è imputabile*» (art. 8, co. 1, lett. a) – delinea un terzo modello

²⁰⁹ C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1242.

²¹⁰ G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza*, in *Riv. it. dir. proc. pen.*, 2005, fasc. 1, p. 56; vd. anche G. DE VERO, *Il nesso causale e il diritto penale del rischio*, cit., p. 695 ss.

²¹¹ Su tale rischio vd. G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 57

²¹² In questo senso A. BERNARDI, *La responsabilità da prodotto nel sistema italiano*, cit., p. 28-29.

imputativo, ulteriore rispetto a quello previsto agli artt. 6 e 7 del decreto (rispettivamente, per i reati commessi da apicali e per i reati commessi da sottoposti), basato sull'omessa auto-organizzazione da parte dell'ente o su un inefficace assetto organizzativo nell'ambito delle singole procedure di contenimento del rischio-reato²¹³. In ogni caso, l'addebito mosso all'ente dovrà pur sempre essere conforme al principio di colpevolezza: bisognerà dunque evitare che venga configurata una responsabilità in capo all'ente per il solo fatto che non sia stato possibile identificare l'autore materiale del reato²¹⁴, dovendosi sempre accertare la c.d. *colpa d'organizzazione*, ossia la mancata predisposizione di «un insieme di accorgimenti preventivi idonei ad evitare la commissione di reati del tipo di quello realizzato»²¹⁵.

De jure condito, l'individuazione delle persone responsabili all'interno delle organizzazioni complesse costituisce uno degli aspetti più problematici della disciplina. Non è di certo questa la sede per ripercorrere le semplificazioni che, anche in questo ambito, si sono registrate in giurisprudenza²¹⁶. Basterà qui rilevare che, in diverse occasioni, si è assistito all'affermazione di un criterio ascrittivo della responsabilità derivato dalla ripartizione dei ruoli prestabilita dal diritto societario²¹⁷: una colpa “da posizione”, che mal si concilia, tra l'altro, con la rinnovata attenzione – frutto di una valorizzazione del principio di colpevolezza – per la misura soggettiva della colpa²¹⁸, da affiancarsi alla ormai consolidata concezione normativa. In particolare, al principio che

²¹³ Così C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1756.

²¹⁴ C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1756, proprio in relazione all'ipotetica applicazione del decreto 231 ai reati da prodotto (intelligente); nello stesso senso C. MINELLI, *La responsabilità “penale” tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 61.

²¹⁵ Cass., sez. VI, 18 febbraio 2010, n. 27735, Scarafia; vd. anche Cass., sez. un., 24 aprile 2014, n. 38343, Espenhahn; Cass., sez. IV, 8 gennaio 2021, 32899, Castaldo.

²¹⁶ Sul tema vd. C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 370; ID., *Danno da prodotto e responsabilità penale*, cit., p. 303 ss.; A. BERNARDI, *La responsabilità da prodotto nel sistema italiano*, cit., p. 25; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 533 ss.

²¹⁷ C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 361; ID., *Danno da prodotto e responsabilità penale*, cit., p. 234, con specifico riferimento al caso *Lederspray*; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., *passim*.

²¹⁸ Vd. per tutti M. DONINI, *Teoria del reato. Una introduzione*, cit., p. 348 ss.; D. CASTRONOVO, *La colpa penale*, cit.; ID., *La colpa “penale”. Misura soggettiva e colpa grave*, cit.; A. CANEPA, *L'imputazione soggettiva della colpa*, cit.; M. GROTTI, *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica*, cit.; A.R. DI LANDRO, *Dalle linee guida e dai protocolli all'individualizzazione della colpa penale nel settore sanitario*, cit.

collega una posizione di garanzia in capo all'amministratore del ramo d'azienda coinvolto nell'evento lesivo – partendo, dunque, logicamente, *dalla violazione realizzatasi*, per giungere *alla funzione* nell'ambito della quale essa è maturata, per poi, infine, individuare il soggetto personalmente deputato a governare la funzione stessa, sulla base dei poteri effettivamente attribuitigli²¹⁹ –, la giurisprudenza in materia di danno da prodotto è spesso andata sostituendo la presunta *competenza totale* del *management* della società per gli eventi derivanti dal prodotto²²⁰.

Nell'ultimo decennio, d'altra parte, si è assistito spesso, in giurisprudenza, alla ricostruzione dell'obbligo di controllo sulla base della predeterminazione delle aree di competenza nella gestione del rischio²²¹ – una predeterminazione, tuttavia, che non tiene conto del carattere relazionale e interattivo della gestione d'impresa, in cui i confini tra le sfere di competenza non sono così netti e vengono spesso a sovrapporsi²²². Il rischio evidente – acutamente sottolineato dalla dottrina più attenta – è che si valorizzino le delimitazioni formalistiche delle aree di competenza, a scapito della sostanziale attribuzione di poteri di gestione del rischio²²³. L'equiparazione tra “posizione di garanzia” e “sfera di responsabilità” risponde, ancora una volta, all'obiettivo pragmatico di semplificare l'ascrizione di responsabilità nelle organizzazioni complesse: a costo, tuttavia, di un'insanabile frizione con il principio di legalità²²⁴.

²¹⁹ Vd. per tutti C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 315.

²²⁰ In questo senso C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, cit., p. 361; C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1241-1242.

²²¹ L. CORNACCHIA, *Posizioni di garanzia nell'ambito della sicurezza agroalimentare*, cit., 460 ss.; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 533 ss., che fa riferimento soprattutto alla sentenza della Cassazione sul caso Thyssenkrupp (Cass., sez. un., 24 aprile 2014, n. 38343).

²²² A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 536-7.

²²³ A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 536-7.

²²⁴ A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 538-539.

4.4.1. La tipizzazione delle “aree di competenza” nella gestione del rischio ad opera dell’AI Act.

Il c.d. *many hands problem* potrebbe acuirsi nell’ambito delle società produttrici di sistemi di i.a.²²⁵. Il dispositivo artificiale può infatti essere un artefatto molto complesso, formato da diverse componenti, spesso prodotte da imprese differenti; il prodotto finale è frutto del lavoro di decine di persone, talvolta centinaia o migliaia, che fanno capo – in ruoli di dipendenza o di direzione – ad enti diversi. Addirittura, come abbiamo già evidenziato (vd. *supra*, cap. II, § 3.1.), può darsi che il prodotto intelligente sia dotato di un *software open source*, che può essere sviluppato e modificato da chiunque lo desideri²²⁶.

Come abbiamo già ribadito più volte, la *black box* tipica dei sistemi di *machine learning* rende già complessa l’individuazione del percorso eziologico all’origine del danno: così, ad esempio, nel caso di incidente che coinvolga una *self driving car*, non sarà agevole stabilire se la *condicio sine qua non* sia da identificarsi in un difetto inerente al *design* del modello algoritmico, o alla raccolta e selezione dei *dataset*, o al *training* dell’algoritmo, al funzionamento dei sensori, all’assemblaggio delle componenti, all’omissione di idonee misure di *cybersecurity* etc. Ancora, il fatto all’origine del danno potrebbe consistere in una condotta – commissiva o omissiva – dell’utilizzatore del veicolo o a un altro utente della strada.

Anche qualora fosse possibile, in futuro, individuare il fatto causativo dell’evento lesivo – ad esempio, attraverso il ricorso ai già citati sistemi di *reverse engineering* (vd. *supra*, cap. II, § 3.5. e, in questo cap., § 4.3.) – resterebbe in ogni caso il problema di individuare, all’interno degli organigrammi societari, la persona alla quale imputare tale

²²⁵ In argomento vd. B. PANATTONI, *Intelligenza artificiale*, cit., p. 338; S. BECK, *Intelligent agents and criminal law*, cit., p. 140; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1754; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 3; F.P. HUBBARD, *Sophisticated Robots: Balancing Liability, Regulation, and Innovation*, cit., p. 1853; ; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 104 ss.; S. AIRES DE SOUSA, “*Não fui eu, foi a máquina*”: teoria do crime, responsabilidade e inteligência artificial, cit., p. 70-71; L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 828-829, che delinea alcuni dei centri di responsabilità che potrebbero venire in rilievo in relazione alla produzione di *autonomous vehicles*.

²²⁶ R. ABBOTT – A. SARCH, *Punishing Artificial Intelligence*, cit., p. 334; F. CONSULICH, *Flash offenders*, cit., p. 1041.

evento. Il rischio, già evidenziato con riferimento a tutti i fenomeni della c.d. criminalità d'impresa, è che la responsabilità venga individuata sulla base della mera posizione che un soggetto riveste all'interno della società coinvolta²²⁷.

Pur serbando silenzio sui profili penalistici della disciplina, la bozza di *AI Act* pare comunque fornire un impulso alla tipizzazione di “aree di competenza” nella gestione del rischio – come, d'altronde, già avvenuto in altri settori produttivi²²⁸. In particolare, l'*AI Act* individua una serie di “sfere di responsabilità”²²⁹ in capo al fornitore (artt. 16-23); al fabbricante (art. 24); ai rappresentanti autorizzati dei fornitori non aventi sede nell'Unione, nel caso in cui non sia possibile identificare un importatore (art. 25); agli importatori di sistemi sviluppati fuori dall'UE (art. 26); ai distributori (art. 27); agli utenti (art. 29)²³⁰.

Perno attorno al quale ruota il sistema della gestione del rischio è senza dubbio il fornitore – definito come «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o che fa sviluppare un sistema di IA al fine di immetterlo sul mercato o metterlo in servizio con il proprio nome o marchio, a titolo oneroso o gratuito» (art. 3, n. 2) – sotto la cui responsabilità sono espletate la gran parte delle attività di valutazione della sicurezza del prodotto (vd. *supra*, cap. II, § 3.4.2), tra cui anche gli obblighi di monitoraggio *post market* (vd. artt. 61 e 62 *AI Act*). In particolare, al fornitore sono affidati, tra gli altri, l'obbligo di verificare la conformità dei sistemi di i.a. ai requisiti di sicurezza di cui al capo 2 *AI Act* (art. 16, co. 1, lett. a), l'obbligo di disporre il sistema di gestione della qualità di cui all'art. 17 *AI Act* (art. 16, co. 1, lett. b), l'obbligo di redigere la documentazione tecnica per i sistemi ad alto rischio (art. 16, co. 1, lett. c), l'obbligo di conservare i *logs* generati automaticamente in caso di incidenti (art. 16, co. 1, lett. d), etc.

²²⁷ F.C. LA VATTIATA, *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, cit., p. 705.

²²⁸ Vd., nel settore dei prodotti agroalimentari, V. MONGILLO, (voce) *Industria alimentare e responsabilità colposa*, cit., p. 710.

²²⁹ Li definisce veri e propri “obblighi di controllo” F.C. LA VATTIATA, *Brevi note “a caldo” sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Dir. pen. uomo*, 30 giugno 2021, p. 13. A parere di quest'ultimo, l'entrata in vigore del regolamento – immediatamente applicabile negli ordinamenti interni – potrebbe dunque determinare l'immediato sorgere di obblighi di impedimento ai sensi dell'art. 40, co. 2, c.p.

²³⁰ Per le definizioni si rinvia all'art. 3 dell'*AI Act*.

Ulteriore perno è costituito dal fabbricante, che, ai sensi dell'art. 24 *AI Act*, «si assume la responsabilità della conformità del sistema di IA al presente regolamento e ha, per quanto riguarda il sistema di IA, gli stessi obblighi imposti dal presente regolamento al fornitore».

Il riferimento alla “persona giuridica” nella definizione di “fornitore” – così come in quella di “rappresentante”, “distributore”, “utente” etc. – è dovuta all’ovvia circostanza che la normativa non è congegnata per soddisfare esigenze di personalizzazione proprie del diritto penale, ma si pone invece obiettivi generali di *risk governance*. Anche qualora l’area di gestione del rischio, all’interno della singola società, sia attribuita ad un organo collegiale, ai fini penalistici sarà ovviamente necessario individuare il soggetto personalmente responsabile.

La tendenza, anche nel settore *de qua*, pare dunque quella di predeterminare le aree di competenza per la gestione del rischio, svincolate dall’individuazione di precisi obblighi di controllo ai sensi dell’art. 40, cpv., c.p. A nostro parere, nella concreta identificazione delle posizioni giuridiche, i principi di legalità e di colpevolezza dovrebbero comunque imporre l’identificazione del soggetto personalmente incaricato dell’obbligo di impedire l’evento, senza che il vaglio giurisprudenziale possa limitarsi alla qualifica formale²³¹.

4.5. La colpa del produttore.

Se anche si riuscisse a dimostrare la sussistenza di un fatto tipico causalmente ascrivibile alle persone “dietro la macchina”, resterebbe il problema della prova dell’elemento soggettivo. La problematicità dell’attribuzione della responsabilità penale per il danno derivante da prodotti intelligenti, infatti, si coglie a pieno soprattutto sul piano dell’accertamento della colpa²³².

²³¹ Per problematiche simili nell’interpretazione del reg. CE 178/2002 in materia di sicurezza alimentare vd. V. MONGILLO, (voce) *Industria alimentare e responsabilità colposa*, cit., p. 710.

²³² In questa sede ci occuperemo del rimprovero colposo, trattandosi della forma di colpevolezza tipica del fenomeno del danno da prodotto. In giurisprudenza, tuttavia, l’imputazione oscilla tra colpa cosciente e dolo eventuale; vd. C. PIERGALLINI, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, cit., p. 1127-1128; ID., *Danno da prodotto e responsabilità penale*, cit., p. 360 ss.

Il *deficit* di determinatezza dell'imputazione colposa è cosa ben nota, e non è questa la sede per soffermarsi²³³. Basterà qui rilevare che, nell'ambito della responsabilità per danno da prodotto, tale indeterminatezza è spesso sfociata, per via giurisprudenziale, in un'estrema flessibilizzazione della categoria colposa, con progressivo affievolimento della rigerosità che il principio di legalità richiederebbe per la dimostrazione delle regole cautelari violate, dell'oggettiva violazione di tali regole, dell'esigibilità dell'osservanza di tali regole da parte del produttore²³⁴.

Sullo sfondo di tale flessibilizzazione giurisprudenziale pare stagliarsi, com'è stato osservato²³⁵, il perenne conflitto – che trova sul terreno della colpa penale un'arena ideale – tra l'aspirazione solidaristica ad istanze di tutela di beni giuridici fondamentali, tipicamente fatta propria dalla giurisprudenza, e il marcato garantismo figlio invece della tradizionale concezione liberale del diritto penale.

Lo sviluppo e la commercializzazione di sistemi di i.a. – caratterizzati da un'imprevedibilità genericamente prevedibile²³⁶ – parrebbe mettere in discussione la stessa possibilità di muovere un rimprovero colposo al produttore, se si considera che, secondo il criterio di copertura del rischio tipico, affinché l'evento lesivo sia rimproverabile all'agente, esso deve costituire realizzazione *tipica* del rischio che la norma cautelare mirava a prevenire. La già evidenziata crisi del modello nomologico (vd. *supra*, § 4.3., in questo cap.), determinata da una tendenziale inesplicabilità dei rapporti causa-effetto che governano l'agire algoritmico, si ripercuote, a cascata, sull'intera struttura del reato e, in particolare, sull'accertamento della colpa in capo al produttore²³⁷.

²³³ Si vd., per tutti, P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 50 ss.; R. BARTOLI, (voce) *Fonti della colpa*, cit., p. 536 ss.

²³⁴ A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, in *Riv. trim. dir. pen. ec.*, p. 30.

²³⁵ A. PERIN, *La crisi del "modello nomologico" fra spiegazione e prevedibilità dell'evento nel diritto penale*, cit., p. 1387; F. STELLA, *Giustizia e modernità*, cit., *passim*; cfr. L. STORTONI, *Angoscia tecnologica ed esorcismo penale*, cit., p. 80, che ritiene che l'obiettivo solidaristico sia in ogni caso destinato a fallire, data l'intrinseca inettitudine del diritto penale a rispondere ad istanze di tutela nei confronti di fenomeni diffusivi e sistemici, tipici della società del rischio.

²³⁶ C. PIERGALLINI, *Intelligenza artificiale*, 1762; vd. anche M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1208-1209; S. BECK, *Intelligent agents and criminal law*, cit., p. 139; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 7: «the behavior of an autonomous agent is predictably unpredictable»; sostanzialmente anche C. IAGNEMMA, *I 'robot medici': profili problematici in tema di alleanza terapeutica e di responsabilità penale*, cit., p. 449.

²³⁷ Sul problema della causalità come "origine" di tutte le altre difficoltà applicative del diritto penale nella materia del danno da prodotto vd. C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., *passim*.

Le regole cautelari, sia nella forma scritta che in quella non scritta, sono infatti regole di esperienza pre-date (ossia, già esistenti al momento della condotta), tratte dal consolidamento di giudizi *causali* di riconoscibilità di determinati rischi e di prevedibilità ed evitabilità di determinati eventi²³⁸. Se uno spazio per la conformazione di regole cautelari è certamente individuabile anche in contesti di incertezza scientifica – nei casi in cui le cautele esistenti si dimostrino obsolete o comunque inadeguate –, l'intrinseca difficoltà di identificare generalizzazioni causali in relazione al comportamento dei sistemi di i.a. rischia di rendere strutturale questa condizione di instabilità delle cautele doverose.

Pur consapevoli della carente caratterizzazione nomologica che le norme cautelari hanno nel settore dell'intelligenza artificiale, nei paragrafi che seguono tenteremo di fare luce sulle cautele (scritte e non scritte) esistenti, volte a minimizzare – e non di certo a neutralizzare – il rischio di eventi algoritmici lesivi (c.d. *cautele improprie*, vd. *infra*, § 4.5.2.1). Come si evincerà nel prosieguo, l'unica efficacia ascrivibile alle cautele, nella produzione di sistemi di i.a., è infatti quella di riduzione del rischio, attraverso l'implementazione di una serie di misure che – in vario modo – possono influire sull'esito del *decision making* algoritmico.

4.5.1. Le regole cautelari scritte: la colpa specifica.

Innanzitutto, ai fini dell'accertamento della colpa in capo all'agente, è necessario individuare la *regola cautelare preesistente che si assume violata*²³⁹. Come già da tempo avviene nel settore della sicurezza sul lavoro, anche con riferimento alla sicurezza dei prodotti e alla tutela della salute dei consumatori si registra un imponente fenomeno di “iperregolazione multilivello”²⁴⁰, attraverso la formalizzazione di regole preventive,

²³⁸ Così, per tutti, D. CASTRONUOVO, (voce) *Colpa penale*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 219.

²³⁹ Sulla necessaria preesistenza della regola cautelare si vd. per tutti, G. FORTI, *Colpa ed evento*, cit., p. 683.

²⁴⁰ L'espressione è di V. MONGILLO, (voce) *Industria alimentare e responsabilità colposa*, cit., p. 713, che si riferisce tuttavia alla sola normativa in materia di sicurezza dei prodotti alimentari.

combinata a forme di "procedimentalizzazione" o "proceduralizzazione" dei processi produttivi e di commercializzazione²⁴¹.

Nel settore della sicurezza dei prodotti, in particolare, si rinvencono due categorie di normazione: da un lato, un modello di *etero-disciplina* che si estrinseca mediante forme di *etero-regolazione* (leggi, regolamenti, *standard* di sicurezza – interni, europei e di natura privata²⁴² –, limiti tabellari, valori soglia), così come di *etero-controllo* da parte di autorità amministrative (sistema di autorizzazioni e controlli, divieti, prescrizioni etc., si vd. ad es. l'art. 105 cod. cons.); dall'altro, un progressivo incremento di paradigmi di *autodisciplina*, fondati sulla tendenza verso una parziale privatizzazione dei compiti legati alla sicurezza, con compiti di controllo rimessi, almeno in parte, in capo agli stessi operatori economici coinvolti nello svolgimento di attività rischiose (c.d. *autonormazione cautelare*)²⁴³. Entrambe le forme di regolazione si caratterizzano per la complessità dei rinvii (espresi o impliciti) e per l'elevatissimo gradiente tecnico-specialistico. Il fenomeno del decentramento dei luoghi di produzione normativa trova proprio in relazione al reato colposo il suo terreno d'elezione, trattandosi di fattispecie "aperta" e porosa per definizione²⁴⁴, che, in un contesto di governo globalizzato del rischio tecnologico, si trova ad interagire con una molteplicità di fonti²⁴⁵.

Come abbiamo accennato, le regole cautelari scritte svolgono un ruolo fondamentale nell'indirizzare la condotta dei consociati, poiché esse, oltre ad avere il

²⁴¹ D. CASTRONUOVO, *La colpa penale*, cit. p. 307; ID., *Responsabilità da prodotto e struttura del fatto colposo*, cit., p. 322; sul crescente processo di positivizzazione della colpa vd. G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., p. 2.

²⁴² Sul costante incremento delle norme cautelari di fonte privata si vd. recentemente R. BLAIOTTA, *Dove va la colpa penale?*, in *Dir. pen. proc.*, 2021, n. 10, p. 1283

²⁴³ D. CASTRONUOVO, *La colpa penale*, cit. p. 307; ID., *Responsabilità da prodotto e struttura del fatto colposo*, cit., p. 322; in generale sul rilievo dei fenomeni di autonormazione nel diritto penale si rinvia a C. PIERGALLINI, *Ius sibi imponere: controllo penale mediante autonormazione?*, in C.E. Paliero e al. (a cura di), *La crisi della legalità. Il «sistema vivente» delle fonti legali*, Edizioni Scientifiche Italiane, 2016, p. 117 ss.; S.B. TAVERRITI, *Autonormazione e prospettive autopoietiche nella gestione della penalità*, in *Riv. it. dir. proc. pen.*, 2020, p. 1931 ss.; D. BIANCHI, *Appunti per una teoria dell'autonormazione*, in *Riv. it. dir. proc. pen.*, 2019, p. 1477 ss. Critico nei confronti della predisposizione di norme cautelari da parte degli stessi destinatari delle condotte doverose è F. STELLA, *La costruzione giuridica della scienza: sicurezza e salute nei luoghi di lavoro*, in *Riv. it. dir. proc. pen.*, 2003, p. 55 ss.

²⁴⁴ Sul carattere "aperto" della fattispecie colposa vd. C. PIERGALLINI, (voce) *Colpa (diritto penale)*, cit., p. 236; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 54 ss.; R. BARTOLI, (voce) *Fonti della colpa*, cit., p. 527.

²⁴⁵ V. TORRE, *Produttori di alimenti o produttori di norme? Gli standard di sicurezza tra fonti pubblicistiche e fonti privatistiche e il loro valore tipizzante nel fatto colposo*, in L. Foffani – A. Doval Pais – D. Castronuovo (a cura di), *La sicurezza agroalimentare nella prospettiva europea*, cit., p. 508.

ruolo di evitare il verificarsi di determinati eventi lesivi, hanno anche una funzione di orientamento e di garanzia nei confronti degli esercenti l'attività pericolosa, determinandone l'affidamento in un'area di "rischio consentito" (vd. *infra*, § 4.5.2. ss.). Insomma, la codificazione delle cautele mira ad evitare che la regola cautelare funga soltanto da parametro di giudizio *ex post*, ponendosi invece l'obiettivo di integrare la norma penale nella descrizione del fatto colposo²⁴⁶.

Attualmente, tuttavia, nello specifico settore dell'intelligenza artificiale, c'è una quasi assoluta carenza di regole cautelari codificate alle quali il produttore possa riferirsi²⁴⁷. Qualora venisse approvato, l'*AI Act* rappresenterebbe il principale riferimento per le regole cautelari in materia di sviluppo e messa in commercio di sistemi intelligenti. Come abbiamo già riportato (vd. *supra*, cap. II, § 3.4.2.), la bozza di regolamento europeo prevede una serie di requisiti per la commercializzazione di sistemi di i.a., specie per quelli c.d. "ad alto rischio", che saranno poi specificati e concretizzati attraverso l'emanazione di norme armonizzate da parte degli enti di normalizzazione a ciò preposti (vd. *supra*, cap. II, § 3.4.1.). Tra i vari requisiti, ci limitiamo a ricordare, per i dispositivi "ad alto rischio", l'adozione di un "sistema di gestione dei rischi" (art. 9), l'utilizzo di *dataset* "di qualità" (art. 10), la predisposizione di idonea documentazione tecnica (art. 11), l'installazione di *event data recorder* sui dispositivi (art. 12), la previsione di meccanismi che garantiscano il principio dell'*human-in-the-loop* (art. 14) etc. Una volta che la normativa sulla sicurezza dei sistemi di i.a. – generale e specifica, per i vari settori coinvolti – sarà entrata in vigore, essa dovrà ovviamente essere presa in considerazione dal giudice nella valutazione della *colpa specifica* del produttore, nel caso in cui dal prodotto derivi un evento lesivo per il consumatore²⁴⁸.

L'emanazione di *standard* tecnici specifici per la messa in commercio di dispositivi intelligenti è auspicata dalla gran parte della dottrina, come misura di attenuazione dell'incertezza regolativa in materia e, di conseguenza, del *deficit* di

²⁴⁶ F. GIUNTA, *La normatività della colpa*, cit., p. 90; V. TORRE, *Produttori di alimenti o produttori di norme?*, cit., p. 509.

²⁴⁷ S. BECK, *Intelligent agents and criminal law*, cit., p. 139; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 108, nota 79; A. MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, cit., p. 113.

²⁴⁸ Sul ruolo della normativa europea nella definizione delle norme cautelari (e sui relativi profili problematici) si vd. l'accurata analisi di C. BUZIO, *Regole cautelari eurolunitarie e colpa penale. Spunti per una riflessione*, in *Discrimen*, 17 aprile 2020.

precisione e determinatezza della fattispecie colposa²⁴⁹. In realtà, come rilevato da parte della dottrina, un guadagno in termini di determinatezza sarebbe conseguibile soltanto qualora la norma positivizzata fornisse *standard* di comportamento rigidi ed esaustivi, dal momento che le cautele elastiche si fondano sul medesimo meccanismo ricostruttivo utilizzato per la colpa generica, ossia il riferimento all'agente modello²⁵⁰. La bozza di regolamento europeo sull'intelligenza artificiale, in questo senso, non pare fornire soluzioni del tutto soddisfacenti, dal momento che la vaghezza con la quale descrive i requisiti minimi per l'accesso dei sistemi di i.a. sul mercato europeo sembra prefigurare l'introduzione di norme cautelari *elastiche*. Si pensi, ad esempio, all'art. 9, § 4, lett. a), dell'*AI Act*, che stabilisce che il sistema di gestione dei rischi debba garantire l'eliminazione o la riduzione dei rischi “*per quanto possibile*”: è ovvio che tale riferimento non può che aprire la strada ad una comparazione con il parametro dell'*homo eiusdem professionis et condicionis*. È quanto accaduto, in passato, tra l'altro, per le norme in materia di diffusione di polveri di qualunque specie negli ambienti di lavoro. L'art. 21 dell'allora vigente d.P.R. 19 marzo 1956, n. 303 stabiliva infatti che «nei lavori che danno luogo normalmente alla formazione di polveri di qualunque specie, il datore di lavoro è tenuto ad adottare i provvedimenti atti ad impedirne o a ridurre, *per quanto è possibile*, lo sviluppo e la diffusione nell'ambiente di lavoro». Come è stato rilevato, la norma si limitava a descrivere una situazione fattuale di pericolo, senza contenere alcuna prescrizione comportamentale specifica; si risolveva, di fatto, in una sorta di (superflua) “positivizzazione” della colpa generica²⁵¹.

Oltre alle già citate norme di armonizzazione in materia di intelligenza artificiale – in corso di elaborazione da parte di CEN e CENELEC²⁵² – un ruolo fondamentale sarà probabilmente svolto dagli *standard* emanati dalla *International Organization for*

²⁴⁹ In generale, sul ruolo delle norme cautelari codificate nell'attenuazione del *deficit* di determinatezza della fattispecie colposa vd. per tutti G. MARINUCCI – E. DOLCINI, *Corso di diritto penale*, Giuffrè, III ed., 2001, p. 41 ss.

²⁵⁰ D. CASTRONUOVO, *Responsabilità da prodotto e struttura del fatto colposo*, cit., p. 328; F. PALAZZO, *Morti da amianto e colpa penale*, in *Dir. pen. proc.*, 2011, n. 2, p. 189.

²⁵¹ In argomento vd. F. PALAZZO, *Morti da amianto e colpa penale*, cit., p. 190.

²⁵² La Commissione Europea, il 5 dicembre 2022, ha pubblicato la bozza di “richiesta di standardizzazione” nei confronti di CEN e CENELEC, disponibile a questo link: <https://artificialintelligenceact.eu/wp-content/uploads/2022/12/AIA---COM---Draft-Standardisation-Request-5-December-2022.pdf>. Una volta adottata la proposta, gli *standard* dovrebbero essere pubblicati entro il 31 gennaio 2025 (dunque, prevedibilmente, prima dell'entrata in vigore dell'*AI Act*, che sarà applicabile dopo 24 mesi dall'approvazione, ai sensi dell'art. 85).

Standardization (c.d. ISO) e dalla *International Electrotechnical Commission* (c.d. IEC). In particolare, ISO e IEC hanno istituito un *focus group* congiunto (ISO/IEC JTC 1/SC 42), finalizzato alla pubblicazione di *standard* per lo sviluppo e la produzione di sistemi di intelligenza artificiale²⁵³: tra 2020 e 2022, tra gli altri, sono stati pubblicati *standard* in materia di discriminazione algoritmica²⁵⁴, *risk governance*²⁵⁵, affidabilità²⁵⁶, valutazione qualitativa dei sistemi di reti neurali artificiali²⁵⁷, etc. Fondamentale sarà tuttavia la pubblicazione di *standard* che disciplinino settorialmente gli specifici campi in cui può essere utilizzata l'intelligenza artificiale.

Ovviamente, la regola cautelare la cui violazione potrà fondare l'elemento soggettivo del reato colposo deve essere finalizzata a prevenire proprio *il tipo di evento* che è stato in concreto causato. Così, ad esempio, nel caso in cui un *healthcare robot* causi una lesione personale ad un paziente, non si potrà individuare, quale norma cautelare violata, l'art. 15 dell'*AI Act*, che prescrive l'adozione di adeguate misure di *cybersecurity* (sempre, ovviamente, che nel caso concreto l'evento lesivo non sia stato la conseguenza proprio di un attacco cibernetico).

In assenza di generalizzazioni scientifiche sui decorsi eziologici del *decision making* algoritmico, potrebbe apparire ontologicamente impossibile la redazione di *standard* tecnici funzionali al contenimento di una certa tipologia di danni²⁵⁸. In realtà, pur non conoscendo le precise modalità con le quali i sistemi di i.a. decidono, sappiamo che gli errori dei sistemi di *machine learning* – se li intendiamo in senso restrittivo –

²⁵³ Informazioni sui lavori del *focus group* – comprensive dell'elenco completo degli *standard* in materia di i.a. pubblicati fino ad oggi – sono consultabili a questo link: https://www.iec.ch/dyn/www/?p=103:22:959712369727:::FSP_ORG_ID,FSP_LANG_ID:21538,25.

²⁵⁴ INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24027:2021 – Bias in AI systems and AI aided decision making*, disponibile a questo link: <https://www.iso.org/standard/77607.html>.

²⁵⁵ INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24368:2022 – Information technology — Artificial intelligence — Overview of ethical and societal concerns*, disponibile a questo link: <https://www.iso.org/standard/78507.html>.

²⁵⁶ INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24028:2020 – Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*, disponibile a questo link: <https://www.iso.org/standard/77608.html>.

²⁵⁷ INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24029-1:2021 – Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*, disponibile a questo link: <https://www.iso.org/standard/77609.html>.

²⁵⁸ Lo notava, in relazione ai prodotti tradizionali, C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 243.

possono sostanzialmente avere a che fare: (i) con un errore nel codice algoritmico; (ii) con l'alimentazione dell'algoritmo con dati erranei o insufficienti; (iii) con un addestramento dell'algoritmo carente o inadeguato²⁵⁹. Uno *standard* cautelare, ad esempio, potrebbe stabilire che, nell'addestramento di sistemi di riconoscimento di immagini, sia necessario, periodicamente, l'utilizzo di tecniche di *supervised learning* (più impegnative, ma – perlomeno per le immagini semplici – più affidabili rispetto a quelle di *unsupervised learning*, vd. *supra*, cap. I, § 2.1.): un incidente automobilistico cagionato, ad esempio, dal mancato riconoscimento di un segnale di stop da parte del sistema di *image recognition* potrebbe costituire, in ipotesi, concretizzazione del rischio che la suddetta norma cautelare mirava ad evitare.

Più in generale, a nostro avviso le norme cautelari finalizzate all'impedire decisioni erranee (e, di conseguenza, lesive) devono riguardare la raccolta e la gestione dei dati²⁶⁰, l'addestramento del sistema di *machine learning*, e, soprattutto, la fase di sperimentazione, che sarà fondamentale per rilevare eventuali falle nel sistema. A tal proposito, meritevole di attenzione pare la proposta, contenuta agli artt. 53 ss. dell'*AI Act*, di introdurre “spazi di sperimentazione normativa per l'IA” (c.d. *regulatory sandboxes*), finalizzati a fornire «un ambiente controllato che facilita lo sviluppo, le prove e la convalida di sistemi di IA innovativi per un periodo di tempo limitato prima della loro immissione sul mercato o della loro messa in servizio conformemente a un piano specifico» (art. 53, § 1, *AI Act*)²⁶¹.

Infine, in linea generale, per la messa in commercio di prodotti intelligenti sono valide le norme cautelari già esistenti previste dalle direttive della c.d. *legislazione*

²⁵⁹ Va tuttavia nuovamente sottolineato che eventi lesivi potranno derivare anche da sistemi di i.a. pienamente conformi alle cautele esistenti, dal momento che i meccanismi probabilistici che ne costituiscono la base di funzionamento sono, per loro natura, *fallibili*.

²⁶⁰ In quest'ottica, è di estremo interesse la Proposta di *Data Governance Act* (Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act), 25-11-2020, COM(2020) 767 final), che ha l'obiettivo di consentire un maggiore flusso di dati (personali e non) dal settore pubblico al settore privato e viceversa. In argomento vd. G. RESTA, *Cosa c'è di 'europeo' nella Proposta di Regolamento UE sull'intelligenza artificiale?*, in *Dir. inf.*, 2022, fasc. 2, p. 335 ss., al quale si rinvia anche per una disamina del rapporto tra tale normativa e le disposizioni previste dal GDPR.

²⁶¹ Le *regulatory sandboxes* erano già state oggetto di attenzione nella Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica. In argomento vd. S. RANCHORDAS, *Experimental Regulations for AI: Sandboxes for Morals and Mores*, University of Groningen Faculty of Law Research Paper No. 7/2021; M. PARACAMPO, *Il percorso evolutivo ed espansivo delle regulatory sandboxes da FinTech ai nuovi lidi operativi del prossimo futuro*, in *Federalismi*, 2022, n. 18, p. 207 ss.

verticale (vd. *supra*, cap. II, § 3.4.1.), a seconda del settore nel quale lo specifico dispositivo rientra²⁶², nonché le norme cautelari previste dal codice del consumo per la sicurezza dei prodotti in generale. Ci concentreremo brevemente su queste ultime.

Il produttore deve innanzitutto immettere sul mercato prodotti sicuri (art. 104, co. 1, cod. cons.)²⁶³ e deve altresì fornire al consumatore «tutte le informazioni utili alla valutazione e alla prevenzione dei rischi derivanti dall'uso normale o ragionevolmente prevedibile del prodotto, se non sono immediatamente percettibili senza adeguate avvertenze, e alla prevenzione contro detti rischi» (art. 104, co. 1, cod. cons.). Si tratta, a parere di chi scrive, di norme cautelari *elastiche*, che possono essere integrate dagli *standard* tecnici previsti per lo specifico settore produttivo – e dagli eventuali atti emanati da parte delle amministrazioni competenti – e che costituiscono il riflesso di una delicata ponderazione tra l'interesse allo svolgimento un'attività pericolosa e l'esigenza di tutelare beni giuridici fondamentali quali la vita, la salute, l'integrità fisica.

In chiave complementare e sussidiaria rispetto alle suddette cautele positive si pongono altresì, come abbiamo già accennato, i modelli di *autonormazione* e *autocontrollo* eventualmente implementati dal produttore, che possono essere valutati dal giudice nel giudizio circa la sicurezza del prodotto (vd. art. 105, co. 3, cod. cons., che fa riferimento, tra i parametri di valutazione, anche ai «codici di buona condotta in materia di sicurezza vigenti nel settore interessato»).

Il produttore che si discosti dal sistema integrato cautelare così delineato, oltre a rispondere penalmente ai sensi delle contravvenzioni previste all'art. 112 cod. cons. (vd. *infra*, § 5), potrà altresì essere chiamato a rispondere degli eventi di morte o lesione derivanti da un malfunzionamento del prodotto. I *rischi residuali* dovrebbero invece

²⁶² Si pensi, ad esempio, alla c.d. direttiva macchine, oggetto, in questo momento, di una proposta di revisione, vd. Commissione Europea (Proposta di Regolamento del Parlamento Europeo e del Consiglio sui prodotti macchina; Bruxelles, COM (2021) 202 final, 21 aprile 2021).

²⁶³ Come abbiamo già riportato precedentemente (vd. *supra*, cap. II, § 3.4.1.) per prodotto sicuro si intende, ai sensi dell'art. 103, co. 1, lett. a), qualsiasi prodotto che «in condizioni di uso normali o ragionevolmente prevedibili, compresa la durata e, se del caso, la messa in servizio, l'installazione e la manutenzione, non presenti alcun rischio oppure presenti unicamente rischi minimi, compatibili con l'impiego del prodotto e considerati accettabili nell'osservanza di un livello elevato di tutela della salute e della sicurezza delle persone», in funzione di vari elementi, tra i quali la norma considera, in via meramente esemplificativa, le sue «caratteristiche» (n. 1), il suo eventuale «effetto» su altri prodotti (n. 2), la sua «presentazione» (n. 3) e le «categorie di consumatori» cui esso è destinato (n. 4), con particolare riguardo a quelle maggiormente vulnerabili (bambini e anziani).

rientrare nell'area del rischio consentito, salvo quanto si dirà circa il possibile (e riconoscibile) fallimento delle norme cautelari codificate (vd. *infra*, § 4.5.2.2.).

4.5.2. Il rapporto tra regole cautelari scritte e regole cautelari non scritte: quale spazio per il rischio consentito?

Poniamo ora il caso che il dispositivo intelligente abbia cagionato un evento lesivo *nonostante l'osservanza*, da parte del produttore, delle norme cautelari codificate. Il rispetto delle cautele positive stabilite dagli *standard* tecnici non potrà infatti evitare il rischio che dal concreto operare dei sistemi di i.a. derivino eventi lesivi per la vita, la salute o l'integrità fisica: nel momento in cui, infatti, si attribuisce una capacità di autonomo *decision making* ad un algoritmo – ancorché parziale e sotto supervisione di un essere umano – non si potrà escludere che da questo si verifichino decorsi eziologici dannosi.

Di qui la questione: il rispetto degli *standard* tecnici di settore – quando saranno introdotti – sarà sufficiente ad escludere la colpa in capo al produttore di un sistema di i.a., o il giudice dovrà invece valutare se la condotta di questi sia conforme a quella dell'*homo eiusdem condicionis et professionis*?

È un problema ben noto, che ha ancora non ha trovato una soddisfacente e condivisa soluzione sul piano dogmatico. Tradizionalmente, si ritiene che il rispetto delle norme cautelari positive non esoneri l'agente dal tenere un comportamento diligente, da identificarsi sulla base del criterio del c.d. agente modello²⁶⁴. Nel caso delle attività produttive, tuttavia, il problema si complica, poiché le norme cautelari, oltre che una funzione di *prevenzione* rispetto al verificarsi dell'evento lesivo, hanno altresì una funzione di *garanzia* e di *orientamento* nei confronti dell'agente, in settori particolarmente rischiosi e ad alta complessità tecnico-normativa²⁶⁵.

²⁶⁴ G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., p. 236; F. MANTOVANI, (voce) *Colpa*, in *Dig. pen.*, II, Utet, 1988, p. 309; G. FORTI, *Colpa ed evento*, cit., 315 ss.; R. BARTOLI, (voce) *Fonti della colpa*, cit., p. 529-530; V. MILITELLO, *Rischio e responsabilità penale*, Giuffrè, 1988, p. 148.

²⁶⁵ Nello specifico settore dei sistemi di i.a., vd. S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 719.

Si apre così la questione circa la sussistenza di un'area di c.d. "rischio consentito" (*erlaubtes Risiko*), da tempo dibattuta in dottrina²⁶⁶ e oggetto di rinnovata attenzione, recentemente, proprio in relazione agli eventi lesivi derivanti da sistemi di i.a.²⁶⁷. In letteratura, in realtà, non vi è unità di vedute circa la collocazione dogmatica del rischio consentito – se debba qualificarsi come fattore di esclusione della tipicità, dell'antigiuridicità o della colpevolezza e, addirittura, se esso debba considerarsi come

²⁶⁶ La letteratura sul tema è ormai di ampiezza considerevole; vd. G. FORTI, *Colpa ed evento*, cit., p. 250 ss.; ID., "Accesso alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione", cit., p. 155 ss.; ID., *La "chiara luce della verità" e l'"ignoranza del pericolo"*, cit., p. 573 ss.; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit.; ID., *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, in *Riv. it. dir. proc. pen.*, 2005, p. 1670 ss.; V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 55 ss. (che preferisce, tuttavia, l'espressione "rischio adeguato"); F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 227 ss.; F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1102 ss.; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 335 ss.; C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione. La c.d. "flessibilizzazione delle categorie del reato"*, in *Criminalia*, 2012, p. 383 ss.; D. CASTRONUOVO, *Principio di precauzione e beni legati alla sicurezza*, in *Dir. pen. cont.*, 21 luglio 2011, p. 1 ss.; ID., *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit.; M. DONINI, *Il volto attuale dell'illecito penale*, cit., p. 119 ss.; A. MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole?*, cit.; D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 647 ss.; C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1743 ss.; R. BLAIOTTA, *Incidenti da aviolancio, problemi causali nuovi, attività rischiose*, in *Cass. pen.*, 2004, n. 7-8, p. 2377 ss.; E. PENCO, *Limiti-soglia e responsabilità colposa*, cit.

²⁶⁷ Vd., con vari accenti, S. GLESS – E. SILVERMAN – T. WEIGEND, *If Robots Cause Harm, Who Is to Blame: Self-Driving Cars and Criminal Liability*, cit., pp. 430-431; S. BECK, *Robotics and Criminal Law. Negligence, Diffusion of Liability and Electronic Personhood*, cit., p. 53 ss.; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1750; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 116 ss.; A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 656 ss.; V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, cit., p. 5; A. CAPELLINI, *Machina delinquere non potest?*, cit., p. 19; ID., *Profili penalistici delle self-driving cars*, cit. p. 4-5; ID., *Reati colposi e tecnologie dell'intelligenza artificiale*, cit., p. 29 ss.; D. PIVA, *Machina discere, (deinde) delinquere et puniri potest*, cit., p. 681 ss.; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 718 ss.; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 73; R. BORGOGNO, *La responsabilità penale nei processi ad elevata automazione*, cit., p. 735; C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 62 ss.; A. MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, cit., p. 113 ss.; R. BORSARI, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, cit., p. 268; L. D'AMICO, *Intelligenza artificiale e auto a guida autonoma. tra prevenzione primaria, colpa penale e rischio consentito*, cit., p. 593 ss.; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 9-10; P.S. GÜNSBERG, *Automated vehicles – is a dilution of human responsibility the answer?*, cit., p. 448 ss.; sostanzialmente anche D. LIMA, *Could AI Agents Be Held Criminally Liable?*, cit., p. 694.

un concetto autonomo o meramente descrittivo e pleonastico²⁶⁸. In generale, la dottrina maggioritaria tende a collocare il rischio consentito nell'ambito degli studi sulla teoria della colpa²⁶⁹. L'ambientazione più immediata dell'istituto è infatti quella inerente alle cautele che presiedono allo svolgimento di attività pericolose socialmente utili, ove il pericolo sopravviva malgrado il rispetto di tutti i doveri di prudenza codificati.

In particolare, secondo un autorevole orientamento dottrinale, possono configurarsi due diverse nozioni di rischio consentito²⁷⁰. Nella sua versione più ampia, all'area del rischio consentito vengono ricondotte tutte quelle attività che, sebbene pericolose, sono legittime, a condizione che vengano svolte sulla base di adeguate misure cautelari, idonee a minimizzare il rischio di eventi lesivi. A questa prima accezione, descrittiva più che precettiva, se ne aggiunge un'altra, che pare conferire una qualche autonomia dogmatica alla nozione di rischio consentito e che, per questo motivo, ci pare preferibile. In questa seconda prospettiva, il concetto di rischio consentito indica «un'area (più ristretta e comunque concentrica rispetto alla prima) di condotte pericolose, ammesse dall'ordinamento nonostante che l'adozione di cautele idonee a contrastare i possibili svolgimenti lesivi sia destinata a residuare un certo grado di pericolosità»²⁷¹.

È quello che viene denominato rischio consentito “in senso stretto” o “specifico” o, ancora, “rischio residuale (*Restrisiko*)”²⁷², intendendosi, con tale designazione icastica, quel *pericolo marginale* che le misure preventive non sono in grado di disinnescare (o che, sulla base di un rapporto costi-benefici, le autorità che “producono” la normativa cautelare *decidono* di non disinnescare) e che, di conseguenza, ricade sulla società.

In quest'accezione, la nozione di “rischio consentito” si innesta sui giudizi di *prevedibilità* ed *evitabilità* dell'evento, tipici della valutazione circa la sussistenza della colpa, e delinea quelle ipotesi in cui il rispetto delle norme cautelari *codificate* – alla base della valutazione circa la sussistenza della *colpa specifica* – impedisce che possa essere

²⁶⁸ Su questi aspetti si vedano vd. C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 10; C. EAD., *La legislazione penale tra “diritto penale dell'evento” e “diritto penale del rischio”*, cit., 120 (in particolare nota 7); V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 62.

²⁶⁹ Per una ricostruzione dei diversi orientamenti dogmatici in argomento si rinvia a F. CONSULICH, (voce) *Rischio consentito*, cit., *passim*.

²⁷⁰ G. FORTI, *Colpa ed evento*, cit., p. 456 ss.

²⁷¹ G. FORTI, *Colpa ed evento*, cit., p. 456; vd. anche C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione*, cit., p. 389.

²⁷² G. FORTI, *Colpa ed evento*, cit., p. 457.

mosso all'imputato un rimprovero per non aver osservato norme di diligenza, prudenza, perizia *non codificate*²⁷³.

Alla base di tale ricostruzione ermeneutica del rischio consentito vi è la constatazione che, nella società in cui viviamo, produzione e rischio sono *grandezze direttamente proporzionali*, sì che non si può rinunciare alla seconda senza perdere anche la prima, con evidenti ricadute sul livello di benessere sociale (in termini, ad esempio, occupazionali o di accesso ad infrastrutture e beni di consumo). È superfluo sottolineare, d'altra parte, che alla crescita dei rischi corrisponde logicamente (e statisticamente) un aumento degli *eventi lesivi che costituiscono la concretizzazione di tali rischi*.

Secondo questa prospettiva, l'agente modello non è per forza di cose colui che, prevedendo il rischio del verificarsi dell'evento lesivo, *si astenga dall'agire*, ma è piuttosto colui che *agisce*, assicurandosi però che il rischio si mantenga entro il livello stabilito a livello legislativo – che per l'appunto, corrisponde al rischio “residuo”, e accettato come tale dalla società²⁷⁴.

Il rapporto tra cautele codificate e criterio dell'agente modello così delineato pare d'altra parte l'unico compatibile con la concezione normativa della colpa, in base alla quale «l'agente modello non è colui che prevede ed evita tutti gli eventi che è possibile prevedere ed evitare, ma colui che prevede ed evita gli eventi che è doveroso prevedere ed evitare»²⁷⁵. Se si consentisse al giudice di sostituire il proprio apprezzamento a quello del legislatore, si finirebbe d'altra parte per ledere l'affidamento del produttore nella liceità del proprio operato, con conseguente violazione dei principi costituzionali di legalità e colpevolezza (in particolare, di *prevedibilità* della sanzione penale)²⁷⁶. La definizione dell'area del rischio consentito è dunque, in tutta evidenza, una questione di bilanciamento tra esigenze divergenti: da un lato, l'esercizio di attività pericolose *utili*,

²⁷³ C. PIERGALLINI, *Attività produttive e imputazione per colpa: prove tecniche di “diritto penale del rischio”*, in Riv. it. dir. proc. pen., 1997, fasc. 4, p. 1492; L. STORTONI, *Angoscia tecnologica ed esorcismo penale*, cit., p. 79; F. PALAZZO, *Morti da amianto e colpa penale*, cit., p. 188.

²⁷⁴ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 358.

²⁷⁵ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 358. Sulla concezione normativa della colpa – ormai dominante nella dogmatica italiana – vd. M. GALLO, (voce) *Colpa penale*, in Enc. dir., VII, 1960, p. 637; G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., spec. p. 113 ss. e 166 ss.; ID., *Il reato come ‘azione’. Critica di un dogma*, Giuffrè, 1971, p. 100 e p. 111-121; G. FORTI, *Colpa ed evento*, cit., p. 125 ss.; F. GIUNTA, *Illiceità e colpevolezza nella responsabilità colposa*, cit., p. 169 ss.; P. VENEZIANI, *Regole cautelari “proprie” ed “improprie”*, cit., p. 10 ss.

²⁷⁶ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 420.

dall'altro, la tutela di beni giuridici penalmente rilevanti, che possono venire lesi dallo svolgimento di tali attività²⁷⁷.

La scelta sulla misura opportuna delle due istanze è prima di tutto rimessa al legislatore, che, sulla base di un'analisi costi-benefici eminentemente *politica*, definisce il livello di rischio socialmente accettabile e, dunque, il grado di tutela che l'ordinamento intende assicurare ai beni giuridici potenzialmente interessati dallo svolgimento dell'attività pericolosa – leggi anche: la quota di eventi lesivi che una società è disposta a tollerare, in cambio dei benefici complessivi derivanti dalla produzione industriale e dal progresso tecnologico. La decisione di delineare un'area di rischio consentito è infatti una scelta che implica l'esercizio (e il non-esercizio) di un potere punitivo, che per esplicita previsione costituzionale (art. 25, co. 2) spetta al legislatore²⁷⁸.

Se non lo si guardasse alla luce del consenso sociale e del bilanciamento politico tra istanze contrapposte, non si capirebbe, ad esempio, perché la produzione e l'impiego di amianto è oggi proibito in Italia, mentre la commercializzazione di tabacco – pur assistita da precisi obblighi informativi²⁷⁹ – sia invece consentita, nonostante la solida e diffusa consapevolezza che il fumo sia causa di *sicuri danni* alla salute dei consumatori²⁸⁰. Ecco che, allora, l'istituto del rischio consentito impedisce che un giudice possa addebitare ai produttori di sigarette la commissione di eventi lesivi, rimproverando loro la concreta *prevedibilità* dei danni, sulla base di una presunta funzione supplente della colpa generica.

Come vedremo più approfonditamente nel prossimo paragrafo, l'unico margine di operatività di una valutazione della colpa generica dovrebbe configurarsi nel caso in cui

²⁷⁷ G. FORTI, *Colpa ed evento*, cit., p. 255; F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1109; J.M. SILVA SÁNCHEZ, *La expansión del derecho penal*, cit., p. 44; C. PIERGALLINI, (voce) *Colpa (diritto penale)*, cit., p. 239; ID., *Danno da prodotto e responsabilità penale*, cit., p. 242; V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 123-124; C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione*, cit., p. 391; F. PALAZZO, *Morti da amianto e colpa penale*, cit., p. 188; E. CORN, *Il principio di precauzione*, cit., p. 158; D. PULITANÒ, *Diritto penale*, cit., p. 308; in generale, sul ruolo di bilanciamento di interessi svolto dalle norme cautelari positive, vd. G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 416.

²⁷⁸ G. FORTI, *Colpa ed evento*, cit., p. 458; F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1112; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 360.

²⁷⁹ Sull'evoluzione della normativa europea vd. S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 412 ss.

²⁸⁰ In questo senso, F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1109-1110; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 418.

la regola cautelari manifesti *segnali univoci* di fallimento, ovverosia di obsolescenza o di inidoneità al raggiungimento degli obiettivi fatti propri dalla regola cautelare scritta.

Anche la Corte Costituzionale, nelle sentenze relative ai c.d. decreti *Salva Ilva*²⁸¹, ha avuto modo di chiarire che la decisione sul bilanciamento tra esigenze produttive (e, in quel caso, lavorative ed occupazionali) e tutela di beni giuridici tutelati dall'ordinamento (in quel caso, salute e ambiente salubre) è rimessa al legislatore e, *purché ragionevole*, non può essere disattesa dall'autorità giudiziaria.

Nessun diritto, nemmeno quello alla vita e alla salute, è “tiranno” rispetto alle altre situazioni giuridiche costituzionalmente riconosciute e protette²⁸². Il bilanciamento tra i beni giuridici deve tuttavia rispondere a «criteri di proporzionalità, e di ragionevolezza, in modo tale da non consentire né la prevalenza assoluta di uno dei valori coinvolti, né il sacrificio totale di alcuno di loro, in modo tale che sia sempre garantita una tutela sistemica e non frammentaria di tutti gli interessi costituzionali implicati»²⁸³. Sulla base di tali principi, la Corte Costituzionale, con la sentenza n. 85 del 9 maggio 2013, era giunta a ritenere legittimo il primo decreto c.d. *Salva Ilva* (d.l. 3 dicembre 2012, n. 207, conv. in legge 24 dicembre 2012, n. 231), che autorizzava il proseguimento dell'attività produttiva, nonostante il sequestro degli stabilimenti industriali, a condizione che fossero rispettate le prescrizioni dettate dall'Autorizzazione Integrata Ambientale (c.d. AIA): in quel caso, la Corte argomentava che la decisione costituiva *ragionevole bilanciamento* tra gli interessi in gioco, costituendo l'esito di un procedimento amministrativo svoltosi in conformità ai principi di prevenzione, precauzione, informazione e partecipazione.

Stessa sorte non è invece toccata al secondo decreto *Salva Ilva* (d.l. 4 luglio 2015, n. 92), che faceva dipendere la prosecuzione dell'attività non da una valutazione proveniente da un soggetto terzo (quale l'autorità amministrativa competente), ma dalla predisposizione di un piano di minimizzazione dei rischi *da parte dello stesso datore di lavoro*, soggetto destinatario dei provvedimenti cautelari. Sulla base di queste e altre motivazioni, i giudici costituzionali giungono a ritenere che «il legislatore ha finito col

²⁸¹ D.l. 3 dicembre 2012, n. 207, conv. in legge 24 dicembre 2012, n. 231, ritenuto legittimo da C. Cost., sent. 9 maggio 2013, n. 85; d.l. 4 luglio 2015, n. 92, ritenuto illegittimo da C. Cost., sent. 7 febbraio 2018, n. 58.

²⁸² C. Cost., sent. 9 maggio 2013, n. 85, § 9.

²⁸³ C. Cost., sent. 7 febbraio 2018, n. 58, § 3.1., che a sua volta cita le sentenze della Corte Cost. n. 63 del 2016 e n. 264 del 2012. Nello stesso senso vd. anche, in dottrina, V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 106.

privilegiare in modo eccessivo l'interesse alla prosecuzione dell'attività produttiva»²⁸⁴, travalicando così i limiti che l'art. 41 Cost. pone all'esercizio dell'attività d'impresa²⁸⁵.

In questa prospettiva, la suddetta concezione del rischio consentito pare del tutto in linea con il principio della separazione dei poteri, impedendo al giudice di sostituire il proprio apprezzamento in merito alla rilevanza degli interessi costituzionali coinvolti al bilanciamento legittimamente effettuato dal legislatore e dagli organi amministrativi²⁸⁶.

In ogni caso, posto che, in prima istanza, il compito di contemperare gli interessi della produzione e la tutela della salute e dell'integrità fisica spetta al legislatore, spesso quest'ultimo non ha le competenze tecniche né le risorse per dettare le regole cautelari *specifiche* per ciascun settore. Ecco che allora sono più spesso gli organismi tecnici, pubblici o privati, o le associazioni di categoria, che assumono la funzione di stabilire con maggiore precisione gli *standard* tecnici di riferimento²⁸⁷: è lo stesso fenomeno, *mutatis mutandis*, che spinge le istituzioni europee a delegare agli enti di normalizzazione la definizione delle norme tecniche rilevanti nei vari settori industriali e produttivi (vd. *supra*, cap. II, § 3.4.1.).

Com'è evidente, il dibattito sui confini della nozione di rischio consentito sembra riprodurre, sul versante penalistico e nel solo settore della responsabilità per il danno cagionato da prodotto difettoso, la *querelle* sulla possibile difettosità del prodotto conforme alle normative in materia di *product safety* (vd. *supra*, cap. II, § 3.4.). In quella sede, siamo giunti alla conclusione – in accordo con la dottrina maggioritaria – che un prodotto, pur conforme alla normativa sulla sicurezza, *può comunque risultare difettoso*, con la conseguenza che il danno da questo cagionato può essere risarcito sulla base della disciplina sulla responsabilità per danno da prodotto. Il rischio consentito, sui cui limiti ora rifletteremo, sembrerebbe invece delineare un'area di impunità nei confronti di colui che eserciti un'attività produttiva nel rispetto delle norme cautelari positivizzate.

²⁸⁴ C. Cost., sent. 7 febbraio 2018, n. 58, § 3.3.

²⁸⁵ Per un approfondimento della complessa vicenda si rinvia a S. ZIRULIA, *Sequestro preventivo e sicurezza sul lavoro: illegittimo il decreto "salva-Ilva" n. 92 del 2015*, in *Riv. it. dir. e proc. pen.*, 2018, fasc. 2, p. 947 ss.; R. DE VITO, *La salute, il lavoro, i giudici*, in *Quest. Gius.*, 24 marzo 2018; G. AMENDOLA, *Ilva e diritto alla salute. La Corte Costituzionale ci ripensa?*, *ivi*, 10 aprile 2018. Vd. anche, nella giurisprudenza di legittimità, il caso relativo alle emissioni elettromagnetiche di Radio Vaticana, Cass., sez. III, 13 maggio 2008, n. 36845, in *Foro it.*, 2009, II, 262, con nota di G. FIANDACA, *Inquinamento elettromagnetico e rilevanza penale: questione approfondita ma non risolta*.

²⁸⁶ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 362-363.

²⁸⁷ F. CONSULICH, (voce) *Rischio consentito*, p. 1112.

La differenza, a nostro parere, si giustifica per il fatto che, in ambito civile, la responsabilità per il danno da prodotto difettoso è imputata oggettivamente al produttore, indipendentemente dalla dimostrazione della colpa: la diligenza che il produttore dimostra rispettando gli *standard* tecnici non sarà dunque rilevante ai fini di un'esclusione di una responsabilità a suo carico; l'osservanza degli *standard* potrà avere soltanto un ruolo indicativo (a dire il vero, piuttosto rilevante) in sede di accertamento dell'oggettiva difettosità del prodotto. Nell'ambito della responsabilità penale, invece, gli *standard* tecnici di sicurezza fungono da norme cautelari e, di conseguenza, integrando la nozione di colpa, hanno altresì una funzione "permissiva" e di *garanzia* nei confronti dell'agente.

4.5.2.1. Individuazione delle aree di rischio consentito e limiti all'operatività dell'istituto.

Definita per sommi capi la nozione e il fondamento del rischio consentito, resta il problema di individuare i casi concreti in cui questo possa delinearsi, *i.e.* i casi in cui la conformità della condotta al dovere di diligenza codificato impedisce che la concreta *prevedibilità* ed *evitabilità* dell'evento da parte dell'agente possa determinare un rimprovero per *colpa generica*, nel caso in cui si verifichino eventi lesivi²⁸⁸.

Deputate per eccellenza al compito di delineare un'area di rischio consentito sono le regole cautelari *improprie* e dotate di *rigidità*.

In base ad una distinzione ormai consolidata in dottrina, per regole improprie si intendono quelle regole che prevedono l'adozione di misure idonee a ridurre il rischio, al contrario di quelle *proprie*, che invece tendono alla sua completa neutralizzazione²⁸⁹. Le norme cautelari che delineano un'area di rischio consentito sono necessariamente *regole improprie*, poiché – qualora la norma cautelare mirasse alla neutralizzazione del rischio

²⁸⁸ Per tale nozione di rischio consentito vd. C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 527; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 360; F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1116; A. PERIN, *Prudenza, dovere di conoscenza e colpa penale*, Editoriale Scientifica, 2020, p. 69.

²⁸⁹ La definizione è stata proposta da P. VENEZIANI, *Regole cautelari "proprie" e "improprie"*, cit., *passim*, ed è ormai comunemente accettata, vd. per tutti C. PIERGALLINI, (voce) *Colpa (diritto penale)*, cit., p. 229; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 364; C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione*, cit., p. 391.

– l’evento lesivo che si genererebbe in caso di violazione sarebbe immediatamente illecito²⁹⁰. La distinzione tra norme cautelari proprie ed improprie tende comunque a scolorirsi, ove si tenga in considerazione che il fatto che tutte le norme cautelari – salvo quelle che impongono *divieti* di esercizio di un’attività pericolosa – possono essere soggette a fallimento, con la conseguenza che anche le regole *proprie* possono lasciare scoperte delle aree marginali di rischio consentito²⁹¹.

I confini del rischio consentito sono poi tracciati con più precisione dalle norme cautelari *rigide*, ovverosia da quelle norme presentano un bilanciamento di interessi integralmente predeterminato dal legislatore, imponendo al destinatario una regola fissa di comportamento²⁹²: paradigmatica di tale tecnica legislativa è la configurazione di limiti-soglia, che definisce preventivamente i livelli di emissione (es. di sostanze tossiche) o di assunzione (es. di sostanze alcoliche) al di sotto dei quali il rischio si considera consentito²⁹³.

Quanto alle norme cautelari elastiche, sebbene vi sia chi ritenga che uno spazio di applicazione del rischio consentito sia configurabile anche in relazione ad esse²⁹⁴, va sottolineato che, come è stato autorevolmente affermato, «quanto più è indeterminata la regola, tanto più la colpa specifica scolora in quella generica»²⁹⁵. Quel giudizio sulla concreta prevedibilità ed evitabilità dell’evento, che l’area del rischio consentito mira ad escludere, si riproporrebbe di fatto nella ricostruzione del contenuto specifico della regola cautelare elastica²⁹⁶.

L’istituto del rischio consentito è strutturalmente legato all’idea che la norma cautelare è *inidonea a prevenire tutti i rischi che è finalizzata a prevenire*. Si tratta di

²⁹⁰ F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1115.

²⁹¹ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 364.

²⁹² Sulla distinzione tra regole *rigide* e regole *flessibili* vd., nella manualistica, G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 430; G. FIANDACA – E. MUSCO, *Diritto penale. Parte Generale*, cit., p. 583; D. PULITANÒ, *Diritto penale*, cit., p. 304.

²⁹³ Sulla problematica qualificazione dei limiti-soglia quali norme cautelari vd., tuttavia, E. PENCO, *Limiti-soglia e responsabilità colposa*, cit., p. 211 ss.; sui valori soglia in generale v. F. D’ALESSANDRO, *Pericolo astratto e limiti soglia*, cit., spec. 267 ss., 333 ss.; nella manualistica vd. G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 286.

²⁹⁴ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 365.

²⁹⁵ F. GIUNTA, *La normatività della colpa penale*, cit., p. 92.

²⁹⁶ F. GIUNTA, *La normatività della colpa penale*, cit., p. 92; D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., 652; C. PIERGALLINI, (voce) *Colpa (diritto penale)*, cit., p.229 e 231; V. TORRE, *Produttori di alimenti o produttori di norme?*, cit., p. 509; F. PALAZZO, *Morti da amianto e colpa penale*, cit., p. 190, che parla di vera e propria “frode delle etichette”;

un'inidoneità *prevedibile e programmata*: il legislatore – e gli altri soggetti che, a vario titolo, sono deputati alla definizione delle regole cautelari – nel momento in cui delineano il bilanciamento tra interessi divergenti *sanno* che la cautela è talvolta destinata a fallire, poiché non sarà in grado di avere efficacia impeditiva nei confronti di *tutti* gli eventi lesivi. È proprio l'accettazione preventiva di una quota di offese a beni giuridici che costituisce il *proprium* del rischio consentito²⁹⁷.

L'area di impermeabilità alla colpa generica è delimitata da questa ipotesi di *ordinario fallimento* della norma cautelare, in qualche modo preventivato dagli organi che la codificano. Al di fuori di tale ristretto ambito – in cui si esplica compiutamente l'attività di definizione politica del bilanciamento tra beni giuridici contrapposti – si ritiene comunemente che non possano esistere zone franche di impunità all'interno dell'ordinamento penale²⁹⁸.

La valutazione della diligenza, prudenza e imperizia dell'agente nel *prevedere* ed *evitare* l'evento lesivo si spingerà dunque al di fuori degli angusti confini della cautela positiva, ogniquale volta vi sia un *fallimento non preventivato* di quest'ultima.

Potrà trattarsi di un fallimento *in concreto*, nel caso in cui la presenza di circostanze anomale renda in concreto inadeguata la norma cautelare – che, tuttavia, mantiene la sua validità a livello generale: si pensi, ad esempio, a condizioni meteorologiche straordinariamente avverse, che possono rendere inadeguata la normativa antinfortunistica che pur venga rispettata da chi organizzi un cantiere edile con gru e ponteggi²⁹⁹.

²⁹⁷ F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1116.

²⁹⁸ F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1116; parlano di un eventuale “fallimento” delle norme cautelari G. FORTI, *Colpa ed evento*, cit., p. 671 ss.; P. VENEZIANI, *Regole cautelari “proprie” ed “improprie”*, cit., p. 61 ss., S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 380; E. PENCO, *Limiti-soglia e responsabilità colposa*, cit., p. 234; vd. anche C. PIERGALLINI, (voce) *Colpa (diritto penale)*, cit., p. 231, che parla di “inadeguatezza” delle cautele.

²⁹⁹ L'esempio è di S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 380; cfr. anche G. FORTI, *Colpa ed evento*, cit., p. 673 ss.; P. VENEZIANI, *Regole cautelari “proprie” ed “improprie”*, cit., p. 61 ss., che avvalorano una concezione parzialmente diversa di “fallimento” in concreto della norma cautelare, che sarebbe rinvenibile nei casi in cui «l'evento si sia verificato attraverso uno dei decorsi causali di cui si era tenuto conto nella costruzione della norma cautelare (che l'agente avrebbe pertanto dovuto prevedere e contrastare) e tuttavia si accerti che la cautela, prospettabile *ex ante*, non sarebbe stata in grado di impedire la verifica dello stesso, rivelandosi appunto, *ex post*, inefficace» (G. FORTI, *Colpa ed evento*, cit., p. 674). Secondo tale concezione, si presuppone che la regola cautelare sia stata violata dall'agente, e che soltanto successivamente emerga che le circostanze del caso concreto avrebbero reso in ogni caso inefficace la norma cautelare violata.

In secondo luogo, potrà anche trattarsi di un fallimento *in astratto*, nel caso in cui la norma cautelare si riveli assolutamente inadeguata allo scopo di bilanciamento che si erano prefissati i soggetti che originariamente l'avevano redatta³⁰⁰; o perché il *rischio reale è più elevato* di quello che era stato previsto in sede di positivizzazione della cautela, o perché si è in presenza di rischi *ulteriori* rispetto a quelli che erano stati presi in considerazione, o, ancora, perché – a parità tra i rischi previsti e quelli reali – vi è un difetto di tecnica normativa che impedisce di raggiungere la riduzione del rischio che costituiva l'obiettivo della codificazione³⁰¹.

In entrambi questi casi – sia che si tratti di fallimento in concreto della norma cautelare, sia che si tratti di fallimento in astratto, sia che, infine, vi sia un difetto nella tecnica normativa – si apre la strada per una valutazione circa la colpa generica dell'agente, *sub specie* di *prevedibilità* dell'inadeguatezza della norma cautelare, o addirittura *conoscenza* da parte dell'agente (qualora egli disponga di conoscenze superiori rispetto all'agente modello)³⁰². Escludere l'accertamento della colpa generica, in questi casi, significherebbe infatti trasformare il rischio consentito in un'ingiustificata e intollerabile forma di immunità³⁰³.

Nell'ottica di un riconoscimento esplicito dell'efficacia “esimente” delle norme cautelari scritte – a meno di un loro riconoscibile fallimento – si esprimeva altresì il Progetto preliminare di riforma del codice penale redatto dalla Commissione Grosso³⁰⁴, che all'art. 28, co. 2, prevedeva che «[i]l rispetto delle regole cautelari specifiche di cui al comma precedente esclude la colpa relativamente agli aspetti disciplinati da dette regole, salvo che il progresso scientifico o tecnologico, nel periodo successivo alla loro emanazione, non le abbia rese palesemente inadeguate».

³⁰⁰ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 380; P. VENEZIANI, *Regole cautelari “proprie” ed “improprie”*, cit., p. 85, p. 61 ss.; F. CONSULICH, (voce) *Rischio consentito*, cit., 1116.

³⁰¹ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 380, nota 208; P. VENEZIANI, *Regole cautelari “proprie” ed “improprie”*, cit., p. 85 ss.

³⁰² D. CASTRONUOVO, *Principio di precauzione*, cit., p. 167.

³⁰³ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 381; vd. anche V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 95 ss.

³⁰⁴ COMMISSIONE GROSSO PER LA RIFORMA DEL CODICE PENALE, *Articolato approvato dalla Commissione Ministeriale per la Riforma del codice penale nella seduta del 26 maggio 2001*, disponibile sul sito del Ministero della Giustizia a questo link: https://www.giustizia.it/giustizia/it/mg_1_12_1.page?contentId=SPS31489&previousPage=mg_1_12_1.

4.5.2.2. Gli obblighi di monitoraggio e il “fallimento” delle regole cautelari positive.

Ai fini dell'identificazione di un eventuale *superamento* del rischio consentito incideranno senza dubbio gli obblighi di monitoraggio *post market*, previsti dalla normativa generale in materia di sicurezza dei prodotti (vd. art. 104, co. 6 ss., cod. cons.), dalla normativa di settore e, infine, dall'*AI Act* (vd. artt. 60 ss.)³⁰⁵.

Partiamo da questi ultimi. L'art. 60 della bozza di regolamento prevede un sistema di registrazione delle applicazioni di i.a. ad alto rischio in una banca dati pubblica a livello dell'UE. Tale registrazione consentirà alle autorità competenti, agli utenti e ad altre persone interessate, di verificare se il sistema di i.a. ad alto rischio sia conforme ai requisiti stabiliti dalla normativa. Al fine di alimentare tale banca dati, i fornitori di i.a. saranno tenuti a fornire informazioni sui sistemi di loro produzione e sulla valutazione di conformità condotta.

La proposta prevede inoltre che i fornitori istituiscano un sistema di monitoraggio successivo all'immissione sul mercato dei sistemi di i.a., che «raccolge, documenta e analizza attivamente e sistematicamente i dati pertinenti forniti dagli utenti o raccolti tramite altre fonti sulle prestazioni dei sistemi di IA ad alto rischio per tutta la durata del loro ciclo di vita e consente al fornitore di valutare la costante conformità dei sistemi di IA ai requisiti» (art. 61, co. 1 e 2). Ai fini dell'implementazione del sistema, sarà tuttavia necessario un atto di esecuzione della Commissione che definisca «un modello per il piano di monitoraggio successivo all'immissione sul mercato e un elenco di elementi da includere nel piano» (art. 61, co. 3)

Ai sensi, poi, dell'art. 62 dell'*AI Act*, i fornitori di sistemi di i.a. ad alto rischio avranno l'obbligo di segnalare qualsiasi incidente grave o malfunzionamento di tali sistemi che costituisca una violazione degli obblighi previsti dal diritto dell'Unione alle autorità di vigilanza del mercato degli Stati membri in cui tali incidenti o violazioni si sono verificati (co. 1). Tale notifica dovrà essere effettuata immediatamente dopo che il fornitore ha stabilito un nesso causale tra il sistema di i.a. e l'incidente o il malfunzionamento o quando stabilisce la ragionevole probabilità di tale nesso e, in ogni

³⁰⁵ In argomento vd. diffusamente C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, cit., p. 430 ss.

caso, non oltre 15 giorni dopo che è venuto a conoscenza dell'incidente grave o del malfunzionamento.

Le autorità nazionali competenti indagheranno sugli incidenti e sui malfunzionamenti, raccoglieranno tutte le informazioni necessarie e le trasmetteranno periodicamente alla Commissione con metadati adeguati. La Commissione si occuperà poi di integrare tali informazioni e, cinque anni dopo l'entrata in vigore dell'*AI Act*, pubblicherà una relazione di valutazione del quadro normativo, eventualmente proponendo modifiche o revisioni alla disciplina. Ai sensi dell'art. 65, co. 2, dell'*AI Act*, l'autorità di sorveglianza che abbia conoscenza di una difformità del sistema di i.a. con i requisiti e le obbligazioni fissati nel regolamento potrà imporre al soggetto interessato l'adozione delle misure appropriate per far cessare la violazione, ritirare il sistema di i.a. dal mercato o richiamarlo per un tempo ragionevole commisurato alla natura del rischio.

Quanto agli obblighi di monitoraggio e di ritiro rivolti alla società produttrice, l'art. 21 dell'*AI Act* stabilisce che «I fornitori di sistemi di IA ad alto rischio che ritengono o hanno motivo di ritenere che un sistema di IA ad alto rischio da essi immesso sul mercato o messo in servizio non sia conforme al presente regolamento adottano immediatamente le misure correttive necessarie per rendere conforme tale dispositivo, ritirarlo o richiamarlo, a seconda dei casi. Essi informano di conseguenza i distributori del sistema di IA ad alto rischio in questione e, ove applicabile, il rappresentante autorizzato e gli importatori».

Si noterà che, anche in questo frangente, la bozza di regolamento non brilla per precisione. Le finalità della disciplina paiono evidenti: evitare o ridurre al minimo la deviazione del prodotto intelligente rispetto allo spazio di “rischio consentito” delineata dall'*AI Act*. Il problema, come abbiamo già evidenziato, è che la suddetta area di *erlaubtes risiko* non pare configurata con sufficiente chiarezza.

In ogni caso, i profili di incertezza che emergono non sono molto diversi da quelli che già caratterizzano la disciplina generale in tema di obblighi *post market* del produttore – che pure sarà applicabile (e già lo è) nel settore dell'intelligenza artificiale. La ricostruzione dei principali orientamenti dottrinali sul punto potrebbe perciò risultare utile anche come modello ermeneutico per un'adeguata implementazione dell'*AI Act*.

Il codice del consumo, così come le norme che disciplinano nello specifico il settore produttivo d'interesse, delineano una serie di obblighi di controllo, monitoraggio,

scambi di informazione con le autorità amministrative (art. 104, co. 6 ss., cod. cons.) ed intervento (*sub specie* di ritiro, di richiamo e di informazione al consumatore, vd. art. 104, co. 3, cod. cons.), nel caso in cui emergano *rischi ulteriori* rispetto a quelli che la cristallizzazione delle norme cautelari mirava a prevenire. È il caso – a cui abbiamo dedicato alcune riflessioni nel precedente capitolo (vd. *supra*, cap. II, § 3.3.1.) – del c.d. *rischio da sviluppo*, che esclude la responsabilità civile del produttore per il danno cagionato dal prodotto se «lo stato delle conoscenze scientifiche e tecniche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso» (art. 118, lett. e, cod. cons., e art. 7, lett. e), 85/374/CEE).

Pur esimendo il produttore dall'obbligo di risarcimento del danno causato dal prodotto difettoso, il rischio da sviluppo, nella prospettiva penalistica oggetto di indagine, costituisce un rischio che fuoriesce dall'area del rischio consentito, intesa come quello spazio di impunità “programmato”, frutto di bilanciamento politico tra istanze contrapposte (vd. *supra*, § 4.5.2.); di conseguenza, l'emersione di un “rischio da sviluppo” determina il l'insorgere di un obbligo di attivazione per *evitare* o *ridurre al minimo* il rischio sopravvenuto.

Gli obblighi previsti dall'art. 104, co. 6 ss., cod. cons. – finalizzati, lo ribadiamo, ad evitare il verificarsi di eventi lesivi non previsti nella originaria area di “rischio consentito” – appaiono tuttavia indeterminati, non chiarendo con precisione né i presupposti né il contenuto dei doveri *post market* del produttore.

In particolare, l'indeterminatezza concerne:

i) i criteri in base ai quali stabilire se e quando un rischio ha travalicato l'area del rischio consentito;

ii) la scelta tra le varie opzioni individuate dalla norma cautelare (e, in particolare, i parametri in base ai quali stabilire quando sia doveroso il *ritiro del prodotto*).

Quanto al profilo *sub i)*, si presenta innanzitutto un'ipotesi non controversa, che concerne i casi in cui vi siano evidenze empiriche circa l'emersione di eventi lesivi che non erano stati presi in considerazione nella definizione di un'area di rischio lecito (o, comunque, non nell'entità effettivamente registratasi). Si pensi, ad esempio, alla pubblicazione di numerosi e autorevoli studi scientifici che accertino l'esistenza di un nesso di causalità tra l'assunzione di un farmaco e l'insorgere di una patologia: è ovvio

che, in un caso di questo tipo, il produttore sarà tenuto ad adottare le misure necessarie all'eliminazione del rischio.

Il quesito rivela invece tutta la sua complessità nei contesti di *incertezza scientifica* – quando, introdotto un prodotto conforme alla normativa cautelare di riferimento, si cominci a sospettare una pericolosità del prodotto *ulteriore* rispetto a quella presa in considerazione dagli *standard* cautelari³⁰⁶. A tal proposito, la dottrina ritiene che un mero sospetto non possa fondare l'obbligo di adottare misure di intervento più stringenti, ma che, a tal fine, sia di converso necessaria la sussistenza di *univoci segnali di incipiente fallimento* della regola cautelare, derivanti eventualmente anche da indici di fatto, non necessariamente dalla dimostrazione scientifica di una provata correlazione tra impiego del prodotto e danno³⁰⁷. Criteri-guida nella valutazione di tale univocità possono essere considerati, in primo luogo, la *frequenza* e la *diffusione* di *eventi lesivi* connessi all'uso del prodotto, in relazione alla *durata della sua commercializzazione* e alla *penetrazione sul mercato*; in secondo luogo, si dovrà verificare se esistano delle spiegazioni causali alternative rispetto ai danni verificatisi³⁰⁸.

Se una simile prospettiva, com'è evidente, si espone ad alcune critiche di non conformità al principio di *determinatezza* della fattispecie colposa³⁰⁹, essa, d'altra parte, consente di non ignorare – ancorandosi ad una pretesa aderenza alla tradizione del diritto penale liberale – le istanze di tutela della salute e dell'integrità fisica dei cittadini nella società del rischio.

Una volta acquisita consapevolezza circa il prevedibile fallimento della regola cautelare codificata, la questione che si pone (*sub ii*) è quella di stabilire quale sia la *misura più adeguata imposta dalla diligenza*: in base a quali criteri dovrà decidere, il produttore, se sia necessario il ritiro o il richiamo del prodotto o se, invece, sia sufficiente

³⁰⁶ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 398.

³⁰⁷ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 398-399.

³⁰⁸ C. PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, cit., p. 538 ss. Tali criteri sono in realtà elaborati dall'Autore come presupposti per la configurazione, *de jure condendo*, del cd. *illecito di rischio*, ossia di un tipo di illecito che prescinde dall'ascrizione nomologica dell'evento, sostituita dalla "propensione al rischio" di una determinata condotta. Tale soluzione consentirebbe di mantenere integre le categorie del diritto penale classico, evitando che queste vengano flessibilizzate – per presunte ragioni di giustizia sostanziale – dalla giurisprudenza. In argomento, per alcuni rilievi critici, vd. G. DE FRANCESCO, *L'imputazione del reato e i tormenti del penalista*, in M. Bertolino – G. Forti (a cura di) *Scritti per Federico Stella*, vol. I, Jovene, 2007, p. 529.

³⁰⁹ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 399.

fornire le informazioni di rilievo al consumatore e alle autorità amministrative di riferimento? Riportandoci alla logica già accennata precedentemente, l'obiettivo del produttore – e il criterio che dovrà utilizzare il giudice nella ricostruzione *ex post* dell'agente modello – dovrà essere quello di «ripristinare il livello di sicurezza al quale mirava la regola fallita»³¹⁰; così, se si trattava di una regola cautelare propria, la condotta diligente consisterà nella neutralizzazione del rischio; se, al contrario, si trattava di una regola impropria, la condotta diligente consisterà nell'adozione di quelle misure che riportano il rischio a quello *consentito* dalla norma cautelare fallita.

Solo in mancanza di cautele idonee, rispettivamente ad azzerare o ridurre adeguatamente il rischio, la condotta diligente consisterà nel ritiro o nel richiamo del prodotto³¹¹.

4.5.2.3. Intelligenza artificiale e rischio consentito: verso un approccio “quantitativo”?

Nel settore della produzione di sistemi di i.a. (ma, in generale, nel settore produttivo *tout court*), un “fallimento” delle cautele positive potrà eventualmente emergere sul piano *astratto*: perché il *rischio reale* è *più elevato* di quello che era stato previsto in sede di positivizzazione della cautela, o perché si è in presenza di rischi *ulteriori* rispetto a quelli che erano stati presi in considerazione, o, ancora, perché – a parità tra i rischi previsti e quelli reali – vi è un difetto di tecnica normativa che impedisce di raggiungere la riduzione del rischio che costituiva l'obiettivo della codificazione.

Quali sono i criteri, tuttavia, per verificare l'avvenuto fallimento delle regole cautelari scritte? Se è vero che la redazione di norme cautelari, con riferimento alle attività autorizzate pericolose, mette preventivamente in conto il verificarsi di una quota di eventi lesivi, non può essere sottovalutato il fatto che tale quota non è praticamente mai esplicitata dal legislatore/decisore pubblico. A partire da quale momento, insomma, il produttore dovrà attivarsi per fornire informazioni al consumatore e alle autorità pubbliche, ed eventualmente ritirare il prodotto?

³¹⁰ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 383.

³¹¹ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 383.

Si tratta di una questione che – salvo gli sparuti criteri che abbiamo indicato nel precedente paragrafo – ancora non ha trovato una risposta univoca, e che contribuisce a rendere sfumati ed imprevedibili i contorni del rischio consentito.

Tradizionalmente, la dottrina ha ritenuto che sulla liceità del rischio incidano la rilevanza dei beni giuridici in gioco, il numero dei soggetti che potrebbero essere colpiti dall'offesa, l'utilità sociale dell'attività svolta, il grado di probabilità dell'evento dannoso³¹². Particolare importanza, nello specifico, riveste quest'ultimo aspetto³¹³, nel senso che l'esercizio di attività pericolose non potrà essere ammesso laddove il rischio di verifica dell'evento sia molto alto³¹⁴.

A tal proposito, in relazione ai prodotti tradizionali, è stato rilevato che «l'indisponibilità di una cornice nomologica idonea a determinare il grado di probabilità dell'evento frustra in radice la possibilità di reperire norme cautelari preventive, funzionali al contenimento di una tipologia di danni, e pregiudica, perciò, la stessa legittimità del riferimento all'istituto del rischio consentito»³¹⁵.

L'inconoscibilità dei meccanismi causali interni alla *black box* si aggrava in relazione ai sistemi di i.a. Tuttavia, la possibilità di registrare le *performance* dei sistemi di *machine learning* – unita all'enorme accrescimento delle capacità computazionale degli artefatti tecnologici – può determinare un incremento nella *calcolabilità* degli errori e, di conseguenza, degli eventi lesivi che possono derivarne. È questa una differenza enorme rispetto ai prodotti tradizionali. La gran parte della fenomenologia in materia di danno da prodotto aveva infatti ad oggetto prodotti dei quali non si poteva stabilire con certezza: (i) la lesività; (ii) il nesso di causalità tra (presunta) lesività del prodotto ed evento lesivo. *Mutatis mutandis*, è tra l'altro quanto accadeva con riferimento ai *toxic cases*.

Dei sistemi di i.a., invece, sappiamo che possono prendere decisioni erranee e, di conseguenza, causare danni (di natura diversa, a seconda del settore in cui operi l'algoritmo); non possiamo, invece, prevedere quali sono gli *input* che potranno determinare un *decision making* fallace, stante l'assenza di leggi scientifiche esplicative

³¹² C. PIERGALLINI, *Danno da prodotto*, cit. p. 243; vd. anche V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 170.

³¹³ C. PIERGALLINI, *Danno da prodotto*, cit., p. 243; G. FORTI, *Colpa ed evento*, cit., p. 255.

³¹⁴ G. FORTI, *Colpa ed evento*, cit., p. 255.

³¹⁵ C. PIERGALLINI, *Danno da prodotto*, cit., p. 243.

(vd. *supra*, § 4.3., in questo cap.). Come abbiamo sottolineato precedentemente, quando abbiamo affrontato il problema della causalità, il decorso causale lesivo, all'interno dell'algoritmo, potrà essere ricostruito *ex post* con molte difficoltà (vd. *supra*, § 4.3., in questo capitolo). Minime sono invece le possibilità che si possa pervenire ad una compiuta generalizzazione del comportamento algoritmico, in grado di spiegare e prevedere con precisione la reazione dell'algoritmo a determinati *input* – per il semplice fatto che i sistemi di *machine learning* sono *unpredictable by design*.

Sebbene l'imprevedibilità del *decision making* algoritmico potrebbe restare a lungo un problema irrisolto – a meno di non voler rinunciare ai benefici dell'intelligenza artificiale, spesso frutto proprio di una sorta di “creatività”³¹⁶ – una prospettiva più vicina (ma pur sempre futura) potrebbe riguardare la possibilità tecnica di registrazione sistematica *ex post* delle decisioni assunte dai sistemi di i.a. Tali rilevazioni statistiche – di cui, previamente, abbiamo osservato l'irrelevanza ai fini dell'accertamento causale, stante l'assenza di qualsivoglia connotato esplicativo generalizzante (vd. *supra*, § 4.3.1.) – potrebbero tuttavia costituire un valido parametro per verificare se le norme cautelari codificate siano o meno “fallite”, nel senso che abbiamo illustrato nel precedente paragrafo. L'implementazione delle tecniche di *Explainable Artificial Intelligence* (c.d. XAI, vd. *supra*, cap. I, § 5.3), inoltre, agevolando la ricostruzione del nesso eziologico *ex post*, potrebbe inoltre rendere ancora più agevole la ricostruzione statistica del tasso di affidabilità ed accuratezza dei sistemi di i.a.

Ci spieghiamo meglio. Abbiamo osservato che il decisore pubblico individua una serie di norme cautelari finalizzate a prevenire il verificarsi di un determinato rischio. Abbiamo altresì osservato che l'agente, qualora emerga un evidente fallimento delle cautele positive (o perché il *rischio reale* è *più elevato* di quello che era stato previsto, o perché si è in presenza di *rischi ulteriori*), deve porre in essere tutte quelle attività necessarie a riportare il rischio al livello che costituiva l'esito del bilanciamento.

Ora, come abbiamo rilevato, uno dei problemi che si pone è quello di comprendere quando gli eventi lesivi costituiscano prevedibile concretizzazione del rischio consentito

³¹⁶ Sul *trade off* tra accuratezza delle previsioni algoritmi e spiegabilità vd. A.J. LONDON, *Artificial intelligence and black-box medical decisions: accuracy versus explainability*, in *The Hastings Center Report*, 49(1), 2019, p. 15 ss.; R. CARUANA E AL., *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission*, in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015

e quando, invece, essi oltrepassino tale limite. Ecco che, a nostro parere, la prospettiva di registrazione algoritmica *ex post* degli incidenti potrebbe rivelarsi uno strumento utile per valutare il superamento quantitativo della soglia di rischio considerato lecito. Non è necessario che tale verifica abbia i crismi di un rigoroso accertamento del nesso di causalità condotto con gli *standard* penalistici: la registrazione dovrebbe essere tutt'al più in grado di *isolare* gli errori algoritmici – senza che, tuttavia, sia necessario individuare il preciso *input* che ha determinato il verificarsi dell'errore algoritmico e, dunque, dell'evento lesivo.

Poniamo il caso, ad esempio, che il legislatore stabilisca un apparato di norme cautelari rigide per lo sviluppo e la produzione di *software* di guida autonoma (es. modelli di raccolta e gestione dati, addestramento, sperimentazione etc), esplicitando che la fissazione di tali norme è funzionale a mantenere il tasso di incidenti al di sotto della soglia dello 0,0001% rispetto all'utilizzazione del veicolo. Nel caso in cui, nonostante il rispetto delle norme cautelari rigide, il produttore verifichi che il tasso di sinistri derivanti dall'utilizzo del *software* sia superiore rispetto alla soglia del rischio lecito, questi dovrà adottare cautele ulteriori per minimizzare i rischi e, eventualmente, richiamare o ritirare il prodotto.

Il superamento del tasso stabilito in via legislativa non determinerebbe l'automatica imputazione degli eventi lesivi "ulteriori" al produttore, ma dovrebbe essere considerato come segnale di *incipiente fallimento delle norme cautelari scritte*, nel senso descritto al paragrafo precedente: dovrebbe, dunque, indurre il produttore ad adottare cautele più pregnanti rispetto a quelle previste dalla normativa esistente. Un simile approccio è stato proposto anche da una parte della dottrina civilistica, con riferimento al tema contiguo dell'accertamento di un difetto di *design* (vd. *supra*, cap. II, § 3.3.3.)³¹⁷. In particolare, tale orientamento ipotizza di valutare la difettosità del prodotto intelligente attraverso la comparazione tra la prestazione complessiva del sistema di i.a. che ha causato il danno e la *performance* di un modello di algoritmo mediamente sicuro. A tal fine, si prospetta l'opportunità di individuare una soglia di riferimento – relativa al rapporto tra eventi lesivi verificatisi e utilizzo complessivo del *software* – al di sopra della quale l'algoritmo dovrebbe essere considerato difettoso.

³¹⁷ Vd. in particolare J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, Baden-Baden, 2019, p. 63 ss.

Ovviamente, in ipotesi, una regolamentazione di questo tipo presupporrebbe, da un lato, che il tasso di accuratezza dell'algoritmo di *machine learning* sia valutato *prima* dell'immissione del prodotto sul mercato (sulla base di dati acquisiti in fase di sperimentazione), attraverso autocertificazione o, preferibilmente, – per i prodotti che possono comportare pericoli per beni giuridici di assoluto rilievo, quali la vita e l'incolumità fisica – attraverso un procedimento di autorizzazione amministrativa; dall'altro, il presupposto per l'efficacia di un sistema simile dovrebbe essere l'obbligo – eventualmente sanzionato penalmente – di condivisione con le autorità competenti dei dati relativi agli eventuali incidenti cagionati dai sistemi di i.a.³¹⁸

Una simile soluzione avrebbe il vantaggio di spostare il *focus* della valutazione sulla sicurezza del prodotto intelligente, dalla prestazione singola – che, come abbiamo rilevato, potrebbe apparire inefficiente e inspiegabile in relazione al particolare episodio lesivo (vd. *supra*, cap. I, § 2.2.) – alla sua *performance* complessiva: una traslazione valutativa che, a nostro parere, è indispensabile se si vuole godere dei benefici derivanti dallo sviluppo e dalla commercializzazione dei sistemi di i.a., senza che singoli eventi lesivi possano mettere in discussione la sicurezza del prodotto.

Analoghe soluzioni “quantitative”, tra l'altro, contribuiscono a definire l'area del rischio lecito in altri settori caratterizzati da elevati gradi di incertezza: ci riferiamo, in particolare, a quello della produzione di farmaci e, nello specifico, di vaccini. Se prendiamo in considerazione la popolazione di vaccinabili (escludendo, dunque, le persone a rischio, a causa di patologie pregresse), è *prevedibile* che si verifichino eventi avversi, ma è *imprevedibile* stabilire quando, come, nei confronti di chi. La farmacovigilanza ha proprio l'obiettivo di raccogliere, valutare e analizzare gli eventi avversi che seguono l'assunzione di farmaci e, nel caso dei vaccini, l'immunizzazione. Nel caso in cui si verifichino eventi avversi in numero eccedente rispetto a quanto emerso nelle fasi sperimentali e di autorizzazione del vaccino (oppure si verifichino eventi avversi *diversi* rispetto a quelli preventivabili), il produttore dovrà fornire ulteriori informazioni nel c.d. bugiardino o, nei casi più gravi, ritirare il prodotto; prima ancora,

³¹⁸ Cfr. G. FORTI, “Accesso” alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione, cit., *passim.*; vd. anche C.E. PALIERO, Responsabilità per danno da prodotto tra persona fisica ed ente collettivo, cit., p. 294.

l'autorità regolatoria potrebbe dover procedere alla sospensione del prodotto³¹⁹. Nel caso in cui, invece, gli eventi avversi si mantengano sotto una soglia “fisiologica”, così come verificata in sede di autorizzazione, questi rientreranno nell'area del rischio consentito e non saranno, dunque, imputabili al produttore³²⁰.

Ovviamente, quello dei vaccini è un settore molto particolare, in cui vi sono questioni di bilanciamento prima di tutto interne al diritto alla salute (benefici dell'immunizzazione/rischio di eventi avversi), e soltanto secondariamente attinenti allo svolgimento di un'attività economica³²¹. I medesimi principi, tuttavia, potrebbero essere traslati nel campo dell'intelligenza artificiale, nei settori in cui lo sviluppo e la messa in commercio di sistemi di i.a. può avere effetti complessivamente benefici sulla tutela del diritto alla vita e all'integrità fisica. Si pensi all'utilizzo, in ambito medico-sanitario, dei CDSS, che potrebbero identificare patologie in fasi molto precoci ed individuare trattamenti ritagliati sulle precise caratteristiche fisico-biologiche del paziente; si pensi, ancora, ai veicoli a guida autonoma, che promettono di ridurre del 90% gli incidenti su strada.

Il principale limite della presente prospettazione – oltre a quello tecnico, che potrebbe tuttavia essere risolto in un futuro prossimo – è evidentemente di natura politica: quale regolatore, infatti, potrebbe mai esplicitare in maniera così diretta il bilanciamento tra esercizio di un'attività economica e beni giuridici di estremo rilievo (quali, in ipotesi, vita ed integrità fisica)? Il rischio, d'altra parte, è che anche qualora si pervenisse all'individuazione di una soglia quantitativa, questa sia ispirata da logiche *iper-*

³¹⁹ In argomento, nella dottrina civilistica, vd. S. FOÀ, *I danni da vaccino Sars-Cov-2 tra obblighi, raccomandazioni e «solidarietà irrinunciabile»*, in *Resp. civ. prev.*, 2022, fasc. 4, p. 1070 ss.

³²⁰ La farmacovigilanza sui vaccini anti-covid offre, tra l'altro, un interessante modello di valutazione della probabilità circa il nesso di causalità tra immunizzazione e verifica dell'evento avverso. Il Comitato Consultivo Globale per la Sicurezza dei Vaccini (GACVS) dell'Organizzazione Mondiale della Sanità ha messo a punto un algoritmo che, tenendo conto di una serie di dati (relazione temporale fra la vaccinazione e la reazione segnalata; presenza di possibili spiegazioni alternative; precedenti evidenze di letteratura; frequenza dell'evento segnalato nella popolazione generale, anche non vaccinata; plausibilità biologica) valuta la *probabilità* della sussistenza del nesso tra assunzione del vaccino ed evento lesivo. Vd. a tal proposito il report pubblicato dalla AGENZIA ITALIANA DEL FARMACO, *Tredicesimo rapporto sulla sorveglianza dei vaccini anti-COVID-19 (27/12/2020 – 26/09/2022)*, disponibile a questo link: https://www.aifa.gov.it/documents/20142/1315190/Rapporto_sorveglianza_vaccini_COVID-19_13.pdf, p. 2.

³²¹ Emerge poi il principio di solidarietà (art. 2), nei casi in cui la vaccinazione sia obbligatoria o comunque fortemente raccomandata (vd. il caso del vaccino per il Covid-19).

cautelative, che determinerebbero la necessità di adottare sempre cautele *ulteriori* rispetto a quelle positive.

Va infine sottolineato che l'ultroneità del rischio rispetto alla soglia consentita può emergere non soltanto sotto il profilo quantitativo, ma anche sotto quello qualitativo. Si pensi, ad esempio, ad un veicolo a guida autonoma che si comporti in maniera assolutamente estranea rispetto a quanto verificato nelle fasi di sperimentazione. È quanto potrebbe essere accaduto – ma si tratta di una mera supposizione avente funzione esplicativa, dal momento che non possediamo informazioni sufficienti a tal proposito – nel novembre 2022, quando un veicolo Tesla, mentre percorreva le strade di una città cinese, ha improvvisamente accelerato, percorrendo quasi tre chilometri al massimo della velocità, causando due morti e tre feriti, prima di scontrarsi contro un edificio³²². Poniamo il caso che, nella fase di *test*, il *software* di guida autonoma avesse mostrato una gamma di fenomenologie di errori (es. mancato riconoscimento dei segnali di stop o dei semafori; mancata concessione delle precedenza dovute; mancato utilizzo della freccia prima di curvare, etc.), nella quale non rientrava l'improvvisa accelerazione. In un caso simile, potrebbe configurarsi un segnale di *fallimento* – ancorché parziale – *delle regole cautelari positive*, che, in ipotesi, non contemplavano la minimizzazione del “rischio da accelerazione”.

Sull'identificazione di un eventuale *superamento* del rischio consentito incideranno senza dubbio gli obblighi di monitoraggio *post market*, previsti dalla normativa generale in materia di sicurezza dei prodotti (vd. art. 104, co. 6 ss., cod. cons.), dalla normativa di settore e, infine, dall'*AI Act* (vd. artt. 61-62): su questi aspetti si rinvia a quanto già rilevato *supra*, al § 4.5.2.2. di questo capitolo.

³²² Il video, piuttosto impressionante, è disponibile sul sito di *Rai News*, a questo link: <https://www.rainews.it/video/2022/11/tesla-lanciata-a-folle-velocit-le-telecamere-a-circuito-chiuso-riprendono-lincidente-mortale-06a60958-096a-4005-9213-50c36fcb7641.html> (Red. Rai News, *Tesla lanciata a folle velocità: le telecamere a circuito chiuso riprendono l'incidente mortale*, 14 novembre 2022).

4.5.3. Le regole cautelari non scritte: la colpa generica.

Come abbiamo cercato di mostrare, la tendenziale impermeabilità dell'area di rischio consentito alla colpa generica è costantemente messa in discussione dalla giurisprudenza, che tende spesso a considerare la colpa generica come «un rinforzo onnipresente di quella specifica, atta a riempire ogni possibile lacuna preventiva che quest'ultima possa presentare»³²³. In ogni caso, anche qualora si aderisse all'impostazione che mira a creare un'area di impunità per il soggetto che si conformi a regole cautelari scritte che siano *esaustive*, uno spazio per la valutazione della *prevedibilità* e dell'*evitabilità* dell'evento sarà pur sempre irrinunciabile, qualora ricorra una delle seguenti circostanze:

a) le norme cautelari codificate siano *elastiche*, rimandando, di fatto, ad una valutazione della colpa generica dell'agente (es. costrutti normativi di questo tipo: *l'agente, per quanto possibile, deve adottare misure idonee a neutralizzare o a ridurre il rischio di verifica di eventi lesivi per la salute*);

b) il “fallimento” (in concreto o in astratto) delle norme cautelari positive sia riconoscibile.

Ecco che allora, in queste due ipotesi, fanno nuovamente capolino le problematiche e le incertezze che l'istituto del rischio consentito mirava a neutralizzare: al di fuori del bilanciamento costi-benefici operato dal legislatore o, comunque, dagli organi rappresentativi degli interessi dei consociati, come individuare compiutamente l'orbita della diligenza? Quali eventi avrebbero dovuto essere evitati e, quali, invece, sono da ricondursi “all'imponderabile”³²⁴?

Bisogna inoltre sottolineare che un'esaustiva formalizzazione delle norme cautelari è opera ormai impensabile, nell'accresciuta complessità della società tecnologica in cui viviamo³²⁵ (salvo, forse, per quanto concerne la tecnica dei valori-soglia), stante la necessità di adeguamento delle norme astratte alle peculiarità del caso

³²³ F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1109. Vd. ad es. Cass., sez. IV, 20 marzo 2000, n. 3567, in *Giur. it.*, 2001, 1709 ss., che riconosceva la colpevolezza del datore di lavoro, nonostante questi avesse rispettato i valori-soglia in materia di esposizione all'amianto; vd. anche Cass., Sez. IV, 4 novembre 2010, n. 38991. Per una rassegna giurisprudenziale sul rapporto tra colpa generica e colpa specifica v. M. GROTTI, *Principio di colpevolezza*, cit., p. 79 ss.

³²⁴ Così V. ATTILI, *L'agente-modello “nell'era della complessità”*, cit., p. 1244-1245.

³²⁵ V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 148-149.

concreto e l'esigenza di districarsi tra norme cautelari provenienti da diverse fonti, talvolta contraddittorie tra loro. Un giudizio sulla valutazione della diligenza generica dell'agente – ancorché “guidata” dalle valutazioni di bilanciamento impresse dal decisore pubblico nelle cautele scritte – sarà dunque spesso indispensabile per il giudice. La codificazione, in questa prospettiva, più che un ruolo di *vincolo insuperabile* alle valutazioni giudiziali, ha la funzione di «definire l'ambito all'interno del quale si muove la problematica della colpa rispetto a un determinato rischio»³²⁶.

Tra l'altro, va nuovamente ribadito che, stando al panorama normativo attuale, nel settore dell'intelligenza artificiale pare ancora lontana la precisa definizione di *standard* cautelari *rigidi* ed *esaustivi*, idonei ad erigere solidi confini dell'area del rischio consentito (vd. *supra*, § 4.5.1.) – con la conseguenza che, perlomeno nel futuro prossimo, la valutazione della colpa in capo al produttore sarà irrimediabilmente legata al parametro dell'agente modello.

4.5.3.1. La prevedibilità dell'evento: è rimproverabile la prevedibilità dell'imprevedibilità algoritmica?

La produzione di sistemi di i.a. costituisce un vero e proprio “shock da modernità” – per utilizzare un'espressione cara a Federico Stella³²⁷ – soprattutto nel settore dell'imputazione colposa. La consapevole messa in commercio di prodotti di per sé imprevedibili ed incontrollabili mette a dura prova la tenuta delle categorie della *prevedibilità* e dell'*evitabilità* dell'evento, che costituiscono i capisaldi dell'accertamento della colpa generica.

Partendo dalla constatazione che il funzionamento dell'intelligenza artificiale è oggi in una certa misura ignoto agli stessi programmatori e progettatori, cercheremo di capire, in questa sede, se gli eventi lesivi derivanti dagli algoritmi di i.a. – *prevedibili* per quanto concerne l'*an*, ma *imprevedibili* con riferimento al *quantum* e al *quomodo* – possano rientrare nell'area di rischio che deve essere necessariamente neutralizzata dal produttore. In particolare, ci soffermeremo su due questioni connesse, già oggetto di

³²⁶ R. BARTOLI, (voce) *Fonti della colpa*, cit., p. 531.

³²⁷ F. STELLA, *Giustizia e modernità*, cit., p. 292.

ampie riflessioni da parte della dottrina: i) quale grado di dettaglio debba avere il giudizio di *prevedibilità* dell'evento; ii) se il principio di precauzione possa avere un ruolo nella concreta individuazione delle cautele costituenti la colpa generica.

La definizione dei confini del giudizio di prevedibilità dell'evento lesivo algoritmico pare uno degli aspetti più rilevanti in relazione all'accertamento della colpa del produttore³²⁸. Anche ai fini dell'accertamento della colpa – così come per la causalità – è necessario fare riferimento a leggi scientifiche. Mentre, tuttavia, la verifica del nesso causale avviene con valutazione *ex post*, in relazione alla colpa deve effettuarsi una valutazione *ex ante*³²⁹: la differenza, com'è stato rilevato, è che, nel primo caso, si ha che fare con una legge *descrittiva*, nel secondo caso con una legge *deontica*³³⁰. Il rischio da scongiurare è che l'opera di integrazione del precetto colposo venga effettuata dal giudice in un'ottica *ex post*, attraverso l'individuazione di una regola cautelare *inesistente* al momento della condotta, sulla base del mero convincimento che l'imputato, nella situazione concreta, avrebbe dovuto comportarsi in una maniera diversa: come sottolineava Giorgio Marinucci, citando un antico proverbio popolare, *del senno di poi son piene le fosse*³³¹.

Quanto all'*oggetto* del giudizio di prevedibilità dell'evento lesivo, la valutazione potrebbe, in teoria, oscillare tra due estremi: da un lato, il giudizio potrebbe parametrarsi sull'evento concreto realizzatosi *hic et nunc*, in tutte le precise e contingenti caratteristiche storico-fattuali (es. morte del pedone investito da *autonomous vehicle*, realizzatasi per complicanze polmonari conseguenti all'incrinatura di una costola, curata con farmaco errato); dall'altro, il parametro potrebbe essere quello di una classe iper-generica di eventi (es. la morte del pedone)³³². La maggiore o minore ampiezza del giudizio di prevedibilità dipende principalmente dai criteri attraverso i quali l'evento

³²⁸ In questo senso L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 834; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1762.

³²⁹ Vd. per tutti G. FORTI, *Colpa ed evento*, cit., p. 210; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 49 ss.

³³⁰ C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione*, cit., p. 395.

³³¹ G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 48; in argomento vd. anche P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 55; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato*, cit., p. 403; V. ATTILI, *L'agente-modello "nell'era della complessità"*, cit., p. 1269; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 319-320.

³³² In questo senso, G. CIVELLO, (voce) *Prevedibilità e reato colposo*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 1017, da cui abbiamo tratto ispirazione per gli esempi citati.

viene descritto (o ri-descritto) da parte dell'organo giudicante, secondo una relazione che potrebbe essere definita di "proporzionalità inversa"³³³: minore è la precisione con cui si seleziona l'evento assunto come parametro, maggiore sarà il rischio di consentire l'accesso a valutazioni "retrospettive" della colpa penale, con conseguente estensione dell'area di responsabilità; maggiore è l'attenzione alle caratteristiche concrete del fatto storico verificatosi, minore sarà lo spazio per il rimprovero penale³³⁴.

Tra i due estremi citati, una mediazione pare indispensabile³³⁵: la complessità nella definizione di un criterio univoco pare tuttavia l'ennesimo sintomo di un *deficit* di determinatezza della fattispecie colposa.

Com'è stato osservato, la tendenza della giurisprudenza, negli ultimi decenni, è stata spesso quella di espandere l'area della responsabilità colposa, soprattutto in relazione all'esercizio di attività pericolose in situazioni di incertezza. In particolare, si è assistito ad un graduale percorso di "astrazione" nella ri-descrizione dell'evento, con uno slittamento dogmatico, in seno al giudizio di prevedibilità, dall'evento di danno alla *situazione* di danno³³⁶. La giurisprudenza è giunta spesso a ricostruire la regola cautelare retrospettivamente, ponendo, alla base del giudizio di prevedibilità, «quelle stesse leggi causali utilizzate per stabilire una relazione di causa ed effetto tra eventi»³³⁷, e confondendo così accertamento concreto del rapporto di causalità e leggi scientifiche disponibili *ex ante*.

³³³ In questo senso, G. CIVELLO, (voce) *Prevedibilità e reato colposo*, cit., p. 1017.

³³⁴ In questo senso, G. CIVELLO, (voce) *Prevedibilità e reato colposo*, cit., p. 1017.

³³⁵ Per un'accurata definizione dell'alternativa che si pone all'interprete vd. G. CIVELLO, (voce) *Prevedibilità e reato colposo*, cit., p. 1017; A. NAPPI, *La prevedibilità nel diritto penale*, Edizioni Scientifiche Italiane, 2020, p. 311; K. SUMMERER, (voce) *Evitabilità dell'evento e comportamento alternativo lecito*, in M. Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, cit., p. 509-510.

³³⁶ In questo senso, G. CIVELLO, (voce) *Prevedibilità e reato colposo*, cit., p. 1017; sulla patologica ridescrizione dell'evento *ex post* vd. S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 322; C. PIERGALLINI, *Il paradigma della colpa nell'età del rischio*, cit., p. 1692.

³³⁷ G. FORTI, *Colpa ed evento*, cit., 208; vd. anche V. ATTILI, *L'agente-modello "nell'era della complessità"*, cit., p. 1269-1270; F. SGUBBI, *Il diritto penale incerto e inefficace*, in *Riv. it. dir. proc. pen.*, 2001, p. 1197; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato*, cit., p. 403; G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 48 ss. Per un interessante studio dell'incidenza dei *bias* cognitivi sui giudizi di prognosi postuma vd. S. GROSSO, *Il giudizio di prevedibilità dell'evento e l'incidenza dell'hindsight bias e dell'outcome bias sul giudizio di responsabilità colposa*, in *Riv. it. med. leg.*, 2016, fasc. 2, p. 583 ss.; D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi*, cit., p. 37 ss.

Nel 1971, ad esempio, nella sentenza sulla vicenda del disastro del Vajont si poteva leggere: «nell'esercizio lecito di attività pericolose, ai fini del giudizio di prevedibilità, deve aversi riguardo alla potenziale idoneità della condotta a dar vita a una *situazione di danno o di pericolo*, e non anche alla specifica rappresentazione ex ante dell'evento dannoso, quale si è concretamente verificato in tutta la sua gravità ed estensione. Perché sorga l'obbligo di una particolare prudenza e cautela – proprio perché si tratta di attività essenzialmente e per definizione pericolose – è, in altri termini, sufficiente la prevedibilità di una particolare situazione di danno o di pericolo, senza che si debba necessariamente estendere agli ulteriori effetti che, in ipotesi estremamente pessimistica, potrebbero scaturire dalla situazione potenziale, in particolare quando, come nella specie, l'attività in questione sia profondamente condizionata dalla incidenza di forze naturali, non sempre conoscibili e talora scatenate ed inarrestabili»³³⁸.

Simili principi furono poi espressi nell'ambito dei casi giudiziari relativi al disastro del Seveso³³⁹ e dello Stava³⁴⁰, per poi confluire nella giurisprudenza sulla responsabilità per esposizione da amianto. Emblematica, in tal senso, è una sentenza emessa dal Pretore di Torino nel 1995, in tema di morte conseguente all'esposizione all'amianto³⁴¹. In quel caso, un datore di lavoro era stato condannato per omicidio colposo per non aver adottato le misure idonee ad evitare l'esposizione dei lavoratori all'amianto, causa della morte per mesotelioma pleurico – a trent'anni dalla condotta – di un dipendente. Ai nostri fini, ciò che qui rileva è che, all'epoca del fatto, la nocività dell'amianto era nota come possibile causa di asbestosi, non di forme tumorali.

Ancora, si pensi alla sentenza di appello³⁴² e di Cassazione³⁴³ sul caso del Petrolchimico di Porto Marghera, in cui i giudici – ribaltando la decisione del Tribunale di Venezia³⁴⁴ – affermavano che, per ritenere sussistente la colpa in capo ai vertici della società, per la morte da tumore di alcuni operai che ivi lavoravano, era sufficiente la conoscibilità del fatto che la sostanza CVM, impiegata negli stabilimenti, fosse idonea a

³³⁸ Cass., sez. IV, 25 marzo 1971, n. 810, Biadone; i corsivi sono aggiunti da chi scrive.

³³⁹ Cass., sez. IV, 23 maggio 1986, in *Cass. pen.*, 1988, p. 1250 ss.

³⁴⁰ Cass., sez. IV, 6 dicembre 1990, n. 4793, Bonetti, in *Cass. pen.*, 1992, p. 2754.

³⁴¹ Pretura Torino, 9 febbraio 1995, in *Riv. it. dir. e proc. pen.*, 1997, p. 1447; sulla vicenda (e sulla successiva sentenza della Corte d'Appello di Torino) si vd. la nota di C. PIERGALLINI, *Attività produttive e imputazione per colpa: prove tecniche di "diritto penale del rischio"*, cit.

³⁴² Corte app. Venezia, sent. 15 dicembre 2004, in *Riv. it. dir. proc. pen.*, 2005, p. 1670 ss.

³⁴³ Cass., sez. IV, 17 maggio 2006, n. 4675, Bartalini.

³⁴⁴ Trib. Venezia, sent. 22 ottobre 2001, Cefis, in *Riv. it. dir. proc. pen.*, 2005, p. 1670 ss.

provocare *malattie al fegato* (acrostevoli), non essendo necessario che fosse prevedibile che la sostanza potesse provocare *tumori al fegato*³⁴⁵. Così la Suprema Corte: «ben da prima dell'epoca della contestazione (che parte dal 1969), dovevano già ritenersi prevedibili *gravi danni alla salute* dei lavoratori esposti al cloruro di vinile e quindi doveva ritenersi sorto l'obbligo (anche se fossero mancate regole cautelari di origine normativa), per il datore di lavoro, di adottare le cautele necessarie per preservare i lavoratori dal rischio in esame, perché la tutela della salute umana costituisce obbligo primario di salvaguardia di un diritto costituzionalmente protetto (art. 32 Cost.)».

L'impostazione fatta propria dalla Corte di Cassazione nella sentenza sul petrolchimico di Porto Marghera è stata criticata dalla dottrina quasi unanime, poiché – nel sostituire, nel giudizio di prevedibilità, un generico evento di danno all'evento *hic et nunc* verificatosi – rende di fatto impossibile l'individuazione di una cautela diversa dall'astensione dall'esercizio dell'attività pericolosa, dal momento che «non esiste – e non potrebbe essere altrimenti – una regola cautelare dotata di un'efficacia preventiva generale»³⁴⁶. L'evaporazione del giudizio di prevedibilità si traduce in un abbraccio mortale della responsabilità oggettiva, dal momento che, di fatto, l'evento viene imputato all'agente sulla sola base della sua causazione materiale³⁴⁷.

Va comunque rilevato che la giurisprudenza di legittimità si è ormai assestata sul principio di diritto in base al quale «in tema di colpa, la necessaria prevedibilità dell'evento – anche sotto il profilo causale – non può riguardare la configurazione dello specifico fatto in tutte le sue più minute articolazioni, ma deve mantenere un certo grado

³⁴⁵ Sulla vicenda giudiziaria che ha coinvolto lo stabilimento petrolchimico di Porto Marghera – la più importante, nel nostro Paese, in materia di danni alla salute e all'ambiente – si rinvia alle approfondite riflessioni di S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 104 ss.; C. PIERGALLINI, *Il paradigma della colpa*, cit., p. 1670 ss.; F. D'ALESSANDRO, *Pericolo astratto e limiti-soglia*, cit., p. 180 ss.; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 132 ss.; A. NAPPI, *La prevedibilità nel diritto penale*, cit., p. 327 ss.

³⁴⁶ C. PIERGALLINI, *Il paradigma della colpa nell'età del rischio*, cit., p. 1692; ID., *Danno da prodotto*, cit., 445; vd. anche G. FORTI, *Colpa ed evento*, cit., 487: «caratteristica della regola cautelare è di prescrivere condotte atte a prevenire un evento descritto assai più concretamente rispetto al modello penale, e dunque caratterizzato da precise modalità o risultato di ben determinati decorsi causali».

³⁴⁷ C. PIERGALLINI, *Danno da prodotto*, cit., 444; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 320; G. CIVELLO, (voce) *Prevedibilità e reato colposo*, cit., p. 1017; C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto "a distanza" tra disvalore d'azione e disvalore d'evento*, cit., p. 236.

di *categorialità*, nel senso che deve riferirsi alla classe di eventi in cui si colloca quello oggetto del processo»³⁴⁸.

A nostro parere, nella valutazione circa la *prevedibilità*, da parte dell'agente modello, dell'evento lesivo derivante da sistema di i.a., l'alternativa sarà, per forza di cose, netta³⁴⁹. Qualora prevalga la concezione "espansiva", maggioritaria oggi in giurisprudenza, in base alla quale è sufficiente la prevedibilità di una determinata classe di eventi, la colpa del produttore di un sistema di i.a. potrebbe essere considerata sempre esistente. Nel caso di produzione di *healthcare robots*, ad esempio, la morte o le lesioni del paziente sottoposto ad operazione chirurgica sono di certo astrattamente prevedibili; allo stesso modo, il decesso del pedone investito da *self driving car* è prevedibile dal produttore di *autonomous vehicles*. Al contrario, qualora prevalga una concezione restrittiva, in cui il giudizio di prevedibilità deve coinvolgere l'evento effettivamente verificatosi, secondo decorsi eziologici noti, il produttore non sarà mai punibile, dal momento che l'evento *hic et nunc* realizzatosi non sarà mai prevedibile da parte del produttore, stante il carattere di intrinseca imprevedibilità del sistema di i.a.³⁵⁰.

4.5.3.2. Principio di precauzione e giudizio di prevedibilità dell'evento.

L'estensione del raggio d'azione del giudizio di prevedibilità dell'evento lesivo è da correlarsi all'affermazione, come principio di politica legislativa e – per quanto qui maggiormente interessa – di politica giudiziaria, del principio di precauzione. In particolare, in giurisprudenza si è spesso giunti a sovrapporre il piano dell'evento *prevedibile* con il piano dell'evento *non suscettibile di essere escluso*, così che sono stati ritenuti prevedibili anche quegli eventi che l'agente *non era in grado di escludere* al momento della condotta³⁵¹.

³⁴⁸ Cass., sez. un., 24 aprile 2014, n. 38343, in *Dir. pen. cont.*, 19 settembre 2014; conf. Cass., sez. IV, 25 giugno 2013, n. 35309.

³⁴⁹ Dello stesso avviso sono C. PIERGALLINI, *Intelligenza artificiale*, cit., 1762; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1208-1209.

³⁵⁰ Così C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1762-3; F. CONSULICH, *Flash offenders*, cit., p. 1054.

³⁵¹ G. CIVELLO, (voce) *Prevedibilità e reato colposo*, cit., p. 1019.

Nonostante le dichiarazioni in senso contrario, la già citata sentenza della Corte di Cassazione sul caso del petrolchimico di Porto Marghera³⁵² – nel ritenere che il *dubbio scientifico* sulla possibilità di eventi avversi *ulteriori* possa fondare un dovere cautelare – è intrisa da logiche precauzionali. Se in quel caso, tuttavia, era quantomeno accertata la dannosità – seppur di minor grado – della sostanza utilizzata negli stabilimenti, le sentenze sugli eventi idrogeologici del Sarno e di Messina spostano l’asticella della prevedibilità ancora più in là, ricomprendendo nel giudizio *tutti i possibili effetti calamitosi* che non si possano escludere³⁵³. Nella pronuncia della Cassazione sul disastro di Sarno si legge: «il giudizio di prevedibilità andava compiuto [...] *ipotizzando quindi la più distruttiva ipotesi che potesse verificarsi* [...]». Se di un fenomeno naturale (o anche cagionato dall’uomo) non si conoscono le caratteristiche fondamentali – in particolare le cause, le possibilità di evoluzione, gli effetti possibili – la cautela che occorre usare nell’affrontarlo per eliminarne o ridurne le conseguenze deve essere ancora maggiore proprio perché non si possono escludere, con valutazione *ex ante* fondata su conoscenze scientifiche affidabili, gli effetti maggiormente distruttivi»³⁵⁴. Principi simili emergono dalla lettura della sentenza sul caso dell’alluvione di Messina si legge: «quanto agli eventi naturali o alle calamità che si sviluppino progressivamente [...] il giudizio di prevedibilità deve tener conto della natura e delle dimensioni di eventi analoghi storicamente già verificatisi, ma valutando altresì *se possa essere esclusa* la possibilità che questi eventi possano avere dimensioni e caratteristiche più gravi o addirittura catastrofiche. Infatti, la valutazione della prevedibilità [...] non può prescindere dalla valutazione su che cosa può avvenire in futuro, a meno che le caratteristiche del fenomeno non siano da sole idonee a fondare un *giudizio di esclusione* di più gravi conseguenze»³⁵⁵.

Ancora, le sentenze di merito e di legittimità sul Terremoto dell’Aquila³⁵⁶ – pur dando atto che la scienza sismologica non è ancora in grado di prevedere con precisione

³⁵² Cass., sez. IV, 17 maggio 2006, n. 4675, Bartalini.

³⁵³ Per un confronto tra la sentenza di legittimità sul caso del petrolchimico di Porto Marghera e quella sul caso del Sarno vd. D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell’incertezza nella struttura del reato*, cit., p. 143 ss.

³⁵⁴ Cass., sez. IV, 11 marzo 2010, n. 16761.

³⁵⁵ Cass., sez. IV, 16 febbraio 2012, n. 17069: nel caso di specie si rimproverava ai pubblici amministratori di non aver impedito la circolazione stradale sul greto asciutto di due fiumi, così che, in occasione dell’alluvione, l’acqua aveva travolto cose o persone che ivi si trovavano.

³⁵⁶ Trib. Aquila, 22 ottobre 2012, n. 380, Barberi, in *Dir. pen. cont.*, con nota di A. GALLUCCIO; C. App. L’Aquila, 10 novembre 2014 n. 3317, Barberi, in *Dir. pen. cont.*, con nota di A. GALLUCCIO; Cass., sez. IV,

né l'*an*, né il *quando*, né la magnitudo degli eventi tellurici – sono comunque pervenute all'accertamento di responsabilità degli amministratori locali, attraverso una traslazione dell'area del giudizio di prevedibilità, dall'evento al *rischio di un evento non escludibile*³⁵⁷. È quella che attenta dottrina ha definito criticamente “colpa eventuale”: l'agente non può concretamente prevedere – e, quindi, evitare – l'evento, ma può soltanto rappresentarsi il rischio di un danno, non potendo escludere il suo verificarsi³⁵⁸.

Talvolta a tali statuizioni si accompagna altresì l'espressa menzione del principio di precauzione (*«le regole che disciplinano l'elemento soggettivo hanno [...] funzione precauzionale, e la precauzione richiede che si adottino certe cautele anche se è dubbio che la mancata adozione provochi eventi dannosi»*³⁵⁹), così come l'identificazione dell'astensione dall'attività come regola cautelare di chiusura, per il caso in cui *«il garante si renda conto di non essere in grado di incidere sul rischio»*³⁶⁰. In questa prospettiva, l'approccio precauzionale muta la regola di *default* da seguire di fronte ad un rischio non scientificamente provato: da *in dubio pro actione* a *in dubio pro omissione*³⁶¹.

D'altra parte, la ricostruzione dell'agente modello, nei casi appena visti, è stata talvolta basata sul principio in base al quale «quando vengono in discussione beni della persona (vita e salute), non va tenuto conto del costo economico necessario per il rispetto delle regole cautelari o addirittura per la rinuncia all'attività»³⁶² – un criterio che appare in evidente contrasto con le legittime istanze di bilanciamento tra beni giuridici concorrenti, già avallate dalla Corte Costituzionale con riferimento ai decreti legge *Salva Ilva* (vd. *supra*, § 4.5.2, in questo cap.). Tale modello precauzionale – che impone di adottare tutte le misure preventive tecnologicamente possibili per evitare il danno – è stato tra l'altro affermato non soltanto in presenza di regole cautelari *elastiche*, ma anche con riferimento alle regole *rigide*. È quanto accade, ad esempio, con riferimento alla già citata tecnica normativa dei limiti-soglia, la cui potenziale utilità – nell'ottica della tutela dell'affidamento dell'agente e della necessaria determinatezza della fattispecie colposa –

19 novembre 2015, dep. 2016, n. 12478, Barberi, in *Dir. pen. cont.*; per un'ampia ricostruzione della vicenda giudiziaria vd. A. NAPPI, *La prevedibilità nel diritto penale*, cit., p. 390 ss.;

³⁵⁷ G. CIVELLO, (voce) *Prevedibilità e reato colposo*, cit., p. 1019.

³⁵⁸ G. CIVELLO, *La “colpa eventuale” nella società del rischio. Epistemologia dell'incertezza e “verità soggettiva” della colpa*, Giappichelli, 2013.

³⁵⁹ Cass., sez. IV, 17 maggio 2006, n. 4675, Bartalini, p. 290.

³⁶⁰ Cass., sez. IV, 14 marzo 2014, n. 22249.

³⁶¹ F. CONSULICH, (voce) *Rischio consentito*, cit., p. 1106;

³⁶² Cass., sez. IV, 17 maggio 2006, n. 4675, Bartalini, p. 290.

è vanificata da quel filone giurisprudenziale che, anche a dispetto dell'osservanza dei limiti quantitativi di esposizione fissati per legge o regolamento, ritiene di poter fondare un rimprovero colposo sulla base della violazione di cautele generiche³⁶³.

La dottrina maggioritaria non ha mancato di criticare l'applicazione del principio di precauzione come base valutativa della colpa generica³⁶⁴. In particolare, la contraddizione con il paradigma tradizionale della colpa sembra consistere primariamente nel fatto che, in assenza di una solida base scientifica che consenta di collegare *condotta* ed *evento*, all'agente rischia di essere rimproverata la verifica di *qualsiasi evento lesivo* derivante dalla condotta, sulla base del mero nesso di causalità e, dunque, in contrasto con il principio di colpevolezza. La "grammatica dell'incertezza"³⁶⁵ che è alla base dell'elaborazione del principio di precauzione mal si concilia, d'altra parte, con l'esigenza di determinatezza che dovrebbe guidare la costruzione e l'interpretazione delle fattispecie incriminatrici. Il richiamo alla logica precauzionale, se non formalizzato in norme cautelari positive, lascia infatti aperta la questione circa l'individuazione della cautela doverosa³⁶⁶: in assenza di un solido apparato nomologico di riferimento, sarà necessario astenersi dalla condotta pericolosa o basterà invece l'adozione di cautele aggiuntive? E, in quest'ultimo caso, *quali*?

Si tratta di scelte, com'è stato osservato, che devono essere necessariamente assunte dal decisore pubblico e che non possono essere scaricate sugli attori privati, anche qualora questi siano soggetti economicamente forti, in grado di sostenere i costi di

³⁶³ Vd. in proposito gli stralci dell'ordinanza pronunciata dal Tribunale di Taranto, in sede di riesame, il 7 agosto 2012, riportati in S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 199 ss., a cui si rinvia per un completo panorama giurisprudenziale in materia. Vd. anche E. PENCO, *Limiti-soglia e responsabilità colposa*, cit., p. 195 ss.

³⁶⁴ C. PIERGALLINI, *Il paradigma della colpa nell'età del rischio*, cit., p. 1694- 1696; F. STELLA, *Giustizia e modernità*, cit., p. 221; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 241 ss.; G.P. ACCINNI, *Criteri di imputazione per colpa tra leggi scientifiche e accertamenti giudiziali*, in *Riv. it. dir. proc. pen.*, 2006, n. 3, p. 928 ss.; V. ATTILI, *L'agente-modello "nell'era della complessità"*, cit., p. 1278 ss.; G. DE FRANCESCO, *"Interpersonalità" dell'illecito penale: un 'cuore antico' per le moderne prospettive della tutela*, in *Cass. pen.*, 2015, n. 2, p. 868-869; A. ALESSANDRI, *Attività d'impresa e responsabilità penali*, in *Riv. it. dir. proc. pen.*, 2005, fasc. 2, p. 565; D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit. p. 651; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 317 ss.; C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto "a distanza" tra disvalore d'azione e disvalore d'evento*, cit., p. 238 ss.; da ultimo, F. D'ALESSANDRO, *Precauzione (principio di)*, cit., p. 1001 ss.

³⁶⁵ C. PIERGALLINI, *"Civile" e "penale" a perenne confronto: l'appuntamento di inizio millennio*, in *Riv. it. dir. proc. pen.*, 2012, n. 4, p. 1320; ID. *Intelligenza artificiale*, cit., p. 1771.

³⁶⁶ S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 321-322.

un'attività istruttoria approfondita³⁶⁷. Piuttosto, in una logica di “democratizzazione” dei processi di valutazione del rischio³⁶⁸, bisognerebbe prendere atto delle conoscenze specifiche esistenti all'interno delle imprese, coinvolgendole nei processi di formazione delle regole cautelari scritte³⁶⁹ ed eventualmente prevedendo un *dovere di diffusione delle informazioni possedute* sanzionato penalmente³⁷⁰.

Se, come già sottolineato, l'orientamento maggioritario in dottrina ritiene che l'impiego del principio di precauzione come base per un giudizio sulla colpa generica sia contrario alla struttura costituzionale della colpa, vi è tuttavia un indirizzo minoritario che ha manifestato una qualche apertura in tal senso³⁷¹. Un parziale ingresso della logica precauzionale nell'individuazione delle cautele generiche sarebbe giustificato, da un lato, dalla specifica menzione del principio, quantomeno in materia ambientale, all'art. 191 TFUE (che potrebbe rientrare nella gamma di regole cautelari che possono rilevare ai sensi dell'art. 43 c.p.)³⁷²; dall'altro, dalla presa d'atto del perenne stato di incertezza che caratterizza le attività umane, che dovrebbe indurre ad una maggiore “responsabilizzazione” nei confronti di coloro che introducono nuovi ed imponderabili rischi nella società – sulla scorta, ci pare, di quella spinta solidaristica che, secondo alcuni, dovrebbe caratterizzare l'applicazione del diritto penale (vd. *supra*, § 4.5., in questo cap.).

Sulla base di tale approccio, alcuni autori propongono una valorizzazione della logica precauzionale nell'ambito del giudizio sulla colpevolezza, nei casi in cui sia presente un dato scientifico di un certo rilievo, seppur non consolidato³⁷³. In quest'ottica, si propone un ampliamento del raggio d'azione della colpa generica, per come tradizionalmente intesa, attraverso una rivisitazione ermeneutica della nozione di *prevedibilità* dell'evento. L'incertezza scientifica al momento della condotta, dunque, non escluderebbe la colpa, che sarebbe così da individuarsi nella «mancata o intempestiva

³⁶⁷ F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 245.

³⁶⁸ G. FORTI, “Accesso alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione”, cit., p. 217.

³⁶⁹ F. GIUNTA, *Il diritto penale e le suggestioni*, p. 245.

³⁷⁰ È la tesi di G. FORTI, “Accesso” alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione, cit., *passim*; per un “aggiornamento” della tesi al contesto produttivo dell'intelligenza artificiale vd. C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1773.

³⁷¹ C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., 1743 ss.; E. CORN, *Il principio di precauzione nel diritto penale*, cit., p. 146 ss.; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 143 ss. e 166 ss.

³⁷² C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1758.

³⁷³ C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1754.

adozione delle cautele suggerite dai *signa facti* e dalle congetture ricavabili dagli studi e dai dati disponibili nel mondo scientifico, seppure a livello di ipotesi non ancora assurte a rango di leggi scientifiche o non ancora condivise dall'intera comunità scientifica»³⁷⁴.

In particolare, il principio di precauzione potrebbe portare ad ampliare lo spettro dei sotto-eventi “prevedibili”, comprendendovi – oltre a quelli noti – anche quelli che presentino “una certa omogeneità” con essi: così, tornando all'esempio dell'esposizione ad amianto, all'agente potrebbe essere rimproverata la morte derivante dal sotto-evento mesotelioma, poiché analogo al sotto-evento prevedibile al momento della condotta, l'asbestosi, nella misura in cui entrambe le patologie coinvolgono l'apparato respiratorio³⁷⁵.

4.5.4. Intelligenza artificiale, rischio consentito ed individuazione delle cautele doverose.

Sulla base di quanto ricostruito, ci pare di poter affermare che quello dell'intelligenza artificiale sia un settore in cui tipicamente si dovrà identificare un'area di rischio consentito. O meglio: la delimitazione di uno spazio d'impunità per i produttori che si conformino agli *standard* tecnici del settore – con i limiti ai quali abbiamo accennato precedentemente (vd. *supra*, § 4.5.2.2., in questo cap.) – ci pare l'unico modo, *de jure condito*, per consentire lo svolgimento dell'attività di ricerca, sviluppo e produzione di sistemi di i.a., garantendo agli operatori che la valutazione sul governo dei rischi effettuata in sede legislativa/amministrativa non sia ribaltata in sede giudiziaria.

L'ingresso del parametro dell'agente modello nella valutazione giudiziale circa la colpevolezza del produttore di intelligenza artificiale rischierebbe infatti – per come correntemente applicato da una parte della giurisprudenza – di ampliare a dismisura la sfera di responsabilità dei vertici delle imprese produttive e dei dipartimenti di progettazione e sviluppo coinvolti.

³⁷⁴ C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1775.

³⁷⁵ C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1770 ss. Vd. anche D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 143 ss. e 166 ss., che si mostra “possibilista” in relazione ad un limitato ampliamento della nozione di “prevedibilità”.

Come abbiamo sottolineato in più occasioni, infatti, una delle caratteristiche più dirompenti dei sistemi di *machine learning* – specie per quanto riguarda l'imputazione penale – è quella di essere *prevedibili* nella loro *imprevedibilità*: il produttore e il programmatore di CDSS sono *consapevoli* che il sistema di supporto alla diagnostica e al trattamento potrà, di tanto in tanto, fornire raccomandazioni sbagliate; lo stesso dicasi per i produttori di sistemi di guida autonoma e di *healthcare robots*. Un *decision-making* basato su analisi stocastiche è *naturalmente destinato a sbagliare*.

Se il giudizio sulla “prevedibilità” dell’evento arriva a comprendere tutti gli eventi appartenenti al *genere di risultato* indesiderato³⁷⁶ (ad esempio, morte derivante da errata diagnosi), l’unica regola cautelare da seguire dovrebbe consistere allora nella *rinuncia all’attività produttiva*. D’altra parte, se dovesse farsi applicazione del principio del *maximum minorum* (*maximin*), vero e proprio fondamento teorico del principio di precauzione, in base al quale *ogni scelta da compiersi in condizioni di incertezza va valutata in base alla peggiore delle sue conseguenze possibili*³⁷⁷, la diligenza imporrebbe di interrompere lo sviluppo e la commercializzazione di sistemi di i.a., per loro “natura” non completamente prevedibili e governabili³⁷⁸.

A nostro parere, l’individuazione di un limite all’espansione del giudizio sulla *prevedibilità generica* dell’evento lesivo è indispensabile, se si vuole creare un quadro di certezza giuridica per le imprese e di incentivo all’innovazione. L’alternativa – pur percorribile, qualora la sensibilità dell’opinione pubblica manifesti un’*avversione al rischio* da intelligenza artificiale – è quella della rinuncia allo svolgimento dell’attività pericolosa: si tratta, tuttavia, di una scelta che dev’essere assunta dal decisore pubblico, e che non può essere surrettiziamente imposta dall’autorità giudiziaria.

³⁷⁶ Così D. CASTRONUOVO, *La colpa penale*, cit., p. 194.

³⁷⁷ A. PERIN, *Prudenza, dovere di conoscenza e colpa penale*, cit., p. 81; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell’incertezza nella struttura del reato*, cit., p. 142, che cita, come espressione di tale orientamento, la pronuncia della Cassazione sul disastro di Sarno, Cass., sez. IV, 11 marzo 2010, n. 16761: «il giudizio di prevedibilità andava compiuto [...] ipotizzando quindi la più distruttiva ipotesi che potesse verificarsi [...]». Se di un fenomeno naturale (o anche cagionato dall’uomo) non si conoscono le caratteristiche fondamentali – in particolare le cause, le possibilità di evoluzione, gli effetti possibili – la cautela che occorre usare nell’affrontarlo per eliminarne o ridurne le conseguenze deve essere ancora maggiore proprio perché non si possono escludere, con valutazione *ex ante* fondata su conoscenze scientifiche affidabili, gli effetti maggiormente distruttivi»; vd. anche E. CORN, *Il principio di precauzione nel diritto penale*, cit., p. 136.

³⁷⁸ Sul punto vd. S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 719.

Il problema principale, in ogni caso, consisterà nell'individuazione di *standard* tecnici *rigidi*, che non lascino spazio alla necessità, per il produttore, di sostituirsi al decisore nel governo del rischio e, di conseguenza, all'ingresso dell'agente modello nei parametri di valutazione giudiziaria³⁷⁹ – posto che, come abbiamo già sottolineato, potranno sempre emergere dei casi in cui la diligenza imponga la disapplicazione delle cautele positive o l'adozione di cautele ulteriori (vd. *supra*, § 4.5.2.2.).

A leggere le disposizioni contenute nell'*AI Act*, la strada da percorrere sembra ancora lunga³⁸⁰. È vero che si tratta di un testo di c.d. “legislazione orizzontale”, che mira a fornire un quadro giuridico per lo sviluppo e la commercializzazione di *tutti i sistemi di i.a.*; esso, tuttavia, è destinato ad essere applicato come normativa sussidiaria, ogniqualvolta non vi siano regolamentazioni specifiche. Sarebbe forse stata necessaria, dunque, una tecnica normativa più precisa. L'auspicio, comunque, è che il processo di standardizzazione inaugurato nel 2019 da CEN e CENELEC sia in grado di fornire presto un apparato organico di norme cautelari *rigide* per i vari settori coinvolti dallo sviluppo dell'intelligenza artificiale.

In ogni caso, data l'intrinseca difficoltà di incidere sulla fase esecutiva del funzionamento dei sistemi di i.a., la strada da seguire pare obbligata ed è quella della regolamentazione preventiva del *design* degli algoritmi di auto-apprendimento. In quest'ottica, un ruolo cardine nella definizione degli *standard* cautelari dovrà essere assunto dalle stesse imprese, spesso dotate di un'*expertise* tecnologica e di una conoscenza delle fonti di pericolo che non ha paragoni nel settore della vigilanza pubblica³⁸¹. Oltre a modelli di *eterodisciplina* e di *eterocontrollo*, dunque, nell'accertamento della colpa per gli eventi lesivi da intelligenza artificiale un ruolo centrale potranno averlo strumenti di *coregolamentazione* e di *autonormazione*, che consentono di adeguare l'intervento preventivo sul *design* alle specifiche esigenze di gestione del rischio dell'impresa³⁸².

³⁷⁹ L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 833.

³⁸⁰ In argomento vd. R. WILLIAMS, *Accountable Algorithms: Adopting the Public Law Toolbox Outside the Realm of Public Law*, in 75 *Current Legal Problems*, 2022, p. 244.

³⁸¹ G. FORTI, “Accesso” alle informazioni, cit., 192 ss.; C. PIERGALLINI, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, cit., p. 1128; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 287.

³⁸² Sulla funzione di *cooperative compliance* che il diritto penale potrà svolgere nell'ambito della regolazione dei sistemi di i.a. vd. G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 73.

D'altra parte, prendendo atto della natura sostanzialmente plurisoggettiva della responsabilità da sviluppo e produzione di intelligenza artificiale, paradigma indispensabile di normativa preventiva sarà quello della responsabilità amministrativa degli enti³⁸³. La colpa che può essere imputata al produttore è infatti intrinsecamente una colpa da *disorganizzazione*, da *mancata gestione* o *scorretta gestione del rischio*. Sebbene, come abbiamo già accennato, il catalogo dei reati-presupposto della responsabilità amministrativa degli enti non contempli le ipotesi di omicidio o lesioni realizzate in violazione delle norme sulla sicurezza dei prodotti – e di certo il principio di tassatività impedisce un'applicazione analogica della normativa –, l'apparato giuridico e concettuale così delineato potrebbe fornire validi strumenti da un punto di vista metodologico³⁸⁴. Le scelte di sviluppo e commercializzazione di un prodotto sono infatti tipicamente scelte che concernono la *politica d'impresa*, che per essere correttamente implementate richiedono l'adozione di modelli organizzativi e di protocolli decisionali.

In un contesto di *incertezza* sulla concreta portata lesiva dello sviluppo e della commercializzazione dei sistemi di i.a., l'individuazione della diligenza concretamente esigibile nelle fasi di *design* necessiterà di forme di procedimentalizzazione. In quest'ottica, la diligenza del *produttore-persona fisica* andrà sempre più assomigliando alla diligenza richiesta ordinariamente *all'ente* – una diligenza che si estrinseca nella *gestione di fattori di rischio spesso imponderabili*. In questo senso pare muoversi la stessa bozza di regolamento europeo sull'intelligenza artificiale, che prevede l'istituzione ed il mantenimento di un sistema di gestione del rischio (art. 17, *AI Act*).

In una prospettiva futura, può inoltre meritare qualche cenno l'orientamento che suggerisce di ricostruire la responsabilità (civile e penale) attraverso l'interazione tra tecnica giuridica, modelli computazionali e strategie organizzative. In particolare, alcuni spunti interessanti possono ricavarsi dalla teoria, introdotta da Mireille Hildebrandt, di *Legal Protection by Design*³⁸⁵, in base alla quale in un futuro prossimo potrebbe essere possibile tutelare i diritti fondamentali – o meglio, in chiave penalistica, i beni giuridici della vita, della salute, dell'integrità fisica – attraverso una loro traduzione dal linguaggio

³⁸³ B. PANATTONI, *Intelligenza artificiale*, cit., p. 362 ss.

³⁸⁴ B. PANATTONI, *Intelligenza artificiale*, cit., p. 362 ss.

³⁸⁵ M. HILDEBRANDT, *Saved by Design? The Case of Legal Protection by Design*, in 11 *NanoEthics*, 2017, p. 307 ss.; EAD, *Law for Computer Scientists and Other Folk*, cit., p. 251 ss.; cfr. A. GARAPON – J. LASSÈGUE, *Giustizia digitale*, Il Mulino, 2021, p. 249.

naturale-giuridico a quello computazionale, mediante la formulazione di requisiti tecnici e impostazioni di *default* per i sistemi di i.a. È la prospettiva del *Law as a code*³⁸⁶.

Tale proposta ha sicuramente il pregio di evidenziare la necessità di un coinvolgimento del diritto fin dalle fasi di *design* dell'intelligenza artificiale; essa, tuttavia, non ci pare risolutiva per il tema che qui ci interessa. Anche qualora si sviluppasse quelle capacità tecnologiche per implementare una *Legal Protection by Design*, questo non impedirebbe che, concretamente, l'algoritmo intelligente sbagli. Ecco, che, allora, si porrebbero i medesimi problemi già visti in punto di dimostrazione dell'elemento soggettivo: come individuare la regola cautelare che si assume violata? L'osservanza della cautela era esigibile? L'evento cagionato dalla condotta anti-doverosa rappresenta la *concretizzazione del rischio* che la regola mirava a scongiurare? E, infine, la condotta doverosa avrebbe avuto *efficacia impeditiva*?

Nell'ottica *reattiva* propria del diritto penale – così come, d'altra parte, della responsabilità aquiliana – la diligente progettazione del codice algoritmico rileva sul piano della valutazione della colpa, ma costituisce soltanto uno degli elementi che devono essere presi in considerazione dal giudice nell'accertamento della responsabilità in capo all'agente.

5. Una tutela anticipata nei confronti dei sistemi di i.a. pericolosi? Prospettive *de jure condito*.

Nei paragrafi precedenti abbiamo illustrato alcune delle difficoltà che potrebbero emergere nell'attribuzione della responsabilità penale in relazione ad eventi lesivi che coinvolgono sistemi di i.a. La crisi del diritto penale d'evento, delineatasi in corrispondenza all'emergere del c.d. *diritto penale del rischio*³⁸⁷, determina la necessità di avvalersi, nella tutela della sicurezza dei consumatori, di forme anticipate di tutela penale. L'imperscrutabile ipercomplessità delle trame causali, l'imprevedibilità del

³⁸⁶ *Code is law* è il titolo di un celebre articolo del giurista statunitense Lawrence Lessig, vd. ID., *Code is Law. On liberty in Cyberspace*, Harvard Magazine, 2000.

³⁸⁷ Vd. per tutti M. DONINI, *Il volto attuale dell'illecito penale*, cit., p. 97 ss.; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato*, p. 403 ss.; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 290 ss.

concreto atteggiarsi del *decision making* algoritmico, l'interattività e l'opacità che caratterizzano l'agire dei dispositivi intelligenti: sono tutti fattori che rendono tortuoso il cammino verso il riconoscimento di una responsabilità penale per il danno derivante dal prodotto.

Di qui la maturata convinzione che – a meno di non voler rinunciare ai principi cardine del diritto penale – una responsabilità del produttore per gli eventi lesivi derivanti dal sistema di i.a. costituirà una rarità. Ecco che, in questo contesto, una tutela penale – indispensabile, a nostro parere, al fine di approntare adeguata salvaguardia di beni giuridici di fondamentale rilievo³⁸⁸ – potrebbe appuntarsi in forma anticipata, attraverso il ricorso a reati colposi di mera condotta, fondati sulla creazione di un pericolo o addirittura di un mero rischio³⁸⁹.

Tale orientamento è tra l'altro da collegarsi a quella posizione che, in dottrina, suggerisce un complessivo ripensamento della responsabilità colposa, attraverso la valorizzazione – da sempre negletta nel nostro Paese, stante il taglio fortemente oggettivistico del nostro ordinamento penale³⁹⁰ – del *disvalore di azione*, rispetto al solo *disvalore di evento*³⁹¹. Le perplessità che scaturiscono dall'idea di una responsabilità colposa nei reati di evento sono note: se l'evento costituisce il discrimine tra la punizione e la non punizione di una condotta, la responsabilità finisce per dipendere dal puro caso, dal momento che la verifica dell'evento, oltre che dall'azione o omissione difformi dallo *standard* cautelare, dipende soprattutto da circostanze esterne e, per l'appunto,

³⁸⁸ G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 56; G. FORTI, "Accesso" alle informazioni sul rischio e responsabilità, cit., p. 180; *contra*, com'è noto, F. STELLA, *Giustizia e modernità*, cit., *passim*, che giunge a sostenere l'impotenza del diritto penale di fronte alla società del rischio.

³⁸⁹ È questa la tesi, in prospettiva *de jure condendo*, di Carlo Piergallini, in relazione ai prodotti tradizionali; vd. ID., *Danno da prodotto e responsabilità penale*, cit., p. 534 ss.

³⁹⁰ Per le ragioni culturali che, nel nostro ordinamento, hanno favorito l'affermarsi del primato del *disvalore* dell'evento – sostanzialmente legate ad un'interpretazione "forte" del principio di offensività – si rinvia a M. MANTOVANI, *Contributo ad uno studio sul disvalore di azione nel sistema penale vigente*, Bononia University Press, 2014, *passim*, spec. 8 ss.; vd. anche N. MAZZACUVA, *Il disvalore di evento nell'illecito penale*, Giuffrè, 1983, p. 2 ss., che, tuttavia, nel suo lavoro, si propone di sottolineare proprio una tendenza evolutiva al ridimensionamento del *disvalore dell'evento*. In argomento vd. anche M. DONINI, *Il volto attuale dell'illecito penale*, cit., p. 197; nello specifico settore dei rapporti tra attività produttiva e principio di precauzione vd. C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto "a distanza" tra disvalore d'azione e disvalore d'evento*, cit., *passim*, spec. p. 240 ss.

³⁹¹ L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., *passim*; L. EUSEBI, *Sistema sanzionatorio e reati colposi*, in M. Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, cit., p. 1212 ss.

sostanzialmente aleatorie³⁹². D'altra parte, lo stesso trattamento sanzionatorio viene calibrato, in maniera del tutto irrazionale, sulla portata lesiva dell'evento stesso, secondo un'ottica di fatto retribuzionista, che prescinde completamente da qualsiasi considerazione relativa all'efficacia preventiva delle norme penali: i soggetti che risponderanno dell'evento lesivo, infatti, sono semplicemente quelli *più sfortunati*, e non necessariamente quelli più negligenti³⁹³.

Tali considerazioni – che ordinariamente non sono tali da scardinare la centralità, nel nostro ordinamento, delle fattispecie colpose d'evento – acquisiscono una maggiore pregnanza nel settore dell'intelligenza artificiale, considerato che l'evento lesivo algoritmico costituisce l'esito finale di una lunga serie di “decisioni” imponderabili ed imprevedibili³⁹⁴. L'introduzione di reati colposi di mera condotta – eventualmente attraverso la tecnica ingiunzionale – pare dunque la strada maggiormente percorribile al fine di realizzare una tutela effettiva nei confronti dei consumatori.

Già *de jure condito*, in realtà, sono rinvenibili alcune norme applicabili. Oltre alle fattispecie di pericolo concreto e astratto previste dal codice penale e poste a tutela dell'incolumità pubblica, la commercializzazione di sistemi di i.a. *pericolosi* potrebbe essere punita sulla base delle fattispecie contravvenzionali di pericolo previste dall'art. 112 cod. cons., commi 1, 2 e 3, che trovano applicazione in via sussidiaria, «*se il fatto non costituisce più grave reato*».

Questione preliminare è la delimitazione delle nozioni di “prodotto” e di “prodotto pericoloso”, elementi costitutivi di tutte le fattispecie previste dall'art. 112 cod. cons. Il concetto di “prodotto” rilevante ai sensi dell'art. 112 cod. cons. merita qualche breve

³⁹² Su questi aspetti vd. L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., p. 1 ss.; M.C. DEL RE, *Per un riesame della responsabilità colposa*, in *Indice penale*, 1985, p. 31 ss., nell'ambito di una critica complessiva alla colpa quale elemento soggettivo rimproverabile all'agente; cfr. D. CASTRONUOVO, *La colpa penale*, cit., p. 105 e ss., per una ricca ricostruzione, anche nella dottrina straniera, di queste posizioni. In argomento vd. anche G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., p. 121 ss.; F. GIUNTA, *Illiceità e colpevolezza nella responsabilità colposa*, cit., p. 332 ss.; A.R. DI LANDRO, *La colpa medica negli Stati Uniti e in Italia. Il ruolo del diritto penale e il confronto col sistema civile*, Giappichelli, 2009, p. 271 ss. Il dibattito sull'opportunità di mantenere una responsabilità penale per colpa è molto acceso negli Stati Uniti; per una lucida sintesi delle varie posizioni vd. L. ALEXANDER – K.K. FERZAN, *Against Negligence Liability*, in P.H. Robinson e al. (eds.), *Criminal Law Conversations*, Oxford University Press, 2011, p. 273 ss.

³⁹³ L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., p. 2; vd. anche L. EUSEBI, *Sistema sanzionatorio e reati colposi*, cit., p. 1201 ss.

³⁹⁴ F. CONSULICH, *Flash offenders*, cit., p. 1053.

considerazione, dal momento che si tratta di una nozione *diversa* da quella fornita dall'art. 115 cod. cons. – che si applica al solo titolo II della parte IV del Codice del consumo (“responsabilità per danno da prodotti difettosi”) – sulla quale ci siamo soffermati nel capitolo precedente (vd. *supra*, cap. II, § 3.1.).

La nozione generale di “prodotto” è fornita dall'art. 3, co. 1, lett. e), cod. cons., in base al quale per “prodotto” si intende «qualsiasi prodotto destinato al consumatore, anche nel quadro di una prestazione di servizi, o suscettibile, in condizioni ragionevolmente prevedibili, di essere utilizzato dal consumatore, anche se non a lui destinato, fornito o reso disponibile a titolo oneroso o gratuito nell'ambito di un'attività commerciale, indipendentemente dal fatto che sia nuovo, usato o rimesso a nuovo; tale definizione non si applica ai prodotti usati, forniti come pezzi d'antiquariato, o come prodotti da riparare o da rimettere a nuovo prima dell'utilizzazione, purché il fornitore ne informi per iscritto la persona cui fornisce il prodotto».

Tralasciando l'evidente problema di circolarità della norma (*un prodotto è... un prodotto!*), la riconducibilità dei sistemi di i.a. alla suddetta definizione sembrerebbe possibile soltanto nei casi in cui l'algoritmo di intelligenza artificiale sia incorporato in un bene *destinato al consumatore o suscettibile di essere da lui utilizzato* (es. sistema di guida autonoma installato su un veicolo)³⁹⁵. Non sembrerebbe, invece, esservi alcuno spazio per un'estensione dell'ambito applicativo della norma ai c.d. *prodotti digitali*, ovvero ai *software* che non necessitano di essere incorporati in un *hardware* – a differenza, invece di quanto potrebbe accadere, con adeguata interpretazione estensiva, in relazione alla definizione fornita dall'art. 115 cod. cons. (vd. *supra*, cap. II, § 3.1.).

Elemento normativo di tutte e tre le fattispecie è poi la nozione di *pericolosità* del prodotto, che, secondo la dottrina dominante, può essere derivata dalla definizione di prodotto sicuro fornita dall'art. 103, co. 1, lett. a)³⁹⁶. Come abbiamo già avuto modo di illustrare nel capitolo precedente (vd. *supra*, cap. II, § 3.4.1.), per “prodotto pericoloso”

³⁹⁵ F. CONSULICH, *Flash offenders*, cit., p. 1039.

³⁹⁶ D. CASTRONUOVO, *Principio di precauzione e beni legati alla sicurezza*, in *Dir. pen. cont.*, 21 luglio 2011, p. 20; *contra* F. CONSULICH, (voce) *Tutela del consumatore*, in F. Palazzo – C.E. Paliero (diretto da), *Commentario breve alle leggi penali complementari*, Cedam, II ed., 2007, p. 2980, che ritiene che la nozione di prodotto pericoloso alla quale fa riferimento l'art. 112 è più ristretta di quella stabilita, in via generale, dall'art. 103, lett. b), cod. cons.; la posizione contraria, infatti, determinerebbe la rilevanza penale di condotte di commercializzazione anche di prodotti che recano rischi minimi, in palese contrasto con i principi di offensività e ragionevolezza.

s'intende «qualsiasi prodotto che non risponda alla definizione di prodotto sicuro» (art. 103, lett. b), cod. cons.), ove quest'ultimo è definito dall'art. 103, co. 1, lett. a), cod. cons. come un prodotto che «in condizioni di uso normali o ragionevolmente prevedibili, compresa la durata e, se del caso, la messa in servizio, l'installazione e la manutenzione, non presenti alcun rischio oppure presenti unicamente rischi minimi, compatibili con l'impiego del prodotto e considerati accettabili nell'osservanza di un livello elevato di tutela della salute e della sicurezza delle persone», in considerazione di una serie di elementi, tra i quali la norma considera, in via esemplificativa, le sue «caratteristiche» (n. 1), il suo eventuale «effetto» su altri prodotti (n. 2), la sua «presentazione» (n. 3) e le «categorie di consumatori» cui esso è destinato (n. 4), con particolare riguardo a quelle maggiormente vulnerabili, tra i quali i bambini e gli anziani.

Il Codice del consumo prevede poi, all'art. 105, due presunzioni di sicurezza *iuris tantum*: la prima, nel caso in cui sia rispettata la legislazione vigente nello Stato membro in cui il prodotto è commercializzato (art. 105, co. 1); la seconda, nel caso in cui vengano osservati gli *standard* tecnici definiti dalle c.d. norme armonizzate o delle equivalenti versioni nazionali (art. 105, co. 2)³⁹⁷. In assenza delle norme citate ai commi 1 e 2, la sicurezza del prodotto non può essere presunta, ma è valutata in base «alle norme nazionali non cogenti che recepiscono norme europee, alle norme in vigore nello Stato membro in cui il prodotto è commercializzato, alle raccomandazioni della Commissione europea relative ad orientamenti sulla valutazione della sicurezza dei prodotti, ai codici di buona condotta in materia di sicurezza vigenti nel settore interessato, agli ultimi ritrovati della tecnica, al livello di sicurezza che i consumatori possono ragionevolmente attendersi» (art. 105, co. 3). Infine, il comma 4 dell'art. 105 cod. cons. stabilisce che le Autorità amministrative competenti possono limitare o impedire l'immissione sul mercato del prodotto, o chiederne il ritiro o il richiamo, qualora questo si riveli, *nonostante la conformità*, pericoloso per la salute e la sicurezza del consumatore.

Il prodotto conforme, dunque, non solo può essere *difettoso* (vd. *supra*, cap. II, § 3.4. ss.), ma può essere altresì *pericoloso*: l'immissione sul mercato di un prodotto

³⁹⁷ Come già osservato nel precedente capitolo, l'art. 105, co. 2, cod. cons. si riferisce alle norme armonizzate come alle «norme nazionali non cogenti che recepiscono le norme europee i cui riferimenti sono stati pubblicati dalla Commissione europea nella Gazzetta Ufficiale delle Comunità europee a norma dell'articolo 4 della direttiva 2001/95/CE, del Parlamento europeo e del Consiglio, del 3 dicembre 2001».

conforme alla normativa vigente, dunque, non esclude, di per sé, la responsabilità penale del produttore ai sensi dell'art. 112 cod. cons.³⁹⁸.

Tornando per un attimo al settore specifico dell'intelligenza artificiale, che qui interessa, pare di potersi affermare che, ad oggi, una valutazione di pericolosità dei sistemi di i.a. potrà effettuarsi principalmente sulla base dei criteri generali di sicurezza forniti dall'art. 103, co. 1, lett. a), cod. cons. – stante la quasi totale assenza di normative specifiche, *standard* e norme armonizzate in materia. Ovviamente, quando il sistema di i.a. sia *componente* di un prodotto facente parte di una categoria merceologica regolata, la normativa rilevante potrà essere utilizzata come parametro per le presunzioni di cui all'art. 105 cod. cons.

Vediamo, ora, le singole fattispecie previste dall'art. 112 cod. cons.

Il comma 1 fa ricorso alla tecnica ingiunzionale nel punire l'immissione di prodotti pericolosi in violazione del divieto imposto dall'autorità ai sensi dell'art. 107, comma 2, lettera e)³⁹⁹. La pena ivi prevista (arresto da sei mesi ad un anno e ammenda da 10.000 a 50.000 euro) è lievemente più alta rispetto a quella stabilita dal comma 2 per l'immissione sul mercato di prodotti pericolosi senza che sia stato emesso un provvedimento amministrativo di divieto (stessa cornice edittale per l'ammenda, ma minimo edittale di 15 giorni per la pena dell'arresto). Il regime sanzionatorio più severo previsto dal primo comma si spiega, com'è ovvio, per la maggiore offensività della condotta ivi punita, poiché, al pericolo per la salute e alla sicurezza dei consumatori, si aggiunge, in quel caso, anche l'offesa al buon andamento della pubblica amministrazione⁴⁰⁰.

Infine, il comma 3 dell'art. 112 prevede il trattamento sanzionatorio più lieve (ammenda da 10.000 euro a 25.000 euro) in caso di mancata ottemperanza, da parte del produttore o del distributore, dei provvedimenti dell'autorità amministrativa competente *diversi* dal divieto di commercializzazione e consistenti, a mero titolo di esempio, nell'ordine di fornire adeguate avvertenze sui rischi che il prodotto può presentare o di

³⁹⁸ F. CONSULICH, (voce) *Tutela del consumatore*, cit., p. 2980; D. CASTRONUOVO, *La colpa penale*, cit., p. 310; A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, cit., p. 16

³⁹⁹ L'art. 107, co. 2, lett. e) stabilisce che le autorità amministrative preposte alla vigilanza sulla commercializzazione dei prodotti possano vietare l'immissione sul mercato dei prodotti pericolosi e adottare le misure necessarie a garantire l'osservanza dei divieti.

⁴⁰⁰ F. CONSULICH, *Tutela del consumatore* (voce), cit., p. 2981.

disporre l'adeguamento di un prodotto già commercializzato agli obblighi di sicurezza previsti dal codice del consumo (vd. art. 107, comma 2, lettere b), numeri 1) e 2), c) e d), numeri 1) e 2).

La struttura delle fattispecie in questione è stata ricostruita, alternativamente, in termini di illeciti di pericolo astratto, illeciti di rischio, illeciti di mera disobbedienza⁴⁰¹. È escluso, invece, che possa trattarsi di reati di pericolo concreto, poiché la pericolosità viene valutata sulla base di caratteristiche astratte del prodotto, senza che sia necessario un accertamento in concreto delle potenzialità lesive della merce⁴⁰². In ogni caso, le fattispecie contravvenzionali sembrano essere ispirate – così come l'intera disciplina generale in materia di sicurezza del prodotto – da logiche precauzionali⁴⁰³ (si pensi, ad esempio, all'art. 107, co. 4, cod. cons., che cita esplicitamente il principio di precauzione come presupposto per l'adozione di misure ingiunzionali).

Come già accennato, tali norme si prestano sicuramente ad essere applicate anche in relazione ai sistemi di i.a. – benché le sanzioni ivi previste siano assolutamente inidonee ad avere un'efficacia di reale deterrenza nei confronti di comportamenti pericolosi per la salute e l'incolumità del consumatore⁴⁰⁴.

Nel caso, poi, in cui i prodotti intelligenti immessi sul mercato rechino indebitamente la marcatura CE, ai produttori saranno ovviamente applicabili anche le fattispecie di cui agli artt. 514 ss. c.p., eventualmente in concorso con quelle previste dall'art. 112 cod. cons., stante la diversità del bene giuridico tutelato (in un caso, la salute e l'incolumità pubblica, nell'altro, la correttezza e lealtà commerciale nei confronti del consumatore)⁴⁰⁵.

⁴⁰¹ F. CONSULICH, *Tutela del consumatore* (voce), cit., p. 2981; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 76.

⁴⁰² F. CONSULICH, *Tutela del consumatore* (voce), cit., p. 2981.

⁴⁰³ D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 76.

⁴⁰⁴ F. CONSULICH, *Flash offenders*, cit., p. 1039.

⁴⁰⁵ Sull'applicabilità degli artt. 514 c.p. ss. nel contesto della responsabilità per danno da prodotto vd. A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, cit., p. 20 ss.

5.1. (segue) Prospettive *de jure condendo*.

Se è vero che le contravvenzioni previste all'art. 112 cod. cons. si prestano ad essere applicate anche in relazione ai sistemi di i.a., non può non rilevarsi come esse rischino di avere un'efficacia di deterrenza molto limitata, stante la cornice edittale molto bassa ivi prevista (la pena per la fattispecie più grave, quella di cui al primo comma, prevede l'arresto da sei mesi ad un anno e l'ammenda da 10.000 a 50.000 euro). Potrebbe dunque essere opportuna l'introduzione di fattispecie incriminatrici *ad hoc*, che diano specifico rilievo alle condotte pericolose inerenti alla produzione e all'utilizzo di sistemi di i.a.

Un'indicazione in questo senso sembrerebbe provenire dall'*AI Act*, che stabilisce che gli Stati Membri debbano introdurre sanzioni «effettive, proporzionate e dissuasive» per il caso di mancato rispetto della disciplina stabilita dal regolamento stesso (art. 71, § 1, *AI Act*). Va comunque sottolineato che – sebbene gli ordinamenti interni possano optare per l'introduzione di apposite norme incriminatrici – l'approccio del legislatore europeo sembra mostrare un *favor* per le sanzioni pecuniarie amministrative, prevedendo – in un'ottica di proporzionalità ed *extrema ratio* – diversi regimi sanzionatori, a seconda della gravità delle violazioni⁴⁰⁶.

Una forma di tutela penale anticipata, in realtà, potrebbe essere opportuna, a nostro avviso, per il caso di mancato rispetto di alcuni dei requisiti stabiliti dall'*AI Act*. In particolare, una parte della dottrina ha proposto l'introduzione di reati colposi di mera condotta, volti a criminalizzare: (i) l'omessa predisposizione, da parte del programmatore, di meccanismi preventivi nell'algoritmo che governa il sistema di i.a.; (ii) la disattivazione, la mancata attivazione o il mancato aggiornamento, da parte dell'utilizzatore, dei presidi di sicurezza di cui invece fosse stato originariamente dotato⁴⁰⁷.

Quanto alla prima proposta, ci limitiamo a sottolineare che tale condotta potrebbe in ogni caso integrare le fattispecie previste dall'art. 112 cod. cons., dal momento che l'omessa predisposizione, da parte del produttore, di idonei meccanismi di prevenzione

⁴⁰⁶ In argomento vd. C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, cit., p. 435 ss.

⁴⁰⁷ Così F. CONSULICH, *Flash offenders*, cit., p. 1038-1039; vd. anche F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 459.

del danno rientra sicuramente nella fattispecie di “pericolosità generica” prevista dall’art. 103, co. 1, lett. a), cod. cons.; si può prevedere, tra l’altro, che misure di questo tipo saranno prescritte dagli *standard* tecnici della materia in corso di elaborazione. In ogni caso, l’introduzione di una norma *ad hoc* potrebbe effettivamente rivelarsi opportuna, laddove la peculiare pericolosità della condotta induca il legislatore a predisporre una cornice edittale più severa di quella attualmente prevista dalle contravvenzioni di cui all’art. 112 cod. cons.

La seconda delle proposte, d’altra parte, potrebbe coprire una lacuna effettivamente esistente nel nostro ordinamento, andando a prevenire le conseguenze dannose all’integrità e all’incolumità fisica che possono derivare dall’utilizzo di sistemi di i.a. in assenza delle dovute precauzioni. La fattispecie potrebbe essere modellata sull’art. 427 c.p. (*Rimozione od omissione dolosa di cautele contro infortuni sul lavoro*) e – in un’ottica di proporzionalità ed *extrema ratio* – potrebbe essere limitata ai soli *embodied system*, ossia a quei sistemi che, agendo direttamente sul mondo reale, possono causare offese ai beni della vita e dell’integrità fisica. La peculiarità dei sistemi di i.a. suggerirebbe, inoltre, ci pare, di introdurre una fattispecie di pericolo astratto: l’imprevedibilità del *decision making* dei sistemi di i.a., infatti, non consentirebbe al giudice di valutare, caso per caso, se si sia verificato un *pericolo* di lesione dei beni giuridici tutelati dalla norma.

Uno spazio, infine, potrebbe residuare, sempre *de jure condendo*, per la valorizzazione del modello ingiunzionale⁴⁰⁸, favorendo così un coordinamento tra sanzione penale e capillare controllo amministrativo⁴⁰⁹. L’ingiunzione, da parte delle autorità competenti, di assumere misure di sicurezza (es. adozione di ulteriori cautele, sottoposizione dei dispositivi intelligenti a *training* o sperimentazione aggiuntiva, fino all’ordine di disattivazione) potrebbe infatti essere presidiata da sanzioni penali, così da fronteggiare – in un’ottica di governo condiviso del rischio – l’eventuale inosservanza,

⁴⁰⁸ Di quest’opinione è C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1773.

⁴⁰⁹ In generale, sulla tecnica ingiunzionale vd. G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 56 ss.; A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), *Manuale di diritto penale dell’impresa*, cit., p. 50 ss.; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 247; C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto “a distanza” tra disvalore d’azione e disvalore d’evento*, cit., p. 252 ss.; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 295.

da parte dell'impresa, di prestazioni di *facere* individualizzate e infungibili⁴¹⁰. Il paradigma ingiunzionale – ispirato a modelli analoghi nel diritto statunitense⁴¹¹ e, quanto all'Italia, alla repressione penale dell'inottemperanza all'ordine giudiziale di cessazione del comportamento antisindacale (art. 28, co. 4, l. 20 maggio 1970, n. 300, c.d. Statuto dei lavoratori)⁴¹² – si segnala per la sua efficienza sul piano della tempestività e dell'individualizzazione delle condotte doverose.

Posto, infine, che le perplessità sul piano della conformità di tale tecnica al principio di riserva di legge ci paiono agevolmente superabili – nella misura in cui la norma penale faccia riferimento ad una *specifica classe* di provvedimenti amministrativi⁴¹³ –, non possiamo, in ogni caso, non rilevare come misure di questo tipo – già presenti nella disciplina generale sulla sicurezza dei prodotti (vd. art. 112, co. 1 e 3, cod. cons.) – siano, ad ora, rimaste soltanto sulla carta.

6. Riepilogo: intelligenza artificiale e responsabilità penale.

Autonomia, interattività, opacità: sono queste le caratteristiche dell'intelligenza artificiale che, nel primo capitolo (vd. § 5.), abbiamo identificato come potenzialmente dirompenti ai fini dell'imputazione penale degli eventi lesivi algoritmici. In questo capitolo, abbiamo tentato di delineare le prospettive di incriminazione dell'utilizzatore e del produttore di un sistema di i.a. per gli eventi lesivi che coinvolgono algoritmi di

⁴¹⁰ È questo uno dei più apprezzabili vantaggi del modello ingiunzionale; vd. a tal proposito A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), *Manuale di diritto penale dell'impresa*, cit., p. 51-52.

⁴¹¹ Sui modelli ingiunzionali tipici del diritto statunitense vd. C. PEDRAZZI, *Odierna esigenze economiche e nuove fattispecie penali*, in *Riv. it. dir. proc. pen.*, 1975, p. 1099 ss., oggi raccolto in ID., *Diritto penale. Scritti di diritto penale dell'economia*, III, Giuffrè, 2003, p. 47 ss.; l'insigne Autore giunge, tra l'altro, ad escludere che simili modelli siano replicabili nel nostro ordinamento, ritenendo piuttosto auspicabili, a determinate condizioni, una valorizzazione della normazione in bianco; sui modelli ingiunzionali statunitense (specificamente in materia ambientale) vd. anche G. AZZALI, *La tutela penale dell'ambiente. Un'indagine di diritto comparato*, Cedam, 2001, p. 245 ss.

⁴¹² A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), *Manuale di diritto penale dell'impresa*, cit., p. 51.

⁴¹³ G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 77; A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), *Manuale di diritto penale dell'impresa*, cit., p. 53.

machine learning – con alcune considerazioni finali circa le possibili tecniche di anticipazione della tutela penale, attraverso il ricorso a reati di mera condotta.

Quanto alla posizione dell'utilizzatore, con riferimento alle fattispecie colpose, i principali problemi emergono in relazione ai sistemi di i.a. parzialmente autonomi, in cui la funzione di governo sul dispositivo è condivisa tra operatore e algoritmo di *machine learning* e potranno sorgere delicati profili di colpa relazionale e di individuazione delle aree di competenza tra i due centri decisionali (§ 3.2.). Questioni di responsabilità più tradizionali sorgeranno invece con riferimento agli eventi lesivi derivanti da sistemi completamente autonomi, in cui l'utente si limita a fruire di un'attività interamente svolta dal dispositivo intelligente (§ 3.1.).

Se il costante incremento dell'autonomia dei sistemi di i.a. determina una contrazione degli eventi lesivi derivanti dall'errore umano, d'altra parte, l'automazione può essere causa di *nuovi rischi*, determinati proprio dall'impatto dell'automazione sulle capacità dell'utilizzatore. In particolare, abbiamo individuato tre fattori, tipicamente legati all'autonomia dei sistemi di i.a., che potrebbero incidere sull'esigibilità della conformazione alle norme cautelari da parte dell'utilizzatore (§ 3.2.2.): a) *l'opacità*, che potrebbe impedire all'operatore di riconoscere l'errore algoritmico e di intervenire/discostarsi dalla raccomandazione ricevuta; b) *la c.d. "distorsione dell'automazione" (technological complacency)*, ossia quella tendenza delle persone ad affidarsi eccessivamente alla tecnologia e a non metterne in discussione i risultati; c) *la ridotta capacità di reazione durante le situazioni di emergenza*, dovuta al fatto che il livello di attenzione di chi si limiti a *monitorare l'attività altrui* è più debole rispetto a quello di chi *agisce* in prima persona.

Nell'ottica di schermare l'operatore di i.a. da quelle istanze di colpevolizzazione che – prevedibilmente – sarà destinato a subire, in quanto centro di accollo della responsabilità "più vicino" rispetto all'evento algoritmico lesivo, abbiamo poi vagliato la possibilità – suggerita da una parte della dottrina – di introdurre una causa di esclusione della responsabilità per gli utilizzatori, sul modello dell'art. 590-sexies c.p., nonché la proposta di limitare la colpa rimproverabile all'agente alla sola forma *grave* (§ 3.2.2.2.).

Tali suggestioni, lungi dal costituire un'isola remota nell'orizzonte del dibattito scientifico sulla colpa, paiono inserirsi perfettamente nell'attuale discussione sull'opportunità di ripensare il modello della responsabilità colposa, in modo da favorire,

da un lato, la determinatezza della fattispecie, già di per sé aperta e porosa, e, dall'altro, la valorizzazione del “momento soggettivo” della colpa, ossia della concreta rimproverabilità dell'agente. Sebbene non sia il caso di giungere a conclusioni affrettate – dovendosi attendere l'emergere di una casistica più significativa, che renda maggiormente trasparente la natura dei rapporti di “cooperazione” tra algoritmi ed esseri umani – ci pare che il settore in esame potrà fungere da valido banco di prova per la sperimentazione di soluzioni maggiormente aderenti al principio di colpevolezza.

Sempre sul versante della posizione dell'utilizzatore, ci siamo inoltre occupati dei profili di responsabilità dolosa. Alle ipotesi non problematiche, in cui il dispositivo intelligente realizzi il piano criminoso voluto dall'agente, potrebbero affiancarsi casi più complessi, in cui si verifichino eventi lesivi *diversi* da quelli programmati dall'utilizzatore del sistema di i.a. (§ 3.3.). Applicando i criteri generali – desunti dal principio di colpevolezza – vi è da ritenere che nei casi di *aberratio ictus* (monolesiva o plurilesiva), *aberratio delicti*, preterintenzione e delitti aggravati dall'evento, l'utilizzatore del sistema di i.a. risponderà dell'evento realizzatosi soltanto qualora versi in una “colpa in attività illecite”, avendo potuto ragionevolmente prevedere, sulla base delle circostanze concrete, una deviazione algoritmica rispetto al piano criminoso. In questa sede conclusiva, ci limitiamo a ribadire quanto già sottolineato: soltanto con estrema difficoltà sarà possibile stabilire se, nel caso concreto, il sistema intelligente potesse essere considerato un esecutore “affidabile” o meno, specie nel caso in cui il dispositivo sia stato acquistato su mercati non autorizzati e, dunque, sia sfuggito ai controlli preventivi alla messa in commercio.

Sempre con riferimento alla responsabilità dell'utilizzatore, poi, la concreta integrazione di fattispecie dolose incontrerà un ostacolo nel caso in cui l'operatore di un sistema algoritmico si limiti a fornire istruzioni *generali* per la commissione del delitto, lasciando al *decision making* algoritmico la “scelta” di come e quando commettere il fatto illecito (§ 3.4.): si pensi, ad esempio, al *trader* che fornisca ai sistemi di *high frequency trading* (HFT) parametri manipolativi del mercato o al programmatore informatico che addestrì *social bots* fornendogli un *dataset* di migliaia di *tweet* ingiuriosi e diffamatori. A differenza delle ipotesi precedentemente esaminate – in cui l'utilizzatore pianifica in concreto la commissione del fatto illecito, ma l'algoritmo si discosta dal progetto criminoso – in questo caso il programmatore si limita a definire il *genus* del reato, senza

fornire indicazioni circa *tempus* e *modus* della commissione. Come abbiamo rilevato precedentemente, quella che è una precisa scelta sistematica di politica criminale – l'irrelevanza del *dolus generalis* – potrebbe rivelarsi, con lo sviluppo e la commercializzazione di sofisticati sistemi di i.a., una lacuna normativa non trascurabile, dal momento che l'agente potrebbe sfruttare l'interposizione del *decision making* algoritmico per commettere impunemente fatti illeciti.

La gran parte delle nostre riflessioni si è tuttavia concentrata – come, d'altra parte, è accaduto nel precedente capitolo (vd. *supra*, cap. II, § 3. ss.) – sulla possibile imputazione al produttore degli eventi lesivi derivanti dai sistemi di i.a. Con la graduale attribuzione di autonomia in capo ai dispositivi intelligenti, infatti, lo spazio di intervento dell'utilizzatore sul funzionamento dei sistemi di i.a. risulterà progressivamente ridotto e, di converso, il produttore – in quanto “creatore” del “soffio vitale” algoritmico – sarà il soggetto maggiormente qualificato per incidere sulla sicurezza dei dispositivi (§ 4. ss.).

Il reato “da intelligenza artificiale” costituisce solo l'ultima tappa di quella tendenziale crisi del delitto d'evento che caratterizza, ormai da decenni, il diritto penale di fronte alla c.d. “società del rischio”. Il danno da dispositivo intelligente ripropone, accentuandoli, alcuni dei profili più problematici già sorti in giurisprudenza – e osservati criticamente in dottrina – in relazione alla responsabilità penale per danno da prodotto: la sostanziale fungibilità tra struttura commissiva e omissiva del reato (§ 4.2.), l'ostica individuazione dei soggetti personalmente responsabili all'interno delle organizzazioni complesse (§ 4.4.), l'accertamento del nesso di causalità (§ 4.3. ss.) e della colpa (§ 4.4. ss.), in situazioni di incertezza scientifica. Se le prime tra le questioni citate – sulle quali pur ci siamo soffermati – riproducono sostanzialmente, aggiornandoli, gli interrogativi già emersi in passato, l'incerto incedere del *decision making* algoritmico porta ad un livello di complessità mai conosciuto prima la verifica circa la sussistenza di un nesso eziologico tra condotta umana ed evento lesivo, nonché la valutazione della colpa in capo all'agente – con particolare riferimento al requisito della *prevedibilità* dell'evento lesivo.

Alla base della problematica riconduzione dei danni da intelligenza artificiale al paradigma dei reati d'evento vi è l'opacità degli algoritmi di *machine learning*. Come già accaduto in passato in relazione a contesti di incertezza scientifica⁴¹⁴, la crisi di

⁴¹⁴ Sul problema della causalità come “origine” di tutte le altre difficoltà applicative del diritto penale nella materia del danno da prodotto vd. C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., *passim*.

comprensione dei processi causali che determinano il verificarsi dell'evento lesivo si ripercuote, a cascata, sull'intera struttura del reato e, in particolare, sull'accertamento dell'elemento soggettivo. Se, tuttavia, con riferimento ai fenomeni di opacità causale presentatisi in passato (es. danno da prodotto tradizionale o da esposizione a sostanze tossiche), la *black box* causale costituisce un problema che – quantomeno in astratto – è transitorio, poiché è suscettibile di essere risolto con l'acquisizione di nuove conoscenze scientifiche, il fenomeno del *machine learning* pare intrinsecamente refrattario alle generalizzazioni causali che costituiscono il fondamento del metodo scientifico. La decisione dei sistemi di auto-apprendimento è basata su modelli probabilistici e stocastici, che impediscono di affermare, seppur a livello statistico, che all'*input* X segue l'*output* Y.

Gli sforzi degli scienziati finalizzati alla messa a punto di paradigmi di *explainability* e *interpretability* (vd. *supra*, cap. I, § 5.3) potrebbero assicurare, in futuro, gli strumenti per rischiarare, *ex post*, il decorso eziologico interno al *decision making* algoritmico, eventualmente attraverso la predisposizione di specifici meccanismi di *reverse engineering*. Nonostante ciò, la possibilità di riconoscere una regolarità nel comportamento algoritmico – che conduca alla formulazione di leggi scientifiche di copertura – non ci pare compatibile con le attuali modalità attraverso le quali i sistemi di i.a. assumono le decisioni⁴¹⁵.

A tal proposito, abbiamo avanzato l'ipotesi – che allo stato attuale delle conoscenze, non può che restare tale – secondo cui l'accertamento della causalità in relazione ai danni cagionati dai sistemi di i.a. possa seguire un paradigma diverso da quello nomologico-deduttivo, così come, di fatto, avviene con riferimento alla causalità psichica, per la quale non sono disponibili leggi scientifiche (§ 4.3.1). Se il sostrato biologico che rende unico il *cogito* umano non è di certo replicabile, né paragonabile al codice algoritmico che sta dietro al funzionamento delle reti neurali artificiali⁴¹⁶, non si

⁴¹⁵ Non si esclude, tuttavia, un possibile ritorno agli schemi simbolico-causali che hanno caratterizzato lo sviluppo dei primi sistemi di i.a., a metà del secolo scorso. In questo caso, tuttavia, potrebbe essere necessario rinunciare alla stupefacente accuratezza degli attuali sistemi che utilizzano le reti neurali, basati su meccanismi decisionali di tipo induttivo.

⁴¹⁶ Come non abbiamo mancato di rilevare (vd. *supra*, § 2.1., in questo capitolo), sono assolutamente da respingere quelle tesi che – in una prospettiva funzionalista – identificano la capacità di svolgere attività in maniera autonoma dei sistemi di i.a. con una sorta di *capacità morale*, facendo derivare da ciò la possibilità di muovere un rimprovero penale diretto ai dispositivi algoritmici.

può non rilevare come il tentativo, da parte dei ricercatori, di imitare alcuni processi decisionali umani abbia avuto un esito stupefacente. I sistemi di *machine learning* sono oggi in grado di imparare dall'esperienza, di migliorare le proprie *performance* sulla base dei *feedback* ricevuti, di comunicare con altri algoritmi, di scegliere le strategie più efficienti per raggiungere un determinato risultato: nel farlo, si basano – proprio come, in relazione a molte attività, fanno gli esseri umani – sull'induzione e sull'analogia, non sull'applicazione deduttiva di principi logico-causali (vd. *supra*, cap. I, § 1.2.3.).

È da questa similarità strutturale tra processi decisionali umani e algoritmici che sorge l'interrogativo circa la possibile estensione di quella “deroga” al paradigma nomologico-deduttivo che, secondo alcuni, è attualmente consentita in relazione alla causalità psichica. Ovviamente, anche qualora si condividesse una tale conclusione, il rigoroso *standard* probatorio dell'oltre ogni ragionevole dubbio imporrebbe un vaglio induttivo ancora più meticoloso sulla *causalità individuale*, attraverso una scrupolosa valutazione delle risultanze probatorie e l'esclusione dei decorsi causali alternativi, così come imposto dal paradigma dell'*alta credibilità razionale* fatto proprio dalle Sezioni Unite della Cassazione nella sentenza Franzese. Qualora, al contrario, si ritenesse inderogabile il criterio della sussunzione sotto leggi scientifiche, l'accertamento della causalità in relazione ai danni da i.a. potrebbe, di fatto, risultare impossibile.

D'altra parte, anche nel caso in cui si riuscisse, nel caso concreto, a provare il nesso di causalità, resterebbe la difficoltosa dimostrazione dell'elemento soggettivo in capo al produttore. In particolare, prendendo atto che l'attività illecita di quest'ultimo è, nella maggior parte dei casi, *involontaria*, tuttalpiù biasimabile sul piano della *negligenza*, nel nostro lavoro ci siamo concentrati sulla possibilità di muovergli un rimprovero colposo – nonostante non sia escluso che la condotta del produttore possa essere caratterizzata da *dolo eventuale*⁴¹⁷.

Ecco che l'*imprevedibilità genericamente prevedibile*⁴¹⁸ dei sistemi di i.a. parrebbe mettere in discussione la stessa possibilità di muovere un rimprovero colposo al produttore, se si considera che, secondo il criterio di copertura del rischio tipico, affinché l'evento lesivo sia rimproverabile all'agente esso deve costituire realizzazione specifica

⁴¹⁷ Vd. per tutti C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 360 ss.

⁴¹⁸ C. PIERGALLINI, *Intelligenza artificiale*, 1762; vd. anche M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1208-1209.

del rischio che la norma cautelare mirava a prevenire (§ 4.5.3.1.). In assenza di conoscenze approfondite circa il concreto funzionamento degli algoritmi di *machine learning*, gli eventi lesivi scaturenti dai sistemi di i.a. possono considerarsi *prevedibili* per quanto concerne l'*an*, ma *imprevedibili* con riferimento al *quantum* e al *quomodo*: troppo poco, ci pare, per determinare un esito positivo nel giudizio di *prevedibilità* dell'evento lesivo.

Va tuttavia rilevato che, in giurisprudenza, si è ormai affermato il principio in base al quale il giudizio di prevedibilità non concerne l'evento *hic et nunc* realizzatosi, quanto, piuttosto, la generica classe di eventi in cui si colloca quello oggetto del processo. L'alternativa, dunque, ci pare secca: qualora prevalga quest'ultimo orientamento, la colpa potrebbe essere considerata sempre sussistente, dal momento che il produttore di un sistema di i.a. potrà sempre prevedere una determinata classe di eventi lesivi algoritmici; al contrario, qualora prevalga una concezione restrittiva, difficilmente potrebbe ritenersi esistente una colpa in capo al produttore, stante il carattere di intrinseca imprevedibilità dell'evento lesivo concretamente realizzatosi.

In ogni caso, un argine all'espansione del giudizio circa la prevedibilità e l'evitabilità dell'evento lesivo potrà essere costituito dalla delimitazione di un'area di rischio consentito, che – d'accordo con la dottrina dominante – abbiamo identificato come quel *pericolo marginale* che le norme cautelari scritte, deputate al bilanciamento politico tra rischi e benefici dell'esercizio delle attività pericolose lecite, non sono in grado di disinnescare e che, di conseguenza, ricade sulla società (§ 4.5.2.). Pur nella difficoltà di concepire la definizione di regole cautelari in un contesto di incertezza scientifica, si ritiene che una minimizzazione del rischio di decisioni erranee da parte dei sistemi di *machine learning* (e, dunque, di incidenti) potrà essere realizzata attraverso misure di corretta raccolta e gestione dei dati, di addestramento, di sperimentazione e, nella fase *post market*, di monitoraggio.

A tal proposito, abbiamo rilevato come i meccanismi di *explainability* in via di implementazione potrebbero avere un ruolo fondamentale nel monitoraggio dell'*adequatezza* delle cautele positive rispetto agli obiettivi di bilanciamento definiti dal decisore pubblico. Gli sviluppi nella registrazione algoritmica degli incidenti potrebbero infatti permettere di verificare – in maniera maggiormente precisa e puntuale rispetto al passato – se e quando emergano rischi *più elevati* o *ulteriori* rispetto a quelli che le cautele

scritte miravano a neutralizzare o minimizzare, consentendo così ai produttori di riconoscere immediatamente quei segnali univoci di “fallimento” delle regole cautelari che fanno scaturire l’obbligo di informazione alle autorità e ai consumatori e, nei casi più gravi, di ritiro o richiamo del prodotto. Così, se l’obiettivo del legislatore – attraverso la predisposizione di norme cautelari rigide ed esaustive – è quello di mantenere il tasso di incidenti algoritmici in un determinato settore al di sotto dell’x%, rispetto all’utilizzo complessivo dei dispositivi, un superamento di tale soglia costituirà un univoco segnale di fallimento delle regole scritte, che dovrà indurre il produttore ad adottare cautele più pregnanti (§ 4.5.2.3.).

Nonostante lo sforzo di ricostruzione sistematica dei tratti salienti di un’ipotetica responsabilità del produttore di dispositivi intelligenti, non possiamo non concordare con quell’autorevole dottrina che, da tempo, segnala come il diritto penale, di fronte alle sfide della società del rischio tecnologico, si trovi di fronte ad una vera e propria “crisi da incontenibilità”⁴¹⁹, determinata dalla difficoltà di ritenere nel “tipo” fenomeni complessi, globalizzati, imperscrutabili. In questo contesto, una tutela penale – la sola in grado, a nostro parere, di esprimere il disvalore insito nella creazione o nel mantenimento di rischi illeciti a beni giuridici di primaria importanza, quali la vita e l’integrità fisica – potrebbe appuntarsi in forma anticipata, attraverso il ricorso a reati colposi di mera condotta, eventualmente basati sulla tecnica ingiunzionale (§ 5.). Da un lato, potranno trovare applicazione le fattispecie contravvenzionali di pericolo previste dall’art. 112 cod. cons., commi 1, 2 e 3 (sebbene le cornici edittali ivi previste non paiano del tutto adeguate allo scopo di deterrenza); dall’altro, *de lege ferenda*, potrebbe essere opportuna l’introduzione di norme *ad hoc* che sanzionino il mancato rispetto di alcuni dei requisiti stabiliti dall’*AI Act* (si pensi, in particolare, all’omessa predisposizione di determinati presidi di sicurezza o, per quanto riguarda la posizione dell’utilizzatore del sistema di i.a., alla loro disattivazione prima dell’utilizzo). In un contesto in cui la messa in commercio (da parte del produttore) e l’attivazione (da parte dell’utilizzatore) di un sistema di intelligenza artificiale conducono a conseguenze non del tutto preventivabili, l’anticipazione della tutela penale può costituire un valido strumento per presidiare, preventivamente, beni giuridici di rilievo primario, impedendo che la complessa ricostruzione di fatto tipico ed elemento soggettivo dei reati causalmente orientati crei intollerabili spazi di impunità.

⁴¹⁹ C.E. PALIERO, *L’autunno del patriarca*, cit., p. 1238.

Maggiore spazio, in quest'ottica, potrebbe poi essere attribuito, sempre *de jure condendo*, alla valorizzazione delle tecniche ingiunzionali, favorendo così un coordinamento tra sanzione penale e capillare controllo amministrativo: non possiamo non rilevare, tuttavia, che misure di questo tipo – già presenti nella disciplina generale sulla sicurezza dei prodotti (vd. art. 112, co. 1 e 3, cod. cons.) – sono, ad ora, rimaste soltanto sulla carta.

In diversi punti della nostra trattazione, infine, è emersa una certa discrepanza tra il carattere necessariamente personalistico della responsabilità penale e i fenomeni criminosi oggetto di disamina, risultato dell'attività di centri politico-decisionali senza dubbio *collettivi*. Con riferimento alla posizione del produttore, sarebbe quantomai opportuna, come già evidenziato dalla dottrina, un'estensione dei reati-presupposto alle fattispecie di pericolo o di danno relative alla sicurezza dei prodotti (§ 4.4): in questo caso, l'art. 8 d.lgs. 231/2001 consentirebbe di muovere un rimprovero all'ente anche nel caso in cui non venisse identificata la persona responsabile all'interno dell'organigramma aziendale. Il medesimo meccanismo potrebbe invece risultare inadeguato laddove la scelta di politica d'impresa non si realizzi attraverso l'attività di un subordinato o di un apicale, quanto, piuttosto, attraverso la mediazione di un *algoritmo* (si pensi all'uso di *high frequency trading* sui mercati finanziari): in questo caso, la mancata identificazione dell'autore non è dovuta a cause di *disorganizzazione* interne all'impresa, ma, piuttosto, alla stessa incerta configurabilità di un *reato*, atteso che tra la scelta di utilizzare il sistema di i.a. e l'evento lesivo si frappone il *decision making* algoritmico (§ 3.4.).

CONCLUSIONI

Nel primo capitolo, descrivendo la *black box* tipica dell'intelligenza artificiale, abbiamo osservato come l'opacità dei sistemi di i.a. sembri costituire la tappa definitiva di quella *crisi di comprensione e controllo* che caratterizza il rapporto post moderno tra uomo e tecnologia (vd. cap. I, § 5.3.1.). Il dominio sul mondo non si realizza più, oggi, attraverso la decrittazione e l'interpretazione di fenomeni naturali oscuri, ma tramite la creazione di "agenti intelligenti", in grado di eseguire accuratamente le funzioni loro delegate, senza che, tuttavia, i processi decisionali che ne regolano il funzionamento siano pienamente comprensibili da parte dei loro stessi ideatori e produttori.

Di tutti gli interrogativi (giuridici ed extragiuridici) che lo sviluppo e la messa in commercio di prodotti "intelligenti" sollevano, nel nostro lavoro ci siamo concentrati sull'impatto che le nuove forme dell'agire algoritmico possono avere sull'imputazione della responsabilità per gli eventi lesivi scaturenti dai sistemi di intelligenza artificiale. Come abbiamo cercato di mostrare attraverso la panoramica sulla casistica (vd. cap. I, § 7), non si tratta di questioni meramente ipotetiche ed astratte, avulse da qualsiasi aderenza alla realtà attuale; si tratta, invece, di problemi che si sono già presentati negli ultimi anni – e che attendono le prime soluzioni giurisprudenziali.

Dopo una prima fase in cui l'impatto dell'intelligenza artificiale sulla responsabilità civile è stata oggetto, da parte delle istituzioni europee, di atti di indirizzo e di *soft law*, tra 2021 e 2022 la Commissione europea ha presentato diverse proposte di regolazione che, a vario titolo, mirano ad adattare la disciplina sulla sicurezza dei prodotti e i regimi di responsabilità aquiliana al nuovo contesto tecnologico. In particolare, il c.d. *AI Act*¹ ha l'obiettivo di dettare i requisiti minimi per lo sviluppo e la messa in commercio dei sistemi di i.a.; la proposta di revisione della direttiva sul danno da prodotto e la proposta di direttiva sulla responsabilità da intelligenza artificiale mirano invece ad alleggerire l'onere probatorio per il ricorrente che intenda ottenere ristoro per danni derivanti dall'utilizzo di un sistema di i.a. (es. presunzione di difettosità e presunzione di

¹ Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione, COM/2021/206 final, 21 aprile 2021.

causalità) e a garantire la *disclosure* degli elementi di prova rilevanti, da parte del produttore.

Nonostante le soluzioni proposte dalla Commissione europea – oggetto, attualmente, di dibattito parlamentare – si discostino, in parte, da quelle suggerite dalla dottrina, nel complesso l'intervento europeo è da salutare con favore. Da un lato, l'*AI Act* mira a bilanciare le esigenze dell'industria e dell'innovazione con la necessità di consentire l'accesso al mercato di prodotti che rispettino determinati requisiti di sicurezza; dall'altro, le due proposte in materia di responsabilità si pongono come obiettivo quello di approntare un rimedio effettivo a favore dei soggetti danneggiati, tale da superare le difficoltà probatorie insite nell'accertamento del nesso di causalità nell'ambito dei danni causati da sistemi intelligenti.

Sebbene non manchino le questioni problematiche (si pensi, ad esempio, alla controversa nozione di “difetto” nella responsabilità per danno da prodotto e all'impatto, sull'interpretazione della medesima nozione, del *machine learning*), la responsabilità extracontrattuale sembra confermarsi uno strumento flessibile, capace di adattarsi ai nuovi fenomeni dannosi e di allinearsi alle decisioni sulla gestione del rischio assunte a livello politico. Le regole sulla responsabilità per fatto illecito, d'altra parte, hanno attraversato la rivoluzione industriale e post-industriale senza subire sostanziali modifiche e senza tradire il proprio ruolo; anzi, si può dire che esse siano andate accentuando la propria centralità nella tutela dei danneggiati (si pensi, a mero titolo di esempio, alla continua espansione della risarcibilità dei danni non patrimoniali).

Il vasto assortimento di soluzioni prospettabili (responsabilità ai sensi dell'art. 2043 c.c.; responsabilità oggettiva, nelle versioni codicistiche e in quella per danno da prodotto; *Manufacturer Enterprise Liability*; personalità elettronica; e, dal punto di vista probatorio, presunzioni) è dovuto ad un modello risarcitorio *atipico*, non vincolato al rispetto di specifiche disposizioni costituzionali, ma delimitato soltanto dagli estesi confini del principio solidaristico (art. 2), del principio di uguaglianza (art. 3), del principio di libera iniziativa economica privata (art. 41).

Salva la necessità di qualche adattamento – e, a nostro parere, l'esigenza di dare nuovo impulso al criterio di imputazione oggettivo – la nostra opinione, che si unisce a quella dominante in dottrina, è che il diritto civile sia ben attrezzato per rispondere alle sfide poste dallo sviluppo e dalla commercializzazione su larga scala dei sistemi di

intelligenza artificiale, complice anche uno *standard* probatorio (quello del *più probabile che non*) meno rigoroso rispetto a quello penalistico. Il settore in esame, anzi, potrebbe costituire un vero e proprio laboratorio per la sperimentazione di soluzioni innovative, così come per l'interpretazione in chiave estensiva di modelli tradizionali di responsabilità oggettiva (es., responsabilità per danno da cose o animali; responsabilità per esercizio di attività pericolosa).

Il diritto penale, per contro, rischia di essere travolto (ancora una volta, verrebbe da dire) dalla fenomenologia del danno derivante da sistemi di i.a. Già messo a dura prova dalla fattispecie del danno da prodotto tradizionale, già piegato ad esigenze extrapenalistiche (rassicurazione sociale, catarsi collettiva nell'individuazione di un capro espiatorio, conforto delle vittime, per limitarci soltanto ad alcune di esse) in relazione a macrofenomeni di esposizione a sostanze tossiche e calamità naturali, il diritto penale potrebbe trovare nei danni da intelligenza artificiale un nuovo ed insidioso terreno di frizione tra bisogno percepito di pena e principi costituzionali.

Se l'ipotesi di una responsabilità penale diretta del sistema di i.a. è da relegare, a nostro parere, alla mera suggestione fantascientifica, i dubbi che circondano la configurabilità di una responsabilità penale per l'utilizzatore e per il produttore dei dispositivi intelligenti appaiono tutt'altro che infondati. Da un lato, c'è il rischio che in capo all'utilizzatore vengano posti obblighi di controllo su sistemi di i.a. il cui comportamento è *intrinsecamente* incomprensibile e imperscrutabile, determinando alcune possibili frizioni con il principio di colpevolezza; dall'altro, per quanto riguarda la posizione del produttore, l'assenza di conoscenze scientifiche sul funzionamento dei sistemi di i.a. mette in crisi il modello nomologico-deduttivo di accertamento della causalità, generando altresì, di conseguenza, un'analoga inadeguatezza della valutazione sulla prevedibilità dell'evento, tipica dell'accertamento sulla colpa.

Posto che le riflessioni esposte in questa sede sono tutt'altro che conclusive, dovendo confrontarsi con un fenomeno in divenire e con una casistica giudiziaria ancora agli albori, ci pare che lo studio dell'impatto dell'intelligenza artificiale sulla responsabilità penale possa costituire un avamposto privilegiato per osservare alcune tendenze del diritto penale di fronte a contesti di iper-complessità tecnologica e relazionale, e per riflettere sull'eventuale necessità di interventi normativi o, più semplicemente, di rinnovati paradigmi interpretativi.

Così è, ad esempio, per quanto riguarda la colpa. Nella nostra trattazione, in particolare, è emersa, da un lato, la necessità di valorizzare la dimensione soggettiva della colpa – specialmente per quanto riguarda la posizione dell'utilizzatore – e, dall'altro, l'opportunità di introdurre regole cautelari rigide ed esaustive, che delineino un'area di *rischio consentito* per il produttore di sistemi di i.a. In sintesi, misura soggettiva e determinatezza della fattispecie colposa: due istanze che sono attualmente al centro del dibattito penalistico e che proprio nella materia *de qua* potrebbero trovare terreno fertile per un'implementazione.

Ancora, in relazione ai danni da intelligenza artificiale parrebbe acuirsi quella crisi del delitto d'evento già preconizzata in relazione ad altri fenomeni tipici della c.d. *società del rischio* e potrebbe aprirsi la strada per una maggiore attenzione al *disvalore d'azione* – attraverso la predisposizione, *de jure condendo*, di forme anticipate di tutela penale. Se, infatti, l'evento lesivo algoritmico non è mai pienamente preventivabile (né evitabile) da parte del produttore e dell'utilizzatore, il nucleo del rimprovero da parte dell'ordinamento dovrà consistere, principalmente, nell'aver commercializzato/utilizzato un prodotto intelligente in assenza delle dovute cautele.

Infine, nel settore della produzione dei sistemi di i.a. emerge plasticamente la disarmonia tra carattere intrinsecamente *plurisoggettivo* degli illeciti e conformazione *personalistica* della responsabilità penale – disarmonia già riscontrata in relazione a tutti i fenomeni tipici della criminalità d'impresa. L'inserimento, nella compagine aziendale, di algoritmi che agiscono *nell'interesse o a vantaggio dell'ente* potrebbe amplificare quelle istanze, provenienti da una parte della dottrina, che sostengono la necessità di favorire una maggiore *autonomia* della responsabilità dell'ente rispetto a quella della persona fisica.

Resta, sullo sfondo, l'interrogativo su quale ruolo il diritto penale potrà concretamente svolgere, nel lungo periodo, in una società retta da meccanismi artificiali, il cui funzionamento non è del tutto conoscibile nemmeno ai loro creatori. Qual è il limite, insomma, oltre il quale la trama ordinamentale del diritto penale, ormai rammendata in più punti, diventerà irriconoscibile rispetto al tessuto originale? La sfida, tornando alla strisciante contrapposizione evidenziata in apertura del terzo capitolo, è di salvaguardare l'insostituibile *utilità sociale* del diritto penale, senza rinunciare alle garanzie che ne

costituiscono il fondamento democratico. Una sfida, ci pare, all'altezza di un' *intelligenza umana*².

² A proposito di intelligenze umane e intelligenze artificiali...Così avrebbe concluso questa tesi *ChatGPT*, il già citato sistema di *natural language processing* accessibile liberamente *online*: «In conclusione, la responsabilità penale per i danni causati dall'intelligenza artificiale rappresenta un tema di grande attualità e rilevanza, sia dal punto di vista teorico che pratico. La crescente diffusione e l'utilizzo sempre più ampio dell'IA in molti ambiti della nostra vita quotidiana, unitamente alla sua capacità di agire in modo autonomo e di prendere decisioni, pone infatti importanti questioni in termini di responsabilità e di imputazione delle conseguenze delle azioni dei sistemi di IA. In questo contesto, è fondamentale che si trovino soluzioni giuridiche adeguate che permettano di individuare in modo chiaro e preciso chi debba rispondere penalmente per i danni causati dall'IA, al fine di garantire una effettiva tutela dei diritti e degli interessi legittimi dei soggetti coinvolti». (Questo l'*input* fornito: "Scrivi una conclusione per tesi sulla responsabilità penale per danni causati da intelligenza artificiale").

BIBLIOGRAFIA

ABBOTT – SARCH, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, in 53 *U. Cal. Davis L.R.*, 2019, p. 323;

ABBOTT, *The Reasonable Robot. Artificial intelligence and the law*, Cambridge University Press, 2020;

ABERSMAN, *Overcomplicated. Technology at The Limits of Comprehension*, Gildan Media Corporation, 2016;

ABRAHAM – RABIN, *Automated vehicles and manufacturer responsibility for accidents: a new legal regime for a new era*, in *Va. L. Rev.*, 2019, p. 127 ss.;

ACCINNI, *Criteri di imputazione per colpa tra leggi scientifiche e accertamenti giudiziali*, in *Riv. it. dir. proc. pen.*, 2006, n. 3, p. 928 ss.;

AFFERNI, *L'analisi economica del diritto*, in Alpa (a cura di), *La responsabilità del produttore*, Giuffrè, 2019, p. 339 ss.;

AGUERRI, *Datos, teoría y causalidad: una breve aproximación al retorno a la teoría*, in F.J. Castro Toledo (coord.), *La transformación algorítmica del sistema de justicia penal*, Thomson Reuters, 2022, p. 77 ss.;

AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, in A.M. Rodrigues (coord.), *Inteligência artificial no direito penal*, 2020, Almedina, p. 59 ss.;

AL MUREDEN, *Event Data Recorder e Advanced Driver Assistance Systems: la "spinta gentile" verso la mobilità del futuro*, in *Contr. Impr.*, 2022, n. 2, p. 390 ss.;

AL MUREDEN, *Il danno da "prodotto conforme". Le soluzioni europee e statunitensi nella prospettiva del "Transatlantic Trade and Investment Partnership" (T.T.I.P.)*, in *Contr. e impr.*, 2015, p. 388;

AL MUREDEN, *La sicurezza dei prodotti e la responsabilità del produttore*, Torino, 2017;

AL MUREDEN, *Sicurezza "ragionevole" degli autoveicoli e responsabilità del produttore nell'ordinamento giuridico italiano e negli Stati Uniti*, in *Contr. e impr.*, 2012, p. 1506;

ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, in *Europa e dir. priv.*, 2019, fasc. 4, p. 995 ss.

ALBUS, *Outline for a theory of intelligence*, in *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 21, Iss. 3, May-June 1991, p. 473 ss.;

ALESSANDRI, *Attività d'impresa e responsabilità penali*, in *Riv. it. dir. proc. pen.*, 2005, fasc. 2, p. 565 ss.;

ALESSANDRI, *Parte generale*, in Pedrazzi e al. (a cura di), *Manuale di diritto penale dell'impresa*, Monduzzi, II ed., 2000;

ALEXANDER – FERZAN, *Against Negligence Liability*, in P.H. Robinson e al. (eds.), *Criminal Law Conversations*, Oxford University Press, 2011, p. 273 ss.;

ALLAIN, *From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems*, in *73 La. L. Rev.*, 2013, p. 1049 ss.;

ALPA, *Ars interpretandi e responsabilità sanitaria a seguito della nuova legge Bianco-Gelli*, in *Contr. impr.*, 2017, n. 3, p. 728 ss.;

ALPA, *Colpa e responsabilità oggettiva nella prospettiva dell'analisi economica del diritto*, in *Politica del diritto*, 1976, p. 431 ss.;

ALPA, *La responsabilità civile. Parte generale*, Utet, 2010;

ALPA, *Quale modello normativo europeo per l'intelligenza artificiale?*, in *Contr. impr.*, 2021, n. 4, p. 1003 ss.

ALPA, *Responsabilità dell'impresa e tutela del consumatore*, Giuffrè, 1975;

AMATO, *Product Liability and Product Security: Present and Future*, in Lohsse e al. (eds.), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 79 ss.;

AMENDOLA, *Ilva e diritto alla salute. La Corte Costituzionale ci ripensa?*, in *Quest. Giust.*, 10 aprile 2018;

AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, in *Giur. it.*, luglio 2019, p. 1715 ss.;

AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 2, cit, p. 281 ss.;

AMISANO, *Prevedere -e non predire- attraverso gli algoritmi e le loro insidie*, in *Arch pen.*, 2022, n. 2;

AMOROSO – TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto: quale ruolo per il controllo umano?*, in *BioLaw Journal*, 2019, n. 1, p. 33 ss.;

ANTOLISEI, *L'azione e l'evento nel reato*, Istituto Editoriale Scientifico, 1928;

ARIFOVIC – HE – WEI, *Machine learning and speed in high-frequency trading*, in *Journal of Economic Dynamics & Control*, 139, 2022, p. 1 ss.;

ASARO, *A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, in Lin e al. (eds.), *Robot Ethics*, cit., p. 169 ss.;

ASARO, *Determinism, machine agency, and responsibility*, in *Pol e soc.*, 2014, n. 2, p. 265 ss.;

ASARO, *How just could a robot war be?*, in Brey – Briggie – Waelbers (eds.), *Current Issues in Computing and Philosophy*, Ios Press, p. 50 ss.;

ASARO, *The liability problem for autonomous artificial agents*, in *AAAI Symposium on Ethical and Moral Considerations in Non-Human Agents*, 2016, p. 190 ss.;

ATHALYE E AL., *Synthesizing Robust Adversarial Examples*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*;

ATTILI, *L'agente-modello "nell'era della complessità": tramonto, eclissi o trasfigurazione?*, in *Riv. it. dir. e proc. pen.*, 2006, fasc. 4, p. 1240 ss.;

AULETTA, *Googled: The End of the World as We Know It*, Penguin Press, 2009;

AZZALI, *La responsabilità penale del produttore per danni alla salute*, in *Riv. trim. dir. pen. ec.*, 1991, p. 849 ss.;

AZZALI, *La tutela penale dell'ambiente. Un'indagine di diritto comparato*, Cedam, 2001;

BAGNOLI, *Teoria della responsabilità*, Il Mulino, 2019;

BALKIN, *The Path of Robotics Law*, in *6 Calif. L. Rev. Circuit*, 2015, p. 49 ss.;

BARONE, *Intelligenza artificiale e processo penale: la linea dura del Parlamento Europeo. Considerazioni a margine della Risoluzione del Parlamento Europeo del 6 ottobre 2021*, in *Cass. pen.*, 2022, fasc. 3, p. 1180 ss.;

BARTOLI, (voce) *Fonti della colpa*, in Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, Giuffrè, 2022, p. 519 ss.;

BARTOLI, *Il problema della causalità penale. Dai modelli unitari al modello differenziato*, Giappichelli, 2010;

BARTOLI, *Riforma Gelli-Bianco e Sezioni Unite non placano il tormento: una proposta per limitare la colpa medica*, in *Dir. pen. cont.*, 2018, n. 5, p. 233 ss.;

BARTOLINI, *Auto a guida autonoma e problemi di responsabilità civile*, in Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 299 ss.;

BARTOLUCCI, *L'art. 8 d.lgs. 231/2001 nel triangolo di Penrose. Tra minimizzazione del rischio-reato d'impresa e 'nuove forme' di colpevolezza*, in *Dir. pen. cont.*, 9 gennaio 2017.

BASHAYREH – SIBAI – TABBARA, *Artificial Intelligence and legal liability: towards an international approach of proportional liability based on risk sharing*, in *Information and Communications Technology Law*, 2020, 30 (2), p. 169 ss.;

BASILE – VALLAR, *Diritto penale e neuroscienze*, in *Giornale Italiano di Psicologia*, n. 4, 2016, p. 799 ss.;

BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in G. Balbi e al. (a cura di), *Diritto penale e intelligenza artificiale. "Nuovi Scenari"*, Giappichelli, 2023, p. 1 ss.;

BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2019, fasc. 10;

BASILE, *La colpa in attività illecita: un'indagine di diritto comparato sul superamento della responsabilità oggettiva*, Giuffrè, 2005;

BASSINI – LIGUORI – POLLICINO, *Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. Pizzetti (a cura di), *Intelligenza artificiale*, cit., p. 349 ss.;

BASTOS – MERCEA, *The Brexit Botnet and User-Generated Hyperpartisan News*, in *Social Science Computer Review*, 2019, Vol. 37(1), p. 38 ss.

BATHAE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *31 Harvard Journal of Law & Technology*, 2018, n. 2, p. 889 ss.;

BAUM E AL., *How long until human-level AI? Results from an expert assessment*, in *78 Technological Forecasting and Social Change*, p. 185 ss.;

BECK S., *Dealing with the diffusion of legal responsibility: the case of robotics*, in F. Battaglia – N. Mukerji – J. Nida-Rümelin (eds.), *Rethinking Responsibility in Science and Technology*, Pisa University Press, 2014, 167 ss.;

BECK S., *Google Cars, Software Agents, Autonomous Weapons Systems – New Challenges for Criminal Law?*, in E. Hilgendorf – U. Seidel (eds.), *Robotics, Autonomics, and the Law*, Nomos, 2017, p. 227 ss.;

BECK S., *Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood*, in 86 *Robotics and Autonomous Systems*, 2016, p. 138 ss.;

BECK S., *The problem of ascribing legal responsibility in the case of robotics*, in 31 *AI & Soc.*, 2016, p. 473 ss.

BECK U., *La società del rischio. Verso una seconda modernità*, Carocci, 2000 (ed. ted. 1986);

BECKERS – TEUBNER, *Three Liability Regimes for Artificial Intelligence*, Bloomsbury, 2022;

BEKEY, *Current Trends in Robotics: Technology and Ethics*, in P. Lin e al. (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, 2012, p. 17 ss.;

BELLISARIO, *Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, in *Eur. dir. priv.*, 2016, fasc. 3, p. 841 ss.;

BELLISARIO, *La rilevanza del criterio presuntivo della conformità alle norme armonizzate*, in *Pers. merc.*, 2012, p. 156 ss.;

BELVEDERE, (voce) *Definizioni*, in *Digesto*, IV ed., vol. V, 1989, p. 149 ss.;

BELVEDERE, *Il problema delle definizioni nel codice civile*, Giuffrè, 1977;

BENJAMIN, *Race After Technology: Abolitionist Tools for the New Jim Code*, Polity, 2019.

BERIAIN, *Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling*, in *Law, Probability & Risk*, 2018, p. 45 ss.;

BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, in *Riv. trim. dir. pen. ec.*, 2003, p. 1 ss.

BERTOLINI – AIELLO, *Robot companions: A legal and ethical analysis*, in *The Information Society*, 2018, vol. 34, no. 3, p. 130 ss.;

BERTOLINI – EPISCOPO, *Robots and AI as Legal Subjects? Disentangling the Ontological and Functional Perspective*, in *Front. Robot. AI*, 2022, vol. 9;

BERTOLINI – PALMERINI, *Regulating Robotics: A Challenge for Europe*, in *Upcoming Issues of EU Law*, 2014, p. 167 ss.;

BERTOLINI, *Artificial Intelligence Does Not Exist! Defying the Technology-Neutrality Narrative in the Regulation of Civil Liability for Advanced Technologies*, in *Europa e dir. priv.*, 2022, n. 2, p. 369 ss.;

BERTOLINI, *Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules*, in *5 Law Innovation & Tech.*, 2013, p. 214 ss.;

BERTOLINI, *Study on Artificial Intelligence and Civil Liability* – Study requested by the JURI Committee, July 2020;

Bertolino – Ubertis (a cura di), *Prova scientifica, ragionamento probatorio e decisione giudiziale*, Atti del Convegno, Milano, 10-11 ottobre 2014, Jovene, 2015;

BESSI – FERRARA, *Social bots distort the 2016 U.S. Presidential Election Online Discussion*, in *First Monday*, vol. 21, n. 11, 2016;

BIANCA, *Diritto civile: Vol. V – La responsabilità*, Giuffrè, 2020;

BIANCHI, *Appunti per una teoria dell'autonormazione*, in *Riv. it. dir. proc. pen.*, 2019, p. 1477 ss.

BICHI, *Intelligenza digitale, giurmetria, giustizia predittiva e algoritmo decisorio. Machina sapiens e il controllo sulla giurisdizione*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 423;

BISHOP, *Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It*, in *Front. Psychol.*, 5 gennaio 2021, disponibile a questo link: <https://doi.org/10.3389/fpsyg.2020.513474>;

BLAIOTTA, *Casualità e colpa: diritto civile e diritto penale si confrontano*, in *Cass. pen.*, 2009, fasc. 1, p. 78 ss.;

BLAIOTTA, *Causalità giuridica*, Giappichelli, 2010;

- BLAIOTTA, *Dove va la colpa penale?*, in *Dir. pen. proc.*, 2021, n. 10, p. 1281 ss.;
- BLAIOTTA, *Incidenti da aviolancio, problemi causali nuovi, attività rischiose*, in *Cass. pen.*, 2004, n. 7-8, p. 2375 ss.;
- BLAUTH E AL., *Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI*, in *IEEE Access* 2022, Vol 10, p. 77110 ss.
- BODI, *The First Amendment Implications of Regulating Political Deepfakes*, in 47 *Rutgers Computer & Tech. L.J.*, 2021, p. 143;
- BÖHM E AL., *New developments on EDR (Event Data Recorder) for automated vehicles*, in *Open Engineering*, 2020, vol. 10, iss. 1, p. 140 ss.;
- BORDOT, *Artificial Intelligence, Robots and Unemployment: Evidence from OECD Countries*, in *Journal of Innovation Economics & Management*, vol. 37, n. 1, 2022, p. 117 ss.;
- BORGES, *New Liability Concepts: the Potential of Insurance and Compensation Funds*, in Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 145;
- BORGHETTI, *Civil Liability for Artificial Intelligence: What Should its Basis Be?*, in *Revue des Juristes de Sciences Po*, 2019, n. 17, p. 94 ss.
- BORGHETTI, *How can Artificial Intelligence be Defective?*, in Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 63 ss.;
- BORGOGNO, *La responsabilità penale nei processi ad elevata automazione*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 727 ss.;
- BORSARI, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *Medialaws*, 2019, n. 3, p. 262 ss.
- BOSCARATO, *Who is Responsible for a Robot's Actions? An initial examination of Italian Law within a European Perspective*, in Van den Berg – Klaming (eds.), *Technologies on the Stand: Legal and Ethical Questions in Neuroscience and Robotics*, Wolf Legal Pubns, 2011, p. 383 ss.;
- BOSTROM, *Superintelligenza*, Bollati Boringhieri, 2014;
- BOVENS, *The quest for responsibility: accountability and citizenship in complex organisations*, Cambridge University Press, 1998;

BRATTBERG –MAURER, *Russian Election Interference. Europe's Counter to Fake News and Cyber Attacks*, Carnegie Endowment for International Peace, Maggio 2018;

BRAYNE, *Predict and Surveil: Data, Discretion, and the Future of Policing*, Oxford University Press, 2020;

BRICOLA, *Responsabilità penale per il tipo e per il modo di produzione*, in AA.VV., *La responsabilità dell'impresa per i danni all'ambiente e ai consumatori*, Milano, 1978, p. 75 ss.;

BRICOLA, *Teoria generale del reato*, in *Nov. dig. it.*, XIX, 1973;

BROWN, *Deepfakes and the weaponization of disinformation*, in *23 Virginia Journal of Law & Technology*, 2020, p. 1 ss.;

BRUNDAGE E AL., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, Future of Humanity Institute, 2018;

BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione. La c.d. "flessibilizzazione delle categorie del reato"*, in *Criminalia*, 2012, p. 383 ss.;

BUITEN E AL., *EU Liability Rules for the Age of Artificial Intelligence*, Centre on Regulation in Europe, 9 aprile 2021;

BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 2019, fasc. 4, p. 1908 ss.;

BURRELL, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, in *Big Data & Society*, January-June 2016, p. 1 ss.;

BUSSANI, *La responsabilità civile al tempo della crisi*, in *Riv. trim. dir. proc. civ.*, 2015, fasc. 2, p. 567 ss.;

BUYERS, *Artificial Intelligence. The practical legal issues*, Law Brief Publishing, 2018;

BUZIO, *Regole cautelari eurounitarie e colpa penale. Spunti per una riflessione*, in *Discrimen*, 17 aprile 2020;

CABRAL, *Liability and artificial intelligence in the EU: Assessing the adequacy of the current Product Liability Directive*, in *Maastricht Journal of European and Comparative Law*, 2020, vol. 27(5), p. 615 ss.;

Cadoppi (a cura di), *Il problema delle definizioni legali nel diritto penale. Ominis definitio in iure periculosa?*, Cedam, 1996;

CAFAGGI, *A Coordinated Approach to Regulation and Civil Liability in European Law: Rethinking Institutional Complementarities*, in Cafaggi (a cura di), *The Institutional Framework of European Private Law*, Oxford University Press, 2006, p. 191 ss.;

CAFAGGI, *Verso un approccio sistematico ai rimedi contro la difettosità dei prodotti*, in Patti (a cura di), *Il danno da prodotti in Italia, Austria, Repubblica Federale di Germania, Svizzera*, Cedam, 1990, p. 83 ss.

CALABRESI – AL MUREDEN, *Driverless Cars*, Il Mulino, 2021;

CALABRESI – HIRSCHOFF, *Toward a Test for Strict Liability in Torts*, in 81 *Yale Law Journal*, 1972, p. 1060 ss.;

CALABRESI., *The Costs of Accidents. A Legal and Economic Analysis*, Yale University Press, 1970, trad. it. *Costo degli incidenti e responsabilità civile. Analisi economico-giuridica*, Giuffrè, 1975;

CALDWELL E AL., *AI-enabled Future Crime*, in *Crime Science*, 9:14, 2020, p. 1 ss.;

CALO, *Artificial Intelligence Policy: A Primer and Roadmap*, in 3 *University of Bologna Law Review*, p. 180–218. p. 399 ss.;

CALO, *Robotics and the Lessons of Cyberlaw*, in 103 *Calif. L. Rev.*, 2015, p. 513 ss.;

CALUDE – LONGO, *The Deluge of Spurious Correlations in Big Data*, in *Foundations of Science*, 22 (3), p. 1 ss.;

CALVERLEY, *Imagining a Nonbiological Machine as a Legal Person*, in *Artificial Intelligence & Society*, 2008, p. 22 ss.;

CAMPOLO – CRAWFORD, *Enchanted Determinism: Power without Responsibility in Artificial Intelligence*, in *Engaging Science, Technology, and Society*, 2020, n. 6;

CANEPA, *L'imputazione soggettiva della colpa. Il reato colposo come punto cruciale nel rapporto tra illecito e colpevolezza*, Giappichelli, 2011;

Canzio – Lupária Donati (a cura di), *Prova scientifica e processo penale*, Cedam, II ed., 2022;

CANZIO, *Intelligenza artificiale, algoritmi e giustizia penale*, in *Sist. pen.*, 8 gennaio 2021;

CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018;

CAPPELLINI, *Profili penalistici delle self-driving cars*, in *Dir. pen. cont. – Riv. Trim.*, 2019, n. 2, p. 326 ss.;

CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, in G. Balbi e al. (a cura di), *Diritto penale e intelligenza artificiale. "Nuovi Scenari"*, Giappichelli, 2023, p. 19 ss.;

CARNEVALI, (voce) *Responsabilità del produttore*, in *Enc. Dir.*, Agg. II, 1998;

CARNEVALI, *Appunti di diritto privato*, Raffaello Cortina Editore, IX ed., 2010;

CARNEVALI, *La norma tecnica da regola di esperienza a norma giuridicamente rilevante. Ricognizione storica e sistemazione teorica. Ruolo dell'UNI e del CEI*, in *Resp. civ. e prev.*, 1997, contenuto in E. Al Mureden (a cura di) *La sicurezza dei prodotti e la responsabilità del produttore. Casi e materiali*, Giappichelli, 2020, p. 50 ss.;

CARNEVALI, *La responsabilità del produttore. Problemi generali*, in Aa.Vv., *La responsabilità dell'impresa per i danni all'ambiente e ai consumatori*, Giuffrè, 1978,

CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, in *Resp. civ. prev.*, fasc. 1, 2005, contenuto in E. Al Mureden (a cura di) *La sicurezza dei prodotti e la responsabilità del produttore. Casi e materiali*, Giappichelli, 2020, p. 61 ss.;

CARUANA E AL., *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission*, in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015;

CARUSO – PARDOLESI, *Per una storia della Direttiva 1985/374/CEE*, in Pardolesi (a cura di), *I 25 anni di products liability*, quaderno speciale di *Danno e resp.*, 2012, p. 9 ss.

CASONATO – MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 2021, 3, p. 415 ss.;

CASTELVECCHI, *Can We Open the Black Box of AI?*, in *Nature*, 5 ottobre 2016, disponibile a questo link: <https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731>;

CASTRONOVO, *La nuova responsabilità civile*, Giuffrè, 2006;

- CASTRONUOVO, *La responsabilità civile*, Giuffrè, 2018;
- CASTRONUOVO, *Sistema e problema del danno da prodotto*, Giuffrè, 1979;
- CASTRONUOVO, (voce) *Colpa penale*, in Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, 2022, p. 200 ss.;
- CASTRONUOVO, *La colpa “penale”. Misura soggettiva e colpa grave*, in *Riv. it. dir. proc. pen.*, 2013, n. 4, p. 1723 ss.;
- CASTRONUOVO, *La colpa penale*, Giuffrè, 2009;
- CASTRONUOVO, *Principio di precauzione e beni legati alla sicurezza*, in *Dir. pen. cont.*, 21 luglio 2011;
- CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell’incertezza nella struttura del reato*, Aracne, 2012;
- CASTRONUOVO, *Responsabilità da prodotto e struttura del fatto colposo*, in *Riv. it. dir. proc. pen.*, 2005, p. 301 ss.;
- CATENACCI, *La tutela penale dell’ambiente. Contributo all’analisi delle norme penali a struttura «sanzionatoria»*, Cedam, 1996;
- CENTONZE, *La normalità dei disastri tecnologici: il problema del congedo dal diritto penale*, Giuffrè, 2004;
- CENTONZE, *Scienza “spazzatura” e scienza “corrotta” nelle attestazioni e valutazioni dei consulenti tecnici nel processo penale*, in *Riv. it. dir. proc. pen.*, 2001, fasc. 4, p. 1232 ss.;
- CERINI, *Responsabilità del produttore e rischio da sviluppo: oltre la lettera della dir. 85/374/CE*, in *Dir. ed econ. assic.*, 1996, p. 29 ss.;
- CHAGAL-FEFERKORN, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, in *30 Stanford Law & Policy Review*, 2019, p. 61 ss.;
- CHAGAL-FEFERKORN, *The Reasonable Algorithm*, in *University of Illinois Journal of Law, Technology & Policy*, 2018, n. 1, p. 111 ss.;
- CHENEY-LIPPOLD, *We Are Data: Algorithms and the Making of Our Digital Selves*, NYU Press, 2018;

CHESINI, *"Terminator Scenario"? Intelligenza artificiale nel conflitto armato: "Lethal Autonomous Weapons Systems" e le risposte del diritto internazionale umanitario*, in *BioLaw Journal*, 2020, fasc. 3, p. 441 ss.;

CHOPRA – WHITE, *A Legal Theory for Autonomous Artificial Agents*, The University of Michigan Press, 2011;

CINGARI, *Causalità psichica e massime di esperienza: un modello differenziato di causalità?*, in *Dir. pen. proc.*, 2009, fasc. 6, p. 767 ss.;

CITRON, *Technological Due Process*, 85 *Wash.U. L. Rev.*, 2008, p. 1249 ss.;

CIVELLO, (voce) *Prevedibilità e reato colposo*, in Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, 2022, p. 1004 ss.;

CIVELLO, *La "colpa eventuale" nella società del rischio. Epistemologia dell'incertezza e "verità soggettiva" della colpa*, Giappichelli, 2013;

CLARKE, *Profiles of the Future: An Enquiry into the Limits of the Possible*, Harper & Row, 1973;

COCA VILA, *Coches autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal*, in *Cuadernos de política criminal*, 2017, 235 ss.;

COECKELBERGH, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, in 26 *Science and Engineering Ethics*, 2020, p. 2051 ss.

COMPORTI, *Esposizione al pericolo e responsabilità civile*, Morano Editore, 1965;

CONFORTINI, *Posizioni della dottrina e della giurisprudenza italiana sul tema della responsabilità civile del produttore per la messa in commercio di prodotti (difettosi e) dannosi*, in *Resp. civ.*, 1977, p. 544 ss.;

CONSOLO, *Riconoscimento di sentenze, specie USA e di giurie popolari, aggiudicanti risarcimenti punitivi o comunque sopracompensativi, se in regola con il nostro principio di legalità (che postula tipicità e financo prevedibilità e non coincide pertanto con il, di norma presente, due process of law)*, in *Corr. giur.*, 2017, p. 1050;

CONSORTE, *Colpa e linee guida*, nota a Cass. pen., sez. IV, 23 novembre 2010, n. 8254, in *Dir. pen. proc.*, 2011, n. 10, p. 1228 ss.;

CONSORTE, *Tutela penale e principio di precauzione*, Giappichelli, 2013;

CONSULICH, (voce) *Rischio consentito*, in Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, 2022, p. 1102 ss.;

CONSULICH, (voce) *Tutela del consumatore*, in Palazzo – Paliero (diretto da), *Commentario breve alle leggi penali complementari*, Cedam, II ed., 2007, p. 2967 ss.;

CONSULICH, *Errare commune est. Il concorrente colposo, il nuovo protagonista del diritto penale d'impresa (e non solo)*, in *Leg. pen.*, 28 marzo 2022;

CONSULICH, *Flash offenders. Le prospettive di accountability penale nel contrasto alle intelligenze artificiali devianti*, in *Riv. it. dir. proc. pen.*, 2022, fasc. 3, p. 1015 ss.;

CONSULICH, *Il nastro di Möbius. intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in *Banca borsa*, 2018, vol. 71, n. 2, p. 195 ss.;

CONSULICH, *Il principio di autonomia della responsabilità dell'ente. Prospettive di riforma dell'art. 8*, in *Rivista 231 – La resp. amm. delle soc. e degli enti*, 2018, n. 4, p. 195 ss.;

CONTISSA – LAGIOIA – SARTOR, *La manopola etica: i veicoli autonomi eticamente personalizzabili e il diritto*, in *Sistemi Intelligenti*, 2017, n. 3, p. 601 ss.;

COPPINI, *Robotica e intelligenza artificiale: questioni di responsabilità civile*, in *Pol. dir.*, 2018, fasc. 4, p. 713 ss.;

CORDIANO, *Sicurezza dei prodotti e tutela preventiva dei consumatori*, Cedam, 2005;

CORN, *Il principio di precauzione nel diritto penale. Studio sui limiti all'anticipazione della tutela penale*, Giappichelli, 2013;

CORNACCHIA, *Concorso di colpe e principio di responsabilità per fatto proprio*, Giappichelli, 2004;

CORNACCHIA, *Il problema della c.d. causalità psichica rispetto ai condizionamenti mentali*, in S. Canestrari – G. Fornasari (a cura di), *Nuove esigenze di tutela nell'ambito dei reati contro la persona*, Clueb, 2001, p. 198 ss.;

CORNACCHIA, *Posizioni di garanzia nell'ambito della sicurezza agroalimentare*, in Foffani – Doval Pais – Castronuovo (a cura di), *La sicurezza agroalimentare nella prospettiva europea*, Giuffrè, 2014, p. 457 ss.;

CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, in *Arch. pen.*, 2022, n. 2;

CORRADO, *Is There an Act Requirement in the Criminal Law?*, in 142 *University of Pennsylvania Law Review*, 1994, p. 1529 ss.;

CORSI, *Le Sezioni Unite: via libera al riconoscimento di sentenze comminatorie di punitive damages*, in *Danno resp.*, 2017, n. 4, p. 429;

COSTANZA, *Impresa robotizzata e responsabilità*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 106 ss.;

COTUGNO, *Intelligenza artificiale e tutela del consumatore. I nuovi obblighi informativi*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 2, il Mulino, 2022., p. 305 ss.

CRAWFORD, *Né intelligente né artificiale. Il lato oscuro dell'IA*, Il Mulino, 2021;

CRESPO, *“Compatibilismo umanista”: una proposta per conciliare le neuroscienze e il diritto penale*, in *Riv. it. dir. proc. pen.*, 2022, fasc. 2, p. 645 ss.;

CUNNEN E AL., *Artificial Driving Intelligence and Moral Agency: Examining the Decision Ontology of Unavoidable Road Traffic Accidents through the Prism of the Trolley Dilemma*, in *Applied Artificial Intelligence - An International Journal*, 2019, Vol. 33, No. 3, p. 267 ss.;

CUPELLI, *L'art. 590-sexies c.p. nelle motivazioni delle Sezioni Unite: un'interpretazione 'costituzionalmente conforme' dell'imperizia medica (ancora) punibile*, in *Dir. pen. cont.*, 2018, n. 3;

D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, in *Dir. pen. cont.*, 2019, 2, 359 ss.;

D'ALESSANDRO, (voce) *Precauzione (principio di)*, in Donini (diretto da), *Reato colposo, Enc. dir. – I Tematici*, 2022, p. 975 ss.;

D'ALESSANDRO, *Pericolo astratto e limiti soglia. Le promesse non mantenute del diritto penale*, Giuffrè, 2012;

D'ALESSANDRO, *Spiegazione causale mediante leggi scientifiche, a dieci anni dalla sentenza Franzese*, in *Criminalia*, 2012, p. 331 ss.;

D'AMICO, *Intelligenza artificiale e auto a guida autonoma. tra prevenzione primaria, colpa penale e rischio consentito*, in *Riv. it. med. leg.*, 2022, fasc. 3, p. 593 ss.;

DANAHER, *Robots, Law and the Retribution Gap*, in 18 *Ethics and Information Technology*, 2016, p. 299 ss.;

DAS E AL., *On Performance Evaluation of Driver Hand Detection Algorithms: Challenges, Dataset, and Metrics*, in 2015 *IEEE 18th International Conference on Intelligent Transportation Systems*;

DAVOLA – PARDOLESI, *In viaggio col robot: verso nuovi orizzonti della r.c. auto (“driverless”)?*, in *Danno e resp.*, 2017, p. 616 ss.;

DAVOLA, *A Model for Tort Liability in a World of Driverless Cars: Establishing a Framework for the Upcoming Technology*, in 54 *Idaho L. Rev.*, 2018, p. 591 ss.;

DE BRUYNE – TANGHE, *Liability for Damage Caused by Autonomous Vehicles: A Belgian Perspective*, in *Journal of European Tort Law*, vol. 8, iss. 3, p. 324 ss.;

DE BRUYNE E AL., *Tort Law and Damage Caused by AI Systems*, in J., De Bruyne (ed.), *Artificial Intelligence and the Law*, Intersentia, 2021, p. 359 ss.;

DE FRANCESCO G., *“Interpersonalità” dell’illecito penale: un ‘cuore antico’ per le moderne prospettive della tutela*, in *Cass. pen.*, 2015, n. 2, p. 854 ss.;

DE FRANCESCO G., *L’imputazione del reato e i tormenti del penalista*, in M. Bertolino – G. Forti (a cura di) *Scritti per Federico Stella*, vol. I, Jovene, 2007, p. 513 ss.;

DE FRANCESCO V., *Sulla misura soggettiva della colpa*, in *Studi Urbinati*, 1977, p. 273 ss.;

DE VERO, *Il nesso causale e il diritto penale del rischio*, in *Riv. it. dir. e proc. pen.*, 2016, p. 670 ss.;

DE VITO, *La salute, il lavoro, i giudici*, in *Quest. Gius.*, 24 marzo 2018;

DEL RE, *Per un riesame della responsabilità colposa*, in *Indice penale*, 1985, p. 31 ss.;

DELFINO, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn’s Next Tragic Act*, in 88 *Fordham L. Rev.*, 2019, p. 887 ss.;

DELITALA, *Il “fatto” nella teoria generale del reato*, Cedam, 1930;

DELLACASA, *Sulle definizioni legislative nel diritto privato. Fra codice e nuove leggi civili*, Giappichelli, 2004;

DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, in C. Piergallini e al. (a cura di), *Studi in onore di Carlo Enrico Paliero*, vol. 1 – Politica criminale e teoria della pena, Giuffrè, 2023, p. 63 ss.;

DESAI – KROLL, *Trust But Verify: A Guide to Algorithms and the Law*, in *Harvard journal of law & technology*, 2017, Vol. 31, n. 1, p. 1 ss.;

DESKUS, *Fifth Amendment Limitations on Criminal Algorithmic Decision-Making*, in *NYU Journal of Legislation and Public Policy*, 2018, p. 237 ss.;

DEVILLÉ E AL., *Basic Concepts f AI for Legal Scholars*, in De Bruyne – Vanleenhove (eds.), *Artificial Intelligence and the Law*, cit., p. 12 ss.;

DHEU E AL., *The European Commission's Approach To Extra-Contractual Liability and AI – A First Analysis and Evaluation of the Two Proposals*, CiTiP Working Paper Series, 6 ottobre 2022, in SSRN, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4239792;

DI CIOMMO, *Valori e funzioni della responsabilità civile nell'epoca del post-turbocapitalismo*, in *Danno resp.*, 2021, n. 2, p. 137 ss.;

DI FLORIO, *La cooperazione nel delitto colposo: una fattispecie con una (problematica) funzione incriminatrice*, in *Arch pen.*, 2021, n. 1.

DI GIOVINE, *Il judge-Bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Cass. pen.*, 2020, 951 ss.;

DI GIOVINE, *Lo statuto epistemologico della causalità penale tra cause sufficienti e condizioni necessarie*, in *Riv. it. dir. proc. pen.*, 2002, fasc. 2, p. 634 ss.;

DI GIOVINE, *Ripensare il diritto penale attraverso le (neuro-)scienze?*, Giappichelli, 2019;

DI LANDRO, *Dalle linee guida e dai protocolli all'individualizzazione della colpa penale nel settore sanitario*, Giappichelli, 2012;

DI LANDRO, *La colpa medica negli Stati Uniti e in Italia. Il ruolo del diritto penale e il confronto col sistema civile*, Giappichelli, 2009;

DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, in Bartoli (a cura di), *Responsabilità penale e rischio nelle attività mediche e d'impresa: (un dialogo con la giurisprudenza)*, Firenze University press, 2010, p. 437 ss.;

DIAKOPOULOS, *Algorithmic Accountability: Journalistic investigation of computational power structures*, in 3 *Digit. Journal*, 2015, p. 398 ss.;

DIAMANTIS, *Algorithms Acting Badly: a Solution From Corporate Law*, in 89 *Geo. Wash. L. Rev.*, p. 1 ss.;

DIAMANTIS, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, in 98 *N.C. L. Rev.*, 2020, p. 894 ss.;

DIAMANTIS, *Vicarious Liability for AI*, Iowa Legal Studies Research Paper No. 27, 2021;

DOMINGOS, *A Few Useful Things to Know about Machine Learning*, in 55 *Communications of the ACM*, nov. 2012, p. 78 ss.;

DOMINIONI, *La prova penale scientifica: gli strumenti scientifico-tecnici nuovi o controversi e di elevata specializzazione*, Giuffrè, 2005;

DONINI, *Il garantismo della condicio sine qua non e il prezzo del suo abbandono. Contributo all'analisi dei rapporti fra causalità e imputazione*, in *Riv. it. dir. e proc. pen.*, 2011, n. 2, p. 494 ss.;

DONINI, *Il volto attuale dell'illecito penale*, Giuffrè, Milano, 2004;

DONINI, *La causalità omissiva e l'imputazione "per l'aumento del rischio"*, in *Riv. it. dir. proc. pen.*, 1999, fasc. 1, p. 32 ss.;

DONINI, *Teoria del reato. Una introduzione*, Cedam, 1996;

DORMEHL, *Thinking Machines: The Quest for Artificial Intelligence and Where It's Taking Us Next*, Tarcherperigree, 2017;

DORMONT, *Quel régime de responsabilité pour l'intelligence artificielle*, in *Contrats, concurrence, consommation*, 2018, études 19;

DOUMA – PALODICHUK, *Criminal Liability Issues Created by Autonomous Vehicles*, in 52 *Santa Clara Law Review*, p. 1157 ss.;

DOWELL, *Fundamental Protections for Non-Biological Intelligences or: How We Learn to Stop Worrying and Love Our Robot Brethren*, in 19 *Minn. J.L. Sci. & Tech.*, 2018, p. 305 ss.;

DRIGO, *L' "Evocativa" integrazione analogica dell'art. 2052 c.c. per i sistemi emergenti di intelligenza artificiale: un'analisi critica*, in *GiustiziaCivile.com*, 2020;

DU BOIS-PEDAIN, *Novus actus and beyond: attributing causal responsibility in the criminal courts*, in *The Cambridge Law Journal*, 2021, Vol. 80, suppl. S1, p. S61 ss.;

EASTERBROOK, *Cyberspace and the Law of the Horse*, in *University of Chicago Legal Forum*, 1996, p. 207 ss.

EBERS, *Civil liability for autonomous vehicles in Germany*, 2022, p. 30, in SSRN, disponibile a questo link: <https://ssrn.com/abstract=4027594>;

EBERS, *Liability For Artificial Intelligence And EU Consumer Law*, in *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 2021, n. 3, p. 204 ss.;

EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, in Ebers – Navas (eds.), *Algorithms and Law*, Cambridge University Press, 2020, p. 37 ss.;

EBERS, *Standardizing AI. The Case of the European Commission's Proposal for an 'Artificial Intelligence Act'*, in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, cit., p. 340 ss.;

EDMONDS, *Uccideresti l'uomo grasso? Il dilemma etico del male minore*, Raffaello Cortina Editore, 2014;

EGAN, *Constructing a European Market: Standards, Regulation, and Governance*, Oxford University Press, 2001;

ELISH – BOYD, *Situating Methods in the Magic of Big Data and Artificial Intelligence*, in *Communication Monographs*, vol. 85, iss. 1, 2018, p. 11.

ELISH – HWANG, *Praise the machine! Punish the human! The Contradictory History of Accountability in Automated Aviation*, in *Comparative Studies in Intelligent Systems*, Working Paper #1 V2, 18 maggio 2015;

ELISH, *Moral crumple zones: Cautionary Tales in Human-Robot Interaction*, in *Engaging Science, Technology, and Society* 5 (2019), p. 40 ss.;

ELLIOTT, *Why Courts? Comment on Robinson*, in 14 *Journal of Legal Studies*, 1985, p. 799 ss.;

ELLUL, *La tecnica. Rischio del secolo*, Giuffr , 1969 (ed. franc. 1954);

ELSON, *The Case for an In-Depth Study of the American Law Institute*, in *Law & Social Inquiry*, 1998, vol. 23, no. 3, p. 625 ss.;

EUSEBI, (voce) *Sistema sanzionatorio e reati colposi*, in M. Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, cit., p. 1200 ss.;

EYKHOLT E AL., *Robust Physical-World Attacks on Deep Learning Visual Classification*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*;

FAIRGRIEVE (ed.), *Product Liability in Comparative Perspective*, Cambridge University Press, 2005;

FARBER, *Coping with Uncertainty: Cost-Benefit Analysis, the Precautionary Principle, and Climate Change*, in *90 Washington Law Review*, 2015, p. 1659 ss.;

FAURE, *Product Liability and Product Safety in Europe: Harmonization or Differentiation?*, in *Kyklos*, 2000, vol. 53, p. 467 ss.;

FERGUSON, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, NYU Press, 2017;

FERRANTE DI RUFFANO E AL., *Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults*, in *Cochrane Database of Systematic Reviews*, 12, 2018;

FIANDACA – MUSCO, *Diritto penale. Parte generale*, Zanichelli, VIII ed., 2019;

FIANDACA, (voce) *Causalità*, in *Dig. Disc. Pen.*, vol. II, 1988, Utet, p. 119 ss.;

FIANDACA, *Il reato commissivo mediante omissione*, Giuffrè, 1979;

FIANDACA, *Inquinamento elettromagnetico e rilevanza penale: questione approfondita ma non risolta*, in *Foro it.*, 2009, n. 2, p. 262 ss.;

FIERENS E AL., *Setting the Scene: On AI Ethics and Regulation*, in De Bruyne – Vanleenhove (eds.), *Artificial Intelligence and the Law*, cit., p. 49 ss.;

FILICETTI, *Sulla definizione di algoritmo (nota a Consiglio di Stato, Sezione Terza, 25 novembre 2021, n. 7891)*, in *Giust. ins.*, 8 febbraio 2023;

FINOCCHIARO G., *Intelligenza artificiale e responsabilità*, in *Contr. impr.*, 2020, n. 2, p. 713 ss.;

FINOCCHIARO G., *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Dir. inf.*, 2022, n. 2, p. 303 ss.;

FINOCCHIARO G., *Riflessioni su intelligenza artificiale e protezione dei dati personali*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, Giuffrè, 2020, p. 237 ss.;

IORELLA, *Il trasferimento di funzioni nel diritto penale dell'impresa*, Nardini, 1985;

IORELLA, *Responsabilità penale* (voce), in *Enc. Dir.*, vol XXXIX, 1988, p. 1289 ss.;

IORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 651 ss.;

IORENTINI, *Machine learning e dispositivi medici: riflessioni in materia di responsabilità civile*, in *Corr. giur.*, 2021, 10, p. 1258 ss.;

FLORES E AL. *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks"*, in *Federal Probation Journal*, vol. 80, n. 2, 2016;

FLORIDI, *Digital's Cleaving Power and Its Consequences*, in *Philosophy & Technology*, 2017, vol. 30, p. 123 ss.

FLORIDI, *The European Legislation on AI: a Brief Analysis of its Philosophical Approach*, in *Philosophy & Technology*, 2021, n. 34, p. 215 ss.;

FOÀ, *I danni da vaccino Sars-Cov-2 tra obblighi, raccomandazioni e «solidarietà irrinunciabile»*, in *Resp. civ. prev.*, 2022, fasc. 4, p. 1070 ss.

FOGEL, *Review of "Computational intelligence: Imitating life"*, in *Proc. of the IEEE*, vol. 83, iss. 11, 1995, p. 1588 ss.;

FORCHIELLI, *Intorno alla responsabilità senza colpa*, in *Riv. trim. dir. proc. civ.*, 1967, p. 1394;

FORNASARI, *Il principio di inesigibilità nel diritto penale*, Cedam, 1990;

FORTI, *"Accesso alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione"*, in *Criminalia*, 2006, p. 155 ss.;

FORTI, *Colpa ed evento nel diritto penale*, Giuffrè, 1990;

FORTI, *La “chiara luce della verità” e l’“ignoranza del pericolo”*. *Riflessioni penalistiche sul principio di precauzione*, in M. Bertolino – G. Forti (a cura di) *Scritti per Federico Stella*, vol. I, Jovene, 2007, p. 573 ss.;

FOSCH-VILLARONGA, *Robots, Healthcare, and the Law. Regulating Automation in Personal Care*, Routledge, 2020;

FRANKS – WALDMAN, *Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions*, in 78 *Maryland Law Review*, 2019, p. 892;

FRANZONI, *Dei fatti illeciti: art. 2043-2059*, Zanichelli, 2020

FRANZONI, *L’illecito*, Giuffrè, 2004;

FREEMAN, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, in *North Carolina Journal of Law & Technology*, vol. 18, 2016, p. 75 ss.;

FREITAS – ANDRADE – NOVAIS, *Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible*, in P. Casanovas – U. Pagallo – M. Palmirani – G. Sartor (eds.) *AI Approaches to the Complexity of Legal Systems*, Springer, 2014, p. 145;

FROOMKIN E AL., *When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning*, in 61 *Ariz. L. Rev.* 33, 2019, p. 33 ss.;

FROSINI, *Il discreto fascino delle forme di governo*, in *DPCE Online*, vol 49, n. 4, apr. 2021;

FUSARO, *I prodotti difettosi e pericolosi: le responsabilità*, cit., in Alpa (a cura di), *La responsabilità del produttore*, Giuffrè, 2019, p. 361 ss.

FUSARO, *Quale modello di responsabilità per la robotica avanzata? Riflessioni a margine del percorso*, in *Nuova giur. civ. comm.*, 2020, n. 6, p. 1344;

FUZAYLOVA, *War torts, autonomous weapons systems, and liability: why a limited strict liability tort regime should be implemented*, in *Cardozo Law Review*, 40, n. 3, 2019;

GALGANO, *Dogmi e dogmatica del diritto*, Cedam, 2010;

GALGANO, *Responsabilità del produttore*, in *Contr. impr.*, 1986, p. 996 ss.;

Galliot – MacIntosh – Ohlin (eds.), *Lethal Autonomous Weapons. Re-Examining the Law and Ethics of Robotic Warfare*, Oxford University Press, 2021;

GALLO E AL., *The new EU AI Act|what do financial services firms need to know?*, in *Deloitte Insights*, disponibile a questo link: <https://ukfinancialservicesinsights.deloitte.com/post/102gxhz/the-new-eu-ai-act-what-do-financial-services-firms-need-to-know>;

GALLO, (voce) *Aberratio delicti, causae*, in *Enc. Dir.*, I, 1958;

GALLO, (voce) *Colpa penale*, in *Enc. dir.*, VII, 1960;

GARAPON – LASSÈGUE, *Giustizia digitale*, Il Mulino, 2021;

GARGANI, *La “flessibilizzazione” giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *Leg. pen.*, 2011, n. 2, p. 397 ss.;

GARGANI, *Posizioni di garanzia nelle organizzazioni complesse: problemi e prospettive*, in *Riv. trim. dir. pen. econ.*, 2017, n. 3-4, p. 508 ss.;

GEISTFELD, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, in 105 *Calif. L. Rev.*, 2017, p. 1611 ss.;

GEISTFELD, *Reconciling Cost-Benefit Analysis with the Principle That Safety Matters More Than Money*, in 76 *New York University Law Review*, April 2001, p. 114 ss.;

GEISTFELD, *The Doctrinal Unity of Alternative Liability and Market-Share Liability*, in 155 *University of Pennsylvania Law Review*, 2006, p. 447 ss.;

GEISTFELD, *The Regulatory Sweet Spot for Autonomous Vehicles*, in 53 *Wake Forest L. Rev.* 2018, p. 337 ss.;

GELLERS, *Rights for Robots. Artificial Intelligence, Animal and Environmental Law*, Routledge, 2021;

GERKE E AL., *Ethical and legal challenges of artificial intelligence-driven healthcare*, in A. Bohr – K. Memarzadeh (eds), *Artificial intelligence in healthcare*, Academic Press, 2020, p. 295 ss.;

GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29 maggio 2019;

GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale: prove d'impatto a livello europeo*, in *Discrimen*, 2022;

GIANNINI, *Lombroso 2.0: On AI and Predictions of Dangerousness in Criminal Justice*, in 92 *International Review of Penal Law*, 2021, p. 179 ss.;

GIFFORD, *Technological Triggers to Tort Revolutions: Steam Locomotives, Autonomous Vehicles, and Accident Compensation*, 2018, 11 *Journal of Tort Law*, p. 71 ss.;

GIFFORD, *The Challenge to the Individual Causation Requirement in Mass Products Torts*, in 62 *Wash. & Lee L. Rev.*, 2005, p. 873 ss.;

GILLESPIE, *The relevance of algorithms*, in Gillespie – Boczkowski – Foot (eds), *Media Technologies: Essays on Communication, Materiality, and Society*, MIT Press;

GIUCA, *Disciplinare l'intelligenza artificiale. La riforma francese sulla responsabilità penale da uso di auto a guida autonoma*, in *Arch pen.*, 2022, n. 2;

GIUFFRIDA E AL., *A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law*, in 68 *Case W. Rsrv. L. Rev.*, 2018, p. 747 ss.;

GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, in *Criminalia*, 2006, p. 227 ss.;

GIUNTA, *Illiceità e colpevolezza nella responsabilità colposa*, Cedam, 1993;

GIUNTA, *La normatività della colpa penale. Lineamenti di una teorica*, in *Riv. it. dir. proc. pen.*, 1999, n. 1, p. 86 ss.;

GIUNTA, *Questioni scientifiche e prova scientifica tra categorie sostanziali e regole di giudizio*, in *Criminalia*, 2014, p. 561 ss.;

GLESS – SILVERMAN – WEIGEND, *If robots cause harm, who is to blame? Self-driving cars and criminal liability*, in *New Criminal Law Review*, 2016, p. 412 ss.;

GLESS, *AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials*, in 51 *Georgetown Journal of International Law*, 2020, p. 195 ss.;

GÖHLER, *Digital challenges to personal autonomy for German criminal law. Do we need a “new” criminal law for autonomous robots?*, in *WWU Münster*, disponibile a questo link: <https://miami.uni-muenster.de/Record/ff3df8a9-0daf-4dd8-a015-af91cb499f1a>;

GRANDI, *Neuroscienze e responsabilità penale: nuove soluzioni per problemi antichi?*, Giappichelli, 2017;

- GRASSO, *Il reato omissivo improprio. La struttura della fattispecie*, Giuffrè, 1983;
- GRIFFIN, *Artificial Intelligence and Liability in Health Care*, in 31 *Health Matrix: Journal of Law-Medicine*, 2021, p. 65 ss.;
- GROSSO, *Il giudizio di prevedibilità dell'evento e l'incidenza dell'hindsight bias e dell'outcome bias sul giudizio di responsabilità colposa*, in *Riv. it. med. leg.*, 2016, fasc. 2, p. 583 ss.;
- GROTTO, *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica*, Giappichelli, 2012;
- GRUT, *The Challenge of Autonomous Lethal Robotics to International Humanitarian Law*, in *Journal of Conflict and Security Law*, vol. 18, iss. 1, 2013, p. 5 ss.;
- GUERINI, *Fake news e diritto penale*, Giappichelli, 2020;
- GUERRA E AL., *Liability for Robots I: Legal Challenges*, in *Journal of Institutional Economics*, 2021, p. 6,
- GUERRA, *Evolving artificial intelligence and robotics in medicine, evolving European law. Comparative remarks based on the surgery litigation*, in 28 *Maastricht Journal of European and Comparative Law*, 2021, p. 805 ss.;
- GUERRA, *Il concetto di difettosità nella realtà che cambia. Un esercizio di microcomparazione*, in *Comparazione e diritto civile*, 2019, fasc. 1, p. 270;
- GUIDOTTI E AL., *A Survey of Methods for Explaining Black Box Models*, in 51 *ACM Computing Surveys*, 2018, p. 1 ss.
- GUNKEL, *Robot Rights*, MIT Press, 2018;
- GÜNSBERG, *Automated vehicles – is a dilution of human responsibility the answer?*, in *New Journal of European Criminal Law*, 2022, Vol. 13(4), p. 439 ss.;
- GURNEY, *Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles*, in *Robot Ethics 2.0*, cit., p. 51 ss.;
- HALLEVY, *"I, Robot – I, Criminal"—When Science Fiction Becomes Reality: Legal Liability of AI Robots committing Criminal Offenses*, in 22 *Syracuse Sci. & Tech. L. Rep.* 1, 2010, p. 1 ss.;
- HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015;

HALLEVY, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in 4 *Akron Intellectual Property Journal*, 2010, p. 171 ss.;

HALLEVY., *Dangerous Robots – Artificial Intelligence vs. Human Intelligence*, 11 febbraio 2018, in SSRN, disponibile a questo link: <https://ssrn.com/abstract=3121905>;

HARRIS, *Deepfakes: False Pornography is Here and the Law Cannot Protect You*, in 17 *Duke Law & Technology Review*, 2019, p. 99 ss.;

HEIDEGGER, *La questione della tecnica*, in Id., *Saggi e discorsi* (a cura di G. Vattimo), Mursia, 1976 (ed. ted. 1957);

HERZOG, *Società del rischio, diritto penale del rischio, regolazione del rischio*, in Stortoni – Foffani (a cura di), *Critica e giustificazione del diritto penale nel cambio di secolo. L'analisi critica della Scuola di Francoforte*, Giuffrè, 2004, p. 357 ss.;

HEVERLY, *Are Existing Tort Theories Ready for AI? An American Perspective*, in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, 2022, p. 89 ss.;

HILDEBRANDT, *Ambient Intelligence, Criminal Liability and Democracy*, in 2 *Crim. Law and Philos.*, 2008, p. 163 ss.;

HILDEBRANDT, *Criminal Liability and 'Smart' Environments*, in R.A. Duff (ed.), *Philosophical Foundations of Criminal Law*, Oxford University Press, 2011, p. 507 ss.;

HILDEBRANDT, *Law for Computer Scientists and Other Folk*, Oxford University Press, 2020;

HILDEBRANDT, *Saved by Design? The Case of Legal Protection by Design*, in 11 *NanoEthics*, 2017, p. 307 ss.;

HILDEBRANDT, *Smart Technologies and the End(s) of law*, Edward Elgar Publishing, 2015;

HILGENDORF, *Automated Driving and the Law*, in E. Hilgendorf – U. Seidel (eds.), *Robotics, Autonomics, and the Law*, cit., p. 171 ss.;

HODGES, *Alan Turing*, in *Stanford Encyclopedia of Philosophy*, 2002, rev. 2013, par. 7;

HOHLOCH, *Prospettive di evoluzione della responsabilità del produttore nel diritto tedesco*, S. Patti (a cura di), *Il danno da prodotti in Italia, Austria, Repubblica Federale di Germania, Svizzera*, cit., p. 268 ss.

HORKHEIMER, *Eclisse della ragione*, Einaudi, 1969;

HORST, *A Native Intelligence Metric for Artificial Systems*, *Performance Metrics for Intelligent Systems, Workshop*, in *Proceedings of the Performance Metrics for Intelligent Systems (PerMIS) Workshop*, 2002, disponibile a questo link: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=824478;

HOWARD – WOOLEY, *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*, Oxford University Press, 2018;

HOWELLS, *The Relationship Between Product Liability and Product Safety - Understanding a Necessary Element in European Product Liability Through a Comparison with the U.S. Position*, in *Wasburn L. J.*, 2000, vol. 29, p. 305 ss.;

HU, *Robot Criminals*, in *52 U. Mich. J. L. Reform*, 2019, p. 487 ss.;

HUBBARD, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation*, in *66 Fla. L. Rev.*, 2015, p. 1803 ss.;

HUQ, *Racial Equality in Algorithmic Criminal Justice*, in *Duke L.J.*, 2019, p. 1043 ss.;

HYLTON, *Preemption and Products Liability: A Positive Theory*, in *16 Supreme Court Economic Review*, 2008, p. 205 ss.;

IAGNEMMA, *I 'robot medici': profili problematici in tema di alleanza terapeutica e di responsabilità penale*, in *Corti supreme e salute*, 2020, n. 2, p. 442 ss.;

ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1, il Mulino, 2022, p. 43 ss.

JACQUEMIN – HUBIN, *Aspects contractuels et de responsabilité civile en matière d'intelligence artificielle*, in Jacquemin – De Streel (eds.), *Intelligence artificielle et droit*, Larcier, 2017, p. 73 ss.;

KABER – ENDSLEY, *The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task*, in *5 Theoretical Issues in Ergonomics Science*, 2004, p. 113 ss.;

KAPLAN, *Artificial Intelligence: What Everyone Needs to Know*, Oxford University Press, 2016;

KARNER – KOCH – GEISTFELD, *Comparative Law Study on Civil Liability for Artificial Intelligence* – Study requested by the Directorate-General for Justice and Consumers (EU Commission), novembre 2020;

KARNER, *Liability for robotics: current rules, challenges and the need*, in Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 117 ss.;

KARNOW, *Liability for Distributed Artificial Intelligence*, in 11 *Berkeley Technol. Law J.*, 1996, p. 147 ss.;

KARNOW, *The application of traditional tort law to embodied machine intelligence*, in R. Calo e al. (eds.), *Robot Law*, Edward Elgar Publishing, 2016, p. 51 ss.;

KEHL E AL., *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Berkman Klein Center for Internet & Society, Harvard Law School, 2017;

KELLEHER – TIERNAY, *Data Science*, MIT Press, 9th ed., 2018;

KING E AL., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in 26 *Science and Engineering Ethics*, 2020, p. 89 ss.;

KNETSCH, *Are Existing Tort Theories Ready for AI? A Continental European Perspective*, in Di Matteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, cit., p. 99 ss.;

KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability – Adapting Liability Rules to the Digital Age and Artificial Intelligence*, in *Journal of European Tort Law*, 2021, vol. 13, iss. 1;

KOCH, *Product Liability 2.0 - Mere Update or New Version?*, in Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 99 ss.;

KORSMO, *High-Frequency Trading: A Regulatory Strategy*, in 48 *U. Rich. L. Rev.*, 2019, p. 523 ss.

KRAFFT E AL., *Defining AI in Policy versus Practice*, in AIES '20: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, p. 72 ss.;

KROLL E AL., *Accountable Algorithms*, in 165 *U. Pa. L. Rev.*, 2017, p. 633 ss.;

KURKI, *A Theory of Legal Personhood*, Oxford University Press, 2019;

KURZWEIL, *The Singularity Is Near: When Humans Transcend Biology*, Viking Penguin, 2005;

LA TORRE, *Un punto fermo sul problema dei “danni punitivi”*, in *Danno resp.*, 2017, n. 4, p. 419;

LA VATTIATA, *Brevi note “a caldo” sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Dir. pen. uomo*, 30 giugno 2021;

LA VATTIATA, *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, in Giordano e al. (a cura di), *Il diritto nell’era digitale. Persona, mercato, amministrazione, giustizia*, cit., p. 695 ss.;

LAGIOIA – CONTISSA, *The Strange Case of Dr. Watson: Liability Implications of AI Evidence-Based Decision Support Systems in Health Care*, in *European Journal of Legal Studies*, vol. 12, n. 2, 2020, p. 245 ss.;

LAGIOIA – SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in *Philosophy & Technology*, 2020, 33, p. 433 ss.;

LAUMONIER, 6|5 – *La rivolta delle macchine*, Nero Editions, 2018;

LAVORGNA – SUFFIA, *La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell’ambito della giustizia penale*, in *Dir. pen. cont. – Riv. Trim.*, 2021, n. 2, p. 88 ss.;

LEE, *Deep learning artificial intelligence and the law of causation: application, challenges and solutions*, in *Information and Communications Technology Law*, 2021, vol. 30, iss. 3, p. 255 ss.;

LEHOULLIER E AL., *The Use of Event Data Recorders in the Analysis of Unintended Acceleration Incidents*, testo della relazione tenuta alla 23e Conférence canadienne multidisciplinaire sur la sécurité routière Montréal, Québec, 26-29 mai 2013, disponibile a questo link: https://www.roadsafetyresearch.ca/papers/edr_unintended_acceleration.pdf;

LEHR – OHM, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, in 51 *University of California Davis Law Review*, 2017, p. 653 ss.;

LELE, *Disruptive Technologies for the Militaries and Security*, Springer, 2019;

LEMLEY – CASEY, *Remedies for Robots*, in 86 *University of Chicago Law Review*, 2019, p. 1311 ss.;

LEONCINI, *Obbligo di attivarsi, obbligo di garanzia e obbligo di sorveglianza*, Giappichelli, 1999;

LEPORE, I.A. *e responsabilità civile. Robot, autoveicoli e obblighi di protezione*, in *Tecnologie e diritto*, 2021, p. 1980 ss.;

LESSIG, *Code is Law. On liberty in Cyberspace*, Harvard Magazine, 2000;

LEVY, *No Need to Reinvent the Wheel: Why Existing Liability Law Does Not Need to Be Preemptively Altered to Cope with the Debut of the Driverless Cars*, in *J. Bus Entrepreneurship & L.*, 2016, 9, p. 355;

LEWIS, *Flash Boys: A Wall Street Revolt*, W W Norton & Co Inc., 2015;

LI – FAURE – HAVU, *Liability Rules for AI-Related Harm: Law and Economics Lessons for a European Approach*, in *European Journal of Risk Regulation*, 2022;

LIMA, *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law*, in *69 South Carolina Law Review*, 2018, p. 680;

LIN, *Artificial Intelligence, Finance, and the Law*, in *88 Fordham L. Rev.*, 2019, p. 531 ss.;

LIN, *The New Market Manipulation*, in *66 Emory Law Journal*, 2017, p. 1253 ss.;

LIPTON, *The Mythos of Model Interpretability*, in *16 Queue*, May-June 2018;

LIU E AL., *Beyond State v Loomis: artificial intelligence, government algorithmization and accountability*, in *International Journal of Law and Information Technology*, vol. 27, iss. 2, 2019, p. 122 ss.;

LLORCA E AL., *Liability regimes in the age of AI: a use-case driven analysis of the burden of proof*, 3 novembre 2022, submitted to *Journal of Artificial Intelligence Research*, disponibile in Arxiv a questo link: <https://arxiv.org/abs/2211.01817>;

LO SAPIO, *Intelligenza artificiale: rischi, modelli regolatori, metafore*, in *Federalismi*, 2022, n. 27, p. 232 ss.;

LOHMANN, *Liability Issues Concerning Self-Driving Vehicles*, in *7 European Journal of Risk Regulation*, 2016, n. 2, p. 335 ss.;

LONDON, *Artificial intelligence and black-box medical decisions: accuracy versus explainability*, in *The Hastings Center Report*, 49(1), 2019, p. 15 ss.;

LU E AL., *The Expressive Power of Neural Networks: A View from the Width*, arXiv preprint, 2017, disponibile a questo link: <http://arxiv.org/abs/1709.02540>;

LUMINOSO, *Certificazione di qualità di prodotti e tutela del consumatore-acquirente*, in *Eur. dir. priv.*, 1999, p. 43 ss.;

Machnikowski (ed.), *European Product Liability. An Analysis of the State of the Art in the Era of New Technologies*, Intersentia, 2016;

MAGNUSON, *Artificial Financial Intelligence*, in 10 *Harvard Business Law Review*, 2020, p. 337 ss.;

MAGRO, *Biorobotica, robotica e diritto penale*, in Provolo e al. (a cura di), *Genetics, Robotics, Law, Punishment*, Padova University Press, 2014, p. 499 ss.;

MAGRO, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in *Leg. pen.*, 10 maggio 2020;

MAGRO, *Robot, cyborg e intelligenze artificiali*, in Cadoppi – Canestrari – Manna – Papa (a cura di), *Cybercrime*, Utet, 2019, p. 1179 ss.;

MAHLER, *Regulating Artificial General Intelligence (AGI)*, in Custers – Fosch-Villaronga (eds.), *Law and Artificial Intelligence*, Asser Press – Springer, p. 521 ss.

MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit.

MANTOVANI F., (voce) *Colpa*, in *Dig. pen.*, II, Utet, 1988;

MANTOVANI F., *Causalità, obbligo di garanzia e dolo nei reati omissivi*, in *Riv. it. dir. proc. pen.*, fasc. 4, 2004, p. 984 ss.;

MANTOVANI F., *Diritto penale. Parte generale*, Cedam, XI ed., 2020;

MANTOVANI F., *Il principio di affidamento nel diritto penale*, in *Riv. it. dir. proc. pen.*, 2009, p. 536

MANTOVANI M., (voce) *Affidamento (principio di)*, in Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, 2022, p. 1 ss.;

MANTOVANI M., *Contributo ad uno studio sul disvalore di azione nel sistema penale vigente*, Bononia University Press, 2014;

MARCHISIO, *In support of “no-fault” civil liability rules for artificial intelligence*, in *SN Social Sciences*, 2021, 1: 54, p. 10 ss.;

MARINO, *Market Share Liability and Economic Efficiency*, in *57 Southern Economic Journal*, 1991, p. 667 ss.

MARINUCCI – DOLCINI – GATTA, *Manuale di diritto penale. Parte generale*, Giuffrè, IX ed., 2022;

MARINUCCI – DOLCINI, *Corso di diritto penale*, Giuffrè, III ed., 2001;

MARINUCCI, *Causalità reale e causalità ipotetica nell'omissione impropria*, in *Riv. it. dir. proc. pen.*, 2009, 523 ss.;

MARINUCCI, *Il reato come ‘azione’. Critica di un dogma*, Giuffrè, 1971;

MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza*, in *Riv. it. dir. proc. pen.*, 2005, fasc. 1, p. 29 ss.;

MARINUCCI, *La colpa per inosservanza di leggi*, Giuffrè, 1965;

MARINUCCI, *Soggettivismo e oggettivismo nel diritto penale. Uno schizzo dogmatico e politico-criminale*, in *Riv. it. dir. proc. pen.*, 2011, p. 1 ss.

MARRA – MCNEIL, *Understanding “The Loop”: Regulating the Next Generation of War Machines*, in *36 Harvard Journal of Law and Public Policy*, 2013, p. 1139 ss.;

MARTINES, *I triloghi: analisi di una prassi decisionale alla luce dei principi di democrazia, equilibrio istituzionale, efficienza e trasparenza*, in *Dir. pubbl. comp. eur.*, 2018, fasc. 2, p. 311 ss.;

MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole?*, in *Dir. pen. cont.*, 2011;

MASSI, *Affidamento sull'intelligenza artificiale e “disimpegno morale” nella definizione dei presupposti della responsabilità penale*, in Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, cit., p. 665 ss.

MASTOP, *Characterising responsibility in organisational structures: The problem of many hands*, in G. Governatore – G. Sartor (eds.), *Deontic logic in computer science*, Springer, 2010, p. 274 ss.;

MATTHIAS, *The responsibility gap: Ascribing responsibility for the actions of learning automata*, in 6 *Ethics and Information Technology*, 2004, p. 175 ss.;

MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence based practices e tutela dei diritti fondamentali*, in *Arch. pen.* (web), 2021, n. 1;

MAYER-SCHONBERGER – CUKIER, *Big data: A revolution that will transform how we live, work, and think*, Houghton Mifflin Harcourt, 2013;

MAZZACUVA, *Il disvalore di evento nell'illecito penale*, Giuffrè, 1983;

MAZZACUVA, *The impact of AI on corporate criminal liability: algorithmic misconduct in the prism of derivative and holistic theories*, in 92 *International Review of Penal Law*, 2021, p. 143 ss.;

MCCARTHY – MINSKY – ROCHESTER – SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 31 agosto 1955, disponibile a questo link: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>;

MCCULLOCH – PITTS, *A logical calculus of the ideas immanent in nervous activity*, in 5 *Bulletin of Mathematical Biophysics*, 1943, p. 115 ss.;

MCGARITY, *The Preemption War: When Federal Bureaucracies Trump Local Juries*, Yale University Press, 2008;

MCGRAW, *Avoiding blame: an experimental investigation of political excuses and justifications*, in 20 *British Journal of Political Science*, 1990, p. 119 ss.;

MCSTAY – ROSNER, *Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies*, in *Big Data & Society*, 2021, p. 1 ss.;

MEYER, *John McCarthy*, in *Communications of the ACM*, 28 ottobre 2011;

MILITELLO, *Rischio e responsabilità penale*, Milano, Giuffrè, 1988;

MILLAR – KERR, *Delegation, relinquishment, and responsibility: The prospect of expert robots*, in Calo e al. (eds.), *Robot Law*, cit., p. 102; ss.;

MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, in *Dir. pen. cont. – Riv. trim.*, 2022, n. 2, p. 50 ss.;

MOBILIO, *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *Biolaw Journal*, 2020, n. 2, p. 401 ss.;

MÖKANDER E AL., *Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation*, in *Minds and Machines*, 2022, vol. 32, p. 245 ss.;

MONATERI, *La responsabilità civile*, in Sacco (diretto da), *Tratt. dir. civ.*, Utet, 1998, p. 1038;

MONATERI, *Le Sezioni Unite e le funzioni della responsabilità civile*, in *Danno resp.*, 2017, n. 4, p.437;

MONGILLO, (voce) *Industria alimentare e responsabilità colposa*, in Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, 2022, p. 697 ss.;

MONGILLO, *Imputazione oggettiva e colpa tra "essere" e normativismo: il disastro di Viareggio*, in *Giur. it.*, aprile 2022, p. 953 ss.;

MOORE, *Act and Crime: The Philosophy of Action and its Implications for Criminal Law*, Oxford University Press, 1993

MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, in *92 International Review of Penal Law*, 2021, p. 109 ss.;

Morse – Roskies (eds), *A Primer on Criminal Law and Neuroscience*, Oxford University Press, 2013;

MULDER – VELLINGA, *Exploring data protection challenges of automated driving*, in *Computer Law & Security Review*, Vol. 40, April 2021;

MÜLLER – BOSTROM, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*, in V.C. Müller (ed.), *Fundamental Issues of Artificial Intelligence*, Springer, 2016, p. 555 ss.;

MULLIGAN, *Revenge Against Robots*, in *South Carolina Law Review*, 2018, vol. 69, p. 579;

MURPHY, *Machine Learning. A Probabilistic Perspective*, MIT Press, 2021;

NAPPI, *La prevedibilità nel diritto penale*, Edizioni Scientifiche Italiane, 2020;

NARECHANIA, *Machine Learning as Natural Monopoly*, 107 *Iowa Law Review*, 2022, p. 1543 ss.;

NEWELL – SIMON, *Computer Science as Empirical Enquiry: Symbols and Search*, in *Communications of the ACM*, 19, 1976, p. 113 ss.;

NEWELL – SIMON, *GPS, A Program that Simulates Human Thought*, in E.A. Feigenbaum – J. Feldman (eds.), *Computers and Thought*, 1963;

NGUYEN E AL., *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 427 ss.;

NINATTI – TANCA, *La natura dei triloghi all'esame del giudice dell'Ue: efficacia "versus" trasparenza del processo legislativo nel caso De Capitani*, in *Quaderni cost.*, 2018, fasc. 2, p. 512 ss.;

NOBLE, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, 2018;

NYHOLM, *Attributing Agency to Automated Systems: Reflections on Human-Robot Collaborations and Responsibility-Loci*, in *Science Engineering Ethics*, 2018, 24, p. 1201 ss.;

O'NEIL, *Armi di distruzione matematica*, Bompiani, 2016;

OHLIN, *Did Russian Cyber Interference in the 2016 Election Violate International Law?*, in *95 Texas Law Review*, 2017, p. 1576;

OHLIN, *Election Interference. International Law and the Future of Democracy*, Cambridge University Press, 2020;

ONORATI, I.A., *politica e reati contro la personalità dello Stato*, in *Intelligenza artificiale e giurisprudizione penale*, Atti del workshop organizzato dalla Fondazione Occorsio il 19 novembre 2021, in *Sist. pen.*, 13 giugno 2022, p. 108 ss.;

OSBORNE, *The European Commission's Artificial Intelligence Act highlights the need for an effective AI assurance ecosystem*, in *Centre for Data Ethics and Innovation Blog*, disponibile a questo link: <https://cdei.blog.gov.uk/2021/05/11/the-european-commissions-artificial-intelligence-act-highlights-the-need-for-an-effective-ai-assurance-ecosystem/>;

OWEN, *A Punitive Damages Overview: Functions, Problems and Reform*, in 39 Vill. L. Rev., 1994, p. 363 ss.;

OWEN, *Rethinking the Policies of Strict Products Liability*, in 33 Vand. L. Rev., 1980, p. 681 ss.;

PADOVANI, *Diritto penale*, Giuffrè, XII ed., 2019;

PAGALLO, *AI and bad robots. The criminology of automation*, in M. R. McGuire – T. Holt (eds.), *The Routledge Handbook of Technology, Crime and Justice*, Routledge, 2017, p. 643 ss.;

PAGALLO, *Saggio sui robot e il diritto penale*, in Vinciguerra – Dassano (a cura di), *Scritti in memoria di Giuliano Marini*, Edizioni Scientifiche Italiane, 2010, p. 595 ss.,

PAGALLO, *The Adventures of Picciotto Roboto: AI & Ethics in Criminal Law*, in Aa. Vv., *The Social Impact of Social Computing. Proceedings of the Twelfth International Conference ETHICOMP 2011*, Sheffield Hallam University, 2011, p. 349 ss.;

PAGALLO, *The Laws of Robots*, Springer, 2015;

PAGALLO, *Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots*, in *Information*, 2018, n. 9, p. 230 ss.;

PALAZZO, *Corso di diritto penale. Parte generale*, Giappichelli, VI ed., 2016;

PALAZZO, *Morti da amianto e colpa penale*, in *Dir. pen. proc.*, 2011, n. 2, p. 185 ss.;

PALIERO, *L'autunno del patriarca. Rinnovo o trasmutazione del diritto penale dei codici*, in *Riv. it. dir. proc. pen.*, 1994, p.1220 ss.;

PALIERO, *La causalità dell'omissione: formule concettuali e paradigmi prasseologici*, in *Riv. it. med. leg.*, 1992, p. 821 ss.;

PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, in G. De Francesco – G. Morgante (a cura di), *Il diritto penale di fronte alle sfide della «società del rischio»*, Giappichelli, 2017, p. 286 ss.;

PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, in *Oss. dir. civ. comm.*, 2020, fasc. 2, p. 445 ss.;

PANATTONI, *AI and Criminal Law: The Myth of 'Control' in a Data-Driven Society*, in 92 *International Review of Penal Law*, 2021, p. 125 ss.;

PANATTONI, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, in *Dir. inf.*, 2021, fasc. 1, p. 317;

PAOLUCCI, *Algoritmi e intelligenza artificiale alla ricerca di una definizione: l'esegesi del Consiglio di Stato, alla luce dell'AI Act*, in *Quest. giust.*, 8 aprile 2022;

PARACAMPO, *Il percorso evolutivo ed espansivo delle regulatory sandboxes da FinTech ai nuovi lidi operativi del prossimo futuro*, in *Federalismi*, 2022, n. 18, p. 207 ss.;

PARASURAMAN E AL., *A Model for Types and Levels of Human Interaction with Automation*, in *30 IEEE Transactions on Systems Man and Cybernetics*, p. 286 ss.;

PASQUALE, *The Black Box Society: The Secret Algorithms that Control Money and Information*, Harvard University Press, 2015;

PATTI, *Il declino della responsabilità oggettiva (a margine dell'art. 2051 c.c.)*, in *Riv. dir. civ.*, 2019, n. 4, p. 977 ss.;

PATTI, *The European Road to Autonomous Vehicles*, in *Fordham International Law Journal*, 2019, vol. 43, p. 125 ss.;

PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen. (web)*, 2022, n. 2;

PEDRAZZI, *Odierna esigenze economiche e nuove fattispecie penali*, in *Riv. it. dir. proc. pen.*, 1975, p. 1099 ss., oggi raccolto in ID., *Diritto penale. Scritti di diritto penale dell'economia*, III, Giuffrè, 2003, p. 47 ss.;

PELLEGATTA, *Automazione nel settore automotive: profili di responsabilità civile*, in *Contr. impr.*, 2019, n. 4, p. 1418 ss.;

PENCO, *Limiti-soglia e responsabilità colposa. Il ruolo incerto delle soglie quantitative, dalla colpa specifica al rischio consentito*, in *Riv. it. dir. proc. pen.*, 2019, fasc. 1, p. 195 ss.;

PENZA, *La responsabilità per danno da prodotti difettosi*, in Demarchi (a cura di), *I diritti del consumatore e la nuova class action*, Zanichelli, 2010, p. 218 ss.;

PERES, *Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook*, in *IEEE Access*, vol. 8, pp. 220121 ss.;

PERIN, *La crisi del “modello nomologico” fra spiegazione e prevedibilità dell'evento nel diritto penale. Note introduttive e questioni preliminari sul fatto tipico colposo*, in *Riv. it. dir. proc. pen.*, 2014, fasc. 3, p. 1371 ss.

PERIN, *Prudenza, dovere di conoscenza e colpa penale*, Editoriale Scientifica, 2020;

PERIN, *Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale*, in *BioLaw Journal*, 2019, n. 1, p. 207 ss.;

PERINI C., *Il concetto di rischio nel diritto penale moderno*, Giuffrè, 2010;

Peris Riera – Massaro (a cura di), *Derecho penal, inteligencia artificial y neurociencias. Diritto penale, intelligenza artificiale e neuroscienze*, Roma Tre Press, 2023;

PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi*, Giappichelli, 2022;

PERROW, *Normal accidents*, Princeton University Press, 1984;

PICOTTI, «Nuovi» crimini cibernetici e possibile rilevanza dell'intelligenza artificiale. Gli Atti Digitali degli Stati Generali del Diritto di Internet, in *Diritto di Internet*, 2022, suppl. fasc. n. 1;

PICOTTI, *Cybercrimes and Criminal Relevance of Artificial Intelligence*, in F.J. Castro Toledo (coord.), *La transformación algorítmica del sistema de justicia penal*, Thomson Reuters, 2022, p. 67 ss.;

PICOTTI, *I primi vent'anni della Convenzione di Budapest nell'ottica sostanzialista e la mancata ratifica ed esecuzione del Primo Protocollo addizionale contro il razzismo e la xenofobia*, in *Dir. pen. proc.*, 2022, n. 8, p. 1028 ss.;

PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, in Catenacci e al. (a cura di) *Studi in onore di Antonio Fiorella*, vol 1, Roma Tre Press, 2021, p. 813 ss.;

PIERGALLINI, “Civile” e “penale” a perenne confronto: l'appuntamento di inizio millennio, in *Riv. it. dir. proc. pen.*, 2012, n. 4, p. 1299 ss.;

PIERGALLINI, (voce) *Colpa (diritto penale)*, in *Enc. dir.*, Annali, X, 2017, 222 ss.;

PIERGALLINI, *Attività produttive e imputazione per colpa: prove tecniche di “diritto penale del rischio”*, in *Riv. it. dir. proc. pen.*, 1997, fasc. 4, p. 1473 ss.;

PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, Giuffrè, 2004;

PIERGALLINI, *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, in *Riv. it. dir. proc. pen.*, 2005, p.1684 ss.;

PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, 2020, n. 4, p. 1745 ss.;

PIERGALLINI, *Ius sibi imponere: controllo penale mediante autonormazione?*, in Paliero e al. (a cura di), *La crisi della legalità. Il «sistema vivente» delle fonti legali*, Edizioni Scientifiche Italiane, 2016, p. 117 ss.;

PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, in *Riv. it. dir. proc. pen.*, 1996, p. 354 ss.;

PIERGALLINI, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, in *Dir. pen. proc.*, 2007, n. 9, p. 1125 ss.;

PINGEN, *EP Resolution on AI in Criminal Law and Policing*, in *Eucrim*, 17 novembre 2021;

PIRAINO, *Il nesso di causalità*, in *Eur. dir. priv.*, 2018, n. 2, p. 399 ss.;

PISANI TEDESCO, *Il nuovo quadro normativo europeo dei dispositivi medici*, in *Dir. comm. internaz.*, 2022, fasc. 3, p. 675 ss.;

PISANI TEDESCO, *Smart mobility e rischi satellitari e informatici: i possibili scenari di allocazione della responsabilità civile*, in *Dir. comm. internaz.*, 2019, fasc. 4, p. 801 ss.;

PIVA, *Machina discere, (deinde) delinquere et puniri potest*, in Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 681 ss.;

PIZZETTI, *La protezione dei dati personali e la sfida dell'Intelligenza Artificiale*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018, p. 5 ss.;

POLI, (voce) *Colpa grave*, in Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*, Giuffrè, 2022, p. 103 ss.;

POLI, *La colpa grave. I gradi della colpa tra esigenze di extrema ratio ed effettività della tutela penale*, Giuffrè, 2021;

PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto “a distanza” tra disvalore d’azione e disvalore d’evento*, in *Riv. trim. dir. pen. econ.*, 2010, p. 225 ss.

PONZANELLI – PARDOLESI, *La responsabilità per danno da prodotti difettosi*, in *Nuove leggi civ.*, 1989, p. 501 ss.;

PONZANELLI, *Causa probabile del danno e responsabilità del produttore*, in *Danno e resp.*, 2004, p. 527 ss.;

PONZANELLI, *La responsabilità civile: profili di diritto comparato*, Il Mulino, 1992;

PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, in Bessone (a cura di), *Casi e questioni di diritto privato*, Giuffrè, IV ed., 2002, p. 510 ss.;

PONZANELLI, *Polifunzionalità tra diritto internazionale privato e diritto privato; Le Sezioni Unite e le funzioni della responsabilità civile*, in *Danno resp.*, 2017, n. 4, p. 435;

POOLE, *Computational Intelligence: A logical approach*, Oxford University Press, 1998;

PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, in R. Giordano e al. (a cura di), *Il diritto nell’era digitale. Persona, mercato, amministrazione, giustizia*, cit., p. 713 ss.

PREZIOSI, *Responsabilità da reato degli enti e intelligenza artificiale*, in *Rivista 231 – La resp. amm. delle soc. e degli enti*, 2020, n. 4, p. 173 ss.;

PRICE II, *Medical Malpractice and Black-Box Medicine*, in I.G. Cohen e al. (eds.), *Big Data, Health Law, and Bioethics*, Cambridge University Press, 2018, p. 295 ss.;

PRICE II, *Regulating Black-Box Medicine*, 116 *Mich. L. Rev.*, 2017, p. 421 ss.;

PRIEST, *La controrivoluzione nel diritto della responsabilità da prodotti negli Stati Uniti d’America*, in *Foro it.*, 1989, IV, p. 119 ss.;

PRIVITERA, *Tecnica, individuo e società: cinque lezioni sulla teoria di Ulrich Beck*, Rubbettino, 2004;

PROCIDA MIRABELLI DI LAURO, *Le intelligenze artificiali tra responsabilità civile e sicurezza sociale*, in P. Perlingieri e al. (a cura di), *Rapporti civilistici e intelligenze artificiali: attività e responsabilità*, Edizioni Scientifiche Italiane, p. 297 ss.;

PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, in *Dir. pen. e proc.*, 2008, p. 647 ss.;

PULITANÒ, *Diritto penale*, Giappichelli, VII ed., 2017;

PUORRO, *High Frequency Trading: una panoramica*, in *Questioni di Economia e Finanza (Occasional Papers)* – Banca d'Italia, n. 198, 2013;

QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion*, Springer, 2020;

QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta Etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *Leg. pen.*, 18 dicembre 2018;

QUEZADA-TAVÁREZ, *Legal challenges in bringing AI evidence to the criminal courtroom*, in *New Journal of European Criminal Law*, Vol. 12(4), p. 531 ss.;

RAGHU E AL., *On the Expressive Power of Deep Neural Networks*, in *Proceedings of the 34th International Conference on Machine Learning*, 2017, n. 2847–54;

RAHWAN E AL., *Machine Behaviour*, in *Nature*, 2019, n. 568, p. 477 ss.;

RAJNERI, *L'ambigua nozione di prodotto difettoso al vaglio della corte di cassazione italiana e delle altre corti europee*, in *Riv. dir. civ.*, 2008, n. 5, p. 624 ss.;

RAJNERI, *The projects of European regulation on liability for damage caused by artificial intelligence. Striking a balance between the interests of consumers, multinationals and SMEs*, in *Riv. dir. banc.*, 2022, fasc. 2, p. 95 ss.;

RANCHORDAS, *Experimental Regulations for AI: Sandboxes for Morals and Mores*, University of Groningen Faculty of Law Research Paper No. 7/2021;

RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente alla luce dell'attuale scenario normativo*, in *Contr. impr.*, 2020, n. 3, p. 1175 ss.;

REED E AL., *Responsibility, Autonomy and Accountability: legal liability for machine learning*, Queen Mary School of Law Legal Studies Research Paper No. 243/2016;

REID, *The Deepfake Dilemma: Reconciling Privacy and First Amendment Protections*, in *23 U. Pa. J. Const. L.*, 2021, p. 209 ss.;

RESTA, *Cosa c'è di 'europeo' nella Proposta di Regolamento UE sull'intelligenza artificiale?*, in *Dir. inf.*, 2022, fasc. 2, p. 323;

RESTUCCIA, *La tutela del consumatore*, in Inzitari (a cura di), *Valutazione del danno e rimedi risarcitori*, Giappichelli, 2016, p. 521 ss.;

REYNOLDS – ISHIKAWA, *Robotic Thugs*, in *2007 ETHICOMP Proceedings*, 2007, p. 487 ss.;

RIONDATO, *Robot: talune implicazioni di diritto penale*, in Moro – Sarra (a cura di), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Franco Angeli, 2017, p. 85 ss.;

RIONDATO, *Robotica e diritto penale (robots, ibridi, chimere e “animali tecnologici”*, in D. Provolo e al. (a cura di), *Genetics, Robotics, Law, Punishment*, Padova University Press, p. 599 ss.;

RISICATO, *Il concorso colposo tra vecchie e nuove incertezze*, in *Riv. dir. proc. pen.*, 1998, p. 132 ss.;

RISICATO, *La causalità psichica tra determinazione e partecipazione*, Giappichelli, 2007;

ROBINSON E AL., *The Role of Healthcare Robots for Older People at Home: A Review*, in *Int. J. Social Robotics*, 2014, 6, p. 575;

ROBINSON, *Probabilistic causation and Compensation for tortious risk*, in *14 Journal of Legal Studies*, 1985, p. 779 ss.;

ROBINSON, *Structure and Function in Criminal Law*, Oxford University Press, 1997;

RODOTÀ, *Il problema della responsabilità civile*, Giuffrè, 1964;

ROE, *Who's Driving That Car?: An Analysis of Regulatory and Potential Liability Frameworks for Driverless Cars*, in *60 B.C. L. Rev.*, 2019, p. 315 ss.;

ROMANO – D'ALESSANDRO, *Nesso causale ed esposizione ad amianto. Dall'incertezza scientifica a quella giudiziaria: per un auspicabile chiarimento delle Sezioni Unite*, in *Riv. it. dir. proc. pen.*, 2016, fasc. 3, p. 1129 ss.;

RONCO, *Aporie scientiste e certezza logico razionale: note in tema di nesso causale*, in *Dir. pen. proc.*, 2007, n. 2, p. 247 ss.;

RONCO, *Le interazioni psichiche nel diritto penale: in particolare sul concorso psichico*, in *Ind. pen.*, 2004, n. 3, p. 815 ss.;

RONCO, *Scritti patavini*, Giappichelli, I, 2017;

ROPPO, *La responsabilità civile di Pietro Trimarchi*, in *Jus Civile*, 2017, n. 6, p. 696 ss.;

ROPPO, *La responsabilità oggettiva*, in Sirena (a cura di), *La funzione deterrente della responsabilità civile alla luce delle riforme straniere e dei Principles of European Tort Law*, Giuffrè, 2011, p. 185 ss.;

ROSENBERG, *The causal connection in mass exposure cases: a «public law» vision of the tort system*, in *97 Harvard Law Review*, 1984, p. 851 ss.;

RUDIN, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, in *1 Nat Mach Intell.*, 2019, p. 206 ss.;

RUFFOLO – AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, in *Giur. it.*, luglio 2019, p. 1704 ss.;

RUFFOLO – AMIDEI, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design, strumenti di certificazione preventive e «autodisciplina» di settore*, in Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1, cit., p. 489 ss.;

RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 2, il Mulino, 2022, p. 229 ss.;

RUFFOLO, *Intelligenza artificiale e automotive: le responsabilità da veicoli self-driving e driverless*, in Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 153 ss.

RUFFOLO, *L'Intelligenza artificiale in sanità: dispositivi medici, responsabilità e «potenziamento»*, in *Giur. it.*, febr. 2021, p. 502 ss.;

RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product: per i fondamenti di un diritto dell'intelligenza artificiale self-learning*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 93 ss.;

RUGA RIVA, *Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*, in Dolcini – Paliero (a cura di) *Studi in onore di Giorgio Marinucci*, II, Giuffrè, 2006, p. 1743 ss.;

RUSSELL – NORVIG, *Artificial Intelligence: A Modern Approach*, Pearson College Div., 4th ed., 2020;

RUTA, *I.a. nei reati economici e finanziari*, in *Intelligenza artificiale e giurisdizione penale*, Atti del workshop organizzato dalla Fondazione Occorsio il 19 novembre 2021, in *Sist. pen.*, 13 giugno 2022, p. 58 ss.;

SALANITRO, *Intelligenza artificiale e responsabilità: la strategia della Commissione europea*, in *Riv. dir. civ.*, 2020, fasc. 6, p. 1246 ss.;

SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, 2021, n. 1, p. 100;

SAMOILI E AL., *AI WATCH. Defining Artificial Intelligence*, EUR 30117 EN – Publications Office of the European Union, 2020, disponibile a questo link: <https://publications.jrc.ec.europa.eu/repository/handle/JRC118163>;

SANTARPIA, *Un'indagine subiettiva sulla responsabilità da custodia*, in *Resp. civ. prev.*, 2018, fasc. 4, p. 1224C ss.;

SANTOSUOSSO – SARTOR, *La giustizia predittiva: una visione realistica*, in *Giur. it.*, 2022, n. 7, p. 1760 ss.;

SARTOR – OMICINI, *The autonomy of technological systems and responsibilities for their use*, in Bhuta e al. (eds.), *Autonomous Weapons Systems. Law, Ethics, Policy*, Cambridge University Press, 2016, p. 39 ss.;

SARTOR, *Gli agenti software: nuovi soggetti del ciberdiritto?*, in *Contr. Impr.*, 2002, p. 465 ss.;

SCAFFARDI, *La medicina alla prova dell'Intelligenza Artificiale*, in *DPCE Online*, vol. 51, n. 1, apr. 2022, p. 349 ss.;

SCHARRE, *Army of None: Autonomous weapons and the future of war*, W.W. Norton&Company, 2018;

SCHERER, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies*, in *Harvard Journal of Law & Technology*, vol. 29, no. 2, 2016, p. 353 ss.

SCHLUETER, *Punitive Damages*, VII ed., vol. I, LexisNexis, 2015;

SCHMITT – THURNER, *Out of the Loop: Autonomous Weapons Systems and the Law of Armed Conflict*, in *Harvard National Security Journal*, 2013, n. 4, p. 231 ss.;

- SCHMITT – WIDMAR, *The Law of Targeting*, Oxford University Press, 2012;
- SCHMITT, *Foreign Cyber Interference in Elections*, in 97 *Int'l L. Stud.*, 2021, p. 739 ss.;
- SCHROLL, *Splitting the bill: creating a national car insurance fund to pay for accidents in autonomous vehicles*, in 109 *Northwestern University L. Rev.*, 2015, p. 803 ss.;
- SCHUETT, *Risk management in the Artificial Intelligence Act*, arXiv preprint, 2022, disponibile a questo link: <https://aps.arxiv.org/pdf/2212.03109.pdf>;
- SCOPINO, *Algo Bots and the Law: Technology, Automation, and the Regulation of Futures and Other Derivatives*, Cambridge University Press, 2020;
- SCOPINO, *Preparing Financial Regulation for the Second Machine Age: The Need for Oversight of Digital Intermediaries in the Futures Markets*, in 2015 *Columbia Business Law Review*, 2015, p. 439 ss.;
- SEARLE, *Minds, brains, and programs*, in 3 *Behavioral and Brain Sciences*, 1980, p. 417 ss.;
- SELBST – BAROCAS, *The Intuitive Appeal of Explainable Machines*, in 87 *Fordham Law Review*, 2018, p. 1085 ss.;
- SELBST, *Negligence and AI's Human Users*, in 100 *Boston University Law Review*, 2020, p. 1315 ss.;
- SEMINARA, *Riflessioni sulla condotta istigatoria come forma di partecipazione al reato*, in *Riv. it. dir. proc. pen.*, 1983, p. 1121 ss.;
- SGUBBI, *Il diritto penale incerto e inefficace*, in *Riv. it. dir. proc. pen.*, 2001, p. 1193 ss.;
- SGUBBI, *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Il Mulino, 2019;
- SGUBBI, *Il reato come rischio sociale. Ricerche sulle scelte di allocazione dell'illegalità penale*, Il Mulino, 1990;
- SGUBBI, *Responsabilità penale per omissio impedimento dell'evento*, Cedam, 1975;
- SHARKEY, *Products Liability Preemption: An Institutional Approach*, in 76 *Geo. Wash. L. Rev.*, 2008, p. 449 ss.

SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc.*, 2020, p. 604 ss.;

SILVA SÁNCHEZ, *La expansión del derecho penal, Aspectos de la política criminal en las sociedades postindustriales*, Civitas, II ed., 2001;

SIMMLER – MARKWALDER, *Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in *Crim. Law Forum*, 2019, n. 30;

SINGH E AL., *Responsibility & Machine Learning: Part of a process*, in 2016, in SSRN, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860048;

SMORTO, (voce) *Certificazione di qualità e normazione tecnica*, in *Dig. disc. priv.*, Agg. II, Utet, 2003;

SOLIS, *The Law of Armed Conflicts*, Cambridge University Press, 2010;

SOLOW-NIEDERMAN, *Emerging Digital Technology and the “Law of the Horse”*, in *UCLA L. Rev.*, 19 febbraio 2019;

SOLUM, *Legal Personhood for Artificial Intelligences*, in *N.C. L. Rev.*, 1992, n. 70, p. 1231 ss.;

SOVRANO E AL., *Metrics, Explainability and the European AI Act Proposal*, in *J*, 2022, n. 5, p. 126 ss.;

SPARROW, *Killer robots*, in 24 *J. Appl. Philos.*, 2007, p. 62;

SPINDLER, *Different Approaches for Liability of Artificial Intelligence – Pros and Cons – the New Proposal of the EU Commission on Liability for Defective Products and AI Systems*, 13 febbraio 2023, in SSRN, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354468;

SPINDLER, *Market processes, standardisation and tort law*, in *European Law Journal*, 2002, vol. 4, iss. 3, p. 316 ss.;

SPINDLER, *User liability and strict liability in the Internet of Things and for Robots*, in Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 125 ss.;

STAPLETON, *Product Liability*, Cambridge University Press, 1994;

STEINZOR, *(Still) "Unsafe at Any Speed": Why Not Jail for Auto Executives?*, in 9 *Harvard Law & Policy Review*, 2015, p. 443 ss.;

STELLA, (voce) *Rapporto di causalità*, in *Enc. Giur. Treccani*, XXV, 1991;

STELLA, *Causalità omissiva, probabilità, giudizi controfattuali. L'attività medico-chirurgica*, in *Cass. pen.*, 2005, p. 1062 ss.;

STELLA, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Giuffrè, III ed., 2003;

STELLA, *La costruzione giuridica della scienza: sicurezza e salute nei luoghi di lavoro*, in *Riv. it. dir. proc. pen.*, 2003, p. 55 ss.;

STELLA, *La nozione penalmente rilevante di causa: la condizione necessaria*, in *Riv. it. dir. proc. pen.*, 1988, p. 1249 ss.;

STELLA, *Leggi scientifiche e spiegazione causale nel diritto penale*, Giuffrè, II ed., 2000;

STORTONI, *Angoscia tecnologica ed esorcismo penale*, in *Riv. it. dir. proc. pen.*, 2004, fasc. 1, p. 71 ss.;

STRAMPELLI, *L'informazione societaria a quindici anni dal T.U.F.: profili evolutivi e problemi*, in *Riv. soc.*, 2014, p. 996 ss.;

SUMMERER, (voce) *Evitabilità dell'evento e comportamento alternativo lecito*, in Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, 2022, p. 489 ss.;

SUNSTEIN, *Laws of Fear. Beyond the precautionary principle*, Cambridge University Press, 2005;

TADDEI ELMI – ROMANO, *Il robot tra ius condendum e ius conditum*, in *Informatica e dir.*, 2016, n. 1, p. 126 ss.;

TADDEI ELMI, *I diritti dell'intelligenza artificiale, tra soggettività e valore: fantadiritto o jus condendum?*, in Lombardi Vallauri (a cura di), *Il meritevole di tutela. Studi per una ricerca coordinata da L. Lombardi Vallauri*, Milano, 1990, p. 685 ss.;

TAMPIERI, *L'intelligenza artificiale: una nuova sfida anche per le automobili*, in *Contr. Impr.*, 2020, n. 2, p. 748;

TARUFFO, *La prova scientifica nel processo civile*, in M. Bertolino – G. Forti (a cura di), *Scritti per Federico Stella*, cit., vol. II, cit., p. 1375 ss.;

TAVERRITI, *Autonormazione e prospettive autopoietiche nella gestione della penalità*, in *Riv. it. dir. proc. pen.*, 2020, p. 1931 ss.;

TENCENT KEEN SECURITY LAB, *Experimental Security Research of Tesla Autopilot*, marzo 2019, disponibile a questo link: https://keenlab.tencent.com/en/whitepapers/Experimental_Security_Research_of_Tesla_Autopilot.pdf;

TEUBNER, *Rights of Non-Humans? Electronic Agents and Animals as New Actors*, in *Journal of Law and Society*, Vol. 33, 2006, p. 497 ss.;

TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, a cura di Pasquale Femia, Edizioni Scientifiche Italiane, 2019;

THOMPSON, *The moral responsibility of public officials: the problem of many hands*, in *74 American Political Science Review*, 1980, p. 905 ss.

TORRE, *Produttori di alimenti o produttori di norme? Gli standard di sicurezza tra fonti pubblicistiche e fonti privatistiche e il loro valore tipizzante nel fatto colposo*, in Foffani – Doval Pais – Castronuovo (a cura di), *La sicurezza agroalimentare nella prospettiva europea*, Giuffrè, 2014, p. 507 ss.;

TRIMARCHI, *Istituzioni di diritto privato*, Giuffrè, XIX ed., 2011;

TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, Giuffrè, III ed., 2021;

TRIMARCHI, *Rischio e responsabilità oggettiva*, Giuffrè, 1961;

TRINCHERA, *Questioni in tema di aberratio ictus*, in *Riv. it. dir. proc. pen.*, 2015, fasc. 3, p. 1412 ss.;

TRIPODI, *Abusi di mercato e trading algoritmico*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 745 ss.;

TRONTI, *La direttiva CEE relativa al riavvicinamento delle disposizioni legislative regolamentari ed amministrative degli stati membri in materia di responsabilità per danno da prodotti difettosi*, in *Giur. merito*, 1988, n. 4, p. 695 ss.;

TURING, *Computing Machinery and Intelligence*, in *Mind*, 49, 1950, p. 433 ss.;

TURNER, *Robot Rules. Regulating Artificial Intelligence*, Palgrave Macmillan Cham, 2019;

TWIGG-FLESNER, *Guiding Principles for Updating the Product Liability Directive for the Digital Age*, ELI Innovation Paper Series – European Law Institute, 2021;

UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Dir. pen. cont. – Riv. Trim.*, 2020, n. 4;

US NATIONAL SCIENCE AND TECHNOLOGY COUNCIL COMMITTEE, *Preparing for the future of Artificial Intelligence*, 12 ottobre 2016, p. 6.

VALLINI, “*Cause sopravvenute da sole sufficienti*” e nessi tra condotte. Per una collocazione dell’art.41, comma 2, c.p. nel quadro teorico della causalità “scientifica”, in *Studi in onore di Franco Coppi*, vol. I, Torino, Giappichelli, 2011, p. 341 ss.

VALLINI, *La causalità nel danno da prodotto alimentare e l’effetto “vasi comunicanti” tra illeciti di rischio, di pericolo, di danno*, in Foffani – Doval Pais – Castronuovo (a cura di), *La sicurezza agroalimentare nella prospettiva europea*, Giuffrè, 2014, p. 485 ss.

VALLOR – BEKEY, *Artificial Intelligence and the Ethics of Self-Learning Robots*, in Lin – Abney – Jenkins (eds.), *Robot Ethics 2.0*, Oxford University Press, 2017, p. 338 ss;

VAN GESTEL – MICKLITZ, *European Integration through Standardization: How judicial review is breaking down the club house of private standardization bodies*, in *50 Common Market Law Review*, 2013, p. 145 ss.;

VAN KOLFSCHOOTEN, *EU regulation of artificial intelligence: Challenges for patients’ rights*, in *59 Common Market Law Review*, 2022, p. 81 ss.;

VAN LEEUWEN, *European Standardisation of Services and Its Impact on Private Law: Paradoxes of Convergence*, Hart Publishing, 2017.

VAN MENSVOOR, *NextNature*, D Editore, 2022;

VEALE, *Demystifying the Draft EU Artificial Intelligence Act*, in *Computer Law Review International*, 2021, n. 4, p. 97 ss.;

VENEZIANI, *Regole cautelari “proprie” ed “improprie” nella prospettiva delle fattispecie colpose causalmente orientate*, Cedam, 2003;

VENTURA, *Radical Choc*, Einaudi, 2020;

VERARDI, *Art. 6 – Esclusione della responsabilità*, in Alpa e al. (a cura di), *La responsabilità per danno da prodotti difettosi*, Giuffrè, 1990, p. 79 ss.;

VIGANÒ, *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, in *Dir. pen. cont. – Riv. Trim.*, 2013, n. 3, p. 380 ss.;

VIGEN, *Spurious Correlations*, Hachette Books, 2015;

VILLANOVA, *La responsabilità per il danno cagionato da cose in custodia*, in *Resp. civ. prev.*, 2022, p. 760 ss.;

VISENTINI, *L'esimente del rischio di sviluppo come criterio di responsabilità del produttore. (l'esperienza italiana e tedesca e la direttiva comunitaria)*, in *Resp. civ. prev.*, 2004, p.1267;

VISINTINI, *Atipicità dei fatti illeciti e danno ingiusto*, in A. Somma e al. (a cura di), *Dialoghi con Guido Alpa. Un volume offerto in occasione del suo LXXI compleanno*, Roma Tre Press., 2018, p. 589 ss.

VISINTINI, *Causalità e danno*, in *Contr. impr.*, 2015, n. 2, p. 262 ss.;

VISINTINI, *Trattato breve della responsabilità civile. Fatti illeciti. Inadempimento. Danno Risarcibile*, Cedam, 2005, p. 748;

VLADECK, *Machines Without Principals: Liability Rules and Artificial Intelligence*, in 89 *Wash. L. Rev.*, 2014, p. 128;

WAGNER, *Robot liability*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, Baden-Baden, 2019, p. 30 ss.;

WAGNER, *Robot, Inc.: Personhood for Autonomous Systems?*, in 88 *Fordham Law Review*, 2019, p. 604;

WALLACH – ALLEN, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, 2009;

WANG E AL., *Deep Learning for Identifying Metastatic Breast Cancer*, arXiv preprint, 2016, disponibile a questo link: <https://arxiv.org/abs/1606.05718>;

WEBER, *La scienza come professione - La politica come professione*, Mondadori, 2006 (ed. ted. 1919);

WEBSTER, *Electronic Health Records: An Asset or a Whole Lot of Hype?*, in 182 *Can. Med. Ass'n J.*, 2010;

WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, in *Journal of European Tort Law*, 2011, vol. 11, iss. 2, p. 150 ss.;

WENDEL, *Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles*, in 56 *California Western Law Review* 129, 2018;

WENDEL, *Technological Solutions to Human Error and How They Can Kill You: Understanding the Boeing 737 Max Products Liability Litigation*, 84 *J. Air L. & Com.*, 2019, p. 379 ss.;

WERBROUCK, *The Hypothesis of Technological Unemployment Caused by AI-Driven Automation and its Impact on Social Security Law*, in De Bruyne – Vanleenhove (eds.), *Artificial Intelligence and the Law*, Intersentia, 2021, p. 317 ss.;

WHEATLEY, *Foreign Interference in Elections under the Non-intervention Principle: We Need to Talk about “Coercion”*, 31 *Duke Journal of Comparative & International Law*, 2020, p. 161 ss.;

WHITTAKER, *Liability for Products. English Law, French Law, and European Harmonisation*, Oxford University Press, 2005.

WHITTAM, *Mind the compensation gap: towards a new European regime addressing civil liability in the age of AI*, in *International Journal of Law and Information Technology*, 2022, XX, p. 1 ss.

WILLIAMS, *Accountable Algorithms: Adopting the Public Law Toolbox Outside the Realm of Public Law*, in 75 *Current Legal Problems*, 2022, p. 237 ss.;

WILLIAMS, *Lords select committee, artificial intelligence committee, written evidence (AIC0206)*, 8 settembre 2017, disponibile a questo link: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/70496.html%23_ftn13#_ftnref7;

WINKLE, *Development and Approval of Automated Vehicles: Considerations of Technical, Legal, and Economic Risks*, in M. Maurer e al. (eds.), *Autonomous Driving*, Springer, p. 589 ss.;

WINNER, *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*, MIT Press, 1977;

WISCHMEYER, *Artificial Intelligence and Transparency: Opening the Black Box*, in T. Wischmeyer – T. Rademacher (eds.), *Regulating Artificial Intelligence*, Cham, 2020, p. 75 ss.

WOLPERT, *Product Liability and Software Implicated in Personal Injury*, in 60 *Def. Couns. J.*, 1993, p. 521 ss.;

WUYTS, *The product liability directive – More than two decades of defective products in Europe*, in *Journal of European Tort Law*, 2014, n. 5, p. 1 ss.

XU, *Adversarial Attacks and Defenses in Images, Graphs and Text: A Review*, in *International Journal of Automation and Computing*, 17(2), April 2020, p. 151 ss.;

YADAV, *How Algorithmic Trading Undermines Efficiency in Capital Markets*, 68 *Vanderbilt Law Review*, 2015, p. 1607;

YADAV, *The Failure of Liability in Modern Markets*, in 102 *Virginia Law Review*, 2016, p. 1031 ss.;

YAMAOKA-ENKERLIN, *Disrupting Disinformation: Deepfakes and the Law*, in 22 *N.Y.U. J. Legis. & Pub. Pol'y*, 2020, p. 725 ss.;

YANG, *Medical robotics-regulatory, ethical and legal considerations for increasing levels of autonomy*, in *Science Robotics*, 2017, 1 ss.;

ZACCARIA, *La responsabilità del “produttore” di software*, in *Contr. impr.*, 1993, p. 302;

ZECH, *Liability for AI: public policy considerations*, in *ERA Forum*, 2021, n. 22, p. 147 ss.;

ZEI, *Tecnica e diritto tra pubblico e privato*, Giuffrè, 2008;

ZERILLI E AL., *Algorithmic decision-Making and the Control Problem*, in 29 *Minds and Machines*, 2019, p. 555 ss.;

ZIRULIA, *Contrasti reali e contrasti apparenti nella giurisprudenza post-Cozzini su causalità e amianto. Riflessioni per un rinnovato dibattito sul c.d. effetto acceleratore nei casi di morte per mesotelioma*, in *Riv. it. dir. proc. pen.*, 2019, fasc. 3, p. 1289 ss.;

ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, Giuffrè, 2018;

ZIRULIA, *Sequestro preventivo e sicurezza sul lavoro: illegittimo il decreto "salva-Ilva" n. 92 del 2015*, in *Riv. it. dir. e proc. pen.*, 2018, fasc. 2, p. 947 ss.;

FONTI ISTITUZIONALI

AGENZIA ITALIANA DEL FARMACO, *Tredicesimo rapporto sulla sorveglianza dei vaccini anti-COVID-19* (27/12/2020 – 26/09/2022), disponibile a questo link: https://www.aifa.gov.it/documents/20142/1315190/Rapporto_sorveglianza_vaccini_COVID-19_13.pdf;

BUREAU D'ENQUÊTES ET D'ANALYSES POUR LA SÉCURITÉ DE L'AVIATION CIVILE (BEA), *Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris*, luglio 2012, disponibile a questo link: <https://bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf>;

BUREAU EUROPÉEN DES UNIONS DE CONSOMMATEURS, *PRODUCT LIABILITY 2.0. How to make EU rules fit for consumers in the digital age*, 7 maggio 2020;

CALIFORNIA CODE OF REGULATIONS, tit. 13, div. 1, ch. 1, art. 3.7, accessibile a questo link: <https://www.law.cornell.edu/regulations/california/title-13/division-1/chapter-1/article-3.7>.

CALIFORNIA STATE ASSEMBLY, *Crimes: deceptive recordings*, AB 1280, Cal. 2019, disponibile a questo link: <https://openstates.org/ca/bills/20192020/AB1280/>;

CALIFORNIA STATE ASSEMBLY, *Elections: deceptive audio or visual media*, AB 730, Cal. 2019, disponibile a questo link: <https://openstates.org/ca/bills/20192020/AB730/>;

CAMERA DEI DEPUTATI, Proposta di legge A.C. 2986, *Introduzione dell'articolo 612-quater del codice penale, in materia di manipolazione artificiale di immagini di persone reali allo scopo di ottenerne rappresentazioni nude*, presentata il 30 marzo 2021, il cui testo è disponibile sul sito della Camera dei deputati, a questo link: <https://www.camera.it/leg18/126?idDocumento=2986>;

CEN-CENELEC, *CEN and CENELEC launched a new Joint TC on Artificial Intelligence*, 3 marzo 2021, disponibile a questo link: <https://www.cencenelec.eu/news-and-events/news/2021/briefnews/2021-03-03-new-joint-tc-on-artificial-intelligence/>;

COMITATO ECONOMICO E SOCIALE EUROPEO, *L'intelligenza artificiale – Le ricadute dell'intelligenza artificiale sul mercato unico (digitale), sulla produzione, sul consumo, sull'occupazione e sulla società*, INT/806, 2017;

COMMISSIONE EUROPEA – GRUPPO INDIPENDENTE DI ESPERTI AD ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, *Orientamenti etici per un'i.a. affidabile*, 8 aprile 2019;

COMMISSIONE EUROPEA, Comunicazione della Commissione al Parlamento europeo, *L'intelligenza artificiale per l'Europa*, COM(2018) 237 final, 25 aprile 2018;

COMMISSIONE EUROPEA, *La guida blu all'attuazione della normativa UE sui prodotti* 2022, 2022/C 247/01, 29 giugno 2022;

COMMISSIONE EUROPEA, *Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale)*, COM (2022) 496 final, 28 settembre 2022;

COMMISSIONE EUROPEA, *Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi*, COM (2022) 495 final, 28 settembre 2022;

COMMISSIONE EUROPEA, *Proposta di Regolamento del Parlamento europeo e del Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali) e che modifica la direttiva 2000/31/CE*, COM(2020) 825 final, 15 dicembre 2020;

COMMISSIONE EUROPEA, *Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale (Legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione*, COM/2021/206 final, 21 aprile 2021;

COMMISSIONE EUROPEA, *Proposta di Regolamento del Parlamento Europeo e del Consiglio sui prodotti macchina*, COM/2021/202 final, 21 aprile 2021;

COMMISSIONE EUROPEA, *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*, COM(2020) 64 final, 19 febbraio 2020;

COMMISSIONE EUROPEA, *The Strengthened Code of Practice on Disinformation* 2022, 16 giugno 2022;

COMMISSIONE GROSSO PER LA RIFORMA DEL CODICE PENALE, *Articolato approvato dalla Commissione Ministeriale per la Riforma del codice penale nella seduta del 26 maggio 2001*, disponibile sul sito del Ministero della Giustizia a questo link: https://www.giustizia.it/giustizia/it/mg_1_12_1.page?contentId=SPS31489&previousPage=mg_1_12_1;

COMMISSIONE PER L'EFFICIENZA DELLA GIUSTIZIA DEL CONSIGLIO D'EUROPA, *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment*, 4 dicembre 2018; una traduzione a cura del Ministero della Giustizia è disponibile a questo link: <https://rm.coe.int/carta-etica-europea-sull-utilizzo-dell-intelligenza-artificiale-nei-si/1680993348>;

CONSIGLIO D'EUROPA, *Algorithms and Human Rights*, DGI(2017)12), 2018;

CONSIGLIO DELL'UNIONE EUROPEA, *Decisione quadro 2002/584/GAI, del 13 giugno 2002, relativa al mandato d'arresto europeo e alle procedure di consegna tra Stati membri*;

CONSIGLIO SUPERIORE DI SANITÀ, *I sistemi di i.a. come strumento di supporto alla diagnostica*, 2021, disponibile a questo link: https://www.salute.gov.it/imgs/C_17_pubblicazioni_3218_allegato.pdf;

CONVENTION ON CERTAIN CONVENTIONAL WEAPONS (CCW), *Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects – Final Report*, 13 dicembre 2019, Annex III, disponibile a questo link: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/343/64/PDF/G1934364.pdf?OpenElement>;

EUROPEAN COMMISSION, *Evaluation of Council Directive 85/374/EEC on the approximation of laws, regulations and administrative provisions of the Member States concerning liability for defective products*, SWD(2018) 157 final, 7 maggio 2018, disponibile a questo link: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0157&from=EN>.

EUROPEAN COMMISSION, *Impact Assessment Report Accompanying the Document Proposal for a Directive of the European Parliament and of The Council on Liability for Defective Products*, SWD(2022) 316 final, 28 settembre 2022, p. 50, disponibile a questo link: https://single-market-economy.ec.europa.eu/document/348b3e35-7d1a-43df-8e9d-296fc09e2c3c_en.

EUROPEAN UNION AGENCY FOR CYBERSECURITY, *IoT and Smart Infrastructures*, disponibile a questo link: <https://www.enisa.europa.eu/topics/iot-and-smart-infrastructures/iot>;

EUROPOL, *Facing reality? Law enforcement and the challenge of deepfakes*, 22 aprile 2022, disponibile a questo link: https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf;

EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, 2019, p. 42-44;

GARANTE DELLA PRIVACY, *Deep fake: il Garante privacy apre un'istruttoria nei confronti di Telegram per il software che "spoglia" le donne*, 23 ottobre 2020; il comunicato stampa è disponibile a questo link: <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9470722>;

GARANTE DELLA PRIVACY, *Deepfake. Il falso che ti «ruba» la faccia (e la privacy)*, dicembre 2020, disponibile a questo link: <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9512226>;

GROUP OF GOVERNMENTAL EXPERTS OF THE HIGH CONTRACTING PARTIES TO THE CONVENTION ON PROHIBITIONS OR RESTRICTIONS ON THE USE OF CERTAIN CONVENTIONAL WEAPONS, *Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapons*, document presentato dal governo statunitense, 28 marzo 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24027:2021 – Bias in AI systems and AI aided decision making*, disponibile a questo link: <https://www.iso.org/standard/77607.html>;

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24368:2022 – Information technology — Artificial intelligence — Overview of ethical and societal concerns*, disponibile a questo link: <https://www.iso.org/standard/78507.html>;

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24028:2020 – Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence*, disponibile a questo link: <https://www.iso.org/standard/77608.html>;

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – INTERNATIONAL ELECTROTECHNICAL COMMISSION, *ISO/IEC TR 24029-1:2021 – Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview*, disponibile a questo link: <https://www.iso.org/standard/77609.html>;

NATIONAL CONFERENCE OF STATE LEGISLATURES, *Autonomous Vehicles State Bill Tracking Database*, disponibile a questo link:

<https://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx>

NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION (NHTSA), *Automated Vehicles for Safety*, disponibile a questo link: <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety>;

NATIONAL TRANSPORTATION SAFETY BOARD (NTSB), *Accident report – Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018*, 19 novembre 2019, disponibile a questo link: <https://www.nts.gov/investigations/accidentreports/reports/har1903.pdf>;

NAZIONI UNITE, Convenzione su certe armi convenzionali, Ginevra, 10 ottobre 1980;

OFFICIAL JOURNAL OF THE EUROPEAN COMMUNITIES, C 114, vol. 32, 8 maggio 1989, disponibile a questo link: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:1989:114:FULL&from=IT>;

PARLAMENTO EUROPEO, *Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts*, giugno 2022, disponibile a questo link: [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU\(2022\)729512_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf);

PARLAMENTO EUROPEO, *Risoluzione del 12 settembre 2018 sui sistemi d'arma autonomi (2018/2752(RSP))*;

PARLAMENTO EUROPEO, *Risoluzione del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL))*, 2018/C 252/25.

PARLAMENTO EUROPEO, *Risoluzione del 20 gennaio 2021 sull'intelligenza artificiale: questioni relative all'interpretazione e applicazione del diritto internazionale nella misura in cui l'UE è interessata relativamente agli impieghi civili e militari e all'autorità dello Stato al di fuori dell'ambito della giustizia penale (2020/2013(INI))*;

PARLAMENTO EUROPEO, *Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, 2020/2014(INL)*;

PARLAMENTO EUROPEO, *Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale (2020/2016(INI))*;

SOCIETY OF AUTOMOTIVE ENGINEERS (SAE), *SAE J3016™ Recommended Practice: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, disponibile al seguente link: https://www.sae.org/standards/content/j3016_202104/;

TEXAS STATE ASSEMBLY, *Relating to the creation of a criminal offense for fabricating a deceptive video with intent to influence the outcome of an election*, Sen. Bill 751, 2019-2020, disponibile a questo link: <https://legiscan.com/TX/text/SB751/id/2027638>;

U.S. CONGRESS, *The SELF DRIVE Act (H.R. 3388_2017)*, accessibile a questo link: <https://www.congress.gov/bill/115th-congress/house-bill/3388/text>;

U.S. CONGRESS, *The SELF DRIVE ACT (H.R. 3711)*, 2021, accessibile a questo link: <https://www.congress.gov/bill/117th-congress/house-bill/3711?s=1&r=55>;

U.S. CONGRESSIONAL RESEARCH SERVICE, *Emerging Military Technologies: Background and Issues for Congress*, novembre 2020;

U.S. DEPARTMENT OF DEFENSE, *Department of Defense Directive (DODD) 3000.09 – Autonomy in Weapon Systems*, 21 novembre 2012, mod. 2017, disponibile a questo link: <https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf>;

U.S. DEPARTMENT OF JUSTICE, *Report on the Investigation into Russian Interference in the 2016 Presidential Election* (c.d. Mueller report), marzo 2019, disponibile a questo link: <https://www.justice.gov/archives/sco/file/1373816/download>;

U.S. GOVERNMENT, *Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapons*, 28 marzo 2018

UNIONE EUROPEA, *Direttiva 2001/95/ce del Parlamento Europeo e del Consiglio del 3 dicembre 2001 relativa alla sicurezza generale dei prodotti*;

UNIONE EUROPEA, *Direttiva 2014/65/UE del Parlamento Europeo e del Consiglio del 15 maggio 2014 relativa ai mercati degli strumenti finanziari e che modifica la direttiva 2002/92/CE e la direttiva 2011/61/UE (c.d. MIFID II)*;

UNIONE EUROPEA, *Direttiva 85/374/CEE del Consiglio del 25 luglio 1985 relativa al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi*;

UNIONE EUROPEA, *Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al*

trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati);

UNIONE EUROPEA, *Regolamento (UE) 2019/2144 del Parlamento europeo e del Consiglio del 27 novembre 2019 relativo ai requisiti di omologazione dei veicoli a motore e dei loro rimorchi;*

UNITED NATIONS ECONOMIC AND SOCIAL COUNCIL (UNECE), *Report of the Global Forum for Road Traffic Safety on its eighty-first session. Addendum. Amendments to Article 1 and new Article 34bis 1968 Convention on Road Traffic*, ECE/TRANS/WP.1/173/Add.1, 14 dicembre 2020, disponibile a questo link: <https://unece.org/sites/default/files/2021-01/ECE-TRANS-WP.1-173-Add1e.pdf>;

UNITED NATIONS, *Convention on Road Traffic Vienna, 8 November 1968 – Acceptance of Amendment to Article 1 and New Article 34 Bis to the Convention*, 21 gennaio 2022, disponibile a questo link: <https://treaties.un.org/doc/Publication/CN/2022/CN.26.2022-Eng.pdf>;

UNITED NATIONS, *Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council* (UN S/2021/229), 8 marzo 2021, disponibile a questo link: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N21/037/72/PDF/N2103772.pdf?OpenElement>;

VIRGINIA STATE ASSEMBLY, *Unlawful dissemination or sale of images of another; penalty*, Virginia House Bill 2678, 8 marzo 2019, disponibile a questo link: <https://legiscan.com/VA/text/HB2678/2019>.

SITOGRAFIA: ARTICOLI, REPORT, VIDEO

Allen, *DOD Is Updating Its Decade-Old Autonomous Weapons Policy, but Confusion Remains Widespread*, in *Center for Strategic & International Studies*, 6 giugno 2022, disponibile a questo link: <https://www.csis.org/analysis/dod-updating-its-decade-old-autonomous-weapons-policy-confusion-remains-widespread>;

Allen, *Russia Probably Has Not Used AI-Enabled Weapons in Ukraine, but That Could Change*, in *Center for Strategic & International Studies*, 26 maggio 2022, disponibile a questo link: <https://www.csis.org/analysis/russia-probably-has-not-used-ai-enabled-weapons-ukraine-could-change>;

Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, in *Wired*, 23 giugno 2008, disponibile al seguente link: <https://www.wired.com/2008/06/pb-theory/>;

Angwin e al., *Machine Bias*, in *ProPublica*, 23 maggio 2016, disponibile a questo link: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>;

Bini, *"L'intelligenza artificiale sta diventando senziente", ingegnere Google lancia l'allarme. E Big G lo sospende*, in *La Repubblica*, 13 giugno 2022, disponibile a questo link: https://www.repubblica.it/economia/2022/06/13/news/google_intelligenza_artificiale_in_gegner_bot-353695551/;

Blakemore, *AI outdoes radiologists when it comes to identifying hip fractures, study shows*, in *Washington Post*, 20 febbraio 2022, disponibile a questo link: <https://www.washingtonpost.com/health/2022/02/20/hip-fractures-ai/>;

Botsman, *Big data meets Big Brother as China moves to rate its citizens*, in *Wired UK*, 21 ottobre 2017, disponibile a questo link: <https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>;

Boudette, *'It Happened So Fast': Inside a Fatal Tesla Autopilot Accident*, in *The New York Times*, 17 agosto 2021, disponibile al seguente link: <https://www.nytimes.com/2021/08/17/business/tesla-autopilot-accident.html>;

Cahlan, *How misinformation helped spark an attempted coup in Gabon*, in *The Washington Post* (online), 13 febbraio 2020, disponibile a questo link: <https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/>;

Chat GPT, disponibile a questo link: <https://openai.com/blog/chatgpt/>;

Clarke, *The EU's leaked AI regulation is ambitious but disappointingly vague*, in *Tech Monitor*, 15 aprile 2021, disponibile a questo link: <https://techmonitor.ai/policy/eu-ai-regulation-machine-learning-european-union>;

Ctrl Shift Face, video *Taxi Driver* starring Al Pacino [DeepFake]; disponibile a questo link: <https://www.youtube.com/watch?v=9NkKj0aNB0s&t=205s>;

Ctrl Shift Face, video *The Shining* starring Jim Carrey: Episode 2 - The Bat [DeepFake], disponibile a questo link: <https://www.youtube.com/watch?v=-ZRUZzZPGto&t=110s>;

Fallows, *Here's What Was on the Record About Problems with the 737 Max*, in *The Atlantic*, 13 marzo 2019, disponibile a questo link: <https://www.theatlantic.com/ideas/archive/2019/03/heres-what-was-on-the-record-about-problems-with-the-737-max/622213/>;

Fisk, *Toyota Settles Oklahoma Acceleration Case After Jury Verdict*, in *Bloomberg*, 25 ottobre 2013, disponibile a questo link: <https://www.bloomberg.com/news/articles/2013-10-25/toyota-settles-oklahoma-acceleration-case-after-jury-verdict>;

Future of Life Institute, *Autonomous Weapons: an Open Letter from Ai & Robotics Researchers*, 9 febbraio 2016, disponibile a questo link: <https://futureoflife.org/open-letter-autonomous-weapons>;

Future of Life Institute, *Killer robots: World's top AI and robotics companies urge United Nations to ban lethal autonomous weapons*, disponibile a questo link: <https://futureoflife.org/2017/08/20/killer-robots-worlds-top-ai-robotics-companies-urge-united-nations-ban-lethal-autonomous-weapons/>;

Giacconi, *Una sorta di magia. Intelligenze artificiali, origami, marionette, serpenti: scoprire l'incanto nei laboratori di robotica*, in *Il Tascabile*, 2 marzo 2023, disponibile a questo link: <https://www.iltascabile.com/scienze/robot/>,

Gorski, *IBM's Watson versus cancer: Hype meets reality*, in *Science-based Medicine*, disponibile a questo link: <https://sciencebasedmedicine.org/ibm-watson-versus-cancer-hype-meets-reality/>;

Griffin, *Facebook's Artificial Intelligence Robots Shut Down After They Start Talking to Each Other in Their Own Language*, in *Independent*, 31 luglio 2017, disponibile a questo link: <https://www.independent.co.uk/life-style/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html>;

Guglielmi, *The next-generation bots interfering with the US election*, intervista a Emilio Ferrara, 28 ottobre 2020, disponibile a questo link: <https://www.nature.com/articles/d41586-020-03034-5>;

Hao, *Training a single AI model can emit as much carbon as five cars in their lifetimes*, in *MIT Technology Review*, disponibile a questo link: <https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/>;

Hern, *Randomly generated tweet by bot prompts investigation by Dutch police*, in *The Guardian* (online), 12 febbraio 2015, disponibile a questo link:

<https://www.theguardian.com/technology/2015/feb/12/randomly-generated-tweet-by-bot-investigation-dutch-police>;

Holt, *Hackers reportedly deepfaked a Binance exec to carry out listing scams*, in *Engadget*, 23 agosto 2022, disponibile a questo link: <https://www.engadget.com/binance-executive-deepfake-listing-scam-154729114.html>;

Human Rights Watch, *Losing Humanity: The Case Against Killer Robots*, 2012, disponibile a questo link: https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf.

ID R&D, *Human or Machine: AI Proves Best at Spotting Biometric Attacks*, 2022, disponibile a questo link: <https://www.idrnd.ai/wp-content/uploads/2022/01/ResearchBrief-HumanVsMachine-JAN2422.pdf>;

International Committee of the Red Cross (ICRC), *Position on Autonomous Weapon Systems*, 2021, disponibile a questo link: <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>;

Jahromi, *Levels of Automation for Autonomous Ground Vehicles*, in *Medium*, disponibile a questo link: <https://medium.com/@BabakShah/levels-of-automation-for-self-driving-cars-d410a4f679b7>.

Kasperkevic, *Swiss police release robot that bought ecstasy online*, in *The Guardian* (online), 22 aprile 2015, disponibile al link: <https://www.theguardian.com/world/2015/apr/22/swiss-police-release-robot-random-darknet-shopper-ecstasy-deep-web>;

Kayser-Bril, *European Council and Commission in agreement to narrow the scope of the AI Act*, in *Algorithm Watch*, 24 novembre 2021, disponibile a questo link: <https://algorithmwatch.org/en/eu-narrow-scope-of-ai-act/>.

Knight, *Russia's Killer Drone in Ukraine Raises Fears About AI in Warfare*, in *Wired US*, 17 marzo 2022, disponibile a questo link: <https://www.wired.com/story/ai-drones-russia-ukraine/>;

Koebler, *In Europe, Robots Can Legally Buy Drugs Online for Art*, in *Vice US*, 15 aprile 2015, disponibile a questo link: <https://www.vice.com/en/article/mgbwg4/in-europe-robots-can-legally-buy-drugs-online-for-art>;

Langewiesche, *The Human Factor*, in *Vanity Fair*, ottobre 2014, disponibile a questo link: <https://www.vanityfair.com/news/business/2014/10/air-france-flight-447-crash>;

Lavalle, *Le persone credono veri più i volti deepfake di quelli reali*, in *La Repubblica online*, 5 marzo 2022, disponibile a questo link: https://www.repubblica.it/tecnologia/2022/03/05/news/le_persone_credono_veri_piu_i_volti_umani_deepfake_di_quelli_reali-338859552/;

Lekach, *Self-flying planes could transport passengers one day—but first, packages*, in *Mashable*, 24 luglio 2021, disponibile a questo link: <https://mashable.com/article/self-flying-airplanes>;

Longhitano, *L'intelligenza artificiale di Facebook parla una lingua incomprensibile, ma niente panico*, in *Wired Italia*, 1 agosto 2017, disponibile a questo link: <https://www.wired.it/gadget/computer/2017/08/01/intelligenza-artificiale-facebook/>;

Lotkowski, *You Are Building a Self Driving AI Without Even Knowing About It*, *Hackernoon*, 25 febbraio 2017, Feb. 25, 2017), disponibile a questo link: <https://hackernoon.com/you-are-building-a-self-driving-ai-without-even-knowing-about-it-62fadbfa5fdf>;

Maruzani, *Are You Unwittingly Helping to Train Google's AI Models?*, in *Towards Data Science*, 26 gennaio 2021, disponibile a questo link: <https://towardsdatascience.com/are-you-unwittingly-helping-to-train-googles-ai-models-f318dea53aee>;

Mattei, *Parliamo di reincantamento*, in *L'Indiscreto*, 8 luglio 2022, disponibile a questo link: <https://www.indiscreto.org/parliamo-di-reincantamento/>.

McKinsey & Company, *Ten ways autonomous driving could redefine the automotive world*, disponibile a questo link: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/ten-ways-autonomous-driving-could-redefine-the-automotive-world>;

Metz, *In Two Moves, AlphaGo and Lee Sedol Redefined the Future*, in *Wired US*, 16 marzo 2016, disponibile a questo link: <https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/>;

Mitchell, *California DMV accuses Tesla of falsely advertising Autopilot and Full Self-Driving features*, in *Los Angeles Times*, 5 agosto 2022, disponibile a questo link: <https://www.latimes.com/business/story/2022-08-05/dmv-false-advertising-tesla>;

Mitchell, *California DMV accuses Tesla of falsely advertising Autopilot and Full Self-Driving features*, in *Los Angeles Times*, 5 agosto 2022, disponibile a questo link: <https://www.latimes.com/business/story/2022-08-05/dmv-false-advertising-tesla>.

Negretti, *Intelligenza artificiale, quali sono le architetture hardware dedicate*, in *AI4Business*, 20 gennaio 2021, disponibile a questo link: <https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-quali-sono-le-architetture-hardware-dedicate/>;

Orpaz, *How Does the Iron Dome Work?*, in *Haaretz*, 19 novembre 2012, disponibile a questo link: <https://www.haaretz.com/2012-11-19/ty-article/.premium/how-iron-dome-works/0000017f-e70b-dc7e-adff-f7af61f70000>;

Pallini, *I flew on a self-flying plane where pilots sat back as the aircraft taxied, took off, and landed on its own and I'm convinced it's the future of aviation*, in *Business Insider*, 8 maggio 2021, disponibile a questo link: <https://www.businessinsider.com/flying-on-self-flying-plane-convinced-its-future-of-aviation-2021-5?r=US&IR=T>;

Perez, *Google Now Using ReCAPTCHA To Decode Street View Addresses*, in *TechCrunch*, 29 marzo 2012, disponibile a questo link: <https://techcrunch.com/2012/03/29/google-now-using-recaptcha-to-decode-street-view-addresses/>.

Porro, *Sul web è caccia alle versioni pirata dell'app per creare finte foto di donne nude*, in *Wired Italia*, 4 luglio 2019, disponibile a questo link: <https://www.wired.it/internet/web/2019/07/04/deepnude-revenge-porn/>;

Price, *Microsoft is deleting its AI chatbot's incredibly racist tweets*, in *Business Insider*, 24 marzo 2016, disponibile a questo link: <https://www.businessinsider.in/Microsoft-is-deleting-its-AI-chatbots-incredibly-racist-tweets/articleshow/51539858.cms>;

Ramey, *Tesla Autopilot Will Now Detect Emergency Lights, but Only at Night*, in *Autoweek*, 22 settembre 2021, disponibile a questo link: <https://www.autoweek.com/news/green-cars/a37694444/tesla-autopilot-will-now-detect-emergency-lights-but-only-at-night/>;

Ray, *A new experiment: Does AI really know cats or dogs -- or anything?*, in *Zdnet*, 28 gennaio 2022, disponibile a questo link: <https://www.zdnet.com/article/what-does-ai-know-of-cats-and-dogs-maybe-very-little/>;

Red. CBS Los Angeles, *Driver to stand trial in Tesla autopilot crash which killed two in Gardena*, 20 maggio 2022, disponibile a questo link: <https://www.cbsnews.com/losangeles/news/driver-to-stand-trial-in-tesla-autopilot-crash-which-killed-two-in-gardena/>;

Red. CBS, *Doctored Nancy Pelosi video highlights threat of "deepfake" tech*, in *CBS news*, 26 maggio 2019, disponibile a questo link:

<https://www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25/>;

Red. Il Post, *Sentirete parlare di video “deepfake”*, in *Il Post*, 24 novembre 2018, disponibile a questo link: <https://www.ilpost.it/2018/11/24/video-deepfake-informazione/>;

Red. Rai News, *Tesla lanciata a folle velocità: le telecamere a circuito chiuso riprendono l'incidente mortale*, 14 novembre 2022, disponibile a questo link: <https://www.rainews.it/video/2022/11/tesla-lanciata-a-folle-velocit-le-telecamere-a-circuito-chiuso-riprendono-lincidente-mortale-06a60958-096a-4005-9213-50c36fcb7641.html>;

Red. Reuters, *UPDATE 1-Germany clears Mercedes-Benz's hands-free drive system*, in *Reuters*, 9 dicembre 2021, disponibile a questo link: <https://www.reuters.com/article/daimler-autonomous-idCNL8N2SU35K>;

Saripalli, *Are Self-Driving Cars the Future of Mobility for Disabled People?*, in *The Conversation*, 5 ottobre 2017, disponibile a questo link: <https://theconversation.com/are-self-driving-cars-the-future-of-mobility-for-disabled-people-84037>;

Schwartz, *Google's Paul Haahr: We Don't Fully Understand RankBrain*, in *Search Engine Roundtable*, 8 marzo 2016, disponibile a questo link: <https://www.seroundtable.com/google-dont-understand-rankbrain-21744.html>.

Smiley, *I'm the Operator': The Aftermath of a Self-Driving Tragedy*, in *Wired US*, 8 marzo 2022, disponibile a questo link: <https://www.wired.com/story/uber-self-driving-car-fatal-crash/>;

Solomon, *World's first certified level 3 autonomous car to hit streets of Japan*, 5 marzo 2021, in *Autox*, disponibile a questo link: <https://www.autox.com/news/car-news/worlds-first-certified-level-3-autonomous-car-to-hit-streets-of-japan-109099/>;

Spector – Levine, *Exclusive: Tesla faces U.S. criminal probe over self-driving claims*, in *Reuters*, 27 ottobre 2022, disponibile a questo link: <https://www.reuters.com/legal/exclusive-tesla-faces-us-criminal-probe-over-self-driving-claims-sources-2022-10-26/>;

Steadman, *IBM's Watson is better at diagnosing cancer than human doctors*, in *Wired UK*, 11 febbraio 2013, disponibile a questo link: <https://www.wired.co.uk/article/ibm-watson-medical-doctor>;

Stupp, *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case*, in *The Wall Street Journal*, 30 agosto 2019, disponibile a questo link: <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>;

Taylor, *German Court Bans Tesla 'Autopilot' Name For Misleading Customers*, in *Forbes*, 14 luglio 2020, disponibile a questo link: <https://www.forbes.com/sites/michaeltaylor/2020/07/14/german-court-bans-tesla-autopilot-name-for-misleading-customers/?sh=7757b3f5c7fb>;

Tiku, *The Google engineer who thinks the company's AI has come to life*, in *The Washington Post*, 11 giugno 2022, disponibile a questo link: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>;

Trager –Luca, *Killer Robots Are Here—and We Need to Regulate Them*, in *Foreign Foreign Policy*, 11 maggio 2022, disponibile a questo link: <https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/>;

Ucla Health, *What Is Robotic Surgery?*, disponibile a questo link: <https://www.uclahealth.org/robotic-surgery/what-is-robotic-surgery>;

Valesini, *Nella guerra in Ucraina scendono in campo le armi autonome*, in *Wired Italia*, 24 marzo 2022, disponibile a questo link: <https://www.wired.it/article/ucraina-guerra-armi-autonome-droni-kamikaze-robot-killer-russia/>;

Wakefield, *Deepfake presidents used in Russia-Ukraine war*, in *BBC.com*, 18 marzo 2022, disponibile a questo link: <https://www.bbc.com/news/technology-60780142>.

RINGRAZIAMENTI

Desidero ringraziare innanzitutto il professor Gian Luigi Gatta, per avermi sempre guidata e sostenuta nel mio percorso di formazione e di ricerca, spronandomi ad esercitare il senso critico in ogni circostanza e a coniugare nitore espositivo e rigore metodologico.

Ringrazio inoltre il procuratore Giovanni Salvi, che ho conosciuto nelle vesti di Presidente del Comitato scientifico della Fondazione Occorsio, con cui ho avuto il privilegio di collaborare negli ultimi due anni. Devo al costante confronto con Lui molti preziosi spunti di riflessione sui rapporti tra intelligenza artificiale e sistema penale.

Durante la stesura della tesi ho avuto l'opportunità di svolgere un periodo di ricerca presso la Cornell Law School (Ithaca, NY): ringrazio il professor Stephen P. Garvey per i numerosi consigli che mi ha fornito sui temi oggetto della mia ricerca e non solo.

Un sincero e profondo ringraziamento va poi a tutte le persone che animano il Dipartimento di Scienze Giuridiche “Cesare Beccaria” dell'Università degli Studi di Milano, vera e propria comunità, non solo scientifica, ma anche di affetti. In particolare, questi anni di dottorato non sarebbero stati così intensi e importanti senza il supporto e l'amicizia – in una parola, la “sorellanza” – delle mie colleghe del XXXV ciclo, Carla, Cecilia e Francesca.

Un enorme grazie, infine, alle persone più care che ho: la mia famiglia, Simone, i miei amici.