



UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Giurisprudenza

Pubblicazioni del Dipartimento di Diritto pubblico italiano e sovranazionale

IA E DISCRIMINAZIONI DI GENERE PROSPETTIVE FUTURE PER UNA REGOLAMENTAZIONE INCLUSIVA

a cura di

MARILISA D'AMICO e NICCOLÒ PANIGADA



G. Giappichelli Editore



UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Giurisprudenza

Pubblicazioni del Dipartimento di Diritto pubblico italiano e sovranazionale

Studi di diritto pubblico

125

La Collana “Pubblicazioni del Dipartimento di Diritto pubblico italiano e sovranazionale” dell’Università degli Studi di Milano raccoglie monografie e altri risultati inediti di ricerche, individuali e collettive, di studiosi che svolgono attività di studio e ricerca nel Dipartimento.

Essa comprende Studi di Diritto costituzionale, di Diritto amministrativo, di Diritto internazionale ed europeo, di Diritto processuale civile, di Diritto comparato, di Storia del diritto, di Politica economica.

La qualità scientifica delle pubblicazioni è assicurata da una procedura di c.d. double blind peer review ad opera di revisori esterni.

IA E DISCRIMINAZIONI DI GENERE
PROSPETTIVE FUTURE
PER UNA REGOLAMENTAZIONE INCLUSIVA

a cura di

MARILISA D'AMICO e NICCOLÒ PANIGADA



G. Giappichelli Editore

© Copyright 2025 - G. GIAPPICHELLI EDITORE - TORINO

VIA PO, 21 - TEL. 011-81.53.111

<http://www.giappichelli.it>

ISBN/EAN 979-12-211-1825-4

ISBN/EAN 979-12-211-6583-8 (ebook-pdf)

ISBN/EAN 979-12-211-6721-4 (ebook-epub)

Il volume è edito con risorse del PRIN 2022 CUP G53D23002160006 - Finanziato dall'Unione europea - Next Generation EU Misura M4C2 – Investimento 1.1.



Stampa: Stampatre s.r.l. - Torino

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume/fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941, n. 633.

Le fotocopie effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi, Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali, Corso di Porta Romana 108, 20122 Milano, e-mail autorizzazioni@clearedi.org e sito web www.clearedi.org.

INDICE

	<i>pag.</i>
<i>Elenco degli Autori e delle Autrici</i>	XV
<i>Note introduttive sul PRIN – AiGeDi (Artificial Intelligence between generating and tackling gender-based discriminations; L'intelligenza artificiale da causa a strumento di contrasto delle discriminazioni di genere)</i> <i>Marilisa D'Amico</i>	XVII
RELAZIONE INTRODUTTIVA	
SFIDE PER COMBATTERE LE DISCRIMINAZIONI DI GENERE NELL'ERA DELL'IA	
<i>Francisco Balaguer Callejón</i>	1
1. Introduzione	1
2. L'uguaglianza di genere come ultima frontiera del diritto costituzionale	3
3. L'era dell'IA e l'algoritmo come nuovo strumento di trasformazione storica	5
4. La parità di genere nell'era dell'intelligenza artificiale	7
4.1. L'orientamento dei diritti verso un mercato discriminatorio	7
4.2. Parità di genere e riproduzione culturale della discriminazione nell'era digitale	10
5. Conclusioni	13

PARTE I

DIRITTI COSTITUZIONALI, IDENTITÀ
E *BIAS* ALGORITMICI: PROSPETTIVE
GIURIDICHE NELL'ERA DELL'IA

CAPITOLO I

LE NUOVE FRONTIERE DELLE DISCRIMINAZIONI DI
GENERE: L'IMPATTO DELL'INTELLIGENZA ARTIFICIALE

Marilisa D'Amico

17

1. Intelligenza artificiale e discriminazioni di genere	17
2. La discriminazione algoritmica e i suoi diversi fattori	19
3. Il linguaggio degli algoritmi e l'impatto discriminatorio	23
4. Lavoro femminile e IA	28
5. AI Act europeo e diritti delle donne	29
6. Considerazioni sul Trattato del Consiglio d'Europa	30
7. Una nuova sfida: l'IA americana e il "free speech" algoritmico	32
8. Postilla: la legge italiana e il riferimento specifico alla "parità di genere"	33

CAPITOLO II

CONCEPTUAL EVOLUTION AND NORMATIVE TENSION
OF PERSONAL IDENTITY BETWEEN AI TECHNOLOGIES
AND BEYOND

Edoardo Carlo Raffiotta

37

1. Background	37
2. The necessary protection of personal identity in the digital world	38
3. The necessary protection of personal identity in the era of AI	44
3.1. From the virtual to the metaverse, opportunities and criticalities of the new dimensions of personal development	47
4. Over-regulation in the EU and the need for fewer, more effective rules	52
5. Between AI and metaverse technology	55
6. A provocation. Conclusions	55

CAPITOLO III

L'INTELLIGENZA ARTIFICIALE DA CAUSA A
STRUMENTO DI CONTRASTO DELLE DISCRIMINAZIONI
DI GENERE E INTERSEZIONALI: CONSIDERAZIONI
A SEGUITO DI INDAGINE "DELPHI"

Elettra Stradella, Silvia Cervia, Gabriele Rugani 57

1. Introduzione: domanda di ricerca e metodologia 57
2. Le discriminazioni di genere e intersezionali derivanti dall'intelligenza artificiale 59
3. Le cause delle discriminazioni algoritmiche 61
4. Gli ambiti in cui si possono verificare gli effetti pregiudizievoli derivanti dall'IA 63
5. Il ruolo degli strumenti giuridici esistenti a livello UE nel prevenire e contrastare le discriminazioni derivanti dall'IA 65
6. Le strategie per addivenire a un'IA che non causi, e addirittura contrasti, discriminazioni di genere e intersezionali, con particolare riferimento al coinvolgimento degli *stakeholders* 69
7. Considerazioni conclusive 71

CAPITOLO IV

CERTIFICAZIONE DI PROVENIENZA DELLE RISPOSTE PER
L'USO PROFESSIONALE DEI *LARGE LANGUAGE MODEL*

Ernesto Damiani 73

1. Introduzione 73
2. Provenienza delle risposte LLM 75
 - 2.1. RAG con marcatori verificabili 76
3. Stato dell'arte tecnologico 79
4. Questioni aperte e direzioni future di ricerca 80
5. Conclusioni 81

CAPITOLO V

I DIRITTI DELLE PERSONE CON DISABILITÀ AL
COSPETTO DEI SISTEMI DI INTELLIGENZA ARTIFICIALE.
UN'ANALISI DELL'AI ACT E DELLA LEGGE N. 132/2025

Giuseppe Arconzo, Gaia Patarini 83

1. Disabilità e sistemi di intelligenza artificiale: opportunità e rischi 83

	<i>pag.</i>
2. Le persone con disabilità come soggetti vulnerabili da tutelare secondo l'AI Act	87
3. (<i>Segue</i>): le altre condizioni richieste per l'applicazione del sistema di tutela nell'AI Act	93
4. Il rilievo, nell'AI Act, dell'accessibilità dei sistemi di IA	95
5. Persone con disabilità e intelligenza artificiale nella legge italiana: sfide e potenzialità di uno strumento in divenire ... ma a costo zero	98
6. Conclusioni	102

CAPITOLO VI

MISOGINIA E *HATE SPEECH ONLINE* NEL DIRITTO DELL'UE

<i>Ilaria Anrò</i>	105

1. Introduzione: la lotta alla misoginia attraverso il contrasto all' <i>hate speech online</i> nel diritto dell'Unione europea	105
2. <i>Hate speech online</i> e strumenti di cooperazione giudiziaria penale	109
3. L'inserimento di <i>hate speech</i> e <i>hate crimes</i> tra i "reati europei"	111
4. La risoluzione del Parlamento europeo del 18 gennaio 2024 a favore dell'estensione della lista <i>ex art.</i> 83 TFUE	115
5. Gli ostacoli tuttora esistenti e il cambio di strategia della Commissione	116
6. Alcune riflessioni conclusive	120

CAPITOLO VII

SISTEMI DI INTELLIGENZA ARTIFICIALE E VIOLENZA DI GENERE: DUE ITINERARI DI RIFLESSIONE TRA LORO CONVERGENTI

<i>Irene Pellizzone</i>	121

1. Considerazioni introduttive sull'impatto dei sistemi di intelligenza artificiale sul fenomeno della violenza di genere: un itinerario a due sensi	121
2. I punti di forza dei Sistemi di intelligenza artificiale: strumenti di contrasto alla violenza di genere (richieste di aiuto ed esame dei fattori di rischio)	125
3. Le donne come vittime predestinate dell'IA: <i>focus</i> sul <i>Deep Fake</i> non consensuale	129
4. Un importante passo avanti: l'intervento normativo attraverso la recente legge n. 132/2025	135
5. Prevenzione della violenza nei confronti delle donne e dell'intelligenza artificiale: un primo bilancio	136

pag.

CAPITOLO VIII

SALUTE DELLE DONNE E REGOLAMENTAZIONE DELL'IA: LA *GENDER-SENSITIVE GOVERNANCE* COME PARADIGMA DI EGUAGLIANZA SOSTANZIALE

Marta Tomasi

139

1. Introduzione. L'irrealizzato obiettivo dell'eguaglianza in salute 139
2. Salute delle donne e discriminazioni strutturali 140
3. L'automatizzazione delle diseguaglianze 142
4. La produzione del dato come spazio di diseguaglianza: la ricerca biomedica tra esclusione e *bias* 144
5. La pratica clinica come spazio di diseguaglianza: diagnosi, cura e dimensione strutturale della salute 148
6. La ricerca dell'eguaglianza di genere nello spazio della regolazione 152
7. Soluzioni normative a confronto: equità sostanziale e *gender-sensitive governance* 156

CAPITOLO IX

I SISTEMI DI IA PER IL CONTROLLO DELL'IMMIGRAZIONE: BENEFICI E RISCHI PER I DIRITTI UMANI E LA PARITÀ DI GENERE

Cecilia Siccardi

161

1. Premessa 161
2. *Smart borders e governance* digitale dell'immigrazione 162
3. L'impatto dei sistemi di IA sulla condizione delle donne migranti 166
4. L'AI Act e la disciplina dei sistemi per il controllo dell'immigrazione 169
5. Le garanzie previste per i sistemi "ad alto rischio" ed alcune eccezioni 172
6. Alcuni ostacoli all'applicabilità delle garanzie previste dall'AI Act 175
7. I sistemi vietati che vengono in rilievo nel contesto dell'immigrazione 177
8. Riflessioni conclusive 178

CAPITOLO X

IL POTERE PERFORMATIVO DEL LINGUAGGIO NELL'ERA DELL'IA

Nannerel Fiano

179

1. L'intelligenza artificiale e il suo linguaggio *vis à vis* con la Costituzione 179

	<i>pag.</i>
2. Il funzionamento dei LLM	181
3. L' <i>hate speech</i> , questo (s)conosciuto: dalla filosofia del linguaggio alla tutela costituzionale	183
4. La regolamentazione europea dell'intelligenza artificiale: quale linguaggio?	185
5. La legge n. 132/2025 e i profili linguistici	186
6. Conclusioni: per un linguaggio umano e tecnologico non discriminatorio	187

CAPITOLO XI

LA PROTEZIONE DEI DATI PERSONALI AL TEMPO DELL'INTELLIGENZA ARTIFICIALE: IL COORDINAMENTO TRA AI ACT E GDPR

Pietro Villaschi 189

1. Introduzione: i dati come linfa vitale della società digitale	189
2. La perdurante attualità di una regolazione che limiti un trattamento "selvaggio" dei dati personali degli utenti della società digitale	190
3. AI ACT e GDPR: punti di contatto e complementarietà delle normative europee	192
4. AI ACT e GDPR: possibili problemi di coordinamento	193
5. Tra proposte di integrazione e rischi di sovra-regolazione	195

CAPITOLO XII

AI ACT E DISCRIMINAZIONI DI GENERE: TUTELA EXTRA-GIUDIZIALE E GIUDIZIALE A CONFRONTO

Paolo Gambatesa 197

1. Introduzione: l'algoritmo in una prospettiva costituzionale quale chiave di lettura per il contrasto delle discriminazioni di genere	197
2. Il contrasto alle discriminazioni di genere nei sistemi vietati e ad alto rischio. Una lettura in filigrana dell'AI Act	201
3. (<i>Segue</i>): le previsioni più esplicite della legge n. 132/2025	205
4. Luci e ombre dello strumento extra-giudiziale della valutazione di impatto ai sensi dell'art. 27 dell'AI Act	206
5. Le possibili prospettive di tutela giudiziale in un'ottica di integrazione dei livelli di protezione dei diritti	209
6. Considerazioni conclusive	212

pag.

CAPITOLO XIII

L'UTILIZZO DI SISTEMI DI IDENTIFICAZIONE BIOMETRICA REMOTA NEL REGOLAMENTO SULL'INTELLIGENZA ARTIFICIALE TRA MECCANISMO DI AUTORIZZAZIONE E MECCANISMO DI SORVEGLIANZA

Giulia Agrati

215

1. Premessa 215
2. Introduzione: il divieto di utilizzo di sistemi di identificazione biometrica remota in "tempo reale" in spazi accessibili al pubblico e a fini di attività di contrasto alla criminalità e le rispettive eccezioni 217
3. Il meccanismo di autorizzazione e il ruolo dell'autorità giudiziaria o autorità amministrativa indipendente 219
4. Il meccanismo di controllo e il ruolo dell'autorità di vigilanza del mercato e dell'autorità nazionale per la protezione dei dati 224
5. Conclusioni 226

CAPITOLO XIV

LE DISCRIMINAZIONI DI GENERE NELL'ACCESSO AL LAVORO GENERATE O AMPLIFICATE DAI SISTEMI DI INTELLIGENZA ARTIFICIALE: PROFILI COSTITUZIONALI

Niccolò Panigada

227

1. I sistemi di IA e la discriminazione di genere nell'accesso al lavoro: origine del problema a partire da alcuni casi concreti 227
2. L'ancoraggio costituzionale e sovranazionale 229
3. La legge italiana sull'intelligenza artificiale in relazione al mondo del lavoro 231
4. Rischi e criticità giuridiche ancora presenti 232
5. I rimedi previsti dall'ordinamento 234
6. Conclusioni 235

CAPITOLO XV

IA GENERATIVA E DIFFUSIONE DEGLI STEREOTIPI ONLINE. LA CONCETTUALIZZAZIONE DI UN AGENTE CONTRO IL LINGUAGGIO DISCRIMINATORIO

Marta Annamaria Tamborini

237

1. Premessa 237

	<i>pag.</i>
2. Le nuove frontiere della discriminazione algoritmica	239
3. <i>Bias</i> nel linguaggio, <i>bias</i> nell'algoritmo e diffusione di odio	241
4. L'IA generativa come alleato e non come minaccia	243
5. Il progetto dell'identificazione degli stereotipi e del linguaggio d'odio attraverso l' <i>Agent for Anti Discriminatory Language</i> (AAL)	243
6. Conclusioni	245

PARTE II

GOVERNANCE INCLUSIVA DELL'INTELLIGENZA ARTIFICIALE: PRATICHE ISTITUZIONALI, FORENSI E AZIENDALI

CAPITOLO XVI

INTELLIGENZA ARTIFICIALE, DISCRIMINAZIONI DI GENERE E PUBBLICHE AMMINISTRAZIONI

Elena Buscemi

249

CAPITOLO XVII

INTELLIGENZA ARTIFICIALE E DISCRIMINAZIONI DI GENERE: IL RUOLO DELLE ISTITUZIONI FORENSI NEL GOVERNO DI UN ECOSISTEMA INCLUSIVO

Antonino La Lumia

253

1. Premessa: perché parlare di IA, genere e avvocatura	253
2. IA e discriminazioni di genere: forme, casi, meccanismi	254
2.1. Dal pregiudizio sociale al <i>bias</i> algoritmico	254
2.2. Occupazione, selezione del personale e valutazione delle performance	254
3. L' <i>AI Act</i> europeo e la centralità del rischio discriminatorio	255
4. IA, genere e responsabilità delle istituzioni forensi	256
4.1. Dall'etica "dichiarata" alla responsabilità "praticata"	256
4.2. IA e dovere di riservatezza: una questione anche di genere	256
5. Parità di genere e accessibilità economica ai sistemi di IA	257
6. IA come strumento di contrasto delle discriminazioni: dal rischio all'opportunità	258
7. Conclusioni: un'agenda condivisa per un'IA inclusiva	258

pag.

CAPITOLO XVIII

INTELLIGENZA ARTIFICIALE RESPONSABILE, UN VANTAGGIO COMPETITIVO: UN FRAMEWORK AZIENDALE PER INNOVAZIONE, ETICA E SOSTENIBILITÀ

Gabriele Secol, Beatrice Zulianello

261

SEZIONE I

L'AI ACT IN AZIENDA

1. Gli impatti delle regolamentazioni europee nelle società di consulenza	261
2. L'iniziativa aziendale	262

SEZIONE II

INNOVAZIONE RESPONSABILE

3. La sostenibilità dell'innovazione	263
4. Algoritmi discriminatori	264
5. Processo di gestione dei <i>bias</i> nei sistemi di intelligenza artificiale	266
6. Sensibilità degli strumenti	267

SEZIONE III

IMPEGNO DI DELOITTE

7. Promozione dell'innovazione per il sociale	269

CAPITOLO XIX

INTELLIGENZA ARTIFICIALE TRA ROTTURA E TRANSIZIONE: DAL “*RED FLAG ACT*” ALLO *HUMAN OVERSIGHT*

Giuseppe Vaciago

271

1. Introduzione: la paura “sbagliata” e il rischio di una nuova grande rottura	271

SEZIONE I

GOVERNARE SENZA FRENARE

2. L' <i>AI Act</i> : ambizione, cornice di tutela e rischio di <i>over</i> -regolazione	273
3. Lo <i>Human Oversight Officer</i> : profilo e ruolo operativo	273

SEZIONE II

POTERI PRIVATI E DIRITTO PUBBLICO

4. Le piattaforme come nuovi regolatori privati	275

	<i>pag.</i>
5. Trasparenza, <i>auditing</i> e tutela effettiva degli utenti	276
SEZIONE III	
UN LESSICO PER IL FUTURO	
6. Dall'informatica giuridica al diritto dell'informatica	277
7. Proposte operative per i corpi intermedi e gli Ordini professionali	278
8. Conclusioni	279
CAPITOLO XX	
L'INTELLIGENZA ARTIFICIALE IN A2A	
<i>Patrick Oungre</i>	281
1. Introduzione: <i>overview</i> sull'utilizzo e sviluppo di sistemi di intelligenza artificiale in A2A	281
2. I pillar di A2A e i progetti	281
2.1. L'obiettivo	281
2.2. Agenti IA a supporto dei team	282
2.3. Tecnologie agentiche a supporto dei processi	282
2.4. L'IA fisica	283
2.4.1. Progetto Prometeo	283
2.4.2. Progetto Pandora: manutenzione predittiva della rete elettrica	284
2.4.3. Progetto Ottimizzazione calore erogato impianti di teleriscaldamento	285
CAPITOLO XXI	
INTELLIGENZA ARTIFICIALE E IMPRESA: TRA INNOVAZIONE TECNOLOGICA E RESPONSABILITÀ CULTURALE	
<i>Alessandro Renna</i>	287
1. Premessa	287
2. Il vero problema non è la tecnologia	287
3. L'IA come leva competitiva delle organizzazioni: la nostra esperienza	288
4. L'IA come abilitatore di una nuova offerta tecnologica in ambito <i>legal</i>	289
5. Questioni interpretative aperte: il concetto di prevalenza	290
6. Conclusioni: un approccio europeo	291

ELENCO DEGLI AUTORI E DELLE AUTRICI

Giulia Agrati, Assegnista di ricerca di Diritto dell'Unione europea, Università degli Studi di Milano.

Ilaria Anrò, Professoressa associata di Diritto dell'Unione europea, Università degli Studi di Milano.

Giuseppe Arconzo, Professore ordinario di Diritto costituzionale, Università degli Studi di Milano.

Francisco Balaguer Calléjon, Professore ordinario di Diritto costituzionale, Università di Granada.

Elena Buscemi, Presidente del Consiglio comunale di Milano.

Silvia Cervia, Professoressa associata di Sociologia dei Processi Culturali e Comunicativi, Università di Pisa.

Marilisa D'Amico, Professoressa ordinaria di Diritto costituzionale, Università degli Studi di Milano.

Ernesto Damiani, Professore ordinario di Informatica, Università degli Studi di Milano.

Nannerel Fiano, Ricercatrice a tempo determinato di Diritto costituzionale, Università degli Studi di Milano.

Paolo Gambatesa, Assegnista di ricerca in Diritto costituzionale, Università degli Studi di Milano.

Antonino La Lumia, Presidente dell'Ordine degli Avvocati di Milano.

Patrick Oungre, Chief Innovation, A.I. and R&D Officer, A2A.

Niccolò Panigada, Dottorando in Diritto costituzionale, Università degli Studi di Milano.

Gaia Patarini, Dottoranda in Immagine, linguaggio, figura, forme e modi della mediazione, Università degli Studi di Milano.

Irene Pellizzone, Professoressa associata di Diritto costituzionale, Università degli Studi di Milano.

Edoardo Carlo Raffiotta, Professore associato di Diritto costituzionale, Università degli Studi di Milano-Bicocca.

Alessandro Renna, Founder e CEO 4cAi.

Gabriele Rugani, Assegnista di ricerca in Diritto dell'Unione europea e Diritto pubblico comparato, Università di Pisa.

Gabriele Secol, Partner, Deloitte officine innovazione.

Cecilia Siccardi, Ricercatrice a tempo determinato di Diritto costituzionale, Università degli Studi di Milano.

Elettra Stradella, Professoressa associata di Diritto pubblico comparato, Università di Pisa.

Marta Annamaria Tamborini, Dottoranda in Immagine, linguaggio, figura, forme e modi della mediazione, Università degli Studi di Milano.

Marta Tomasi, Professoressa associata di Diritto pubblico comparato, Università degli Studi di Trento.

Giuseppe Vaciago, Avvocato Cassazionista esperto in diritto penale societario e delle nuove tecnologie e fondatore di LT42.

Pietro Villaschi, Ricercatore a tempo determinato di Diritto costituzionale, Università degli Studi di Milano.

Beatrice Zulianello, Analyst, Deloitte officine innovazione.

NOTE INTRODUTTIVE SUL PRIN – *AiGeDi*
(*ARTIFICIAL INTELLIGENCE BETWEEN GENERATING
AND TACKLING GENDER-BASED DISCRIMINATIONS*;
L'INTELLIGENZA ARTIFICIALE DA CAUSA A STRUMENTO
DI CONTRASTO DELLE DISCRIMINAZIONI DI GENERE)
Marilisa D'Amico

Questo volume raccoglie i frutti di un lavoro di riflessione multidisciplinare e interdisciplinare sui temi dell'intelligenza artificiale. Il progetto "*Artificial Intelligence between generating and tackling gender-based discriminations (AiGeDi) – L'intelligenza artificiale da causa a strumento di contrasto delle discriminazioni di genere*" è stato promosso dall'Unità Capofila dell'Università degli Studi di Milano, che ho coordinato, composta, da un lato, da costituzionalisti e costituzionaliste del Dipartimento di diritto pubblico, italiano e sovranazionale e, dall'altro, informatici e informatiche del dipartimento di informatica "Giovanni degli Antoni", e due Unità Satellite, l'Università Milano-Bicocca, con il team del professor Edoardo Raffiotta, e l'Università degli Studi di Pisa, con il team della professoressa Elettra Stradella.

L'attività svolta è partita dalla considerazione che se, da un lato, l'IA rappresenta una delle più promettenti frontiere dell'innovazione tecnologica, dall'altro essa costituisce una sfida crescente sul piano dell'uguaglianza e dei diritti fondamentali. All'interno di questo scenario critico si è inserito il presente progetto, che proponeva un'indagine interdisciplinare sulle potenzialità discriminatorie delle tecnologie algoritmiche, con particolare attenzione al genere e alle dimensioni intersezionali della discriminazione.

Una sfida rilevante affrontata dal progetto è stata la necessità di analizzare un contesto normativo e tecnologico in divenire, che si è evoluto, e si evolve, a due velocità. Dal 2022, anno di presentazione del progetto, ad oggi, infatti, siamo passati, sul piano tecnologico, dall'interessarci agli algoritmi e poi ai sistemi di fondazione di IA generativa, i Large Language Models, per arrivare

all'era dell'intelligenza artificiale agentica, ossia macchine che autonomamente interagiscono con altre macchine in un flusso di lavoro. Dal punto di vista della *governance*, l'Unione Europea, attraverso l'approvazione dell'AI Act (regolamento (UE) 2024/1689) e delle sue linee guida attuative (in particolare il "General Purpose AI Code of Practice", alla cui redazione ho contribuito in qualità di esperta indipendente), ha inteso intraprendere un percorso, forse ambizioso, di inquadrare e governare l'uso e le applicazioni dell'intelligenza artificiale anche sotto la lente dell'impatto sui diritti fondamentali; percorso che l'Italia ha peraltro seguito con la recente legge in materia di intelligenza artificiale (legge n. 132/2025). Ciononostante, l'intelligenza artificiale generativa ha avuto un impatto dirompente, integrando l'IA nelle attività quotidiane di una vastissima platea di utenti, rendendo necessaria una ulteriore riflessione sulle linee guida e sui correttivi previsti, con il rischio di rendere già obsolete norme che, peraltro, non sono ancora entrate pienamente in vigore¹.

Questo ha richiesto un costante aggiornamento scientifico della nuova fonte regolamentare europea, al fine di valutarne l'efficacia nel prevenire e contrastare le discriminazioni di genere e intersezionali, nonché la sua coerenza con i principi del diritto antidiscriminatorio dell'Unione Europea.

Elemento distintivo del progetto è stato l'approccio integrato tra diritto costituzionale, comparato e del lavoro, informatica e sociologia. Attraverso un'analisi incrociata della normativa, dei modelli algoritmici e delle percezioni sociali (mediante il metodo Delphi), *AiGeDi* ha inteso delineare *best practices* e proposte concrete di *policy*. Il mondo del lavoro rappresenta il campo privilegiato d'indagine, dove l'uso di IA per il reclutamento e la gestione del personale comporta rischi significativi per la parità di genere.

Il progetto ha inteso contribuire al dibattito normativo e culturale sull'IA e la discriminazione, proponendo soluzioni capaci di trasformare le tecnologie in leve per una democrazia realmente inclusiva.

L'obiettivo centrale del lavoro è stato comprendere come e perché l'IA riproduca o amplifichi disuguaglianze di genere, spesso già radicate nei costrutti sociali, e individuare strategie giuridiche, sociologiche e informatiche per contrastarle.

Partendo da evidenze empiriche consolidate, dai noti casi Amazon alle

¹ Si pensi, ad esempio alla recente proposta del c.d. regolamento "Digital Omnibus" che si pone l'obiettivo di semplificare e armonizzare il quadro normativo digitale dell'UE proponendo un'ampia revisione tecnica e razionalizzazione di diverse fonti, tra cui il GDPR, l'AI Act e il Data Act, per ridurre sovrapposizioni e oneri burocratici per imprese e cittadini, pur mantenendo gli obiettivi di fondo.

analisi emerse in letteratura come “Gender Shades”, il progetto ha indagato il ruolo che i *bias* nei dati, la carenza di *diversity* nei team di sviluppo e la non trasparenza degli algoritmi o del loro *training*, nel caso dell’IA Generativa, hanno nel generare trattamenti differenziati e svantaggiosi per le donne, soprattutto per quelle appartenenti a gruppi vulnerabili.

Con riferimento ai risultati attesi, il progetto si prefiggeva di contribuire alla ricerca multidisciplinare in varie direzioni, incrementando le conoscenze sulle implicazioni del ricorso all’intelligenza artificiale rispetto alle ricadute discriminatorie di genere e intersezionali, attraverso le analisi giuridiche e sociologiche condotte dalle Unità coinvolte.

I risultati sono stati tre filoni di iniziative che hanno dato vita a uno spazio di confronto stabile e interdisciplinare sui rapporti fra intelligenza artificiale, diritti e disuguaglianze. In primo luogo, si è tenuto un ciclo di incontri, con il Prof. Francisco Balaguer Callejón, che ha approfondito in chiave comparata e transnazionale le implicazioni etico-giuridiche dell’IA a partire dal volume *La Constitución del Algoritmo*, riunendo costituzionalisti e informatici per discutere le criticità sistemiche dell’IA rispetto al principio di eguaglianza e ai diritti fondamentali. In secondo luogo, sono poi seguiti altri incontri che hanno accompagnato la presentazione del progetto Human Hall for AI, volto a mettere in dialogo accademia, istituzioni, imprese e società civile su etica dell’IA, governance dei dati e prevenzione delle discriminazioni algoritmiche, con un primo incontro dedicato a “Nuovi diritti, compliance e sfide dell’AI” e un secondo, “Human AI literacy”, incentrato sulle ricadute dell’IA su lavoro, formazione e professioni. Infine, è stato dedicato un terzo ciclo di seminari all’intelligenza artificiale in una prospettiva intersezionale, con attenzione specifica al nesso tra genere, salute e disabilità, mettendo in luce i rischi che le tecnologie intelligenti rafforzino disuguaglianze preesistenti, in particolare nella sanità, nella violenza di genere digitale e nell’accessibilità per le persone con disabilità, e sottolineando la necessità di soluzioni tecnologiche sensibili alle differenze, orientate alla giustizia sociale e alle pari opportunità.

Il secondo risultato atteso era l’elaborazione di azioni concrete, grazie al supporto degli informatici e sulla base dei prodotti della ricerca sociologica condotta, capaci di intervenire sulle metodologie di intelligenza artificiale già esistenti allo scopo di ridurre l’impatto discriminatorio ai danni delle donne e di elaborare *best practices*. Il frutto di questo lavoro ha contribuito, come anticipato in precedenza, alla nascita di “Human Hall 4 AI”, un centro di ricerca all’interno dell’*hub* sull’inclusione di Human Hall dell’Università degli Studi di Milano, che ha sviluppato una metodologia di valutazione di impatto dei

sistemi di intelligenza artificiale sui diritti umani². Human Hall 4 AI ha supportato le attività del presente progetto, permettendone un maggiore e più ampio approfondimento delle tematiche trattate.

Il terzo obiettivo era lo sviluppo di un software “pilota” da testare nel settore del diritto del lavoro, con il contributo sinergico dei team di ricerca facenti capo a tutte le Unità. Questo obiettivo si è tradotto in un approccio all'avanguardia di applicazione dell'intelligenza artificiale generativa per riconoscere il linguaggio discriminatorio nei testi tipici del contesto lavorativo (ad esempio, gli annunci di lavoro), con particolare attenzione alla discriminazione intersezionale. Ciò è stato possibile attraverso la progettazione sinergica tra informatici e giuristi di un agente pilota in grado di riconoscere il linguaggio discriminatorio e di analizzarne la radice stereotipata e intersezionale per proporre una qualificazione giuridica fondata sugli esiti delle ricerche di diritto costituzionale e diritto antidiscriminatorio al fine di testare l'agente pilota rispetto alle categorie classiche del diritto antidiscriminatorio e del controllo di ragionevolezza delle scelte selettive³.

In questa prospettiva, i contributi raccolti nel presente volume intendono offrire una chiave di lettura multidisciplinare delle trasformazioni in atto, mantenendo al centro il nesso tra innovazione tecnologica, eguaglianza sostanziale e tutela dei diritti fondamentali. L'attenzione al genere e alle discriminazioni intersezionali non rappresenta un capitolo specializzato o separato, ma il banco di prova attraverso cui verificare la capacità dell'ordinamento di governare l'intelligenza artificiale senza rinunciare alle promesse, ancora attuali, dell'art. 3 della Costituzione. Il volume si articola lungo un percorso che rispecchia la struttura del progetto: viene ricostruito il quadro teorico e normativo, europeo e nazionale, entro cui si colloca l'uso dell'IA, soffermandosi in particolare sui profili di eguaglianza, non discriminazione e responsabilità. L'analisi si sposta successivamente sul terreno dei casi d'uso, con particolare riguardo al lavoro e agli spazi digitali in cui si manifestano nuove forme di violenza e di esclusione,

² P. CERAVOLO, E. DAMIANI, M.E. D'AMICO, B. DE TEFFÉ ERB, S. FAVARO, N. FIANO, P. GAMBATESA, S. LA PORTA, S. MAGHOOL, L. MAURI, N. PANIGADA, L.M. RATTO VAQUER, M.A. TAMBORINI, *HH4AI: A Methodological Framework for AI Human Rights Impact Assessment under the EU AI Act*, in *HHAI-WS 2025: Workshops at the Fourth International Conference on Hybrid Human-Artificial Intelligence (HHAI 2025)*, Pisa, CEUR Workshop Proceedings, vol. 4074, 2025, 251-263.

³ M.A. TAMBORINI, F. MOHAMMADI, P. CERAVOLO, S. MAGHOOL, M.E. D'AMICO, C. SICCARDI, *Agent for Anti-Discriminatory Language (AAL): A Human-Centered AI Approach to Discriminatory Speech in Italian*, in *IEEE International Conference on Cyber Humanities*, 2025, Proceedings, Firenze, 8-10 settembre 2025.

mostrando come i *bias* algoritmici si intreccino con disuguaglianze preesistenti. I contributi del presente volume si concentrano anche sugli strumenti: dalle metodologie di valutazione di impatto sui diritti umani sviluppate nell’ambito di Human Hall 4 AI alle soluzioni tecnico-giuridiche sperimentate con il software pilota per il riconoscimento del linguaggio discriminatorio nei contesti lavorativi. L’auspicio è che il dialogo tra diritto costituzionale, diritto antidiscriminatorio, informatica e scienze sociali che ha animato il progetto “AiGeDi” possa proseguire oltre i confini del progetto, alimentando pratiche di regolazione, di *design* e di utilizzo dei sistemi di intelligenza artificiale più consapevoli e responsabili. Se anche solo una parte delle riflessioni qui raccolte contribuirà a orientare le scelte del legislatore, delle istituzioni, delle imprese e degli operatori del diritto verso un’IA capace di ridurre, anziché amplificare, le disuguaglianze di genere e le discriminazioni intersezionali, il lavoro svolto dalle Unità coinvolte potrà dirsi, almeno in parte, riuscito.

RELAZIONE INTRODUTTIVA

SFIDE PER COMBATTERE LE DISCRIMINAZIONI
DI GENERE NELL'ERA DELL'IA

Francisco Balaguer Callejón

SOMMARIO: 1. Introduzione. – 2. L'uguaglianza di genere come ultima frontiera del diritto costituzionale. – 3. L'era dell'IA e l'algoritmo come nuovo strumento di trasformazione storica. – 4. La parità di genere nell'era dell'intelligenza artificiale. – 4.1. L'orientamento dei diritti verso un mercato discriminatorio. – 4.2. Parità di genere e riproduzione culturale della discriminazione nell'era digitale. – 5. Conclusioni.

1. Introduzione.

Uno degli aspetti che accomuna i gruppi accademici di Granada e Milano, e che emerge in questo convegno, è la partecipazione di ricercatori di varie generazioni. Come ha affermato Marilisa D'Amico nell'introduzione a questo convegno, dai giovani dobbiamo imparare. Sono loro a conoscere meglio di noi la realtà dell'intelligenza artificiale e rappresentano anche una generazione con modelli culturali differenti che gli permettono di comprendere la realtà in modo diverso da noi. Dobbiamo sempre imparare dai giovani. Il mio maestro tedesco, Peter Häberle, sosteneva l'importanza del dialogo intergenerazionale e del confronto con tutte le generazioni per comprendere la realtà¹. Ricordiamo il proverbio arabo che paragona la verità a uno specchio caduto dal cielo, che

¹ Come sottolinea P. Häberle, lo Stato costituzionale si oppone agli Stati totalitari e alle pretese fondamentaliste di verità assoluta, in quanto «si caratterizza per la consapevolezza di non essere in possesso di precostituite verità eterne, ma di essere invece destinato a una mera ricerca della verità». In modo tale che «[l']immagine dell'uomo dello stato costituzionale sottintende un cittadino che necessita di verità e che è in possesso degli strumenti per ricercarla e "conquistarla"», P. HÄBERLE, *Wahrheitsprobleme im Verfassungsstaat*, Baden-Baden, 1995, versione italiana di G. ZAGREBELSKY, J. LUTHER, *Diritto e verità*, Torino, 2000, 85 e 110.

lascia a ciascuno di noi solo un frammento, un pezzo di verità. Dobbiamo mettere insieme tutti questi frammenti per conoscere la verità, perché essa è una costruzione collettiva (anche intergenerazionale) e nessuno ne possiede più di una parte².

Ecco perché, per caratterizzare la democrazia pluralista, mi piace sempre utilizzare questa metafora araba dello specchio. Se conoscere la verità e capire la realtà richiede di mettere insieme tutti i frammenti che ognuno di noi possiede, allora tutti questi frammenti hanno la stessa importanza e la stessa dignità. Non possiamo ignorare o rifiutare nessuno semplicemente perché non è in linea con la nostra comprensione della verità³. La verità di Elon Musk, che possiede un pezzo enorme di questo specchio, ha la stessa dignità di quella di un contadino di un piccolo paese italiano o spagnolo. Tutti i pezzi sono necessari per la costruzione sociale della verità. Tuttavia, i pezzi dello specchio che vengono condivisi e che sono rilevanti per la costruzione sociale della verità dipendono ora da fattori di potere che non possiamo controllare attraverso la democrazia pluralista, perché si trovano al di fuori del circuito statale e, negli ultimi anni, hanno ampiamente promosso una regressione nella lotta per la parità di genere⁴.

Le sfide per promuovere la parità di genere sono molteplici e la battuta d'arresto degli ultimi anni è evidente, almeno per quanto riguarda la lotta culturale per l'uguaglianza, rispetto a qualche anno in cui vi era un clima molto più favorevole. Per esempio, in Spagna, che è sempre stata all'avanguardia nella promozione dell'uguaglianza di genere. Prima della pandemia, la Spagna si trovava in una situazione molto positiva, al punto che molte persone, che mai avremmo immaginato, si sono dichiarate a favore dell'uguaglianza di genere⁵.

Ora, invece, ci troviamo in una situazione difficile in molti Paesi. In particolare, negli Stati Uniti si sta verificando un'involuzione culturale, costituzionale e di civiltà che sta influenzando anche le rivendicazioni a favore della parità di genere. In Europa, inoltre, ci sono movimenti politici populistici che hanno

² F. BALAGUER CALLEJÓN, *La costituzione dell'algoritmo*, Milano, 2023.

³ F. BALAGUER CALLEJÓN, M.L. BALAGUER CALLEJÓN, *Verdad e interpretación en la sociedad digital*, Cizur Menor, 2023.

⁴ F. BALAGUER CALLEJÓN, *La trasformazione dei diritti nella società digitale e il suo impatto sulla parità di genere*, in M. D'AMICO, B. LIBERALI (a cura di), *I diritti delle donne. Problematiche attuali e prospettive future*, Torino, 2024.

⁵ Pensiamo alla presidente della Banca di Santander (una delle più grandi banche del mondo) che si è dichiarata pubblicamente femminista nel 2018. A. BOTÍN, *Por qué me considero feminista y tú también deberías*, in *El País*, 20 agosto 2018.

avuto un grande sviluppo negli ultimi anni e che si sono manifestati contro l'idea della promozione dell'uguaglianza di genere.

2. *L'uguaglianza di genere come ultima frontiera del diritto costituzionale.*

La parità di genere rappresenta l'ultima frontiera del diritto costituzionale. Si tratta certamente di una grande trasformazione, perché non riguarda solo le donne, ma l'intera società e l'emancipazione di tutti. In questo senso, si tratta di una trasformazione essenziale, che definisce l'ultima tappa del costituzionalismo. Può sembrare strano, ma non lo è se si considera l'evoluzione del costituzionalismo nell'arco di duecento anni. Il costituzionalismo ha avuto inizio come un movimento non democratico a suffragio censitario, dal quale le donne erano escluse. Erano esclusi anche molti uomini che non possedevano i requisiti richiesti, a causa dell'istruzione, della proprietà e di altri fattori.

L'evoluzione costituzionale è stata inizialmente incentrata sull'idea di trasformare l'economia e la società attraverso la forza della legge, la liberazione delle forze produttive e la costruzione dei mercati nazionali. Questa è stata la grande rivoluzione del mondo moderno e da essa sono derivate ulteriori trasformazioni: dallo Stato legale di diritto si è passati allo Stato costituzionale di diritto⁶. Si è verificata anche una trasformazione del quadro costituzionale stesso, con una nuova tappa del costituzionalismo partita dal *laissez-faire* in ambito economico e poi evoluta nell'idea di costituzione economica nazionale e, infine, nel processo di integrazione europea⁷. Tutti questi sono processi di transizione all'interno dell'evoluzione del processo rivoluzionario del costituzionalismo.

L'ultima trasformazione del movimento costituzionalista è proprio quella della parità di genere, molto recente dal punto di vista culturale, ma che ha registrato un forte sviluppo negli ultimi quarant'anni. È possibile far risalire alcune idee fondamentali alla lotta contro la discriminazione di genere. In primo luogo, il suo legame con il costituzionalismo: maggiore è la densità costituzionale nella società, maggiori sono le possibilità di fare progressi a favore della parità di genere. Questa è una caratteristica molto importante del movimento a favore dell'uguaglianza di genere. Si tratta di un movimento

⁶ F. BALAGUER CALLEJÓN, *Fuentes del Derecho*, Madrid, 1991/1992 (nuova edizione nel CEPC, Madrid, 2022).

⁷ F. BALAGUER CALLEJÓN, *Costituzione economica e globalizzazione*, in *federalismi.it*, 5, 2019, 43 ss.

costituzionale che ha bisogno della costituzione e che si sviluppa meglio nell'ecosistema costituzionale ⁸.

Ci sono ulteriori motivi per cui l'uguaglianza di genere dovrebbe essere inserita anche nella prospettiva economica, perché la parità di genere non è solo un movimento culturale o costituzionale, ma ha anche una sua funzionalità dal punto di vista economico. In particolare, come ultima fase della liberazione delle forze produttive e del lavoro, l'uguaglianza di genere ha permesso alle donne, che erano rimaste nell'ambito domestico, di raggiungere una condizione di maggiore parità sul piano sociale, politico ed economico. Questa funzionalità è diversa da quella costituzionale, in quanto dipende dalle condizioni dell'economia e del mercato. Tuttavia, il *background* costituzionalista è molto importante. Ciò significa che, quando ci sono delle sfide per il costituzionalismo, ci sono anche delle sfide per la parità di genere. Quindi, le due sfide sono collegate.

Una caratteristica molto importante da considerare nella promozione dell'uguaglianza di genere è la capacità di trasformazione culturale, oltre a quella sociale. È possibile ottenere una trasformazione sociale che sia una trasformazione culturale minore. Infatti, ci sono più difficoltà per la trasformazione culturale rispetto a quella sociale. Il movimento per l'uguaglianza di genere ha ottenuto grandi successi sul piano economico, politico e sociale, ma non ha avuto lo stesso impatto sul piano della trasformazione culturale. Un esempio chiaro è il linguaggio non discriminatorio. Il linguaggio è uno strumento essenziale per la configurazione delle strutture e delle relazioni di potere, in quanto ne facilita la formazione e la riproduzione, che avvengono in modo impercettibile, ostacolando l'opposizione e il conflitto ⁹. Il linguaggio ha una dimensione performativa dell'ordine sociale, generando una riproduzione culturale permanente che tende a consolidare le posizioni occupate nella società, rendendo difficile la lotta per l'uguaglianza di genere nella misura in cui è stata storicamente costruita su modelli maschili ¹⁰.

⁸ F. BALAGUER CALLEJÓN, *La trasformazione dei diritti nella società digitale e il suo impatto sulla parità di genere*, cit.

⁹ Opera di riferimento su questo tema è "Lingua Tertii Imperii" di Victor Klemperer che analizza come la lingua tedesca sia stata plasmata per adattarsi all'ideologia del regime totalitario nazista, trasformando il significato di alcuni termini, creandone altri, escludendo quelli che non interessavano e dando una maggiore presenza pubblica a quelli che erano importanti per il consolidamento del nuovo sistema di potere. V. KLEMPERER, *LTI-Lingua Tertii Imperii. Notizbuch eines Philologen*, Berlino, 1947. Versione spagnola: *La lengua del Tercer Reich*, Barcellona, 2001.

¹⁰ F. BALAGUER CALLEJÓN, *La construcción del lenguaje jurídico en la Unión Europea*, in

C'è una grande difficoltà per favorire lo sviluppo di un linguaggio non discriminatorio e non ci sono molti che hanno compreso l'importanza della trasformazione del linguaggio. Modificare il linguaggio è importante soprattutto sul piano giuridico, come sostiene Marilisa D'Amico. La trasformazione del linguaggio è importante anche sul piano personale, nell'ambito delle nostre relazioni sociali. Ma sul piano giuridico ci sono delle norme e la discriminazione continua a rimanere a livello normativo ¹¹.

Secondo me, c'è un ostacolo culturale che impedisce di trasformare la società in senso egualitario. La riproduzione culturale della disuguaglianza si manifesta in grande misura nei piani subliminali, attraverso stereotipi e abitudini che non vengono messi in discussione perché non siamo consapevoli della loro esistenza. Comprendere questa capacità di ritardo culturale è fondamentale per promuovere l'uguaglianza di genere attraverso la trasformazione culturale. Infatti, quando si incontrano ostacoli sia per il costituzionalismo sia per la trasformazione culturale, ci si trova di fronte a delle sfide per l'uguaglianza di genere e per la lotta contro le discriminazioni di genere.

3. *L'era dell'IA e l'algoritmo come nuovo strumento di trasformazione storica.*

In questo contesto entra in gioco il profilo dell'intelligenza artificiale. Per comprendere il nuovo mondo della società digitale, è utile distinguere tra *transizioni* e *rottture*. La Rivoluzione francese, anche se continua a essere concettualmente rilevante nel quadro di questa nuova "tappa" della promozione dell'uguaglianza di genere, ha segnato la transizione dallo stato assoluto alla modernità. La dimensione culturale del mondo moderno proviene da una rivoluzione, da una rottura. Una rottura causata per mezzo della costituzione e il diritto (la *Force de Loi*) come strumento. Questa rottura è diversa dalle transizioni, anche dell'uguaglianza di genere compresa nel costituzionalismo come la sua ultima frontiera. La rottura è lo stabilimento del costituzionalismo e dell'universo concettuale del nuovo mondo che, però, non ha definito il suo volto se non

Revista de derecho constitucional europeo, 1, 2004, 307-322 e ID., *Las formulaciones normativas de la promoción de un lenguaje igualitario*, in *Artículo 14, una perspectiva de género: Boletín de información y análisis jurídico*, 31, 2009, 30-31.

¹¹ Se si considera che il linguaggio giuridico è fissato in modo immutabile nelle disposizioni normative, a differenza del linguaggio comune che si nutre in modo permanente dello sviluppo sociale e culturale. M. D'AMICO, *Parole che separano. Linguaggio, costituzione, diritti*, Milano, 2023, 46-47.

attraverso le diverse transizioni originate in più di duecento anni al suo interno¹².

La transizione, invece, non è così. La transizione ti consente di vedere cosa c'è dietro la trasformazione, qual è il suo futuro. Nella transizione dallo Stato legale a quello costituzionale di diritto, dallo Stato liberale allo Stato sociale c'era già una idea anche sul piano teorico. Pensiamo nel primo dopoguerra, al contributo dei costituzionalisti tedeschi, ma anche italiani: da Franz Neumann con la costituzione economica¹³, Hermann Heller con lo Stato sociale di diritto¹⁴, Hans Kelsen con la configurazione della Costituzione come norma che fornisce unità all'ordinamento giuridico, la democrazia pluralista e la giurisdizione costituzionale¹⁵, a Costantino Mortati con "La Costituzione in senso materiale"¹⁶. Tutti grandi giuristi che hanno dato un contributo nella costruzione dello Stato costituzionale¹⁷.

È facile comprendere la differenza tra transizione e rottura. È come quando si apre una porta: quando c'è una transizione, è possibile immaginare più o meno cosa ci sarà dall'altra parte. Non sappiamo tutto, perché si tratta di una transizione, ma possiamo vedere alcune cose. Per quanto riguarda l'eguaglianza di genere c'è una dinamica che non ci consente di vedere tutto il futuro, ma questo è normale nella transizione. La rottura è diversa: in questo caso, quando apriamo la porta, non vediamo nulla. La situazione dell'intelligenza artificiale è così. Non sappiamo cosa c'è dietro la porta, perché, in realtà, l'intelligenza

¹² F. BALAGUER CALLEJÓN, *Crisi, transizioni e rotture. Sul futuro del costituzionalismo*, in *Passaggi Costituzionali*, 2, 2024, 45 ss.

¹³ F. NEUMANN, *Über die Voraussetzungen und den Rechtsbegriff einer Wirtschaftsverfassung*, 1931, versione spagnola in *Revista de Derecho Constitucional Europeo*, 9, 2008.

¹⁴ H. HELLER, *Rechtsstaat oder Diktatur?*, 1929, versione spagnola nella raccolta dello stesso autore *Escritos políticos*, Madrid, 1985.

¹⁵ H. KELSEN, *Reine Rechtslehre*, II ed., Verlag Franz Deuticke, Vienna, ristampa 1967, ID., *Vom Wesen und Wert der Demokratie*, 1920, versione spagnola della seconda edizione del 1929, *Esencia y valor de la democracia*, Labor, Barcellona, 1934 e ID., *La garantie juridictionnelle de la Constitution (La Justice constitutionnelle)*, 1928, versione spagnola, *Anuario Iberoamericano de Justicia Constitucional*, 1, 2011, Madrid.

¹⁶ C. MORTATI, *La Costituzione in senso materiale*, Milano, 1940, ristampa del 1998. Sul rapporto tra Mortati e Kelsen, si veda F. BALAGUER CALLEJÓN, *Potere costituente e limiti alla revisione costituzionale visti dalla Spagna*, in F. LANCHESTER (a cura di), *Costantino Mortati. Potere costituente e limiti alla revisione costituzionale*, Padova, 2017, 85-112.

¹⁷ In particolare, per quanto riguarda il contributo dei costituzionalisti italiani alla costruzione dello Stato costituzionale, F. BALAGUER CALLEJÓN, *La teoria delle fonti del diritto nello Stato costituzionale*, in AA.VV., *Scritti in onore di Antonio Ruggeri*, Napoli, 2021.

artificiale si muove ogni giorno. Abbiamo pensato che l'ultima frontiera fosse l'intelligenza artificiale generativa. Invece, ci hanno detto che è necessario cambiare metodologia, perché tutti questi modelli hanno incorporato, almeno in inglese, tutti i dati raggiungibili su internet. Non è possibile, quindi, aggiungere più dati per migliorare. Non è aumentando i dati a nostra disposizione che otterremo un'intelligenza generale; è necessario ripensare la metodologia.

Non sappiamo ancora quale sarà questa nuova intelligenza artificiale. Non sappiamo come sarà, quale nome avrà né quale sarà la sua influenza sulle relazioni sociali e sulla politica. Si tratta di una trasformazione che avviene al di fuori del diritto e della Costituzione. La rottura costituzionale che ha dato vita al mondo moderno è avvenuta grazie alla Costituzione e ai giuristi. Sono stati i giuristi, con la forza della legge, a cambiare le regole dell'*Ancien Régime*, creando il mondo moderno. Oggi, però, la forza trasformatrice del mondo non è la Costituzione. Al contrario, viene considerata un ostacolo, perché il mondo giuridico regola e la regolamentazione può ostacolare lo sviluppo della tecnologia. Oggi, al posto del *costituzionalismo* c'è la *tecnologia* e al posto della *costituzione*, *l'algoritmo*. Queste sono le due nuove forze trasformatrici del mondo. L'intelligenza artificiale è una rivoluzione, una rottura affidata all'algoritmo come strumento e alla tecnologia come ideologia e cultura ¹⁸.

4. *La parità di genere nell'era dell'intelligenza artificiale.*

4.1. *L'orientamento dei diritti verso un mercato discriminatorio.*

In che modo questa rivoluzione fondata sull'intelligenza artificiale influisce sulla parità di genere o sulla lotta contro la discriminazione di genere? Come dicevo all'inizio, più ci allontaniamo dalla Costituzione, più la parità di genere si indebolisce. Allo stesso tempo, se non siamo in grado di promuovere una trasformazione culturale, viene meno la possibilità di un'evoluzione a favore della parità di genere.

È necessario, dunque, analizzare questi due profili. Il primo profilo riguarda la posizione del costituzionalismo nella società digitale. È una trasformazione che si è verificata anche con la terza globalizzazione, e che va oltre la tecnologia, ma che trova in essa un grande fondamento. Si tratta di una trasformazione che riguarda

¹⁸ F. BALAGUER CALLEJON, *Constitution, démocratie et mondialisation. La légitimité de la Constitution face à la crise économique et aux réseaux sociaux*, in *Mélanges en l'honneur du Professeur Dominique Rousseau. Constitution, justice, démocratie*, Parigi, 2020.

i diritti fondamentali e la loro struttura. Qui è possibile vedere, come affermato da Augusto Aguilar, la trasformazione dei diritti dei cittadini in diritti dei consumatori¹⁹. Si è verificata una evoluzione per cui i diritti dei cittadini si sono gradualmente trasformati in diritti dei consumatori. Questa trasformazione non si riscontra nel testo delle Costituzioni, che continuano a dare poco o nessun ruolo ai diritti dei consumatori²⁰. D'altra parte, la globalizzazione ha allontanato il processo produttivo dall'economia, che non rimane sempre all'interno dello Stato, ma può estendersi al di fuori di esso con grandi compagnie e aziende globali.

Il mercato continua a essere molto vivace e oggi si registrano più conflitti tra i diritti dei consumatori e degli utenti. È difficile pensare che la maggior parte della società italiana abbia oggi un conflitto permanente con lo Stato riguardo alla difesa dei propri diritti. Ci sono, però, molti conflitti tra i consumatori per l'acquisto di qualsiasi prodotto o tra gli utenti nell'utilizzo dei nuovi servizi presenti nella struttura economica.

Ciò significa che i diritti fondamentali si allontanano dalla Costituzione, perché la loro protezione non è più legata all'idea della dignità della persona e alle condizioni costituzionali, ma alle condizioni del mercato. Ciò può garantire una protezione più efficace, come è accaduto in Spagna durante la crisi finanziaria. In quel caso, era praticamente impossibile per i tribunali evitare la lesione del diritto all'abitazione per chi non aveva pagato l'ipoteca. Invece, la Corte di Giustizia dell'Unione Europea ha protetto i diritti dei consumatori e ha bloccato le espulsioni dalle abitazioni applicando il diritto dei consumatori. Il diritto dei consumatori si è rivelato così più efficace nel garantire un diritto costituzionale come quello all'abitazione²¹.

¹⁹ A. AGUILAR CALAHORRO, *El sujeto de derecho en la sociedad del consumo: el ciudadano como consumidor*, in M.A. GARCÍA HERRERA, J. ASENSI SABATER, F. BALAGUER CALLEJÓN (a cura di), *Constitucionalismo crítico: liber amicorum Carlos de Cabo Martín*, Valencia, 2016.

²⁰ Come sottolinea Raffaele Perna, in relazione alla Costituzione italiana, «[i]l baricentro dell'impianto culturale della costituzione economica è rappresentato dai produttori (lavoratori, imprese pubbliche, imprese private) mentre è del tutto assente la categoria del consumatore, che viceversa rappresenta l'architrave dell'economia di mercato», R. PERNA, *L'Italia ha bisogno di una nuova costituzione economica?*, Incontro alla fondazione Magna Carta, 15 giugno 2010, reperibile al seguente link http://www.astrid-online.it/static/upload/protected/Pern/Perna-R_Incontro_fondazione-Magna Carta_15_06_10.pdf.

²¹ Il diritto all'abitazione, nonostante fosse proclamato all'art. 47 della Costituzione spagnola, non è stato tutelato, ma è stato protetto dalla Corte di Giustizia UE in relazione alla violazione della legislazione europea sui diritti dei consumatori (direttiva 93/13/CEE del Consiglio del 5 aprile 1993 relativa alle clausole abusive nei contratti stipulati con i consumatori), Corte Giust. UE, 14 marzo 2013, causa C-415/11, Mohamed Aziz c. Catalunya caixa.

Nonostante questo aspetto positivo dell'effettività del diritto dei consumatori, non si può nascondere che questo tipo di diritto è meno legato alla Costituzione e alla dignità della persona. La logica del diritto è garantire le condizioni di mercato e non la dignità della persona. Qui si evidenzia una particolare debolezza del costituzionalismo con riguardo ai nuovi diritti.

Ciò è più evidente nel diritto alla protezione dei dati. In questo caso, infatti, non cambia solo il soggetto, che non è più il cittadino ma il consumatore. Cambia anche l'oggetto della protezione, che sono i dati, considerati il "petrolio" della nuova economia. Ciò significa che i dati, nei diritti alla loro protezione, si trasformano in una merce. Il diritto non è più solo un diritto soggettivo, ma diventa anche un oggetto, una parte del mercato ²².

Questi due diritti (il diritto dei consumatori e il diritto alla protezione dei dati) sono i grandi diritti della globalizzazione e della rivoluzione tecnologica. In essi si verifica una trasformazione della natura dei diritti, che non sono vincolati ai principi costituzionali come la dignità della persona nello stesso modo. Sono vincolati al mercato e hanno oggi una grande proiezione nel mercato stesso, al fine di garantirne la stabilità e il corretto funzionamento.

Questo orientamento dei diritti verso il mercato pone un problema in relazione all'uguaglianza di genere, in quanto la posizione della donna nel mercato è più debole rispetto a quella dell'uomo. Ciò pone una prima sfida riguardo al legame tra la lotta contro la discriminazione di genere e il diritto costituzionale, perché il mercato è discriminatorio.

Questo è evidente nella concezione di chi progetta i sistemi algoritmici. Non controllano i *bias* contro le donne presenti negli algoritmi, perché i *bias* riproducono il mercato e la realtà sociale. Tuttavia, la riproduzione della realtà sociale non ha nulla a che vedere con la Costituzione, in quanto quest'ultima impone un obbligo di trasformazione a favore dell'uguaglianza. Non è quindi possibile accettare la legittimità dei *bias* dal punto di vista costituzionale ²³.

²² F. BALAGUER CALLEJÓN, *Los derechos constitucionales en el contexto global y digital. Transformación del sujeto y conversión en objeto*, in W.C. ROTHENBURG (a cura di), *Direitos fundamentais, dignidade, constituição: estudos em Homenagem a Ingo Wolfgang Sarlet*, Londrina, 2021.

²³ Come indicato nel rapporto del 23 maggio 2018 "Algorithms in decision-making", Science and Technology Committee (House of Commons): «Algorithms, in looking for and exploiting data patterns, can sometimes produce flawed or biased 'decisions' – just as human decision-making is often an inexact endeavour. As a result, the algorithmic decision may disproportionately discriminate against certain groups, and are as unacceptable as any existing 'human' discrimination», rinvenibile al link di seguito indicato: <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/35102.htm>.

4.2. *Parità di genere e riproduzione culturale della discriminazione nell'era digitale.*

La necessità di trasformare la realtà ed eliminare le discriminazioni non è sostenuta solo dal diritto costituzionale. Rappresenta anche un elemento fondamentale della lotta culturale per l'uguaglianza. Tuttavia, questa lotta culturale sta attualmente subendo una regressione in molti Paesi, a partire dagli Stati Uniti. L'offensiva contro la regolamentazione dell'intelligenza artificiale si inserisce in questo contesto regressivo, in quanto mira a lasciare campo libero alle aziende tecnologiche, che impongono le proprie regole basate su un modello di *business* che antepone i loro profitti a qualsiasi altro criterio ²⁴.

Qui entra in gioco il profilo dell'intelligenza artificiale e dei social networks. È necessario distinguerli.

L'intelligenza artificiale ha una dimensione generale e, se pensiamo al suo utilizzo nei processi amministrativi, giudiziari, di concessione di aiuto e di sovvenzione, nonché in altri processi, il problema fondamentale è quello del *bias*. Questo problema non viene normalmente riconosciuto come tale dalle aziende tecnologiche. Vi è anche un problema strutturale: la mancanza di donne sia tra gli operatori che progettano gli algoritmi, sia nelle aziende che controllano l'intelligenza artificiale. Manca, quindi, una sensibilità per evitare la discriminazione ²⁵. Questo spiega perché i programmi e le applicazioni sono orientati al maschile e perché è necessario adottare misure per evitare questa tendenza.

La situazione è così squilibrata che sia l'ONU, che l'UNESCO che l'Unione Europea hanno richiamato l'attenzione sulla necessità di promuovere l'inserimento delle donne nel settore tecnologico e anche sul piano educativo, al fine di favorire la trasformazione ²⁶. Se continua a esserci una dimensione maschile

²⁴ F. BALAGUER CALLEJÓN, *La regolamentazione digitale europea nel nuovo contesto globale*, in *lceonline*, 2, 2025.

²⁵ Nelle parole di Marilisa D'Amico, «risulta evidente che algoritmi costruiti da uomini, che introducono elementi o dati di per sé discriminatori, sono a loro volta discriminatori», M. D'AMICO, *Una parità ambigua. Costituzione e diritti delle donne*, Milano, 2020, 317.

²⁶ A tal proposito si è pronunciata l'UNESCO indicando che «*Member States should encourage female entrepreneurship, participation and engagement in all stages of an AI system life cycle by offering and promoting economic, regulatory incentives, among other incentives and support schemes, as well as policies that aim at a balanced gender participation in AI research in academia, gender representation on digital and AI companies' top management positions, boards of directors and research teams*», UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 23 novembre 2021, 32. Nello stesso senso, lo ha fatto anche la Commissione delle Nazioni Unite sulla condizione delle donne (CSW), *Innovation and technological change, and education in the digital age*

egemonica nella configurazione dell'intelligenza artificiale, è difficile pensare che si possa avanzare verso la parità di genere. È quindi necessario ripensare le condizioni delle aziende che si occupano di intelligenza artificiale, così come l'educazione in questo campo. Questa è una sfida fondamentale.

D'altro canto, il problema della discriminazione non diminuisce quando aumenta l'autonomia della IA, anzi, è più importante nell'IA generativa. Quando entrano in gioco sistemi di intelligenza artificiale più autonomi, come quelli generativi, ovviamente ci sono più possibilità di discriminazione. La IA generativa utilizza modelli linguistici e tende a riprodurre la realtà esistente in tutte le sue dimensioni, comprese quelle discriminatorie.

Ora, rivolgiamo la nostra attenzione ai social network, che hanno una grande capacità di riproduzione culturale. Per combattere la discriminazione di genere, è necessario un obbligo di trasformazione culturale nei processi comunicativi che, invece, sui social network si trasforma in un processo più favorevole alla riproduzione culturale. Non è un problema di volontà: non si tratta semplicemente di dire alle aziende tecnologiche di favorire la trasformazione culturale.

Questo perché c'è un problema strutturale nel processo comunicativo. Quando c'è un processo comunicativo con i mass media tradizionali, si verifica una segmentazione della popolazione. Pensiamo ai lettori del Corriere della Sera, della Stampa o della Repubblica: ognuno di loro ha la propria ideologia, il proprio profilo ideologico. In una democrazia pluralista, tutte queste ideologie e le linee editoriali dei mass media si confrontano nello spazio pubblico, in uno spazio aperto, al fine di ottenere consensi e, in qualche modo, una percezione condivisa della realtà. L'apertura è molto importante perché favorisce l'evoluzione e la trasformazione culturale.

Invece, il social network ha una struttura molto diversa. Il social network ha un pubblico universale²⁷. Non è suddiviso in base alle linee ideologiche. Il pubblico italiano di Facebook o Instagram è potenzialmente costituito da tutti gli italiani: da chi legge La Stampa, La Repubblica, Il Corriere della Sera o qualsiasi altro giornale, da chi ha un'ideologia di destra, di sinistra o non ha un'ideologia

for achieving gender equality and the empowerment of all women and girls, Commission on the Status of Women, United Nations, 2023, 32. Ugualmente, l'Unione Europea nella sua Strategia per l'uguaglianza di genere 2020-2025, 5 marzo 2020. A tal proposito, viene indicato che «Women only represent 17% of people in ICT studies and careers in the EU and only 36% of STEM graduates, despite the fact that girls outperform boys in digital literacy», 10.

²⁷ F. BALAGUER CALLEJÓN, *La costituzione dell'algoritmo e le trasformazioni culturali nella società digitale*, in *lceonline*, 3, 2024.

precisa, è populista o ha qualsiasi altro orientamento ideologico. Questo crea un problema molto importante per le aziende tecnologiche. Come è possibile mantenere un pubblico universale quando il pubblico è diverso e c'è questa pluralità? La risposta è chiara: si fa con gli algoritmi. Gli algoritmi consentono di mantenere questa pluralità ideologica offrendo a ciascun utente opinioni e visioni del mondo compatibili con il suo profilo personale.

Loro utilizzano l'intelligenza artificiale in senso commerciale, per il modello di business che possiedono, per ottenere dati da utilizzare per la pubblicità. Facebook dipende praticamente al 100% dalla pubblicità come fonte di entrate²⁸. Ciò significa che i social devono fornire a ciascuno ciò che corrisponde ai propri gusti. Se hai un orientamento ideologico particolare e i social hanno il tuo profilo ideologico (grazie a dati, cronologia e attività su Internet), i social ti restituiscono la tua stessa visione del mondo. In questo senso, il social network ha una capacità di riproduzione culturale enorme, perché fornisce a ciascuno ciò che ciascuno pensa. Ovviamente, questo provoca anche la frammentazione e la radicalizzazione dello spazio pubblico.

La frammentazione significa che ognuno continua a pensare ciò che ha sempre pensato, perché i social network alimentano le posizioni ideologiche²⁹. All'interno della stessa applicazione, tutti gli utenti si trovano in una bolla³⁰ in cui non conoscono o non hanno interesse a conoscere le posizioni degli altri, perché ritengono che la propria sia corretta. E proprio perché i social network alimentano le loro posizioni, finiscono per ritenere che la loro sia l'unica verità. Questo ovviamente provoca la radicalizzazione, perché il processo comunicativo è strutturato in modo da favorirla.

Sia la riproduzione culturale che la radicalizzazione sono elementi molto disfunzionali per l'uguaglianza di genere. La riproduzione culturale ostacola la trasformazione culturale a favore dell'uguaglianza di genere. La radicalizzazione introduce un discorso irragionevole rispetto alla democrazia pluralista e al costituzionalismo, indebolendoli. Questo discorso è promosso dagli algoritmi dei social media perché attira la massima attenzione e interazione da parte degli utenti, consentendo alle aziende tecnologiche di ottenere più dati e, di conseguenza, generare maggiori entrate pubblicitarie. Il risultato è un'involuzione

²⁸ Il 98% dei suoi ingressi proviene dalla pubblicità, C. GALINDO, *Las grandes tecnológicas consolidan su liderazgo tras dos años de pandemia*, in *El País*, 5 febbraio 2022.

²⁹ Per quanto riguarda l'apparente eccezione a questo modello rappresentata dal social network X dal 2024 in poi, si veda F. BALAGUER CALLEJÓN, *La regolamentazione digitale europea nel nuovo contesto globale*, cit.

³⁰ E. PARISER, *The Filter Bubble. What the Internet Is Hiding from You*, Londra, 2011.

costituzionale e culturale che mette in difficoltà la promozione dell'uguaglianza di genere.

5. Conclusioni.

È possibile far risalire alcune idee fondamentali alla lotta contro la discriminazione di genere. In primo luogo, il suo legame con il costituzionalismo: maggiore è la densità costituzionale nella società, maggiori sono le possibilità di fare progressi a favore della parità di genere. Questa è una caratteristica molto importante del movimento a favore dell'uguaglianza di genere. Si tratta di un movimento costituzionale che ha bisogno della costituzione e che si sviluppa meglio nell'ecosistema costituzionale.

Il movimento per l'uguaglianza di genere ha ottenuto grandi successi sul piano economico, politico e sociale, ma non ha avuto lo stesso impatto sul piano della trasformazione culturale. La riproduzione culturale della disuguaglianza si manifesta in larga misura a livello subliminale, attraverso stereotipi e abitudini che non vengono messi in discussione perché non siamo consapevoli della loro esistenza. Comprendere questa capacità di ritardo culturale è fondamentale per promuovere l'uguaglianza di genere attraverso la trasformazione culturale. Infatti, quando si incontrano ostacoli sia per il costituzionalismo sia per la trasformazione culturale, ci si trova di fronte a delle sfide per l'uguaglianza di genere e per la lotta contro le discriminazioni di genere.

Abbiamo visto alcune delle sfide che il nuovo contesto generato dall'era digitale pone alla lotta contro la discriminazione di genere. Ci sono sfide che derivano dall'allontanamento dalla Costituzione e dalla nuova struttura di un mondo che sta cambiando e che non conosciamo bene perché l'era digitale non rappresenta una semplice transizione, ma una vera e propria rottura, come lo fu la Rivoluzione francese nel suo tempo. È in corso una trasformazione i cui contorni non conosciamo.

Oggi, però, la forza trasformatrice del mondo non è la costituzione, come avvenne durante la Rivoluzione francese. Al contrario, la costituzione viene considerata un ostacolo perché il mondo giuridico regola e la regolamentazione può ostacolare lo sviluppo della tecnologia. Oggi, al posto del costituzionalismo, c'è la tecnologia e, al posto della Costituzione, l'algoritmo. Queste sono le due nuove forze trasformatrici del mondo. L'intelligenza artificiale è una rivoluzione affidata all'algoritmo come strumento e alla tecnologia come ideologia e come cultura.

Poiché la globalizzazione e la società digitale orientano i diritti verso il

mercato, allontanandoli dai loro fondamenti costituzionali, la lotta contro la discriminazione di genere diventa più difficile, dal momento che la posizione delle donne nel mercato è più debole. C'è anche la sfida della riproduzione culturale. La pressione a favore della riproduzione culturale è molto forte sia nel disegno dell'intelligenza artificiale sia nella struttura dei social network. L'attuale configurazione dei social network favorisce la riproduzione culturale, così come produce una frammentazione e una radicalizzazione dello spazio pubblico. Tutti questi elementi ostacolano la lotta contro la disuguaglianza di genere.

PARTE I
DIRITTI COSTITUZIONALI, IDENTITÀ
E *BIAS* ALGORITMICI:
PROSPETTIVE GIURIDICHE NELL'ERA DELL'IA

CAPITOLO I

LE NUOVE FRONTIERE DELLE DISCRIMINAZIONI DI
GENERE: L'IMPATTO DELL'INTELLIGENZA ARTIFICIALE

Marilisa D'Amico

SOMMARIO: 1. Intelligenza artificiale e discriminazioni di genere. – 2. La discriminazione algoritmica e i suoi diversi fattori. – 3. Il linguaggio degli algoritmi e l'impatto discriminatorio. – 4. Lavoro femminile e IA. – 5. AI Act europeo e diritti delle donne. – 6. Considerazioni sul Trattato del Consiglio d'Europa. – 7. Una nuova sfida: l'IA americana e il “*free speech*” algoritmico. – 8. Postilla: la legge italiana e il riferimento specifico alla “parità di genere”.

1. Intelligenza artificiale e discriminazioni di genere.

L'intelligenza artificiale (IA) sta impattando su quasi ogni aspetto della vita delle persone e delle istituzioni, ponendo enormi interrogativi sui diritti fondamentali e sulla necessità di strumenti efficaci per governarne l'impatto. Come ogni rivoluzione tecnologica, quella dell'Intelligenza artificiale contiene enormi potenzialità, ma anche rischi innumerevoli per la garanzia dei diritti fondamentali. Infatti, grazie alla possibilità di analizzare enormi quantità di dati, prendere decisioni autonome e apprendere autonomamente, i sistemi di intelligenza artificiale stanno rapidamente penetrando in settori come la sanità, la mobilità, l'efficienza amministrativa e l'accessibilità dei servizi pubblici. Proprio queste caratteristiche possono comportare violazioni sistematiche o involontarie dei diritti umani, specialmente nel campo della privacy, della non discriminazione, della libertà di espressione e del diritto a un equo processo.

Così si sono già posti interrogativi sulla possibile manipolazione della decisione giudiziale che utilizzi algoritmi predittivi o sulla invasione della privacy creata da sistemi di sorveglianza. La natura spesso complessa dei modelli IA rende difficile identificare chiaramente le responsabilità in caso di violazioni di diritti, facendo emergere la necessità di nuovi approcci regolamentari e

giuridici, incentrati sulla trasparenza e la tracciabilità dei percorsi, in modo da poter mantenere, in qualsiasi momento, il pieno controllo umano.

Negli ultimi anni, la necessità di un approccio etico e rispettoso dei diritti fondamentali da parte degli algoritmi e delle ultime frontiere dell'IA, dai sistemi di LLM all'ultima frontiera dell'IA generativa, è stato al centro delle riflessioni e preoccupazioni delle moderne democrazie occidentali, con un approccio che si è negli ultimi tempi divaricato fra le scelte "regolatorie" europee e le decisioni trumpiane di totale indifferenza ai problemi di rispetto dei diritti umani e della trasparenza dei sistemi, come vedremo.

Nella nostra ricerca, ci siamo concentrati su un aspetto specifico, ma centrale e cioè del modo in cui l'IA possa diventare fortemente discriminatoria soprattutto nei confronti delle donne.

Il punto di partenza sono stati proprio gli studi sugli impatti degli algoritmi sulle persone, che hanno evidenziato la capacità fortemente selettiva e discriminatoria nei confronti delle donne, potenziata anche quando al profilo del genere viene sommato quello razziale.

Partiamo dalla vicenda diventata ormai famosa, nella quale una dottoranda all'MIT Media lab, Joy Buolamwini, scopre e denuncia gli algoritmi di riconoscimento facciale di società come IBM, Microsoft e altre grandi aziende, i quali risultavano molto meno accurati sui volti delle persone di pelle scura e ancora meno sulle donne, in generale, e sulle donne di pelle scura, in particolare, rispetto a quanto non lo fossero per gli uomini, in particolare bianchi. Questo progetto, nato con il nome di *Aspire Mirror*, una volta chiarito che i sistemi discriminavano in quanto allenati su maschi bianchi, cambia nome in "*Gender Shades*" e Buolamwini decide di fondare la *Algorithmic Justice League*, un gruppo che intendeva portare alla luce il fenomeno della discriminazione algoritmica, chiamata "*the Coded Gaze*", lo sguardo dell'algoritmo, espressione che intendeva fare riferimento al "*Male Gaze*", e quindi al sistema di cultura patriarcale e maschilista che influenza non soltanto il modo di rappresentare le donne, ma la stessa rappresentazione della realtà.

Il percorso di Buolamwini è influenzato anche da Cathy O'Neil, una matematica e attivista, che nel suo volume *Weapons of Math Destruction* aveva denunciato come i sistemi algoritmici di modelli matematici per orientare alcune decisioni future non fossero per nulla neutrali, ma, al contrario, incorporando dati storici esattamente come i sistemi di database di riconoscimento facciale, fossero spesso sbilanciati, anche inconsapevolmente, in favore dei gruppi sociali più privilegiati.

Il mondo dell'IA, non sottoposto a regole, riproduce in modo veloce e incontrollato quel "monopolio maschile" che fin dalla storia antica rappresenta

un ostacolo culturale fortissimo alla costruzione di una democrazia paritaria e del pieno rispetto del principio di uguaglianza.

2. *La discriminazione algoritmica e i suoi diversi fattori.*

Diversi sono i motivi della discriminazione algoritmica, che ci rendono pienamente consapevoli che gli algoritmi non siano affatto “neutri” e che l’intelligenza artificiale, riproducendo quella “umana”, ne incorpora le profonde differenze esistenti, differenze e disuguaglianze che purtroppo discriminano prima di tutto le donne.

In primo luogo, il “sesso” di chi programma, essendo i programmatori in moltissima parte uomini; in secondo luogo, l’incorporazione di *bias* esistenti nella società, ancora fortemente, anche se spesso a livello inconscio, maschilisti; da ultimo, il modo di funzionamento degli algoritmi, che, raggruppando e selezionando, producono impatti ancora più discriminatori e pervasivi di quelli dei dati che hanno assorbito e che incorporavano pregiudizi.

Il primo aspetto, ben presente ormai anche a livello istituzionale, è costituito dalla prevalenza di programmatori uomini, che riproducono la visione maschile anche nel mondo dell’intelligenza artificiale.

I dati evidenziano come il *gender gap* coinvolga anche i settori più innovativi e all’avanguardia in tema di tecnologia e intelligenza artificiale, in cui attualmente solo il 22% dei professionisti su base mondiale è di sesso femminile contro il restante 78% maschile, mentre alla luce dei dati del *Global gender gap* del 2025, in generale la situazione va migliorando e si attesta a un 68% per gli uomini e a un 32% per le donne.

Va anche evidenziato che le donne rappresentino solo il 18% nei principali convegni scientifici in tema di IA e il 20% dei docenti di IA, come che femminile sia soltanto il 15% e il 10% del personale di ricerca dei colossi informatici come Facebook e Google, principali sviluppatori oggi delle ricerche innovative fuori dalle aule universitarie.

Nelle Università, comunque, alcuni settori scientifici, come in particolare Ingegneria, sono ancora dominati dagli uomini: in Italia, solo il 27% dei laureati ingegneri è donna, un dato che presenta comunque miglioramenti rispetto al 23% del 2020, e anche in Paesi come gli Stati Uniti nell’ultimo decennio si è avuta una flessione del 10% delle lauree STEM da parte delle donne. Si tratta infatti di una palese “segregazione orizzontale”, che si verifica quando in un’area vediamo una prevalenza di uno dei due sessi. Sul problema si ipotizzano cause

diverse, a partire da aspetti psicologici e dall'esistenza di stereotipi che pur conosciuti non sono ancora oggetto di politiche pubbliche efficaci.

Ancor oggi il modello di esperto informatico viene rappresentato come un individuo isolato, con scarse capacità relazionali e pochi contatti umani, provocando un allontanamento delle donne dalla scelta di questo tipo di studi, meno attrattivi per una popolazione femminile più incline a prediligere lavori in team o nei quali l'impatto relazionale ed emotivo sia presente. Su questa rappresentazione pesa ancora lo stereotipo che, a partire dal mondo greco, ritiene le donne meno adatte a calcoli o a dimensioni matematiche rispetto agli uomini.

Tale situazione condiziona e condizionerà in futuro, in negativo, la stessa situazione lavorativa delle donne, alla luce della circostanza che oggi come in futuro il possesso di abilità informatiche e tecnologiche consente non solo di avere maggiori possibilità di accesso, ma soprattutto di ambire a posizioni lavorative migliori e meglio retribuite.

La circostanza che a concepire gli algoritmi siano quasi esclusivamente uomini porta quindi con sé il concreto rischio che essi immagazzinino stereotipi di genere e che tali stereotipi si riproducano e assumano dimensioni più ampie nel momento stesso in cui vengano utilizzati.

D'altronde i dati del *Global Gender Gap Report*, pubblicato ogni anno dal *World Economic Forum* e basato sul confronto fra 149 Paesi in relazione ai progressi sulla parità di genere certificano questa situazione preoccupante in modo inequivocabile. I pregiudizi implementati da algoritmi costruiti da uomini stanno già discriminando le donne, ad esempio, sul reclutamento dei candidati per determinate posizioni di lavoro. Questo è accaduto, ad esempio, nel caso dell'algoritmo che Amazon ha dovuto abbandonare, proprio perché favoriva sistematicamente candidati uomini nella selezione dei *curricula* in quanto era stato "allenato" e programmato tramite l'immissione di *curricula* prevalentemente maschili; lo stesso problema si è verificato anche nel campo del c.d. *platform work*, come nel caso di Uber.

Anche la piattaforma nota come LinkedIn ha dovuto fronteggiare problemi di "discriminazione automatizzata" quando le offerte lavorative per profili altamente retribuiti venivano visualizzate maggiormente nelle ricerche effettuate da utenti di sesso maschile a discapito di quelli di sesso femminile; lo stesso vale per le piattaforme di assistenza virtuale come Siri sono state tacciate di sessismo per l'associazione tra femminilità e servilismo che contribuisce a rafforzare gli stereotipi di genere, come evidenziato dal *World Economic Forum*.

Come abbiamo appena visto, risulta evidente che algoritmi costruiti da uomini, che introducono elementi o dati di per sé discriminatori, si presentino a loro volta discriminatori.

L'intelligenza artificiale presenta, infatti, un potenziale e rilevante problema di disegualianza di genere. Gli algoritmi funzionano, infatti, aggregando i dati che l'uomo fornisce loro: se tali dati risultano condizionati da squilibri o pregiudizi è inevitabile che la macchina discrimini determinate categorie.

Sui rischi, in termini di lesione dei diritti fondamentali, derivanti dall'utilizzo di algoritmi predittivi in ambito giudiziario, con particolare riferimento alla materia penale, emblematico fu negli Stati Uniti il c.d. "caso *Loomis*", del 2016, inquadrabile, probabilmente, come *leading case*. In quella circostanza l'imputato era stato infatti condannato a sei anni di reclusione dal Tribunale del Wisconsin sul mero risultato fornito da un algoritmo predittivo denominato *Compas*, concepito per valutare il rischio di recidiva di un soggetto e di cui tuttavia non era dato conoscere il funzionamento. L'imputato, sostenendo che fosse stato leso il proprio diritto ad un equo processo, decise così di ricorrere alla Corte Suprema del Wisconsin, che tuttavia rigettò il ricorso, affermando, pur con alcune cautele, come fosse legittimo l'utilizzo dell'algoritmo menzionato nel processo di autodeterminazione del giudice.

Al contrario, nel caso *K.W. v. Armstrong*, del 2023, la Corte distrettuale dell'Idaho ha sottolineato la necessità di garantire, in corrispondenza al canone – fondamentale – del giusto processo, l'accesso alle informazioni relative alla formula algoritmica che governava l'erogazione dell'assistenza sanitaria per persone con disabilità. Allo stesso modo, nel più risalente caso *Houston Federation of Teachers, Local 2415 v. Houston Independent School District*, del 2017, i giudici hanno affermato che l'utilizzo di algoritmi segreti per l'adozione di decisioni in merito all'assunzione di decisioni gravanti sulla carriera degli insegnanti da parte della *Houston Independent School District* contrastava con i principi del giusto processo. Diverso è il caso *People v. Wakefield*, del 2022, dove è stato negato alla difesa di accedere alle formule algoritmiche utilizzate per l'interpretazione del DNA.

La riproduzione degli stereotipi può avvenire non solo con l'immissione "intenzionale" di informazioni affette da pregiudizi, ma anche, in via indiretta e spesso inconsapevole, tramite informazioni a prima vista "neutre", che, però, in un secondo momento, rielaborate dalla macchina stessa tramite meccanismi di *machine learning* e *deep learning*, producono un effetto discriminatorio "in uscita" sulla base del genere.

Moltissimi studi dimostrano, ad esempio, come gli stereotipi si annidino anche già nel solo nel linguaggio e che ciò ha notevoli implicazioni sul funzionamento degli algoritmi, i quali, avendo associato nel corso dell'apprendimento di una lingua un determinato genere a determinate caratteristiche, tenderanno a perpetuare tale assimilazione al momento della loro utilizzazione pratica, con effetti potenzialmente discriminatori.

Le ragioni che determinano il funzionamento discriminatorio delle tecniche di intelligenza artificiale sono da ricondurre al ruolo assolto dai dati.

L'intelligenza artificiale discrimina, infatti, perché si fonda e si alimenta di dati, talvolta non completi, volontariamente oppure involontariamente parziali, errati; dati che omettono di riflettere la realtà sulla quale la tecnologia è destinata a riverberare i suoi effetti.

Le conseguenze deteriori dell'intelligenza artificiale si sono dimostrate, però, particolarmente evidenti sul versante specifico della violazione del principio di eguaglianza di genere. Questo aspetto è emerso in modo chiarissimo di fronte all'analisi del funzionamento dei sistemi di riconoscimento facciale, incapaci di riconoscere volti di donna, ma particolarmente accurati rispetto a volti di uomini, specie se bianchi. I dati, in questo caso, si rivelano parziali perché sono forniti alle tecnologie di intelligenza artificiale da programmatori uomini, che omettono di assicurare la costruzione di un *data-set* completo con effetti deteriori sul funzionamento, discriminatorio, degli algoritmi.

Non si tratta, però, soltanto di una discriminazione che deriva dal funzionamento delle tecnologie di intelligenza artificiale perché i dati sono, come detto, non adeguatamente rappresentativi oppure affetti da vizi e parziali. In altre parole, la discriminazione non nasce solo nel momento in cui i programmatori individuano e, poi, selezionano i dati da fornire ai sistemi di intelligenza artificiale. A questa origine o causa, se ne affiancano altre che ripetono le fasi in cui si costruisce l'algoritmo e quelle, particolarmente complesse, in cui se ne articola il funzionamento successivo. In tutti questi "momenti" o passaggi, le tecnologie di intelligenza artificiale hanno dimostrato una attitudine più o meno elevata a produrre effetti discriminatori, proprio perché tendenti a selezionare e a raggruppare, e questa operazione meccanica, se compiuta su un gruppo di dati selezionato in origine in modo differente e quindi spesso discriminatorio, conduce la macchina a moltiplicarne a dismisura gli effetti.

Va anche sottolineato che la ricerca del "dove" si annida il rischio dell'effetto discriminatorio, del "quando" esso può verificarsi e del "come" quest'ultimo si realizza, si rivela particolarmente complessa per tecnologie di intelligenza artificiale di largo impiego e il cui funzionamento diviene autonomo sino ad essere, talvolta, addirittura sconosciuto al programmatore iniziale. In questi casi, la tecnologia agisce senza che chi ne abbia costruito inizialmente l'algoritmo sia in grado di prevedere gli effetti concreti derivanti dal suo funzionamento su larga scala. Ci si riferisce alla c.d. *black box theory* che evoca l'idea per la quale, una volta programmato, istruito, messo in funzione, l'algoritmo sarebbe capace di agire secondo modalità che divengono progressivamente ignote al programmatore. L'"apprendimento" che pure la persona impartisce alla tecnica per

assicurare un funzionamento corrispondente alle finalità assumerebbe, in altre parole, una connotazione progressivamente distante ed imprevedibile, rendendo oscuri i passaggi che consentono a quella tecnica di raggiungere un risultato specifico. Detto altrimenti, il funzionamento delle tecnologie di intelligenza artificiale non sarebbe conoscibile e ricostruibile *ex post*, nemmeno da chi le abbia programmate, essendo viceversa ignoto come questi sistemi producono gli effetti originariamente voluti e sempre che questi ultimi corrispondano poi effettivamente a ciò che era stato originariamente desiderato.

Alcuni sistemi di intelligenza artificiale sono, inoltre, “chiusi” rispetto ad interventi successivi esterni, cioè ad azioni del programmatore, precludendo correzioni al malfunzionamento delle tecnologie ed ostacolando comportamenti distorsivi e discriminatori. Ed è per queste ragioni che l’aggiunta di dati ad un *data-set* iniziale oppure la previsione del controllo umano, il c.d. *human in the loop*, non è sempre sufficiente a correggere il funzionamento del sistema, escludendo discriminazioni derivanti dall’intelligenza artificiale. Ciò, soprattutto, come detto, perché quando non si conosce esattamente come il sistema raggiunge l’obiettivo desiderato, diventerà praticamente impossibile adottare comportamenti che rimedino al malfunzionamento del sistema. Tanto più, quindi, l’intelligenza artificiale agisce in modo autonomo rispetto alla programmazione della persona e tanto più le fasi entro cui si snoda il suo funzionamento è oscuro, tanto più l’individuazione del rischio che si producano effetti discriminatori diviene importante, di complessa prevenzione, di difficile contrasto.

3. Il linguaggio degli algoritmi e l’impatto discriminatorio.

Come è noto, il problema del linguaggio è molto importante per il suo impatto sui diritti. Il linguaggio non solo esprime la realtà, ma la può anche trasformare e condizionare, per la sua natura performativa. In Italia, negli anni recenti, a livello istituzionale si è molto discusso, soprattutto per l’iniziativa del dizionario Treccani, che ha adottato un linguaggio c.d. binario, rigettando l’universale maschile come scelta necessaria per la lingua italiana. La natura performativa del linguaggio, poi, può incidere molto negativamente sulla realtà, soprattutto per ciò che riguarda il c.d. hate speech, specie se utilizzato con strumenti informatici, in quanto potenzialmente più lesivi a causa della loro “virilità”.

Rispetto a tali tematiche, il campo di indagine rispetto alle tecnologie di IA più interessante è senza dubbio quello dei sistemi di c.d. *natural language processing* (NLP).

Il *natural language processing* si occupa di studiare e di costruire meccanismi di intelligenza artificiale in grado di permettere che i computers comprendano testi e parole, replicando abilità umane.

Si tratta, quindi, di algoritmi che si vorrebbero in grado di analizzare, rappresentare e comprendere il linguaggio naturale “delle persone”, sia esso parlato oppure scritto. Come segnalato anche nel Report UNESCO del 2024, dal titolo *Challenging systematic prejudices: an investigation into bias against women and girls in large language models* emerge, allora, in questa prospettiva, il problema dell'effetto che i pregiudizi e gli stereotipi che caratterizzano il linguaggio, scritto e parlato, riflettono sui meccanismi di intelligenza artificiale, tra cui non fanno eccezione i citati sistemi di *natural language processing*.

Un esempio delle ripercussioni discriminatorie che caratterizzano questi peculiari sistemi di intelligenza artificiale è costituito dagli strumenti di traduzione automatica.

Cercando di adattare un testo redatto in una lingua neutra sotto il profilo del genere (si pensi al caso dell'inglese), ad una lingua connotata, viceversa, dalla esistenza della doppia declinazione maschile/femminile (l'italiano oppure il francese e, ancora più complesso, il tedesco per la presenza del neutro), tali sistemi assai di frequente tendono ad oscurare la dimensione di genere per favorire l'impiego del già citato c.d. maschile universale, altrimenti definito “maschile neutro”. Il “maschile neutro”, cioè, viene considerato idoneo a riflettere le differenze di genere, che sono ritenute superflue da riprodurre in sede di traduzione del testo. Si tratta di una semplificazione linguistica che, però, come già detto, nega la rappresentazione della realtà, fatta di uomini e di donne.

Il fenomeno, peraltro, riproduce anche quanto è in tutto identico avviene ancor oggi, almeno in Italia, nel linguaggio comune dove nella maggior parte dei casi si ricorre al maschile poiché ritenuto presuntivamente idoneo a “coprire” anche il femminile.

Nel simulare il linguaggio comune o non mediato artificialmente, gli strumenti di *natural language processing*, come sono appunto i *softwares* di traduzione automatica, riflettono così i *bias* e, sarebbe meglio dire, gli “errori” del linguaggio ordinario.

I sistemi di *natural language processing* ripropongono gli usi della lingua, discriminatori o sessisti di cui si è detto e, tuttavia, possono anche essere utilmente impiegati anche per finalità di tipo diverso. Si tratta, cioè, di un diverso e opposto utilizzo di tali sistemi allo scopo di promuovere e diffondere un linguaggio che si dimostri, viceversa, sensibile alle differenze di genere. Sul punto, un esempio significativo è offerto da Google, che, nel 2017, ha scelto di ridefinire il proprio programma di traduzione automatica allo scopo di inserire una

gamma di traduzioni alternative capaci di rispondere alle specificità di genere.

L'esempio di Google Translate è molto interessante perché dimostra che, laddove sussista una volontà umana in questo senso, l'intelligenza artificiale è in grado di passare dall'essere un mezzo per discriminare al rappresentare uno strumento, viceversa, capace di contrastare la diffusione e l'irrobustimento di stereotipi. In questo senso, un utilizzo consapevole e trasparente dell'intelligenza artificiale, rispettoso dei principi costituzionali, potrebbe permettere di invertire la rotta favorendo un impiego della innovazione tecnologica che non sia "nemica" dei diritti delle persone, ma sappia invece contribuire alla loro tutela.

La doppia natura dell'intelligenza artificiale – da un lato, mezzo (in)consapevole di discriminazione e, dall'altro, potenziale alleato per assicurarne il contrasto – traspare in modo evidente se si considera il fenomeno del c.d. odio online.

Se l'intelligenza artificiale assolve ad un ruolo importante nella creazione e nella diffusione dell'odio online, al tempo stesso essa può, però, atteggiarsi anche a strumento efficace per il suo contrasto agendo in modo analogo a quanto già osservato trattando dei sistemi di traduzione automatica.

Se parliamo di odio online, d'altra parte, occorre sempre considerare che le prime vittime sono proprio le donne.

È, infatti, innegabile che il web, lasciato nella piena disponibilità degli utenti e sprovvisto di una regolamentazione, costituisca un terreno fertile per la genesi di comportamenti discriminatori, che assumono la forma dei c.d. discorsi d'odio. La circostanza che il web sia luogo facilmente dominabile dai c.d. "odiatore" (*haters*) non coincide con l'esistenza di una responsabilità dell'intelligenza artificiale quale creatrice e promotrice dell'odio online, se non nella misura in cui essa si sottragga all'apprestamento di argini sufficienti alla diffusione di contenuti discriminatori.

È noto che anche l'intelligenza artificiale contribuisce fortemente al fenomeno dell'odio online, a causa della c.d. indicizzazione dei contenuti impiegati sul web: le piattaforme online e, in particolare, i social networks funzionano, infatti, sulla base del meccanismo della c.d. profilazione attraverso cui vengono monitorati e analizzati i comportamenti degli utenti al fine di prevederne comportamenti, preferenze e scelte, sì da fornire contenuti personalizzati e, nelle intenzioni, corrispondenti ai desideri, ai gusti di coloro che fruiscono delle piattaforme.

Si è infatti dimostrato che contenuti viziati da sentimenti d'odio sono messi a disposizione in modo proporzionalmente maggiore in favore di chi si è già dimostrato ad essere più incline a ricorrere a contenuti discriminatori, favorendo la creazione di quello che la dottrina chiama la c.d. "camera dell'eco",

cioè l'amplificazione e il rafforzamento dell'opinione che si forma all'interno di un gruppo con il successivo orientamento dello stesso verso posizioni progressivamente più estreme e radicalizzate.

Secondo questo meccanismo, un contenuto discriminatorio oppure un fenomeno d'odio può essere amplificato da algoritmi, volutamente programmati per assicurare che i contenuti mostrati all'utente siano in linea con le sue preferenze.

A questo, si affianca un secondo aspetto meritevole di considerazione.

Come già nella realtà, più un contenuto è "d'impatto" e dirompente, tanto più tende a suscitare interesse. Ne discende che, in tanto gli algoritmi sono programmati per dare massima diffusione a ciò che presenta più interesse, nel momento in cui il contenuto che suscita più interazioni tra gli utenti è espressione di odio e di intolleranza, se la piattaforma non vi pone argini, sarà questo stesso contenuto ad essere diffuso fra gli utenti grazie all'azione dei sistemi di intelligenza artificiale

Si spiegano, così, fenomeni purtroppo all'ordine del giorno come il c.d. *revenge porn*, lo *stalking* digitale e, più in generale, il sessismo e la discriminazione che vengono amplificati dalla rete e destinati a diventare virali in poche ore. La "viralità" è, del resto, caratteristica intrinseca alla rete e obiettivo a cui contestualmente ambiscono le piattaforme. Quando però virale è un contenuto negativo e di odio, occorre immaginare antidoti efficaci alla diffusione di tali contenuti: per questo è importante ragionare sulla regolamentazione e sulla moderazione dei contenuti disponibili sul web, una condizione necessaria perché la rete e, dunque, l'intelligenza artificiale che la governa, passi dall'essere solo creatrice di odio oppure, anche, un'alleata efficace per il suo contrasto.

Come osservato, se non si interviene su algoritmi "viziati", questi agiranno in modo automatico e ne diffonderanno i contenuti, anche, quelli negativi: le tecnologie di intelligenza artificiale non sono però soltanto al servizio dell'odio online, dal momento che esistono e possono essere agilmente sviluppate modalità e strategie che consentono all'intelligenza artificiale di arginare la diffusione dell'odio sul web.

L'intelligenza artificiale può, così, diventare un valido alleato nella lotta alla diffusione dell'*hate speech online* soprattutto mediante l'utilizzo di strumenti di moderazione automatizzata, che si stanno progressivamente sostituendo alle attività prima assolute in via esclusiva dai moderatori umani.

Attualmente, le piattaforme online dispongono di una pluralità di strumenti algoritmici attraverso cui filtrare le c.d. *bad information* diffuse online, siano esse costituite da immagini, video oppure testi. Il filtraggio dei contenuti negativi o discriminatori avviene attraverso le già più volte citate tecniche di *machine*

learning, ossia mediante la programmazione di algoritmi capaci di riconoscere i contenuti lesivi per poi, successivamente, limitarli oppure, nei casi più gravi, eliminarli.

Nonostante la necessità di utilizzare sistemi automatizzati per arginare efficacemente la diffusione di odio on line, occorre considerare che tale manipolazione contiene un rischio, molto rilevante per i valori delle democrazie occidentali, che si reggono sulla libertà di manifestazione del pensiero e del rifiuto di qualsiasi forma di censura: il rischio infatti è quello di colpire in modo indiscriminato vere espressioni o pensieri di odio, insieme invece a espressioni o contenuti che magari sono soltanto di critica, limitando quindi e standardizzando il pensiero delle persone e il dibattito pubblico.

Gli strumenti di moderazione automatica rischiano, quindi, da una parte, di essere troppo permissivi e di non raggiungere l'obiettivo desiderato, cioè la tutela dei diritti degli utenti; dall'altra, di essere troppo stringenti, limitando il diritto alla libertà di espressione anche di chi non commette alcun tipo di abuso.

Una garanzia in questa direzione può certamente derivare dalla possibilità di conoscere a priori i criteri sulla base dei quali l'intelligenza artificiale opera quando monitora i contenuti diffusi sulle piattaforme online a cui dovrebbe affiancarsi la capacità di conservare il controllo sul funzionamento successivo dell'algoritmo. Ad oggi, i sistemi di moderazione, così come più in generale i sistemi di intelligenza artificiale, non sempre rispondono al principio di trasparenza, anche poiché gli algoritmi sono spesso coperti da garanzie di proprietà intellettuale. La non trasparenza del funzionamento dei sistemi di intelligenza artificiale ne rende, così, particolarmente arduo un controllo esterno impedendo agli utenti di conoscere in ultima istanza come opera la piattaforma anche per quanto attiene alla moderazione dei suoi contenuti.

In considerazione della difficoltà di operare un controllo diretto sugli algoritmi, una seconda strategia di intervento in grado di bilanciare diritti e libertà potrebbe essere costituita da un intervento *ex ante*, individuando a monte i principi a cui i programmatori e le programmatrici devono obbligatoriamente conformare le rispettive mansioni.

Proprio per rispondere a queste problematiche, rese evidenti da una serie di attentati terroristici alimentati anche dall'odio sul web, nel 2016 la Commissione europea, insieme con alcune delle più influenti aziende informatiche, ha adottato il Codice di condotta per lottare contro le forme illegali di incitamento all'odio online.

Nel Codice di condotta vengono responsabilizzate le piattaforme online sia con riguardo all'eliminazione di contenuti lesivi dei diritti fondamentali, sia quanto alla opportunità di istruire gli utenti in ordine alle conseguenze legate

alla diffusione di contenuti discriminatori e potenzialmente pericolosi. Il Codice di condotta mira così ad agire su un duplice fronte: la rimozione coattiva dei contenuti e l'“educazione” degli utenti mediante il coinvolgimento attivo dei *providers* digitali.

In questa prospettiva, è anche determinante il ruolo degli utenti, che possono a loro volta segnalare contenuti potenzialmente discriminatori, attivando strumenti di moderazione delle piattaforme.

Questa modalità di contrasto dell'odio online, poggiando sul ruolo assolto dalle persone che si interfacciano con le piattaforme digitali, non fa che dimostrare ancora una volta come la via per il contrasto dell'odio online non possa che passare da una alleanza fra la persona e i sistemi di intelligenza artificiale che si riveli attenta garante dei diritti costituzionali degli utenti del web.

Il 25 gennaio 2025 il codice di condotta riveduto è stato integrato nel quadro normativo della legge sui servizi digitali e ad esso hanno aderito numerose aziende informatiche, come Facebook, Microsoft, Youtube, Instagram.

4. *Lavoro femminile e IA.*

Con l'IA il mercato del lavoro si sta trasformando profondamente, consentendo da una parte modalità organizzative più agili e positive – come lo smart working – per chi, come soprattutto le donne, si trova a conciliare i tempi del lavoro con i tempi della cura familiare, dall'altra, introducendo nuove figure di lavoratori, con competenze tecniche molto richieste e potenzialmente avvantaggiate.

Questa trasformazione, in ogni caso, presenta anche molti rischi, soprattutto per le donne che sono a livello mondiale meno occupate, più dedite a lavori esecutivi e meno competenti, come abbiamo detto, a livello tecnologico e informatico.

Alcuni impieghi che richiedevano mansioni esecutive sono stati depotenziati o eliminati, colpendo in modo rilevante le categorie di persone in età più avanzata, difficili da istruire sulle nuove competenze digitali richieste.

Ancora una volta, l'impatto sulle donne è potenzialmente molto più forte: secondo un Rapporto del 2025 dell'Organizzazione Internazionale del Lavoro, nei Paesi ad alto reddito l'IA mette a rischio il 9,6 % dei posti di lavoro occupati da donne, rispetto al 3,5% di quelli occupati da uomini. Il dato è impressionante e non fa altro che confermare quanto gli esperti avevano già previsto, alla luce delle caratteristiche del lavoro femminile, meno tecnologico e, in generale, più precario.

La trasformazione digitale costituisce dunque una profonda sfida per la nostra società e, soprattutto, per le donne, e il pericolo che nelle grandi trasformazioni epocali queste ultime rimangano ancora in secondo piano, che vengano escluse, o addirittura, che perdano in modo velocissimo e irreversibile parte delle conquiste raggiunte nel giro di poco più di un secolo, è troppo grande per non affrontarlo con consapevolezza e incisività.

Di questo, sul piano mondiale ci si sta rendendo conto e, per fortuna, da anni ci sono state proposte sul piano generale, come quelle evidenziate in un documento risalente, ma ancora molto attuale: un report intitolato *Artificial intelligence: open questions about gender inclusion*, Policy Brief – W20 Argentina 2018, nel quale si incoraggiavano i governi di tutti gli Stati a coinvolgere le donne nella costruzione degli algoritmi, con programmi specifici di formazione, adottando linee guida che ne favorissero la trasparenza e la tracciabilità, con attenzione anche agli impatti presenti e futuri della trasformazione digitale sul mondo del lavoro.

Di tutte queste problematiche è stata consapevole soprattutto l'Unione Europea, nella quale fin dal 2018, è stata adottata una strategia per lo sviluppo di un'intelligenza artificiale antropocentrica, seguita da un piano coordinato in materia, da linee guida sull'etica nella materia, fino all'adozione di un vero regolamento basato sul criterio del rischio algoritmico e sulla verifica dell'impatto sui diritti fondamentali nell'ambito dei settori più rilevanti nella vita delle persone, come quello della sicurezza, della giustizia, della sanità.

5. *AI Act europeo e diritti delle donne.*

Con un percorso complesso, l'Unione Europea ha creato un quadro normativo chiaro e uniforme per l'utilizzo dell'intelligenza artificiale, ispirato a principi per garantire la sicurezza, la trasparenza e il rispetto dei diritti fondamentali. Il "Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i Regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le Direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828", noto come Regolamento sull'intelligenza artificiale o "Artificial Intelligence Act", è entrato in vigore il 1° agosto 2024, stabilendo una graduale entrata in vigore dei differenti articoli: l'art. 5 sui sistemi vietati è entrato in vigore il 1 febbraio del 2025, mentre le disposizioni relative ai sistemi a rischio elevato dovrebbero diventare pienamente applicabili dal 2 agosto 2026, al termine del periodo di adeguamento biennale.

Un aspetto molto discusso dal Regolamento è stato proprio quello dell'impatto che i sistemi di IA possono avere sui diritti di genere, proprio per evitare quei *bias* e quelle discriminazioni che colpiscono soprattutto le donne.

Nel Preambolo e nel testo dell'art. 27 sui sistemi "ad alto rischio" si fa riferimento alla *compliance* rispetto ai diritti fondamentali, quali contenuti anche nelle fonti europee e, in particolare, nei Trattati e nella Carta dei diritti fondamentali: la parità di genere costituisce un pilastro dell'Unione, tanto da essere oggetto non solo di specifici regolamenti, direttive e atti, ma anche di una intensa giurisprudenza della Corte di Giustizia, che vede una dimensione molto significativa ed estesa del principio di "non discriminazione di genere", applicato in ambiti vasti e non limitati al diritto del lavoro.

Oltre a questo aspetto, va sottolineato che lo stesso art. 5 sui sistemi vietati costituisce una formulazione generale di aspetti discriminatori soprattutto nei confronti delle donne: ci riferiamo alla parte che riguarda il divieto dei sistemi di profilazione, che, come è noto, erano stati utilizzati soprattutto per la selezione del personale e hanno costituito casi di evidente discriminazione di genere.

6. *Considerazioni sul Trattato del Consiglio d'Europa.*

Nel panorama internazionale troviamo anche un altro importante fondamentale documento regolatorio: si tratta del primo Trattato al mondo sul tema, approvato dal Consiglio d'Europa, frutto del lavoro di oltre due anni di un comitato intergovernativo che ha coinvolto attivamente i 46 Stati membri del Consiglio d'Europa, l'Unione Europea e altre 11 nazioni non appartenenti all'organizzazione (Argentina, Australia, Canada, Costa Rica, Santa Sede, Israele, Giappone, Messico, Perù, Stati Uniti d'America e Uruguay).

Il Trattato, aperto anche alla firma di nazioni non europee, delinea un quadro normativo completo che abbraccia l'intero ciclo di vita dei sistemi di IA. Non solo ne definisce i criteri di progettazione, sviluppo e dismissione, ma affronta anche i potenziali rischi connessi, promuovendo al contempo un'innovazione responsabile.

Emerge, fin dal Preambolo, una chiara consapevolezza delle opportunità offerte dai sistemi di intelligenza artificiale, per lo sviluppo del benessere dell'umanità, e anche dei diritti fondamentali, ma al tempo stesso dei rischi che essi possono arrecare alla stessa dignità umana, ai diritti, alla democrazia e allo stesso Stato di diritto.

Nel Preambolo viene evidenziata la necessità di evitare che gli stessi sistemi

di intelligenza artificiale possano essere utilizzati a fini repressivi, attraverso arbitrarie e illegittime pratiche di sorveglianza o di censura, lesive del diritto di privacy individuale.

Il Trattato individua un quadro normativo completo che abbraccia l'intero ciclo di vita dei sistemi di IA: non solo ne definisce i criteri di progettazione, sviluppo e dismissione, ma affronta anche i potenziali rischi connessi, promuovendo al contempo un'innovazione responsabile.

Da una prima analisi, sembra che la Convenzione adotti un approccio analogo all'AI Act in termini di regolamentazione delle tecnologie di IA in base al rischio e con riferimento all'opportunità di identificare in chiave preventiva i rischi associati alle tecnologie di IA e, successivamente, di identificare strategie di mitigazione e contenimento dei danni derivanti dal ricorso all'IA.

Rispetto però al regolamento europeo, la Convenzione insiste nella necessità di una piena padronanza dei sistemi di intelligenza artificiale allo scopo (art. 3) di garantire il rispetto dei «diritti umani, democrazia e Stato di diritto». Da ciò, l'ampiezza del testo, che si rivolge sia alle autorità pubbliche, che ai soggetti privati che li utilizzano.

Un intero capitolo (cap. III) è dedicato ai principi che dovrebbero governare il ricorso alle tecnologie di IA: l'impostazione è sicuramente simile a quella ricavabile dall'AI Act, ma, come si è già detto, si riscontra una più esplicita attenzione ai diritti umani soggetti ad essere negativamente pregiudicati dalle tecniche di IA.

Fra i principi fondamentali della Convenzione troviamo, come nell'AI Act europeo, la trasparenza e la supervisione. Il testo definisce requisiti specifici in base ai contesti e ai rischi potenziali, come la necessità di una identificazione dei contenuti generati dai sistemi di IA, aspetto importante per la lotta ai *deep fake*.

Inoltre, il Trattato prevede l'obbligo per gli Stati ratificanti di adottare misure per identificare e valutare i rischi (chiedendo un'analisi preventiva per individuare i potenziali pericoli associati all'utilizzo dei sistemi di IA); prevenire e mitigare i rischi (esigendo l'implementazione di strategie per ridurre al minimo le conseguenze negative); garantire responsabilità e obblighi di rendicontazione (individuando i responsabili in caso di impatti negativi e l'adozione di misure per prevenirli); promuovere l'equità e la non discriminazione (dovendo essere assicurato da parte dei sistemi di IA il rispetto dei principi di uguaglianza, inclusa la parità di genere, e il divieto di discriminazione); proteggere la privacy (un aspetto che ricorre e sempre collegato anche al rischio di abuso da parte dei poteri pubblici).

Va infine sottolineato che in questo documento, ancor più che nell'AI Act

dell'Unione, è contenuto un riferimento esplicito alla parità di genere, e non soltanto al principio di uguaglianza.

7. Una nuova sfida: l'IA americana e il "free speech" algoritmico.

Il 23 luglio 2025, la Casa Bianca ha pubblicato un nuovo documento strategico intitolato *America's AI Action Plan*. Si tratta di un testo lungo, articolato e denso di proposte operative il quale, prima ancora che per i contenuti, colpisce per il linguaggio. Sin dalle prime righe si capisce che non si tratta di una semplice *roadmap* tecnologica ma di una dichiarazione di principio, frutto di una precisa scelta politica ideologica e identitaria.

Il Governo americano dichiara, infatti, che l'intelligenza artificiale è lo strumento attraverso cui gli Stati Uniti devono riaffermare il proprio «dominio tecnologico incondizionato» sul piano globale. Il lessico è muscolare, competitivo, quasi bellico, si parla di corsa, supremazia, infrastrutture da costruire a ogni costo e in tutto il discorso spicca un'espressione ricorrente: "American values".

Opposta la visione rispetto a quella adottata fino a quel momento negli Stati Uniti, dove il centro della riflessione era occupato, come in Europa, da principi universali come i diritti umani, la non discriminazione, la protezione dei soggetti vulnerabili: ora il baricentro si sposta sui "valori americani" e sulla "libertà di parola", da intendere però come opposizione a ogni forma di moderazione, filtraggio o ingegneria sociale. In modo esplicito, il piano chiede di rimuovere dalle linee guida federali ogni riferimento a disinformazione, inclusione, cambiamento climatico.

Come per altri Decreti presidenziali aventi ad oggetto vere e proprie inversioni di tendenza rispetto a temi come il genere, la *diversity*, il *climate change*, la strategia sull'IA utilizza la leva economica.

Il piano prevede, infatti, che gli stati federati che adottano regolazioni sull'IA considerate troppo "onerose" possano essere esclusi dai finanziamenti pubblici. Una misura che suona come un chiaro avvertimento: chi tutela troppo, resta indietro. A essere penalizzati non sarebbero solo gli stati più progressisti, ma anche le aziende e le università che operano in ambienti regolati.

Ancora più delicato è il passaggio dedicato alla protezione delle innovazioni commerciali e governative, che lascia intravedere un potenziale ampliamento del controllo statale sulla diffusione di tecnologie AI, con logiche protezionistiche e di segretezza difficilmente compatibili con i principi di apertura, interoperabilità e trasparenza promossi fino a quel momento sul piano globale all'interno del mondo occidentale.

Sotto il profilo del lavoro, l'*AI Action Plan* parla esplicitamente di «potenziare la *workforce* americana», concentrandosi quasi esclusivamente su aspetti tecnici, come il riferimento a persone formate da impiegare nell'espansione della manifattura e nella costruzione di *data center* e impianti energetici. Nulla si dice, invece, sull'impatto dell'automazione cognitiva nei servizi, nell'amministrazione, nei settori ad alto impiego impiegatizio. Il rischio di perdita massiccia di posti di lavoro viene in questo caso esplicitamente ignorato.

Infine, un capitolo centrale riguarda il ruolo del Dipartimento della Difesa, che viene spinto ad adottare su larga scala tecnologie AI per scopi operativi e decisionali, in un'ottica di automazione crescente dei processi militari e di intelligence. Anche qui, nessun riferimento a vincoli di diritto internazionale, a salvaguardie etiche, a limiti d'uso.

Ma l'aspetto che più interessa nella nostra prospettiva è, ancora una volta, il riflesso sui diritti delle donne: nel documento esplicitamente si vieta, come già era stato fatto nei decreti presidenziali di inizio mandato, di utilizzare prospettive legate alla *diversity, women's rights, human rights, climate change*. E si minaccia anche qualsiasi ente o istituzione, a partire dalle Università, di fare adozione di questa prospettiva e impostazione. Ciò significa che, ad esempio, negli Stati Uniti dovrebbero essere rimossi (alcune piattaforme lo stanno già facendo) quei meccanismi di filtro dell'odio online o dell'adozione indiscriminata di algoritmi contenenti *bias*, come abbiamo detto pregiudizievole soprattutto per le donne.

Emerge allora in modo chiarissimo che fra Stati Uniti ed Europa si sta alzando una vera e propria barriera, che attiene al modo di valutare i diritti fondamentali delle persone: ma è significativo che, per le prospettive più avanzate, come quella del mondo *onlife*, la battaglia veda al centro proprio e ancora i diritti delle donne.

8. *Postilla: la legge italiana e il riferimento specifico alla "parità di genere".*

Nelle more della pubblicazione di questo lavoro è entrata in vigore la legge italiana n. 132/2025 ("Disposizioni e deleghe al Governo in materia di intelligenza artificiale"), un elaborato ampio e complesso composto di 28 articoli che toccano non soltanto i principi e il metodo di utilizzo, ma una serie di materie e ambiti strategici, dalla sanità, al lavoro, alla *cybersecurity*, nonché aspetto essenziali di *governance*.

Per dare un giudizio complessivo occorrerà attendere i decreti attuativi, ai quali, nello stesso testo, è delegato il Governo.

Sulla carta, la disciplina si presenta di grande impatto: essa contiene innanzitutto, come abbiamo detto, una serie di principi e di regole nell'ambito della cornice europea, a cui la legge fa puntuale riferimento, a partire dall'art. 2 (che si occupa di definire i sistemi di IA) e dall'art. 3 (che introduce i principi generali).

Va osservato che quest'ultimo è chiarissimo nella piena adesione all'*AI Act* europeo, richiamandosi alla necessità del rispetto dei diritti fondamentali, valorizzando la Costituzione italiana e il diritto dell'Unione Europea. Sempre sul solco delle norme europee, si ribadiscono come centrali i principi di «trasparenza, proporzionalità, sicurezza, protezione dei dati personali, riservatezza, accuratezza, non discriminazione, parità dei sessi e sostenibilità» (art. 3, comma 1).

Viene anche ribadito, ai commi 2 e 3 della stessa disposizione, la necessità di trasparenza, correttezza, qualità e appropriatezza dei processi di sviluppo per finalità generali, sottolineando la necessità del «rispetto dell'autonomia e del potere decisionale dell'uomo» e dell'assicurazione di possibilità della «sorveglianza e dell'intervento umano».

Rispetto all'*AI Act*, però, il comma 4 dell'art. 3 si sofferma sulle implicazioni e sui pericoli dell'uso dell'intelligenza artificiale a livello democratico, richiedendo che venga garantito «lo svolgimento con metodo democratico della vita istituzionale e politica (...)» e che non venga pregiudicato «la libertà del dibattito democratico da interferenze illecite, da chiunque provocate, tutelando gli interessi della sovranità dello Stato (...)».

Inoltre, il testo contiene un riferimento esplicito anche alla parità di genere.

Un altro aspetto molto significativo e originale rispetto alle regole europee, che pure ad esso fanno riferimento, è il comma 7 dell'art. 3, che introduce la necessità di garantire alle «persone con disabilità il pieno accesso ai sistemi di intelligenza artificiale e alle relative funzionalità ed estensioni», alla luce del principio di eguaglianza e di non discriminazione.

Molto interessante risulta l'art. 5 della legge (principi in materia di sviluppo economico) che affida allo stato e alle altre autorità pubbliche il compito di sviluppare il mondo produttivo e l'innovazione attraverso la promozione e l'utilizzo dell'intelligenza artificiale come «strumento per migliorare l'interazione uomo-macchina nei settori produttivi» e per «avviare nuove attività economiche» (lett. a), nonché a favorire la «creazione di un mercato dell'intelligenza artificiale innovativo» (lett. b), dove lo Stato dovrebbe facilitare anche la comunità scientifica (lett. c), privilegiare la creazione e utilizzo di data center nazionali (lett. d), nonché favorire la «ricerca collaborativa tra imprese, organismi di ricerca e centri di trasferimento tecnologico in materia di intelligenza artificiale» (lett. e).

L'insistenza sull'autonomia nazionale dei *data center*, in un quadro completamente diverso dal punto di viste della relazione con i principi fondamentali che tutelano le persone, richiama però la strategia americana, fondata proprio sulla difesa del proprio mondo virtuale.

Molte sono le norme dedicate al mondo della sanità e agli sviluppi della robotica (artt. 7-10), con riferimenti alla disabilità (art. 7), alla ricerca e sperimentazione in ambito scientifico (per finalità di prevenzione, diagnosi e cura di malattie, sviluppo di farmaci, terapie e tecnologie riabilitative, realizzazione di apparati medicali, salute pubblica (art. 8), al trattamento dei dati personali (art. 9) e alle implicazioni sullo stesso fascicolo sanitario (art. 10).

Importanti le disposizioni in materia di lavoro (artt. 11-12), di professioni intellettuali (art. 13), della pubblica amministrazione (art. 14) e della giustizia (art. 15).

Nel suo approccio all'intelligenza artificiale, in un mondo diviso sulla strategia europea "di prevenzione e difesa", a cui si contrappone quella americana, che si presenta come una strategia "di guerra", l'Italia propone una terza visione: quello dell'intelligenza artificiale come strategia "di sviluppo". Rispetto agli altri testi esistenti, quello italiano si caratterizza per una consapevolezza anche degli ambiti che verranno maggiormente trasformati dall'utilizzo dell'IA, con un approccio significativo, che oscilla fra riconoscimento delle garanzie e valorizzazione degli ambiti in una chiave di sviluppo.

CAPITOLO II

CONCEPTUAL EVOLUTION AND NORMATIVE TENSION OF PERSONAL IDENTITY BETWEEN AI TECHNOLOGIES AND BEYOND

Edoardo Carlo Raffiotta

SOMMARIO: 1. Background. – 2. The necessary protection of personal identity in the digital world. – 3. The necessary protection of personal identity in the era of AI. – 3.1. From the virtual to the metaverse, opportunities and criticalities of the new dimensions of personal development. – 4. Over-regulation in the EU and the need for fewer, more effective rules. – 5. Between AI and metaverse technology. – 6. A provocation. Conclusions.

1. Background.

The advent of artificial intelligence (AI) in the realm of image creation and manipulation has opened new frontiers in digital artistry and communication. However, it also raises significant questions concerning the right to personal identity, as AI-generated images undeniably offer both opportunities and risks in how identities are constructed, perceived, and protected in the digital age.

As everyone knows, AI solutions capable of translating text into images have evolved just as rapidly in a very short time. On 1 July 2015, DeepDream was introduced, one of the first instances in which it was possible to use neural networks to generate image patterns (although not images tout court, from scratch); while a few months later (on 9 November 2015) AlignDRAW made its appearance: trained on Microsoft's 'COCO' dataset, AlignDRAW was probably the first text-to-image model, and – although capable of generating only a few low-quality images – it probably paved the way for AI-generated images. In 2016, solutions based on generative adversarial networks¹ were

¹ Models based on generative adversarial networks include StackGAN; CycleGAN; AttnGAN;

deployed for the first time, while shortly thereafter – in 2021 – it was the turn of DALL-E, the model released by OpenAI that – it is reasonable to say – definitively brought text-to-image to the knowledge of the wider lay public. From DALL-E (and subsequent updates, up to DALL-E3) to Midjourney, the evolution of AI models took only a few years to replace low-quality images with the creation of deepfakes indistinguishable (or almost) from reality (so much that the European Union was forced to provide accordingly: as it is well known, the recent AI Act requires for certain AI contents that their AI-generated origin must be explicitly indicated) ².

So, if – as is true – the possibility (and ability) to represent the world through images is also one of those essential factors that, like language, characterize the person as such (and, by extension, the community) ³ we should not be surprised by the growing attention paid by the public debate to personal identity, whose fundamental coordinates are – as we shall see – considerably stressed and challenged by AI technologies.

2. *The necessary protection of personal identity in the digital world.*

When it comes to investigating the points of intersection between personal rights and digital technologies, too often the attention of the legal scholar – and the constitutionalist – is devoted to the right to privacy. Certainly, one cannot doubt that privacy has become of primary interest over the years (as its evolution, also in case law, testifies), yet we too often forget how much digital technologies impact (and will increasingly impact) other aspects of the person as well. While we cannot stop here on the origins and significance for public law of the right to personal identity ⁴, it is necessary to clarify the extent to which the right to personal identity has undergone a gradual formal and substantive expansion: starting from John Locke's first theorizations on the importance of personal identity ⁵, today personal identity has transcended the boundaries of

BigGAN and GauGAN, the latter released in 2019 by NVIDIA and capable of generating realistic landscape backgrounds. GauGAN2, currently the basis of NVIDIA Canvas, was released in 2021.

² See regulation UE 1689/2024, arts. 52 et seq.

³ G. DEMURO, *Identità linguistica e forma di Stato*, in *Riv. AIC*, 1, 2023.

⁴ See J. MARSHALL, *Human Rights Law and Personal Identity*, London, 2014.

⁵ J. LOCKE, *Essay Concerning Human Understanding* (see in particular the chapter “Of Identity and Diversity”).

sociology and the social sciences⁶, raising questions even in everyday life: in an age where “digital swarms”⁷ dominate, one’s personal identity undeniably undergoes new pressures and faces new tests, from being part of an online community to attempting to develop one’s individual personality as empowerment and self-determination whilst acknowledging our inter-subjective identities⁸.

In other words, while it is right to talk about privacy when it comes to digital technologies, we must also remember that «the concept of private life extends to aspects relating to personal identity such as “a person’s name”»⁹ or “a person’s picture”¹⁰, as the European Court of Human Rights has well reminded us. Moreover, it is the European Convention on Human Rights itself that – in Article 8 – expressly protects a form of “right to identity”¹¹, whose content, over the years, has progressively expanded to include a *quid pluris* with respect to the ‘mere’ name¹².

Given the above, even if there is no unitary definition today – at a global level – of what contributes to representing personal identity (and indeed, on closer inspection, there is not even a univocal definition of personal information)¹³, the need for protection of the right to personal identity arises precisely from technology and, in particular, from the technology of dissemination of images or thoughts through the press or even through advertising: one may recall the first decisions of the Italian Judge of Rome¹⁴ in defense of two citizens who, having participated in a photographic course, had later seen themselves represented as two farmers, although they were not. Not only that: those images

⁶ H.W. NOONAN, *Personal Identity*, London, 2019.

⁷ B.C. HAN, *Im Schwarm. Ansichten des Digitalen*, Berlin, 2013.

⁸ J. MARSHALL, *Human Rights Law*, cit.

⁹ European Court of Human Rights, 22 February 1994, *Burgharts v. Switzerland*.

¹⁰ European Court of Human Rights, 21 February 2022, *Schüssel v. Austria*.

¹¹ European Court of Human Rights, 28 January 2003, *Peck v. United Kingdom*. See also C. SULLIVAN, *Digital identity, privacy and the right to identity in the United States of America*, 4, 2013.

¹² European Court of Human Rights, 25 September 2012, *Godelli v. Italy*; 7 February 2002, *Mikulić v. Croatia*; 13 February 2003, *Odièvre v. France*; 5 June 2018, *Relu Adrian Coman and Others v. Inspectoratul General pentru Imigrări and Ministerul Afacerilor Interne*. See also E. MICHALKIEWICZ-KADZIELA, E. MILCZAREK, *Legal boundaries of digital identity creation*, in *Internet Policy Review*, 1, 2022.

¹³ P.M. SHWARTZ, D.J. SOLOVE, *Reconciling personal information in the United States and European Union*, in *California Law Review*, 2014.

¹⁴ *Pretore di Roma*, 6 May 1974.

had then been used for propaganda purposes against the divorce referendum, regardless of the will (rectus: contrary to the will) of the protagonists: who were not farmers, were not married, and were – above all – in favor of the divorce law.

More recently, we can recall the *Peck v. the United Kingdom* (2006)¹⁵ case, in which the applicant, who was suffering from depression, complained about the broadcast in the media of footage, captured from a CCTV camera mounted in the street, showing him walking alone with a kitchen knife in his hand. He further complained of the lack of an effective domestic remedy in that regard, and the Court found that the disclosure of the footage by the municipal council had not been accompanied by sufficient safeguards and constituted disproportionate and unjustified interference with the applicant's private life, in breach of Article 8 (right to respect for private life) of the EU Convention. Similarly, in the *Reklos and Davourlis v. Greece* (2009)¹⁶ case, the Court stated that photographs of a new-born baby taken in a private clinic without the parents' prior consent, and the retention of the negatives violated art. 8 of the Convention (the overriding consideration was the fact that the photographer had kept them without obtaining the applicants' consent, and that the baby's image had thus been retained in the hands of the photographer in an identifiable form with the possibility of subsequent use against the wishes of the child and/or his parents). Again, in 2021, after publishing an advertisement in a nationwide daily newspaper aimed at finding a surrogate mother, someone found himself the subject of a television report by an investigative reporter who had recorded her meetings with him covertly whilst pretending to be a potential surrogate mother: the Court held that there had been a violation of article 8 because of the violation of "personal integrity"¹⁷.

These are simple cases but extremely significant from a legal standpoint: it is from this desire to be represented according to one's real identity – the result of personal convictions and beliefs – that this right has begun to assert itself, having a content of novelty compared to the traditional abusive use of someone else's image or name¹⁸. Representative, in other words, of that heritage of

¹⁵ <http://hudoc.echr.coe.int/sites/eng-press/pages/search.aspx?i=003-687182-694690>.

¹⁶ <http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-90617>.

¹⁷ European Court of Human Rights, 1 July 2021, *Hájovský v. Slovakia*.

¹⁸ The Rome praetor wrote, also ordering the publication of the measure in the local edition of three national newspapers, that «It constitutes a violation of the right to personal identity, understood as the right not to have one's individual personality misrepresented, to affix a poster for propaganda in favour of the repeal of the divorce law, in which the image of persons who,

political and cultural identity, even religious¹⁹, which completes the person, representing him or her. What, in other words, allows a faithful reconstruction of the person and thus a true and truthful public representation of him or her, characterized by having to be as faithful as possible to reality. Indeed, even the first, embryonic theorization of the right to be left alone partly owes its origins to the fear that, with the new technologies, someone could have been photographed without their consent for those specific purposes²⁰.

Of course, as Pirandello teaches, our identity is not immutable: on the one hand, it is subject to the passage of time, thus to the changes and contaminations that – on closer inspection – constitute an integral part of it; on the other hand, and at the same time, it is highly dependent on the point of view adopted by the observer and thus susceptible as such to being perceived differently based on, for example, different cultural parameters²¹. The digital phenomenon stresses this ideal (and normative) concept, to such an extent that the idea of the person that we all have today corresponds to how that particular subject is represented (also) through computer and digital tools²².

although supporters of the institution of divorce, are made to appear as exponents of repeal is portrayed» (*Pretore di Roma*, 6 May 1974).

¹⁹ Religious identity has been defined as «an extreme case of a constitutive factor of personal identity, characterized by a strong expansiveness and voracity toward other profiles of individual identity, and combativeness on the public sphere». G. PINO, *Identità personale, identità religiosa e libertà individuali*, in *Quaderni di diritto e politica ecclesiastica*, 1, 2008.

²⁰ «Warren and Brandeis were not satisfied with confidentiality because they had in mind the candid photographer, a stranger who did not have a relationship with the subject of the photo. They observed that with earlier photographic technology, “one’s picture could seldom be taken without his consciously ‘sitting’ for the purpose”», N.M. RICHARDS, D.J. SOLOVE, *Privacy other’s path: recovering the law of confidentiality*, in *Georgetown Law Journal*, 123, 2007; and C. SULLIVAN, S. STALLA-BOURDILLON, *Digital identity and French personality rights. A way forward in recognising and protecting an individual’s rights in his/her digital identity*, in *Computer Law & Security Review*, 31, 2015.

²¹ The impact of the cultural dimension on the formation of a person’s identity becomes manifest in its “pathological” dimension, with reference to so-called culturally oriented crimes: one may recall the recognized excusability in favour of a mother accused of abusive exercise of the medical profession for having practiced – for ritual reasons, thus culturally oriented and determined – circumcision on her son, *Cass. pen., Sez. VI*, 22 June 2011, n. 45646.

²² Studies straddling law, philosophy, and sociology on the influence of social networks on the construction of personal identity are now numerous: *ex multis* may recall those who assert that such technology alters «constantly, and without encountering resistance, sensory reactions or forms of perception», resulting in a reversal of the «personal point of view», Q. FIORE, M.

Yet, it should not be forgotten that, whenever dealing with personal identity, the fundamental categories of constitutional law come into play, as repeatedly affirmed by case law and, in particular, by the Italian Constitutional Court. Still intact today is the legacy of the notorious²³ Italian “Veronesi case”: personal identity is a constitutional right that rests on Article 2 Const. and that – as a “new right”²⁴ – deserves protection appropriate to the context in which the regulation of the “new world” is inserted²⁵.

Above all, we cannot forget that while our identity is not immutable (and we are therefore subject to various kinds of changes, including aesthetic ones), it is also true that, today, technology is able to reflect our changes and – even – anticipate them (e.g. through AI-generated images). In this scenario, the advent and spread of the Internet have played and still play the role of game-changer, making possible and as easy as ever the circulation of an almost infinite amount of information, capable of placing under the spotlight not only public figures, who have always been accustomed to being the object of dissemination of more or less accurate information and news, but also ordinary people, exposing anyone to the myriad of data that can be found today on social networks (or extracted from them)²⁶ or through search engines: data that – we certainly cannot forget – circulate increasingly faster, posing risks of generating confusion and of providing the outside observer with distorted,

MCLUHAN, *Il medium è il messaggio*, Mantova, 2011. From this follows, unsurprisingly, a marked malleability of personal identity, unable to consolidate autonomously, extremely permeable to external and contingent influences (a “liquid identity”).

²³ Cass. civ., Sez. I, 22 June 1985, n. 3769.

²⁴ The very concept of the right to personal identity-and its modern declinations-makes it all the more evident that rights are «historical products [that] arise from needs when they historically emerge», according to N. BOBBIO, *L'età dei diritti*, Torino, 1990.

²⁵ A. D'ALOIA (ed. by), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Milano, 2020.

²⁶ Exemplary is the famous Cambridge Analytica case, well demonstrative of the potential related to the exploitation of big data for profiling activities (defined by the Italian Data Protection Authority, in its Guidelines on the subject, as «the analysis and processing of information relating to users or customers, in order to divide the interested parties into “profiles”, that is, into homogeneous groups for increasingly specific behaviours or characteristics, with the aim of reaching the unequivocal identification of the single user (cd. single out) or of the terminal and, through it, also of the profile, precisely, of one or more users of that device» – GPDP, Guidelines on the *Processing of Personal Data for Online Profiling*, 2015. On the (undesirable) effects of the implications of the “algorithmic society” on the individual see A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e futuro delle libertà*, in *BioLaw Journal-Riv. BioDiritto*, 1, 2019.

or out-of-date information²⁷, of that identity (information having to be understood as not only the image or name but – precisely – rather the heritage of cultural elements that classify a person as such).

And while it is true that not every piece of information constitutes sensitive data susceptible of protection (one may recall the principle, enshrined in US jurisprudence, that «A general level of discomfort from knowing that people can readily access information about us does not necessarily rise to the level of a substantial state interest, for it is not based on an identified harm. Our names, addresses, types of cars we own, and so on are not intimate facts about our existence, certainly not equivalent to our deeply held secrets or carefully guarded diary entries. In cyberspace, most of our relationships are more like business transactions than intimate interpersonal relationships»)²⁸ it is also true that one's image is certainly an indispensable asset for the protection of the right to one's personal identity: so much so that, as it has been well specified, «[A] person's image constitutes one of the chief attributes of his or her personality, as it reveals the person's unique characteristics and distinguishes the person from his or her peers. The right to the protection of one's image is thus one of the essential components of personal development. It mainly presupposes the individual's right to control the use of that image, including the right to refuse publication thereof ...»²⁹.

Certainly, the legal system has provided answers in recent years: in addition to the attention given to the abovementioned cases³⁰, it is not possible to forget the contribution provided by the data protection regulations – the GDPR above all – which has first and foremost introduced the politicization of the right to correct or remove inaccurate information, or to remove illegal content (one thinks in this regard of the effects and positive reverberations on the specific disciplines against child pornography, or the discipline on revenge porn, or even the right to be forgotten and the possible fields of application of the

²⁷ Recall how a piece of news «originally complete and true, becomes out of date, thus turning out to be partial and inaccurate, and therefore substantially untrue», *Cass. civ., Sez. III*, 5 April 2012, n. 5525.

²⁸ 10th Circ, 1999, *U.S. West, Inc. v. Federal Communications Commission*, 182 F.3d 1224.

²⁹ European Court of Human Rights, 7 February 2021, *von Hannover v. Germany* (no. 2), § 96, in which the European Court of Human Rights held that there had been a violation of Article 8 (right to respect for private life) of the European Convention on Human Rights.

³⁰ See also European Court of Human Rights, *Right to the protection of one's image*, November 2023.

latter to protect individual identity)³¹. If, in short, information about us is destined to remain in the ether, through the digital world, indefinitely (or potentially indefinitely), the genesis of the need to correct or remove that information that is incorrect or no longer current becomes highly understandable: as mentioned above, the person changes over time and needs to update his or her identity so that it can be said to be always current and corresponding to the truth.

3. The necessary protection of personal identity in the era of AI.

The image-generating possibilities currently offered by AI are helping to redefine the concept of artistic creation and visual communication, offering professionals and the general public hitherto unknown opportunities. Similarly to what happened with the arrival of the first blogs and websites (as a result of which everyone could become his or her own ‘editor’), today AI allows anyone to create hyper-realistic images or even to go further towards the creation of new identities: this is the case, for instance, of digital influencers, created by means of AI programs and nevertheless widely followed, on a par with ‘real’ influencers (e.g. the case of Aitana Lopez, an AI influencer created by a Spanish communication agency, who currently has 326.000 followers on Instagram)³².

New opportunities may of course arise from such possibilities, allowing us to go beyond traditional art forms or bring to light hidden or underlying aspects of our existence (as Refik Anadol, a pioneer of new algorithmic visual art by means of AI, openly proposes to do)³³, but they also raise important ethical and legal issues, concerning, in particular, the right to personal identity. As an

³¹ Since the case of “Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González” (c.d. “Google Spain”; Court of Justice of the European Union, 2014, C-131/12), the right to be forgotten has received specific positivity in Reg. EU 679/16 and, finally, in Article 1 paragraph 25, of Law 134/2021 (as a result of which Article 64b of the Implementation, Coordination and Transitory Rules of the Code of Criminal Procedure now enshrines the «Right to be forgotten of defendants and persons under investigation»).

³² Also J. LOOI, L.A. KAHLOR, *Artificial Intelligence in Influencer Marketing: A Mixed-Method Comparison of Human and Virtual Influencers on Instagram*, in *Journal of Interactive Advertising*, 2, 2024.

³³ R. ANADOL, *Space in the Mind of a Machine: Immersive Narratives*, in *Architectural Design. Machine Hallucinations: Architecture and Artificial Intelligence*, 3, 2022.

example, it has been found ³⁴ that when talking about digital avatars, introverts are more likely to select attractive avatars than their extravert counterparts are, which «can be seen as an effort to reach out, to attract more social interaction» ³⁵. While this may have positive implications for personal self-esteem, the risks (also psychological) arising from the person's over-identification with an identity – the avatar – that does not, however, correspond to reality are evident.

Other uses of AI also directly raise additional legal problems worth a more thorough analysis. This is the case of models for the generation of 'undress-images', which exploit AI to digitally remove clothes from real photos, thus creating the illusion of being in front of a – legitimate – nude photograph: the phenomenon has grown exponentially in the last two years (especially among teenagers and in school contexts) and has started to arouse the interest of specialized news ³⁶, scholars ³⁷ and Big Techs. Evident are the problems that may arise in terms of the personal identity of those involved, all the more so considering that they are often minors (the case *People of the State of California vs. Sol Ecom Inc.* is currently pending in the San Francisco Superior Court, in which it is stated that «nonconsensual intimate images ('NCII') are used to bully, threaten, and humiliate women and girls»: among the charges, violation of California Penal Code section 311.3(a) prohibiting the knowing development of nonconsensual obscene images of persons under the age of 18 years, or aided and abetted violations of California Penal section 311.3(a) in violation of California Penal Code section 31 by the acts and practices set forth herein ³⁸. These are issues that appear to be more and more difficult to pursue through ordinary regulatory instruments (such as the US National Defense Authorization Act – NDAA; or the European DSA and GDPR) ³⁹ and that therefore require a careful reflection on a possible and necessary self-regulation of AI technologies through soft law (e.g. associations of experts and practitioners).

³⁴ R.A. DUNN, R.E. GUADAGNO, *My avatar and me – Gender and personality predictors of avatar-self discrepancy*, in *Computers in Human Behavior*, 28, 2012.

³⁵ *Ibidem*, 104.

³⁶ M. BURGESS, *Harmful "Nudify" Websites Used Google, Apple, and Discord Sign-On Systems*, in *Wired*, 29 August 2024.

³⁷ N.G. BRIGHAM *et al.*, *Violation of my body: Perceptions of AI-generated non-consensual (intimate) imagery*, in *Proceedings of the 20th Symposium on Usable Privacy and Security*, 2024.

³⁸ https://www.sfcityattorney.org/wp-content/uploads/2024/08/Complaint_Final_Redacted.pdf.

³⁹ C. OKOLIE, *Artificial Intelligence-Altered Videos (Deepfakes), Image-Based Sexual Abuse, and Data Privacy Concerns*, in *Journal of International Women's Studies*, 25, 2023.

That of the NCII is one of the clearest and most striking cases on the possible critical issues related to the use of AI in relation to personal identity, but it is certainly not the only one: from the possible effects of the empathic relationship with AI on a person's identity⁴⁰, to the possible effects on the professional capacity (and identity) of workers who – in order to perform tasks that they used to perform autonomously – now rely on AI⁴¹.

Besides, one shall not forget the risks connected to the creation of digital identities tout court or, in any case, to the generation of fundamental traits of a person's identity: the recent case of an AI model that presented the same voice as the actress Scarlett Johansson is well known, while the most notorious cases certainly concern the use of deepfakes, especially for political propaganda or fraud purposes: cases that, in all evidence, can severely damage the identity of the subjects represented, violating that necessary correspondence – mentioned above – between what one is and how one wants to be represented externally and where the possibility of recreating the voice and likeness of a person, thus creating real “artificial subjects”, inevitably opens up critical issues – ethical and normative above all – unprecedented to date⁴².

Obviously, not all uses of AI for the creation of digital identities are per se harmful: for instance, one may recall cases of digital identity creation (so-called human honeypots) which, by adopting a proactive approach to cybersecurity, make it possible to limit the proliferation of online threats, at the same time without ‘harming’ ‘real’ human counterparts⁴³. If such uses do not raise particular problems, neither from an ethical nor from a regulatory point of view (including with regard to European regulations, including GDPR and the AI Act),

⁴⁰ To comply or to react, that is the question: the roles of humanness versus eeriness of AI-powered virtual influencers, loneliness, and threats to human identities in AI-driven digital transformation, S. VENUS JIN, *Computers in Human Behavior: Artificial Humans*, Elsevier, 1, 2023.

⁴¹ F. STRICH, *What Do I Do in a World of Artificial Intelligence? Investigating the Impact of Substitutive Decision-Making AI Systems on Employees' Professional Role Identity*, in *Journal of the Association for Information Systems*, 2, 2021.

⁴² Questions have been raised—for example—about the constitutional compatibility of an absolute ban on the creation and dissemination of *deepfakes* and thus about the necessary balancing of opposing interests and equal rights that ensues; an obvious demonstration of the problematic nature of the topic. Thus M. CAZZANIGA, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes*, in *MediaLaws*, 1, 2023.

⁴³ C. GONZALEZ *et al.*, *Adaptive Cyberdefense with Deception: A Human – AI Cognitive Approach*, in T. BAO *et al.* (ed. by), *Cyber Deception. Techniques, Strategies, and Human Aspects*, Berlin, 2022.

such reference is however useful to remember that one should not forget that – as it is well known – the same technological solution – e.g. the creation of non-existing AI identities, either through images or through even more advanced systems – can be used both with beneficial effects (as in the case of cybersecurity) and with malicious purposes (as we have seen with pornographic images). The – distinctly⁴⁴ – individual dimension of modern constitutionalism, in which therefore personal identity finds expression, is the digital dimension.

3.1. *From the virtual to the metaverse, opportunities and criticalities of the new dimensions of personal development.*

It should not be forgotten, then, that we are currently in a context in which so-called “non-things” have taken space away from “things”⁴⁵, where “non-things” must mean information and data: it could be said, by way of example, that where once were the sky and the earth are now the Cloud and Google Earth. The simplest daily habits, moreover, are proof of the pervasiveness of the digital in the individual and collective spheres⁴⁶: from the weather (for which people now turn to a digital assistant or smartphone rather than to their senses), to the regulation of daily routines, it is now evident how the category of person increasingly responds to a value paradigm inseparable from the digital component and, as a result, to the challenges arising from the tension between two historically distant conceptual orders (as is the case of the deepfakes). Thus, if personal identity is already under conceptual and normative stress in the digital world, such a context is inevitably destined to find new ferment (and new challenges) when the same artificial intelligence technologies will make it possible not only to create virtual identities *ex novo* (as is already the case now) but even to create an entire virtual new reality where AI identities could find a place.

⁴⁴ M. AINIS, *Il regno dell'Urobora. Benvenuti nell'era della solitudine di massa*, Milano, 2018.

⁴⁵ B.C. HAN, *Le non cose. Come abbiamo smesso di vivere il reale*, Torino, 2022.

⁴⁶ Among the most evident effects of the penetrative attitude of technologies towards everyday life must certainly be remembered the conceptual transition in which the concept of (right to) security (appropriately defined as a “superprimary value”) is the protagonist, today largely absorbed by its own digital declination in terms of cybersecurity. Writes of a right to security as a superprimary value G. CERRINA FERONI, *Il diritto alla sicurezza: un valore superprimario*, in *Percorsi costituzionali*, 1, 2008. See also on the topic T.F. GIUPPONI, *Le dimensioni costituzionali della sicurezza*, Bologna, 2010 and A. PACE, *La funzione di sicurezza nella legalità costituzionale*, in *Quad. cost.*, 4, 2014, as well as recently M. BARONI, *Intelligenza artificiale e cybersicurezza in una prospettiva costituzionale*, in G. CERRINA FERONI, C. FONTANA, E.C. RAFFIOTTA, *AI Anthology*, Bologna, 2022.

That would be the case of the metaverse or – as it has been correctly put ⁴⁷ – the “metaverses”, since these are multiple digital contexts within which as many virtual societies are created and where the concept of truth takes on a meaning “beyond” the digital as we have known it up to now: if digital today is synonymous with a two-dimensional reality – which can be accessed through computer desktops or smartphone apps – the metaverse will probably have a quid plus identifiable in the claim to make the virtual become real by immersing the physical person in it. Certainly, the current version of the metaverse – summarized by the accessories required to enter it (the visors, plug-ins for hands or skin) – is nonetheless still extremely linked to the real world: it is possible to use such tools, by way of example, to explore the interior of a car, to drive it perhaps, even to touch its fabrics, remotely and without having to go to the dealer; yet still to enable – and promote – the purchase of the good in the real world. And yet, as is well known, the ultimate innovative lead of the metaverse will be to go beyond the real: one need only look at the data available ⁴⁸ to see the investments that companies such as Nike, Prada, and Volkswagen have made over the past few years: amounts in the millions of euros that – to the outside observer – are symptomatic of how (and especially to what extent) the bet that the major private powers ⁴⁹ are making on metaverses is relevant, identifying these spaces as the next ones within which to build their own – recognizable – identity (digital and commercial) ⁵⁰.

In short, the impression is that the largest and most heavily capitalized corporations on the planet believe in the possibility of shaping a (virtual) reality in which the real is virtual (and vice versa). A reality in which the person immerses himself in the metaverse and finds therein the dimension where he can express his personality: the subject will no longer use the metaverse to see or try out the car to buy in the real world, but – rather – will decide to buy through and – what matters most – for his avatar. The paradigm of reality is subverted, and here, therefore, the real revolution must be identified: the pleasure of buying something will not be intended to satisfy the user in reality but rather will be

⁴⁷ See on this point the speeches – preceding that of the writer – at the Conference organized by the University of Messina *Un diritto virtuale? La regolamentazione giuridica del metaverso*, 7 July 2023.

⁴⁸ *What Long-Term Investors Need to Know About the Metaverse*, in *Forbes*, for which the market-cap of the metaverse will reach a total value of \$800 billion by 2027.

⁴⁹ M.R. FERRARESE, *Poteri Nuovi*, Bologna, 2022; writes of a “soft, informal, indirect” power.

⁵⁰ G. BAIMA, L.M. ZAGNI, *I brand e la costruzione della propria identità digitale nel Meta-verso*, Università di Torino, 2022.

designed to satisfy the avatar of the physical person. There can be no doubt that such an economic and conceptual model still lacks credibility in today's society (in which, among other things, broadband distribution is still insufficient⁵¹, and in which digital services in PA are struggling to revolutionize the relationship between offices and users)⁵², yet it is already possible to turn one's gaze to other social realities – characterized by greater and more effective digital penetration – to understand immediately that this is a phenomenon not to be underestimated, and indeed to be cautiously watched⁵³.

What place will the person be able to occupy in this world? Will he or she retain ownership (and to what extent) over his or her AI identity? On closer inspection, this is the potential Gordian knot connected with the advent of the metaverse: will it be possible, for example, to allow other individuals to create an avatar in my image and likeness? If, after all, we have doppelgangers in reality, what can prevent (me, or others) from imagining an avatar in my likeness? And again, if the avatar can be a reconstruction of my person within the digital world that is faithful to reality or can – on the contrary – resolve itself into an exaltation of the characteristics that the 'real' person wishes to emphasize (at the expense of those that he or she wishes instead to conceal, or smooth out), where can the line be drawn between the "technically possible" and the "legally and ethically acceptable"⁵⁴ in the construction of a "bespoke" identity?

It is fair to ask, moreover, who will be the subject who animates the avatar. In particular, there are two major sources of criticality facing the technician and the constitutionalist: on the one hand, the possibility for anyone who controls

⁵¹ See the (not comforting) data published by the Communications Guarantee Authority, available at <https://www.agcom.it/mappatura-delle-reti-di-accesso-ad-internet>.

⁵² The subject is notoriously rich and would deserve much more space. Let it suffice here to recall the necessary balancing of opposing needs, well represented by the different answers provided by administrative jurisprudence (*ex multis* TAR Lazio, Sez. III, n. 4409/2021, for which the human component in the PA decision-making process is inescapable; and the jurisprudence formed within the Council of State for which instead even a «constrained» administrative activity would be an expression of «authoritative activity carried out in the pursuit of the public interest», so Cons. Stato, Sez. VI, n. 8472/2019).

⁵³ Paolo Benanti likens modern Companies owners of social platforms and, in the future, of metaverses, to "демиurge" the new reality, «master of a society that becomes medieval where I as lord offer spaces to vassals and valvassors who produce value for me» (among others P. BENANTI, *Metaverso, dimensione utopica e opportunità di business*, in *L'Osservatore Romano*, 19 October 2021).

⁵⁴ G. CERRINA FERONI, *Intelligenza artificiale e ruolo della protezione dei dati personali*, in *Key4biz*, 14 February 2023.

an avatar to use a pseudonym and the forms of anonymization of the digital to act – in the virtual world – in a different way from how he would instead do in reality, with the responsibilities that would inevitably follow⁵⁵. On the other hand – and even more profoundly – it is necessary to question the cognitive advancement of AI: while it is almost certain that (for now)⁵⁶ the avatar will not enjoy its own ‘thoughts’ it is already clear, however, that the person’s digital counterpart will also be able to perform actions and movements in a sphere – albeit limited – of substantial autonomy⁵⁷.

The upcoming metaverse requires the matter of responsibility (legal and ethical) for the actions of avatars to be addressed and treated as one of the most delicate and relevant issues: while it is not possible today to predict whether (and when) in the future there will be a perfect correspondence between the digital and virtual worlds, it is however already necessary to take note of the current state of the art, in which – as we said – technology allows (*rectius*: now begins to allow anyone) to model their virtual life differently from the real one. Precisely such a possibility is, moreover, creating – and will increasingly create – more than a few difficulties for the legal world: certainly the metaverse(s) undoubtedly represents, in positive terms, an agora where the full and free manifestation of thought is expressed, yet the study of law and of constitutional law in particular requires that we do not forget how inherent in the concept of law is the concept of a limit⁵⁸ (to the manifestation of that right): the possibility of free and unconstrained manifestation of one’s thought must necessarily be counterbalanced by an assumption of responsibility, the ability – in other words – to take responsibility for any consequences (e.g., criminal) arising from the manifestation of one’s thought (e.g., when it has resulted in an injury to others’ equally constitutionally protected rights)⁵⁹.

⁵⁵ M. BASSINI, O. POLLICINO, *Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. PIZZETTI (ed. by), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018. Of the issues related to the use of anonymized data by AI writes T.E. FROSINI, *La privacy nell’era dell’intelligenza artificiale*, in *DPCE online*, 1, 2022, as well as in T.E. FROSINI, *L’orizzonte giuridico dell’intelligenza artificiale*, in *BioLaw Journal-Riv. BioDiritto*, 1, 2022.

⁵⁶ Wonders extensively about the future evolutions of the relationship between the human brain, AI and the metaverse G. SCORZA, *Neuroverso. Il cervello è nudo. Quale impatto sulle nostre vite, diritti e libertà*, Milano, 2023.

⁵⁷ On the dystopian implications of *autonomous agent actions* see among others P.A. HANCOCK, *Avoiding adverse autonomous agent actions*, in *Human-Computer Interaction*, 2021.

⁵⁸ *Corte cost. italiana*, 2 July 1956, n. 1.

⁵⁹ Returning to the point is the distinction between *intra-value* and *inter-value* conflicts, by

Certainly in the digital world everything is tracked or at least traceable (at least for the average user, not making use of Tor or other anonymization services), and therefore – in today's Internet – it is relatively easy to link the activity of a given network user with the real person to whom a given account refers, nevertheless it can already be said that with the metaverse there will instead be a radical paradigm shift: the first cases of harassment and violence perpetrated against an avatar are already known⁶⁰, and it is therefore legitimate – if not obligatory – to ask how to configure the hypothesis of violence-harassment in the metaverse. Whether then, it is possible to prosecute someone in the real world (not for a comment or an online 'post' but) for the actions enacted by one's virtual alter-ego. What happens, in terms of liability, when the digital is no longer a mere projection of the real, having instead become a reality in its own right? What when, even the acting subject of the virtual world is endowed with its own 'individuality'⁶¹, or its own will? The latest jurisprudence on the subject, even in Italy, seems to affirm the possibility of sanctioning a contact that constitutes harassment even when it has occurred in the virtual world, based on the assumption that such an avatar would still be 'guided' by the physical person who controls its will and action⁶².

In this regard, when it comes to defining the regulatory issues of the 'world of tomorrow' two fundamental considerations must be taken as a starting point: first, the technology of the metaverse is built and designed for social models different from the European standard; in particular, it has been conceived on the basis of and for models of societies such as the U.S. or Asian (Chinese, in particular), which, as well known, are built upon values extremely different from those of the Old Continent⁶³. Suffice to think – as a demonstration of

A. MORRONE, *Bilanciamento (giustizia costituzionale)*, in *Enc. dir.*, vol. II, tomo II, *Annali*, 2008.

⁶⁰ SumOfUs, *Metaverse: another cesspool of toxic content*, 2022.

⁶¹ The case of Sophia, the first robot to be granted – by Saudi Arabia – citizenship, is well known. Among others see A. CELOTTO, *I robot possono avere diritti?*, in *BioLaw Journal-Riv. BioDiritto*, 1, 2019; as well as National Committee for Bioethics, National Committee for Biosafety, Biotechnology and Life Sciences, *Sviluppi della robotica e della roboetica*, July 17, 2017.

⁶² Depose for the irrelevance, for the purposes of the integration of the crime, of the lack of physical contact between the agent and the victim *Cass. pen.*, Sez. III, 30 October 2018, n. 17509; *Cass. pen.*, Sez. III, 2 July 2020, n. 25266 (as also cited in M. MARTORANA, Z. SICHI, *Il Metaverso e il reato di molestie sessuali nella realtà virtuale*, in *Altalex*, 23 February 2022).

⁶³ Albeit in the well-known, respective differences between the U.S. (freer and more market-oriented) approach to the technology issue, on the one hand; and China's state capitalism, with a strong surveillance-oriented bias, on the other. Further demonstrating the Union's distinctly

the different consideration of the right to personal identity overseas – of the recent first (U.S.) case of tokenization of a person, capable of digitizing himself and selling parts of himself: a single example, which nonetheless clearly raises unpostponable ethical questions. Starting from this premise then becomes crucial in order to hypothesize a regulatory framework that is truly capable of governing the future digital world, starting with the individual, constitutional rights and personal freedoms.

In all likelihood, it will – above all – be a matter of seeking the balance between the regulatory will (aimed at avoiding improper uses of digital) on the one hand; and – on the other – the necessary freedom of development that the Internet and technology in general certainly cannot do without.

4. *Over-regulation in the EU and the need for fewer, more effective rules.*

When it comes to delving into the role played by the person within the digital sphere, the interest of the jurist⁶⁴ – but, probably, of every scholar – is twofold: first of all, because that very special constitutional freedom that is personal identity has already been tested – and greatly stressed – by the digital phenomenon and even more so by artificial intelligence⁶⁵. Secondly, then, it is easy to assume that, in all likelihood, similar peculiarities will playfully recur beyond AI itself. And this is so even if we are talking about a society in the

anthropocentric approach is the *International Outreach for human-centric Artificial Intelligence* initiative, promoted by the European Commission's Foreign Policy Instruments Service (FPI) and the Directorate-General for Networks, Content and Communications Technology (DG CONNECT), in collaboration with the European External Action Services (EEAS) (<https://digital-strategy.ec.europa.eu/en/policies/international-outreach-ai>).

⁶⁴ The approach of which, whenever one dwells on the intersection between the legal discipline and what intersects with it (such as the digital phenomenon, ed.), must nevertheless be aimed at describing and systematizing the normative datum, determining the extent of the constraint that derives from it, M. LUCIANI, *Governo (forme di)*, in *Enc. dir., Ann. III*, 2010.

⁶⁵ Let us refer on this point to E.C. RAFFIOTTA, M. BARONI, *Intelligenza artificiale, strumenti di identificazione e tutela dell'identità*, in *BioLaw Journal-Riv. BioDiritto*, 1, 2022. In particular, it may be recalled how a brief analysis of the conceptual evolution that has affected the traditional models of identification and identity, altering their discipline, has helped to reveal the methodological difficulties with which law is necessarily confronted during any attempt to interpret a society in perpetual – and very rapid – movement. For this reason, too, the distinctly *risk-based*, Brussels-made approach employed in the drafting of the AI Act is to be welcomed: only an adaptive AI regulation can indeed withstand the definitional stress that artificial evolution imposes and will impose on the aforementioned categories.

making, where it is still possible to raise questions about what could be the most relevant criticalities that will emerge from new technologies, also in the light of the context of conceptual and theoretical (thus, as a result, normative) stress⁶⁶ in which personal identity is already immersed.

Europe is, at present, in an uneasy position, the victim – on closer inspection – of over-regulation: and, indeed, over-regulation is evident even when we set our minds to the regulation of personal identity protection. One thinks of the GDPR⁶⁷, or the provisions – which recently came into force – aimed at removing content that is harmful to constitutional rights.

Consider, again, the competencies that come into play in the cases at hand: on the one hand, the Data Protection Authority, which is entrusted with overseeing the application of and compliance with the GDPR; on the other hand, the AGCOM, which enforces the Digital Services Act; and on the other hand, again, the AGCM, which is naturally entrusted with protecting against practices harmful to consumers: one may recall the (often excessively intricate) relations between data protection and the competence of the Italian Data Protection Authority; consumer protection and the regulation of the digital services market (also affected by the Digital Market Act) within the competence of AGCM; the fight against illegitimate digital content – also under the Digital Services Act – within the competence of AGCOM; and again the banking and financial matters, which in the field of artificial intelligence the AI Act leaves to the competence of the ECB and in the national field to the competence of the Bank of Italy and Consob⁶⁸. To this already composite framework could

⁶⁶ *Ex multis* see C. IRTI, *Dato personale, dato anonimo e crisi del modello normativo dell'identità*, in *Juscivile*, 2, 2020 and U. RUFFOLO, *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Milano, 2020.

⁶⁷ The system of protection introduced by the GDPR is complemented by the Data Protection Enforcement Directive (directive no. 680/2016) and regulation no. 1725/2018, as a result of which regulation (EC) no. 45/2001 is repealed and where the duties and powers of the European Data Protection Supervisor are defined. Recall also regulation no. 1724/2018, which follows EU regulation no. 1024/2012 and establishes the One-stop Digital Shop, and the Data Governance Act (EU regulation no. 868/2022), which, by amending regulation no. 1724/2018, establishes the conditions for the reuse, within the Union, of certain categories of data held by public bodies, as well as introducing a notification and control framework for the provision of data brokering services (in fact expanding what has already been regulated in 2019 through directive 2019/1024). Not forgetting the Data Act, currently at a mere proposal level. And it could go on.

⁶⁸ For a more in-depth discussion of the topic, let us refer to E.C. RAFFIOTTA, *Dalla self-regulation alla over-regulation in ambito digitale: come (e perché) di un necessario cambio di prospettiva*, in *Osservatorio sulle fonti*, 2, 2023.

then even be added the designated Authorities for the compliance of the AI Act. An overlap of disciplines aimed – on this point there remains no doubt – at protecting and safeguarding the individual. But – it seems reasonable to ask – from what? From big companies, the so-called “sovereign powers” of this digital world ⁶⁹? Or, rather, from the misuse of technologies?

The node is central to the upcoming regulatory issue that is imposing itself on the world of law. If it is well known to everyone the ease with which today personal data, information and images are disseminated online (as it is equally well known that the dissemination on the network coincides with an entry into the digital world having almost permanent effects – even given the chronicle inadequacy of instruments aimed at reducing *ex post* their circulation, and the difficulty of enforcing the discipline of revenge porn is the most striking example of this) ⁷⁰; in the near future, AI and the metaverse will allow for a processing of such data, information and images that is still unprecedented today, capable – it was said – of creating deep fakes or multiple avatars of a single subject, making it more difficult – as well as more problematic in terms of law, and ethics – to trace such digital alter-egos in the real world (with obvious criticalities on the side of security and the right to security) ⁷¹.

This is why it must be strongly affirmed that the first article of the ideal constitutional charter for the protection of the person, and therefore of his identity and constitutional freedoms in the society of the future (whether based on AI, augmented or virtual reality), should be represented by awareness, culture and knowledge: the users of tomorrow must be educated so that the familiarity that undeniably the ‘Internet generation’ has with the new technologies does not turn into an excess of security, in presumption (and, therefore, in the absence of the necessary precaution). As repeatedly affirmed by the Italian Data Protection Authority itself, whose activity continues tirelessly in this direction.

⁶⁹ Writes of «new powers» M.R. FERRARESE, *Poteri nuovi*, cit.

⁷⁰ So much so that the Italian Data Protection Authority has set up a special online reporting service (<https://www.garanteprivacy.it/temi/revengeporn>), while Meta recently launched “Take it Down”, to recognize and delete sexually explicit content on the Internet.

⁷¹ Note the argument that “guaranteeing the right to security” means guaranteeing “the security of rights”, A. BARATTA, *Diritto alla sicurezza o sicurezza dei diritti?*, in M. PALMA, S. ANASTASIA, (eds. by), *La bilancia e la misura*, Milano, 2001. Software for the detection of AI-generated images are already spreading: see for example D.C. EPSTEIN *et. al.*, *Online Detection of AI-Generated Images*, in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2023.

5. *Between AI and metaverse technology.*

Both AI models and their foreseeable evolutions will significantly contribute to shaping and influencing personal identity in digital contexts. Both of them will facilitate virtual identity construction, allowing individuals to even cultivate aspects of their identities in immersive environments (even without the fear of being judged): this capability will probably allow individuals to experiment with different identity expressions, which might not be possible or acceptable in physical reality. It is correct to say that AI technology is already influencing personal identity by creating and altering data that mirrors individual preferences, actions, and relationships; and it is entirely probable that the same thing will happen (with even greater impact) thanks to technologies that will probably go beyond AI as we know today.

Obviously, this will raise concerns over data privacy and control, with particular regards to risks of identity theft or manipulation: thus, as individuals increasingly live parts of their lives online, ensuring data privacy is crucial to protecting the integrity of personal identities.

Moreover, it should not be forgotten that a further obstacle to regulation is offered (in addition to the speed at which new technologies evolve) by the countless possible implications of the application of the technologies (it has already been mentioned above how with AI we can see that the same artificial intelligence can be used in both an attacking and a defensive function). Not only that, but different technologies also often imply different problems. Thus, the jurist who turns his attention to the regulation of these issues cannot forget a fundamental difference: while the use of AI is – at least for the moment – circumscribed within certain coordinates (those given to the model by the programmers), this is not – and will not be – the case for the metaverse, through which one will probably be able to take decisions that will also directly affect one's 'real' life (this also poses problems of 'control': as we have seen, in the metaverse it will probably be necessary to deal with deviations from the lawful, even malicious ones).

6. *A provocation. Conclusions.*

The rapid integration of AI and virtual identities into daily life poses profound implications for the right to personal identity. As these digital landscapes continue to evolve, legal and regulatory frameworks must adapt themselves, safeguarding these rights in order to reach a user-centered approach to digital identity management.

In sum, as society continues to embrace these technologies, careful consideration must be given to fostering environments where individuals can safely and authentically express their identities in the digital realm.

Finally, a cue from which relevant ethical implications could well arise: the avatars of the metaverse will be built by users through AI tools: it is clear to all that the metaverse will be closely related to artificial intelligence-and it is no coincidence that the AI Act provides for a special discipline, instituting transparency burdens as well as impact assessment on the protection of individuals⁷². This will, certainly, be an inevitable transformation: it has already been mentioned how even today technology potentially allows anyone to create an avatar that has their own likeness and voice. But not only that: the avatars of the future, in fact, will be able to “inherit” from their creator in the real world even ideas, political preferences, cultural opinions, sexual orientation, culinary tastes (this is, after all, information that companies such as META already possess for every user).

It will then be possible to ‘build’ a person (albeit a virtual one) in one’s own image and likeness, eventually giving him or her a certain sphere of autonomy (decision-making and otherwise). Almost, it could be said, a conscience. Which, on closer inspection, is a way to acquire what man has always desired: immortality. By means of an identity that goes beyond the person and will continue to speak in that tone of voice, to manifest those kinds of ideas, feelings and passions. Yet, while certainly all of this will be an opportunity – by way of example in the medical field⁷³, or business⁷⁴ or education⁷⁵ – we must not forget how human living is also synonymous with seeing and relating to each other, and how – ultimately – we cannot allow ourselves to be completely taken over by the digital world.

⁷² As is well known, the AI Act was formally approved in March 2024.

⁷³ For a recent overview of the possibilities related to the application of the metaverse in healthcare see R. CHENGODEN *et al.*, *Metaverse for Healthcare: A Survey on Potential Applications, Challenges and Future Directions*, in *IEEE Access*, 11, 2023.

⁷⁴ So much so that new examples of ownership and contractual *synallagma* formation are on the horizon, resulting in a transformation of the individual-goods relationship, as well as of the ways in which rights are determined between the parties. On this point see R. BOCCHINI, *Nuovi beni digitali e mondi dematerializzati. Il metaverso*, in *European Journal of Privacy Law & Technologies*, 1, 2023.

⁷⁵ It is possible, in fact, that the metaverse will represent in the future «the horizon of possibilities within which to work out a new balance in the relationship between experience, self-education, education and, consequently, between school and the world» as F. BUCCINI, *Apprendere nel metaverso: per un uso didattico degli ambienti ibridi*, in *QTimes – Journal of Education – Technology and Social Studies*, 3, 2023.

CAPITOLO III

L'INTELLIGENZA ARTIFICIALE DA CAUSA
A STRUMENTO DI CONTRASTO DELLE
DISCRIMINAZIONI DI GENERE E INTERSEZIONALI:
CONSIDERAZIONI A SEGUITO DI INDAGINE “DELPHI”
Elettra Stradella, Silvia Cervia, Gabriele Rugani

SOMMARIO: 1. Introduzione: domanda di ricerca e metodologia. – 2. Le discriminazioni di genere e intersezionali derivanti dall'intelligenza artificiale. – 3. Le cause delle discriminazioni algoritmiche. – 4. Gli ambiti in cui si possono verificare gli effetti pregiudizievoli derivanti dall'IA. – 5. Il ruolo degli strumenti giuridici esistenti a livello UE nel prevenire e contrastare le discriminazioni derivanti dall'IA. – 6. Le strategie per addivenire a un'IA che non causi, e addirittura contrasti, discriminazioni di genere e intersezionali, con particolare riferimento al coinvolgimento degli *stakeholders*. – 7. Considerazioni conclusive.

1. Introduzione: domanda di ricerca e metodologia.

Il presente contributo mira a interrogarsi su quali possano e debbano essere le strategie non solo per prevenire le discriminazioni di genere e intersezionali derivanti dall'impiego dell'intelligenza artificiale (IA), ma anche per addivenire a sistemi di IA che siano in grado di giocare un ruolo positivo nel contrastare dette discriminazioni.

Al fine di rispondere alla suddetta domanda di ricerca, dal punto di vista metodologico è stata adottata la tecnica c.d. “Delphi”¹, un sondaggio

¹ Le autrici e l'autore dell'articolo ringraziano sentitamente gli esperti che hanno partecipato allo studio, che si riportano di seguito: Carrozza Maria Chiara (Università degli Studi di Milano-Bicocca); Casonato Carlo (Università degli Studi di Trento); Catelani Elisabetta (Università di Pisa); Chiarello Filippo (Università di Pisa); Enrichens Arianna (Avvocata esperta di discriminazioni di genere, violenza e molestie); Fabris Adriano (Università di Pisa); Federici Chiara (Avvocata esperta di discriminazioni di genere e consigliera di fiducia del CNR); Giannotti Fosca (Scuola

qualitativo multistadio autocorreggente², fondato sulla somministrazione di domande ad un panel di esperti in due distinti *round*.

Questa tecnica è particolarmente indicata nei percorsi empirici indirizzati a focalizzare questioni che ruotano attorno ad uno specifico tema e ad ipotizzare scenari futuri³. Non bisogna infatti dimenticare che la Delphi ha origine negli anni Cinquanta del XX secolo presso la Rand Corporation, al fine di rispondere a una domanda della Difesa americana in merito alle conseguenze di possibili attacchi nucleari sovietici. La previsione di tali conseguenze non era estrapolabile con metodi tradizionali quali l'analisi dei trend, poiché il tipo di situazione non presentava dati preesistenti: dal momento che all'epoca non c'erano mai stati (né, a onor del vero, ci sarebbero mai stati) bombardamenti nucleari sovietici, nessuno poteva affidarsi a un precedente. Olaf Helmer, Norman Dalkey e Nicholas Resker, pertanto, idearono la tecnica di cui trattasi, chiamata "Delphi" per la funzione previsionale che aveva. In seguito la Delphi, con diverse varianti, è stata utilizzata fino ai giorni nostri per costruire scenari possibili di breve e medio raggio su un'ampia varietà di tematiche civili. Tale tecnica, tuttavia, può ben essere usata anche prescindendo da una simile funzione predittiva e come semplice strumento di analisi dell'attualità fondata sul giudizio di esperti⁴.

La tecnica "Delphi", pertanto, è parsa particolarmente adatta a rispondere alla domanda di ricerca di cui sopra: essa, infatti, ha permesso non solo di analizzare approfonditamente il tema delle discriminazioni di genere e intersezionali derivanti dall'impiego dell'IA, ma anche di formulare previsioni circa le possibili strategie attraverso cui prevenire e contrastare detto fenomeno.

Più nel dettaglio, durante il primo *round*, i partecipanti – appartenenti a numerose categorie, tra cui accademici in materie STEM, in materie giuridiche e in altre materie SSH, avvocati, giudici, rappresentanti del mondo delle imprese,

Normale Superiore Pisa) con Marchiori Manerba Marta (Università di Pisa); Giomi Elisa (Università Roma Tre); Longo Erik (Università di Firenze); Malgieri Gianclaudio (Università di Leiden); Mancini Emanuele (giudice penale); Matera Maristella (Politecnico di Milano); Neri Veronica (Università di Pisa); Pistone Bruno (Amazon Web Service Italy); Poggio Barbara (Università degli Studi di Trento); Praticchizzo Domenico (Università degli Studi di Siena); Pronzato Riccardo (Università di Bologna); Scialdone Marco (Avvocato); Tarquini Elisabetta (giudice del lavoro); e Torriani Paola Maria (Università di Torino). Le autrici e l'autore dell'articolo si assumono comunque direttamente la responsabilità delle considerazioni e delle analisi in esso contenute.

² Sul punto v., *inter alios*: A. MATURO, *Governance ed opzioni deliberative*, in C. CIPOLLA (a cura di), *Manuale di sociologia della salute. Ricerca*, vol. 2, Milano, 2004.

³ Sul punto v., *inter alios*: H.J. STRAUSS, L.H. ZEIGLER, *The Delphi technique and its uses in social science research*, in *The Journal of Creative Behavior*, 9, 1975, 253 ss.

⁴ C. BEZZI, *Fare ricerca con i gruppi*, Milano, 2013, 144.

nonché esperti di questioni di genere – hanno compilato una scheda con cinque domande aperte riguardanti i seguenti aspetti: (1) consapevolezza della problematica rappresentata dalle discriminazioni di genere e intersezionali derivanti dall'intelligenza artificiale; (2) cause delle discriminazioni algoritmiche; (3) ambiti in cui si possono verificare gli effetti discriminatori derivanti dall'IA; (4) ruolo degli strumenti giuridici esistenti a livello UE nel prevenire e contrastare le discriminazioni derivanti dall'IA; (5) strategie per addivenire a un'IA che non causi, e addirittura contrasti, le discriminazioni di genere e intersezionali, con particolare riferimento al coinvolgimento degli *stakeholders*. Nel secondo *round*, invece, gli esperti sono stati invitati a integrare le risposte fornite durante la prima fase attraverso domande più specifiche, a risposta chiusa, sui medesimi temi.

Il sondaggio è stato distribuito via *e-mail*. La rilevazione si è svolta tra giugno 2024 e febbraio 2025.

2. Le discriminazioni di genere e intersezionali derivanti dall'intelligenza artificiale.

È ormai da tempo ampiamente dimostrato – da analisi statistiche (quali, ad esempio, quelle del *World Economic Forum* ⁵ e dell'UNESCO ⁶), studi accademici (tra cui il progetto “*Gender Shades*” sulle tecniche di riconoscimento facciale ⁷), nonché note vicende (si pensi a quella dell'algoritmo progettato da Amazon per il reclutamento ⁸, di cui si dirà meglio *infra*) – che l'intelligenza artificiale incorpora *bias* di carattere culturale che producono discriminazioni strutturali ai danni delle donne – alla luce della natura non *gender-neutral* dei modelli, tra gli altri, di *machine learning* e *deep learning* – nonché di vari gruppi minoritari e vulnerabili ⁹:

⁵ WORLD ECONOMIC FORUM, *The Global Gender Gap Report*, Cologny-Ginevra, 2018.

⁶ UNESCO (UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION), *Artificial Intelligence and Gender Equality*, Parigi, 2020.

⁷ J. BUOLAMWINI, T. GEBRU, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in *Proceedings of Machine Learning Research*, 81, 2018, 77 ss.; J. BUOLAMWINI, *Hearing on: Artificial Intelligence: Societal and Ethical Implications – United States House Committee on Science, Space and Technology*, Washington DC, 2019.

⁸ Sul punto v., *inter alios*: J. DUSTIN, *Amazon scraps secret AI recruiting tool that showed bias against women*, in *Reuters*, 11 ottobre 2018; J. LAURET, *Amazon's sexist AI recruiting tool: how did it go so wrong?*, in *becominghuman.ai*, 16 agosto 2019.

⁹ Sul punto v., *inter alios*: C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, in *Riv. “Gruppo di Pisa”*, 3, 2021, 9 ss.

ci si riferisce a persone, ad esempio, appartenenti a minoranze etnico-razziali, con diverse abilità, non eterosessuali, oppure non cisgender¹⁰.

Particolarmente penalizzato risulta chi presenta contestualmente più fattori di discriminazione, e.g. una donna appartenente a una minoranza etnico-razziale. Come ormai noto, si parla in questi casi di “discriminazioni intersezionali”¹¹: tale concetto, di origine nord-americana¹², è stato successivamente preso in considerazione anche a livello di Nazioni Unite¹³, nonché di istituzioni dell’Unione Europea¹⁴.

¹⁰ Secondo la definizione dell’Enciclopedia Treccani, con l’espressione “cisgender” si intende una «persona nella quale collimano il sesso biologico e l’identità di genere». Pertanto, “non cisgender” è una persona in cui non vi è tale corrispondenza.

¹¹ Sul punto v., *inter alios*: C.A. MACKINNON, *Intersectionality as method: A note*, in *Signs: Journal of Women in Culture and Society*, 38.4, 2013, 1019 ss.; H. LUTZ, *Intersectionality as method*, in *DiGeSt. Journal of Diversity and Gender Studies*, 2.1-2, 2015, 39 ss.; J. BAILEY, V. STEEVES, J. BURKELL, L.R. SHADE, R. RUPARELIA, P. REGAN, *Getting at equality: Research methods informed by the lessons of intersectionality*, in *International Journal of Qualitative Methods*, 18, 2019, 1 ss.; S. GRABE, *Research methods in the study of intersectionality in psychology: Examples informed by a decade of collaborative work with majority World Women’s Grassroots Activism*, in *Frontiers in Psychology*, 11, 2020; C. HAYNES, N.M. JOSEPH, L.D. PATTON, S. STEWART, E.L. ALLEN, *Toward an understanding of intersectionality methodology: A 30-year literature synthesis of black women’s experiences in higher education*, in *Review of Educational Research*, 90.6, 2020, 751 ss.; A.J. PETROZZIELLO, *Intersectionality as method for human rights research: Identifying who is made stateless and how through UN treaty body reviews*, in *Journal of Human Rights*, 24.2, 2025, 182 ss.

¹² K. CRENSHAW, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, in *University of Chicago Legal Forum*, 1, 1989, 139 ss.

¹³ CEDAW (UNITED NATIONS COMMITTEE ON THE ELIMINATION OF DISCRIMINATION AGAINST WOMEN), *General recommendation No. 28 on the core obligations of States parties under article 2 of the Convention on the Elimination of All Forms of Discrimination against Women*, CEDAW/C/GC/28, 16 dicembre 2010, par. 18, in base a cui: «Intersectionality is a basic concept for understanding the scope of the general obligations of States parties [...]. The discrimination of women based on sex and gender is inextricably linked with other factors that affect women, such as race, ethnicity, religion or belief, health, status, age, class, caste and sexual orientation and gender identity. Discrimination on the basis of sex or gender may affect women belonging to such groups to a different degree or in different ways to men. States parties must legally recognize such intersecting forms of discrimination and their compounded negative impact on the women concerned and prohibit them. They also need to adopt and pursue policies and programmes designed to eliminate such occurrences, including, where appropriate, temporary special measures [...]».

¹⁴ PARLAMENTO EUROPEO, *Risoluzione sulla discriminazione intersezionale nell’Unione europea: la situazione socioeconomica delle donne di origine africana, mediorientale, latinoamericana e asiatica (2021/2243(INI))*, 6 luglio 2022, par. A, in base a cui: «per discriminazione intersezionale si

Orbene, pressoché la totalità dei partecipanti afferma non solo di essere a conoscenza delle discriminazioni di genere e intersezionali derivanti dall'IA, ma anche di ritenere tali effetti una priorità da affrontare. Una simile urgenza dipenderebbe da varie ragioni, individuate dagli esperti e ordinate dalla netta maggioranza dei medesimi in termini di rilevanza – dalla più rilevante alla meno rilevante – nel modo che si andrà di seguito a riportare.

Innanzitutto, ragioni di equità, giustizia sociale, inclusività e pari opportunità: gli algoritmi che perpetuano o amplificano discriminazioni esistenti, infatti, rafforzano le disuguaglianze sociali a danno di alcune fasce della popolazione in svariati contesti: lavoro, istruzione, salute, frontiere, *social media* e molti altri.

In secondo luogo, ragioni di affidabilità e trasparenza delle tecnologie: riconoscere e affrontare i *bias* è un passo fondamentale per le aziende e le istituzioni che sviluppano e implementano tecnologie di IA, le quali – se non dimostrano affidabilità e trasparenza – avranno serie difficoltà a creare fiducia nel pubblico.

In terzo luogo, anche ragioni di efficacia delle tecnologie medesime, dal momento che gli algoritmi che presentano *bias* risultano pure meno accurati ed efficaci.

In ultima istanza, vi è anche la necessità di evitare danni economici, con conseguente perdita di competitività, per le imprese e il sistema. Non bisogna infatti dimenticare che, nel breve/medio periodo, aziende ed organizzazioni che discriminano subiscono danni di diverso tipo: in particolare, possono soffrire conseguenze legali, derivanti dalla violazione di norme anti-discriminazione, nonché danni reputazionali.

Dalle ragioni di cui sopra, deriva quindi l'esigenza di affrontare prioritariamente la questione delle discriminazioni di genere e intersezionali derivanti dall'IA.

3. *Le cause delle discriminazioni algoritmiche.*

Le cause per cui l'IA può produrre effetti discriminatori possono risiedere in diversi fattori che, secondo autorevole dottrina¹⁵, sono specialmente i seguenti:

intende una situazione nella quale intervengono diversi motivi di discriminazione che interagiscono tra loro, per esempio il genere con altri motivi di discriminazione quali razza, colore della pelle, status etnico o socioeconomico, età, orientamento sessuale, identità ed espressione di genere, caratteristiche sessuali, caratteristiche genetiche, religione o credo, nazionalità, status di residente, provenienza da un contesto migratorio o disabilità, tra gli altri, in modo tale da essere indissociabili e da produrre tipologie specifiche di discriminazione».

¹⁵ Sul punto, v. *inter alios*: S. BAROCAS, A.D. SELBST, *Big data disparate impact*, in *California Law Review*, 104, 2016, 671 ss.; F.Z. BORGESIU, *Discrimination, artificial intelligence, and*

l'individuazione e le relazioni istituite tra la caratteristica che il sistema ricerca (c.d. "*target variable*") e la categoria che le viene associata (c.d. "*class label*")¹⁶; il "*data training*", i.e. la raccolta e la selezione dei dati, poiché più essi sono parziali o frutto di pregiudizi, più il modello ne rifletterà la parzialità, agendo in modo discriminatorio (si usa parlare di regola del "*garbage in, garbage out*")¹⁷; la selezione delle caratteristiche individuali rilevanti per la costruzione del modello (c.d. "*feature selection*")¹⁸; la scelta del c.d. "*proxy*", l'elemento, cioè, in base al quale la macchina differenzia¹⁹; nonché la discriminazione intenzionale, cioè consapevolmente posta in essere dal programmatore, che discenda da una volontaria e parziale selezione di dati che si riverbera negativamente ai danni di una o più classi protette ("*masking*")²⁰.

I partecipanti all'indagine Delphi, tendenzialmente, ritengono che tutti i fattori di cui sopra possano trovarsi alla base degli effetti discriminatori dell'IA,

algorithmic decision-making, Strasburgo, 2018, 10 ss.; P. ZUDDAS, *Intelligenza artificiale e discriminazioni*, in *Consulta Online*, 2020, 1 ss.; C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, cit., 17 ss.

¹⁶Le tecniche di IA, infatti, funzionano sulla base di associazioni tra i due parametri suddetti: al ricorrere dell'uno, la "*target variable*", la tecnica di IA vi associa una determinata categoria, la "*class label*", operando una scelta. Il problema sorge quando si chiede di individuare e selezionare caratteristiche da collegare in via automatica a categorie di difficile definizione (come il "buon lavoratore" o il "creditore affidabile"). Le scelte fatte in tal senso nella fase di elaborazione del modello possono essere irragionevoli e discriminatorie, provocando pertanto effetti deteriori ai danni delle classi protette. Sul punto, v. *inter alios*: S. BAROCAS, A.D. SELBST, *Big data disparate impact*, cit., 677 ss.; T.B. GILLIS, J.L. SPIESS, *Big Data and Discrimination*, in *The University of Chicago Law Review*, 86.2, 2019, 459 ss.

¹⁷Sul punto v., *inter alios*: G. RESTA, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Pol. dir.*, 2, 2019, 199 ss.; A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal-Riv. BioDiritto*, 1, 2019, 63 ss.

¹⁸Una delimitazione parziale che ometta di considerare tutti i profili, oppure ne accentui taluni a discapito di altri, si traduce infatti in un funzionamento potenzialmente discriminatorio della macchina.

¹⁹La scelta del *proxy*, che non dovrebbe ricadere su fattori di discriminazione classici, talvolta ne risulta ugualmente predittivo: tanto più la scelta del *proxy* investe caratteri che accomunano i membri di una categoria protetta, tanto maggiori saranno le probabilità che la macchina, distinguendo in base a quel *proxy*, farà oggetto coloro che appartengono a quel gruppo di un trattamento discriminatorio.

²⁰Sul punto v., *inter alios*: F.Z. BORGESIU, *Discrimination, artificial intelligence, and algorithmic decision-making*, cit., 13 ss.; J.A. KROLL, J. HUEY, S. BAROCAS, E.W. FELTEN, J.R. REIDENBERG, D.G. ROBINSON, H. YU, *Accountable algorithms*, in *University of Pennsylvania Law Review*, 165, 2017, 633 ss.

ancorché in varia misura. In particolare, pressoché la totalità dei rispondenti concorda nell'ordinare tali fattori – dal più rilevante al meno rilevante – come di seguito descritto.

In primis, quello che desta maggior preoccupazione è indubbiamente il *data training*. A fini di prevenzione e contrasto, sono dunque ritenuti necessari: *dataset* di qualità e rappresentativi; *audit* per identificare e correggere i *bias*; utilizzo di tecniche di *de-biasing*. Vi è poi chi suggerisce correttivi promotori di uguaglianza anche a costo di alterare la rappresentatività e chi propone di affrontare il tema a livello di *business model* con un *impact assessment*. Tuttavia, vi è pure chi esprime scetticismo sulla possibilità di controllare efficacemente la raccolta e la selezione dei dati.

La scelta del *proxy* è poi vista come il secondo elemento più allarmante. Sono pertanto considerate auspicabili misure che assicurino che i suddetti *proxy* non siano correlati a caratteristiche protette.

Il terzo fattore che riceve più attenzione è invece la *feature selection*. È dunque suggeribile scegliere *feature* non correlate a categorie protette e usare tecniche per minimizzare l'impatto di *feature* potenzialmente discriminatorie.

Successivamente, figura la scelta di *target variable* e *class label*. A tal proposito, viene pertanto proposta un'analisi critica delle variabili affinché non riflettano pregiudizi esistenti.

Infine, un commento a parte lo merita il fattore rappresentato dalla discriminazione intenzionale: esso è ritenuto preoccupante dalla quasi totalità dei partecipanti; andando però ancor più nello specifico, la leggera maggioranza ritiene tale fattore allarmante, ancorché meno di altri, ma vi è anche un numero significativo di partecipanti che lo reputa il fattore più preoccupante in assoluto. Per prevenire la discriminazione intenzionale, pertanto, vengono suggeriti: normative e standard etici rigorosi; *audit* esterni e indipendenti sui modelli; trasparenza nello sviluppo dell'IA.

4. *Gli ambiti in cui si possono verificare gli effetti pregiudizievoli derivanti dall'IA.*

Anche in ragione di quanto appena esposto, appare ormai chiaro come gli algoritmi possano produrre effetti assai pericolosi in un gran numero di ambiti differenti, che richiedono pertanto un intervento.

Si è già fatto cenno *supra* alla nota vicenda del funzionamento discriminatorio dell'algoritmo progettato per il reclutamento da Amazon.

Quest'ultima società aveva predisposto uno strumento che usava l'IA per

fornire punteggi ai candidati, da una a cinque stelle, e quelli con il punteggio più alto sarebbero stati assunti. Nel 2015, tuttavia, Amazon ha scoperto che il sistema, quando attribuiva punteggi ai candidati per ruoli di sviluppatori di *software* e altre posizioni ad alto contenuto tecnologico, non lo faceva in modo neutrale rispetto al genere. Questo perché il modello era stato allenato con i dati relativi ai processi di reclutamento di candidati/e dei dieci anni precedenti e, dato che in quel periodo erano stati assunti soprattutto uomini, l'algoritmo aveva operativizzato una implicita preferenza di genere (a favore dei candidati uomini)²¹. Quello appena riportato rappresenta, come detto, solo uno dei casi più noti, ma si potrebbero menzionare vari altri effetti in un'ampia gamma di contesti.

Sul tema, i partecipanti all'indagine Delphi sono tendenzialmente concordi nel ritenere che l'IA possa cagionare effetti pericolosi di diverso tipo. In primo luogo, ovviamente, può perpetuare e amplificare discriminazioni contro determinati gruppi; ma vengono evidenziati anche altri possibili pregiudizi: tra questi, la mancanza di trasparenza, l'erosione della *privacy*, i problemi di sicurezza, la perdita di posti di lavoro dovuta all'automazione, nonché certi rischi per le aziende.

Quanto agli ambiti in cui i suddetti effetti possono verificarsi – e in relazione ai quali si rende dunque opportuno un intervento – pressoché la totalità dei partecipanti ritiene che sia rilevante e prioritario intervenire nell'ambito del lavoro (e.g. selezione, progressione, condizioni di lavoro e retributive, ecc.). La stragrande maggioranza dei rispondenti reputa inoltre rilevante e prioritario un intervento negli ambiti di giustizia, ordine pubblico e sicurezza (e.g. previsione del rischio di recidiva, determinazione della pericolosità sociale, ecc.), nonché di migrazioni e frontiere (e.g. esame delle domande di asilo, visto o permesso di soggiorno, ecc.). Un'ampia maggioranza, poi, considera rilevante e prioritario un intervento negli ambiti di sanità (e.g. accesso alle cure, qualità delle stesse, ecc.), istruzione (e.g. accesso a istituti di istruzione, valutazione dell'apprendimento, ecc.) e media (e.g. diffusione di disinformazione e propaganda, ecc.). Una leggera maggioranza ritiene altresì rilevante e prioritario un intervento in ambito bancario (e.g. valutazione dell'affidabilità creditizia per l'accesso al credito, ecc.). Quanto all'accesso ad altri servizi (e.g. valutazione dell'ammissibilità alle prestazioni, determinazione dei prezzi dei servizi, ecc.), un'ampia maggioranza dei rispondenti ritiene che un intervento sia rilevante, ma non prioritario.

²¹ Sul punto v., *inter alios*: J. DUSTIN, *Amazon scraps secret AI recruiting tool that showed bias against women*, cit.; J. LAURET, *Amazon's sexist AI recruiting tool: how did it go so wrong?*, cit.

5. *Il ruolo degli strumenti giuridici esistenti a livello UE nel prevenire e contrastare le discriminazioni derivanti dall'IA.*

Giunti a questo punto, occorre valutare l'adeguatezza del quadro giuridico esistente a livello di Unione Europea a prevenire e contrastare le discriminazioni derivanti dall'IA.

A tal proposito, non si può non evidenziare che, con specifico riferimento alle discriminazioni di genere e intersezionali, gli strumenti classici del diritto antidiscriminatorio, come le due epocali direttive CE del 2000 (direttiva 2000/78/CE «che stabilisce un quadro generale per la parità di trattamento in materia di occupazione e di condizioni di lavoro»²² e direttiva 2000/43/CE «che attua il principio della parità di trattamento fra le persone indipendentemente dalla razza e dall'origine etnica»²³, nonché le nozioni ivi contenute, come quelle di discriminazione “diretta” e “indiretta”, mostrano una certa inadeguatezza a ricomprendere le peculiarità della discriminazione derivante dal ricorso a tecniche di intelligenza artificiale²⁴.

Tuttavia, più recentemente, l'Unione Europea si è concentrata in modo più specifico sulla disciplina della c.d. “*AI-derived discrimination*”. In prima battuta, è possibile menzionare il regolamento (UE) 2016/679 (regolamento generale sulla protezione dei dati o GDPR)²⁵, che contiene alcuni riferimenti alla discriminazione e dedica ampio spazio a concetti strettamente connessi come quello di “profilazione”²⁶.

²² Direttiva 2000/78/CE del Consiglio, del 27 novembre 2000, «che stabilisce un quadro generale per la parità di trattamento in materia di occupazione e di condizioni di lavoro», *GUL* L 303 del 2 dicembre 2000, 16-22.

²³ Direttiva 2000/43/CE del Consiglio, del 29 giugno 2000, «che attua il principio della parità di trattamento fra le persone indipendentemente dalla razza e dall'origine etnica», *GUL* L 180 del 19 luglio 2000, 22-26.

²⁴ In tal senso v. C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, cit., 42 ss.

²⁵ 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)», *GUL* L 119 del 4 maggio 2016, 1-88.

²⁶ La “profilazione” è definita, dall'art. 4, n. 4), del GDPR, come «qualsiasi forma di trattamento automatizzato di dati personali consistente nell'utilizzo di tali dati personali per valutare determinati aspetti personali relativi a una persona fisica, in particolare per analizzare o prevedere aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze personali, gli interessi, l'affidabilità, il comportamento, l'ubicazione o gli spostamenti di detta persona fisica». L'art. 22 del GDPR, poi, è dedicato proprio al «Processo decisionale

Soprattutto, però, occorre far riferimento all'ancora più recente regolamento (UE) 2024/1689²⁷, meglio noto come "AI Act" (AIA) o "Legge sull'intelligenza artificiale"²⁸, che prende assai sul serio la discriminazione algoritmica, fin dai Considerando²⁹, ponendosi l'obiettivo di minimizzare i rischi legati alla medesima in relazione a tutte le fasi della ideazione e della messa in funzione delle tecniche di IA.

Tale strumento, come noto, adotta un approccio basato sul rischio o "*risk-based approach*"³⁰, assoggettando le tecniche di IA a regole differenziate in dipendenza del rispettivo grado di rischio: se un sistema di IA crea un rischio "inaccettabile", vi è un divieto totale (Capo II AIA); se il rischio è "alto", sono previsti rigorosi obblighi per fornitori³¹ e *deployers*³² (Capo III AIA); se il

automatizzato relativo alle persone fisiche, compresa la profilazione» e prevede quanto segue: «1. L'interessato ha il diritto di 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»; sono previste alcune eccezioni.

²⁷ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, «che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)», *GUL* 2024/1689 del 12 luglio 2024.

²⁸ Sull'*AI Act*, v. *inter alios*: C.N. PEHLIVAN, N. FORGÓ, P. VALCKE, *The EU Artificial Intelligence (AI) Act: A Commentary*, Wolters Kluwer, Alphen aan den Rijn, 2025.

²⁹ Ad esempio, il Considerando 58 dell'*AI Act* recita come segue: «[...] I sistemi di IA [...] possono portare alla discriminazione fra persone o gruppi e possono perpetuare modelli storici di discriminazione, come quella basata sull'origine razziale o etnica, sul genere, sulle disabilità, sull'età o sull'orientamento sessuale, o possono dar vita a nuove forme di impatti discriminatori [...]».

³⁰ Sul "*risk-based approach*", v. *inter alios*: G. DE GREGORIO, P. DUNN, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age*, in *Common Market Law Review*, 59.2, 2022, 473 ss.

³¹ Un "fornitore" è definito dall'art. 3, n. 3), dell'*AI Act* come «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o un modello di IA per finalità generali o che fa sviluppare un sistema di IA o un modello di IA per finalità generali e immette tale sistema o modello sul mercato o mette in servizio il sistema di IA con il proprio nome o marchio, a titolo oneroso o gratuito».

³² Un "*deployer*" è definito dall'art. 3, n. 3), dell'*AI Act* come «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale».

rischio è “basso o minimo”, sono previsti solo alcuni obblighi in determinati casi (Capo IV AIA).

Giova evidenziare che i sistemi che presentano i rischi di discriminazione più gravi rientrano tra quelli vietati, mentre i sistemi da cui possono derivare pericoli comunque elevati sono classificati ad alto rischio e sottoposti ad obblighi, numerosi dei quali finalizzati proprio a ridurre i rischi di discriminazione: tra questi, soprattutto quelli concernenti i dati e la loro *governance*, che impongono pratiche tra cui un esame atto a valutare le possibili distorsioni suscettibili di incidere sulla salute e sulla sicurezza delle persone, di avere un impatto negativo sui diritti fondamentali o di comportare discriminazioni vietate dal diritto dell'Unione (art. 10, par. 2, lett. f), AIA), nonché misure adeguate per individuare, prevenire e attenuare le possibili distorsioni di cui sopra (art. 10, par. 2, lett. g), AIA); nonché quelli in materia di accuratezza, robustezza e cbersicurezza, che prescrivono *inter alia* di sviluppare sistemi di IA in modo tale da eliminare o ridurre il più possibile il rischio di output potenzialmente distorti (art. 15, par. 4, AIA)³³. Non deve essere poi trascurato l'obbligo, in capo ad alcuni *deployers*³⁴, di effettuare una valutazione dell'impatto sui diritti fondamentali (“*fundamental rights impact assessment*” o FRIA) che l'uso del sistema di IA può produrre (art. 27)³⁵.

L'AI Act tenta altresì di regolamentare, con una disciplina *ad hoc*, i “modelli di IA per finalità generali” o “*general-purpose AI models*”³⁶, tra cui i grandi

³³ Sul punto v., *inter alios*: K. MEDING, *It's complicated. The relationship of algorithmic fairness and non-discrimination regulations in the EU AI Act*, in arxiv.org, 2025.

³⁴ L'art. 27 dell'*AI Act*, dedicato alla “Valutazione d'impatto sui diritti fondamentali per i sistemi di IA ad alto rischio”, recita come segue: «[...] Prima di utilizzare un sistema di IA ad alto rischio di cui all'articolo 6, paragrafo 2, [...] i deployer che sono organismi di diritto pubblico o sono enti privati che forniscono servizi pubblici e i deployer di sistemi di IA ad alto rischio di cui all'allegato III, punto 5, lettere b) e c), effettuano una valutazione dell'impatto sui diritti fondamentali che l'uso di tale sistema può produrre». Si ricorda che i sistemi di IA ad alto rischio di cui all'allegato III, punto 5, lettere b) e c) sono rispettivamente «i sistemi di IA destinati a essere utilizzati per valutare l'affidabilità creditizia delle persone fisiche o per stabilire il loro merito di credito, a eccezione dei sistemi di IA utilizzati allo scopo di individuare frodi finanziarie» e «i sistemi di IA destinati a essere utilizzati per la valutazione dei rischi e la determinazione dei prezzi in relazione a persone fisiche nel caso di assicurazioni sulla vita e assicurazioni sanitarie».

³⁵ A. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, in *Computer Law & Security Review*, 54, 2024.

³⁶ Un “modello di IA per finalità generali” è definito dall'art. 3, n. 63), dell'*AI Act* come «un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato da una generalità significativa e

modelli di IA generativi ³⁷ come ChatGPT, Gemini e DeepSeek (Capo V AIA): essi, non avendo una finalità predeterminata, mal si collocano all'interno delle categorie di rischio, ma sono comunque in grado di discriminare e perpetuare stereotipi e *bias* ³⁸.

Orbene, in riferimento alla normativa sopra descritta, la quasi totalità dei partecipanti all'indagine Delphi si definisce o completamente d'accordo o comunque parzialmente d'accordo con l'affermazione in base a cui la legislazione dell'Unione Europea sul tema, con particolare riferimento all'AI Act, rappresenta un importante passo avanti, *inter alia* perché viene riconosciuto esplicitamente il problema della discriminazione algoritmica e perché viene adottato un approccio basato sul rischio, vietando i sistemi inaccettabili e stabilendo per i sistemi ad alto rischio obblighi, tra cui la valutazione di impatto sui diritti fondamentali *ex art. 27*. Qualcuno dei rispondenti, inoltre, ritiene di dover contestare la tesi secondo cui l'AI Act sarebbe un freno all'innovazione, dal momento che il diritto deve assumersi il compito di orientare il mercato nella giusta direzione.

Tuttavia, la quasi totalità dei partecipanti all'indagine Delphi si definisce o completamente d'accordo o comunque parzialmente d'accordo anche con il fatto che permangano, nella legislazione di cui sopra, alcune criticità che meriterebbero interventi mirati. Tra gli aspetti da migliorare, vengono citati i seguenti: definire discriminazione diretta e indiretta e includere disposizioni più specifiche su discriminazioni di genere e intersezionali; coinvolgere maggiormente la società civile e i gruppi vulnerabili, magari con una FRIA più robusta; più in generale, migliorare il ruolo della FRIA; adeguare la normativa regolarmente in modo veloce e flessibile; insistere di più sulla formazione di sviluppatori, aziende, enti regolatori, ecc.; assicurare implementazione, monitoraggio e *audit*, anche nei confronti dei *big players*; promuovere politiche di investimento culturale ed educativo; modificare in senso restrittivo le eccezioni dell'art. 6.3, che prevede deroghe alla classificazione ad alto rischio.

sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato».

³⁷ Il considerando 99 dell'*AI Act* recita come segue: «I grandi modelli di IA generativi sono un tipico esempio di modello di IA per finalità generali, dato che consentono una generazione flessibile di contenuti, ad esempio sotto forma di testo, audio, immagini o video, che possono prontamente rispondere a un'ampia gamma di compiti distinti».

³⁸ Sui modelli di IA per finalità generali, v. *inter alios*: N. HELBERGER, N. DIAKOPOULOS, *ChatGPT and the AI Act*, in *Internet Policy Review*, 2023.

Quanto al GDPR, viene ricordato dai rispondenti all'indagine che esso contiene disposizioni che possono indirettamente aiutare a contrastare la discriminazione algoritmica, ma non è specificamente progettato a tal fine e risulta dunque limitato.

Infine, non si può tacere come vari partecipanti evidenzino l'insufficienza degli strumenti giuridici a causa di loro difetti intrinseci: non riescono a stare al passo con gli sviluppi tecnologici; la regolamentazione deve essere completata da una riflessione di ordine etico; l'approccio trasgressione-punizione non è sempre valido come deterrente; occorrerebbe preliminarmente cambiare il possesso di chi detiene le tecnologie di IA e modificare la cultura di partenza.

6. Le strategie per addivenire a un'IA che non causi, e addirittura contrasti, discriminazioni di genere e intersezionali, con particolare riferimento al coinvolgimento degli stakeholders.

Dopo aver preso in considerazione il quadro giuridico attualmente esistente a livello di UE, l'indagine si concentra sulle strategie da predisporre per il prossimo futuro. Una delle principali sfide, infatti, è senza alcun dubbio quella di addivenire a sistemi di IA che non solo non producano discriminazioni di genere e intersezionali, ma che addirittura – come si anticipava in apertura – giochino un ruolo positivo nel contrastare tali discriminazioni. Un'intelligenza artificiale che, in buona sostanza, non debba essere più vista solo come una minaccia, ma che possa diventare un utile strumento verso l'uguaglianza di genere (e non solo).

Orbene, in tale ottica la stragrande maggioranza dei partecipanti ritiene che, per ridurre o addirittura eliminare il rischio che i sistemi di IA producano discriminazioni di genere e intersezionali e trasformarli in uno strumento di contrasto e superamento di tali fenomeni, possa essere utile predisporre meccanismi di partecipazione e consultazione degli *stakeholders*, coinvolgendo minoranze e gruppi svantaggiati. Ciò consentirebbe innanzitutto di rappresentare diverse prospettive e necessità, riducendo i *bias* e promuovendo l'inclusività; ma avrebbe pure altri vantaggi: e.g. aumentare trasparenza e fiducia nell'IA, stimolare soluzioni creative ed efficaci, valorizzare il pluralismo e la democrazia partecipativa.

La strategia consistente nel coinvolgimento degli *stakeholders*, tuttavia, presenta anche criticità e limiti, che a parere dei rispondenti dovrebbero essere affrontati, ancorché con differente grado di priorità. In particolare, l'assoluta totalità dei partecipanti ritiene necessario occuparsi delle difficoltà per i gruppi

con meno risorse e/o visibilità, e un'assai ampia maggioranza ritiene ciò addirittura prioritario. Una netta maggioranza ritiene prioritario anche fronteggiare le difficoltà nel rendere i gruppi autonomi, autosufficienti e indipendenti, nonché quelle nello scegliere gli interessati che parteciperanno. Una significativa percentuale dei rispondenti reputa altresì che debba essere presa in considerazione la questione dell'attribuzione alla consultazione di un effetto vincolante. Tra gli altri fattori su cui concentrarsi figurano poi anche i costi più elevati, i tempi più lunghi e i conflitti e le complicazioni nel processo decisionale.

Per quanto riguarda più nello specifico le migliorie da apportare alla legislazione dell'Unione e segnatamente all'AI Act al fine di meglio disciplinare il dialogo con i portatori di interessi, l'assoluta totalità di coloro che ritengono utile tale dialogo si è dichiarata d'accordo con interventi mirati che possano garantire un maggiore coinvolgimento della società civile e dei gruppi vulnerabili, ad esempio potenziare ulteriormente istituti già presenti come la valutazione di impatto sui diritti fondamentali *ex art. 27* dell'AI Act.

I rispondenti, comunque, fanno riferimento pure a strategie alternative per addivenire a un'IA che non causi, e addirittura contrasti, discriminazioni di genere e intersezionali, anche in virtù delle criticità e dei limiti del coinvolgimento degli *stakeholders* di cui si è detto sopra.

In particolare, pressoché la totalità dei partecipanti è completamente d'accordo sulla necessità che i *team* di sviluppo dei sistemi di IA siano multidisciplinari (composti cioè da tecnologi, esperti di etica, sociologi, giuristi, ecc.), diversificati (in termini di genere, etnia, *background* socio-economico, ecc.) e adeguatamente formati sui temi delle discriminazioni e delle differenze. Inoltre, anche chi non è completamente d'accordo sul punto, lo è quantomeno parzialmente.

L'assoluta totalità dei rispondenti, poi, è completamente d'accordo sulla necessità di prevedere un'adeguata formazione e educazione sui temi delle discriminazioni e delle differenze anche di decisori, aziende ed utilizzatori.

Con riferimento a ulteriori misure da adottare per garantire che i sistemi di IA non producano discriminazioni e/o permettere che possano giocare un ruolo positivo nel contrastarle, la quasi totalità dei rispondenti ritiene molto utile condurre *audit* regolari sui modelli di IA per identificare e correggere *bias*. I partecipanti, inoltre, reputano tutti che sia molto conveniente, o comunque abbastanza conveniente, implementare tecniche di *de-biasing* durante il *pre-processing* dei dati, nonché garantire il più possibile la spiegabilità e la possibilità di contestare le decisioni dell'IA.

La maggioranza dei rispondenti ritiene altresì funzionale creare incentivi per le aziende che sviluppano e utilizzano IA etiche e non discriminatorie.

Può essere anche citata, ancorché in subordine, la misura consistente nel prescindere per quanto possibile dalle caratteristiche che generano discriminazione (genere, età, provenienza geografica, ecc.).

7. *Considerazioni conclusive.*

In conclusione, è possibile evidenziare i principali aspetti su cui gli esperti che hanno partecipato all'indagine Delphi – ancorché appartenenti a differenti categorie, tra cui accademici in materie STEM, in materie giuridiche e in altre materie SSH, avvocati, giudici, rappresentanti del mondo delle imprese, nonché esperti di questioni di genere – si sono trovati a convergere.

I partecipanti non solo sono a conoscenza delle discriminazioni di genere e intersezionali derivanti dall'IA, ma ritengono anche tali effetti una priorità da affrontare: ciò soprattutto per ragioni di equità, giustizia sociale, inclusività e pari opportunità, ma anche per motivi di affidabilità e trasparenza, e pure di efficacia, delle tecnologie, nonché per la necessità di evitare danni economici per le imprese e il sistema.

A detta dei rispondenti, le cause delle discriminazioni di cui sopra possono risiedere in diversi fattori, tra cui quello che desta maggior preoccupazione è indubbiamente il *data training*, ma non è certo l'unico. Quanto agli ambiti in cui i suddetti effetti possono verificarsi – e in relazione ai quali si rende dunque opportuno un intervento – tra di essi figurano soprattutto lavoro; giustizia, ordine pubblico e sicurezza; migrazioni e frontiere; ma anche sanità, istruzione, media e ambito bancario.

Con riferimento all'adeguatezza dell'attuale legislazione UE a prevenire e contrastare le discriminazioni derivanti dall'IA, vi è ampio consenso sul fatto che gli attuali strumenti, con particolare riferimento al GDPR e – ancor di più – all'AI Act, rappresentano importanti passi avanti, ma anche sul fatto che permangono criticità che meriterebbero interventi mirati.

Infine, i partecipanti si sono concentrati sulle strategie da predisporre per il prossimo futuro per addivenire a sistemi di IA che non solo non producano discriminazioni di genere e intersezionali, ma che addirittura giochino un ruolo positivo nel contrastare tali discriminazioni. Particolarmente suggerita è la strategia rappresentata dalla predisposizione di meccanismi per la partecipazione e consultazione degli *stakeholders*: ciò consentirebbe di rappresentare diverse prospettive e necessità, e avrebbe pure altri vantaggi. Tale soluzione presenta però anche criticità e limiti, tra cui la difficoltà nel dare voce ai gruppi con meno risorse e/o visibilità, ma non solo.

Pertanto, si rende necessario sia apportare miglione alla legislazione UE – soprattutto all’AI Act – al fine di meglio disciplinare il dialogo con i portatori di interessi, sia implementare strategie alternative. Tra queste, fare in modo che i *team* di sviluppo dei sistemi di IA siano multidisciplinari, diversificati e adeguatamente formati; formare ed educare anche decisori, aziende ed utilizzatori; condurre *audit* regolari sui modelli di IA; implementare tecniche di *de-biasing* durante il *pre-processing* dei dati; garantire il più possibile la spiegabilità e la possibilità di contestare le decisioni dell’IA; nonché creare incentivi per le aziende che sviluppino e utilizzino IA etiche e non discriminatorie.

La maggiore questione che, tuttavia, resta ancora aperta riguarda quale possa e debba essere il ruolo del diritto nel perseguimento delle strategie in questione.

Infatti, in parte, i partecipanti ritengono che gli strumenti giuridici – in quanto tali – non siano sufficienti a perseguire le finalità di cui trattasi, proprio a causa di *deficit* intrinseci che li contraddistinguono: non riescono a stare al passo con gli sviluppi delle tecnologie; non sempre l’approccio trasgressione-punizione è valido come deterrente; sarebbe più utile una riflessione di ordine etico; sarebbe preliminarmente necessario cambiare il possesso di chi detiene le tecnologie, e così via; ciò, in alcuni casi, porta ad uno “scetticismo” rispetto all’utilità di tali strumenti.

In parte, al contrario, i rispondenti reputano che il diritto sia invece chiamato ad assumersi il compito di orientare il mercato nella giusta direzione. In questa cornice, le critiche rivolte agli strumenti giuridici – e.g. il fatto che rappresenterebbero un freno all’innovazione – sono trascurabili, in quanto rappresentano una costante della storia che si presenta ogniqualevolta si cerchi di regolamentare un nuovo settore.

Non si può tuttavia trascurare come la maggioranza dei rispondenti, in realtà, propenda più per la seconda tra le posizioni di cui sopra, quella in base a cui il diritto è chiamato ad assumersi il compito di orientare il mercato nella giusta direzione.

CAPITOLO IV

CERTIFICAZIONE DI PROVENIENZA
DELLE RISPOSTE PER L'USO PROFESSIONALE
DEI *LARGE LANGUAGE MODEL*

Ernesto Damiani

SOMMARIO: 1. Introduzione. – 2. Provenienza delle risposte LLM. – 2.1. RAG con marcatori verificabili. – 3. Stato dell'arte tecnologico. – 4. Questioni aperte e direzioni future di ricerca. – 5. Conclusioni.

1. *Introduzione.*

La natura “a scatola nera” dei *Large Language Model* (LLM) e la loro propensione agli errori fattuali (le cosiddette allucinazioni) hanno sollevato preoccupazioni riguardanti la verificabilità delle fonti usate dai modelli LLM per generare le loro risposte. Questo articolo esplora il problema della certificazione della provenienza delle risposte dei modelli LLM, sottolineando il ruolo della certificazione nel garantire trasparenza e fiducia nelle risposte dei modelli di intelligenza artificiale, attribuendo correttamente la responsabilità del loro utilizzo professionale. Il nostro contributo si concentra sui meccanismi di certificazione adatti alle tecniche di *Retrieval Augmented Generation* (RAG), identificando gli aspetti critici, come la scalabilità e l'integrazione multimodale, e proponendo direzioni future per i sistemi RAG certificati.

L'avvento dei *Large Language Models* (LLM), come GPT-5, Grok e LLaMA, ha reso accessibile ai professionisti di tutti i settori la generazione avanzata di contenuti in linguaggio naturale. I modelli LLM, addestrati su grandi basi documentali (i “*corpora*”), eccellono nei compiti dialogici, come la generazione di risposte alle domande degli utenti, la sintesi e la traduzione di testi. L'impiego di modelli LLM in settori come l'assistenza legale pro bono rivolta a utenti

“deboli” come minoranze e anziani¹, la sanità e il giornalismo, richiede, oltre alla garanzia della qualità delle risposte dal punto di vista della competenza linguistica, la tracciabilità delle fonti di informazione che hanno contribuito alla generazione delle risposte del modello, garantendo la verificabilità delle risposte rispetto agli errori fattuali (le “allucinazioni”)².

Le allucinazioni degli LLM derivano soprattutto da dati di addestramento obsoleti, che inducono i modelli a produrre risultati plausibili al momento dell’addestramento, ma errati al momento della richiesta³. Un esempio classico di allucinazione di questo tipo riguarda un modello LLM di ambito medico, addestrato su informazioni raccolte prima del 2021. Il modello può raccomandare la prescrizione di idrossiclorochina (o cloroquina) come trattamento per il Covid-19, fornendo una risposta non aggiornata rispetto ai risultati degli studi successivi al momento del suo addestramento.

Oggi, gli LLM sono corredati di sistemi per il *Retrieval Augmented Generation* (RAG), che aggiungono dati e conoscenze esterne alle richieste degli utenti, aiutando il modello LLM a generare output di qualità. Un sistema RAG recupera dati e documenti da Internet o interroga una base di conoscenza esterna per stabilire il contesto di una domanda indirizzata al modello LLM, ma non attribuisce con precisione i risultati alle rispettive fonti. L’effetto è quello di un “telefono senza fili”, in cui le informazioni di contesto intermedio vengono ignorate o riportate parzialmente⁴.

La spinta verso la certificazione della provenienza deriva dall’impiego degli LLM in scenari reali in cui gli errori si propagano in modo dannoso.

Il quadro normativo dell’EU AI Act (2024) impone ai sistemi di intelligenza

¹ C.V. CHIEN, M. KIM, *Generative AI and legal aid: Results from a field study and 100 use cases to bridge the access to justice gap*, in *Loyola of Los Angeles Law Review*, 2, 2024, 383-466, disponibile su: <https://ssrn.com/abstract=4921370>.

² P. LEWIS, E. PEREZ, A. PIKTUS, F. PETRONI, V. KARPUKHIN, N. GOYAL, H. KUTTLER, M. LEWIS, W.T. YIH, T. ROCKTASCH, S. RIEDEL, D. KIELA, *Retrieval-augmented generation for knowledge-intensive NLP tasks*, in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. XXXIII, 2020, 9459-9474, disponibile su: <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Paper.pdf>.

³ Y. GAO, Y. XIONG, X. GAO, K. JIA, J. PAN, Y. BI, Y. DAI, J. SUN, Q. GUO, M. WANG et al., *Retrieval-augmented generation for large language models: A survey 2023*, disponibile su: <https://arxiv.org/abs/2312.10997>.

⁴ Y. GAO, Y. XIONG, X. GAO, K. JIA, J. PAN, Y. BI, Y. DAI, J. SUN, M. WANG, H. WANG, *Retrieval-Augmented Generation with Explanations: A New Paradigm for Explainable AI*, 2024, disponibile su: arXiv preprint Sep. 2024 arXiv:2401.00396 [cs.CL].

artificiale ad alto rischio di garantire la tracciabilità delle decisioni, sottolineando la tracciabilità etica.

La certificazione affronta questo problema fornendo i meccanismi per (i) annotare le integrazioni RAG con le fonti (autentiche e aggiornate) utilizzate per produrle e (ii) preservare tali attribuzioni nelle risposte del modello LLM⁵. L'inclusione delle attribuzioni alle fonti nelle risposte dei LLM limita la responsabilità dei professionisti che utilizzano i modelli LLM nella loro attività professionale nei confronti di terzi⁶. Nei settori altamente regolamentati, come la finanza o la medicina, la certificazione della provenienza deve essere conforme anche alle normative GDPR (art. 5, par. 2 e art. 30) e HIPAA (*Security Rule* 45 CFR 164.316), che impongono la trasparenza sulla provenienza dei dati⁷. Questo contributo discute lo stato dell'arte delle tecniche di certificazione della provenienza delle informazioni presenti nelle risposte dei modelli LLM, concentrandosi sulla certificazione dei sistemi RAG.

Il paragrafo 2 ricorda le informazioni di base; il paragrafo 3 esamina brevemente le tecniche disponibili e i nostri contributi; infine, il paragrafo 4 delinea le nuove sfide e le direzioni di ricerca.

2. Provenienza delle risposte LLM.

Negli LLM, la provenienza di una risposta indica perché il modello linguistico l'ha generata, collegandola alle parole e alle frasi (i “*token*”) di ingresso, al contesto e ai valori dei parametri interni del modello. Le tecniche tradizionali di elaborazione del linguaggio naturale (*Natural Language Processing* – NLP) si basavano su modelli “a scatola nera”, la cui conoscenza era interamente statica e acquisita al momento dell'addestramento⁸. La tecnica RAG, introdotta nel 2020⁹,

⁵ S. ES, J. JAMES, L. ESPINOSA-ANKE, S. SCHOCKAERT, *Ragas: Automated evaluation of retrieval augmented generation*, 2023, disponibile su: <https://arxiv.org/abs/2309.15217>.

⁶ H. KO, X. CHEN, Y. ZHANG, J. LI, L. WANG, *Who bears the burden? LLM usage in professional settings and third-party liability*, in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 2024, 1234-1246.

⁷ G. ZYSKIND, O. NATHAN, A.S. PENTLAND, *Decentralized data provenance and compliance with GDPR and HIPAA*, in *IEEE Transactions on Network and Service Management*, 3, 2022, 2456-2469.

⁸ Y. GAO, Y. XIONG, X. GAO, K. JIA, J. PAN, Y. BI, Y. DAI, J. SUN, Q. GUO, M. WANG *et al.*, *Retrieval-augmented generation for large language models*, cit.

⁹ P. LEWIS, E. PEREZ, A. PIKTUS, F. PETRONI, V. KARPUKHIN, N. GOYAL, H. KUTTLER, M.

fornisce al modello LLM informazioni esterne suppletive, estratte al momento della generazione della risposta, mitigando le limitazioni derivanti dall'uso esclusivo della conoscenza statica. Il sistema RAG recupera informazioni rilevanti da una base documentale tramite identificatori di passaggi terminologicamente e semanticamente “densi”. Questi passaggi vengono poi concatenati alla domanda inviata a un modello LLM per generare la risposta. La certificazione estende la tecnica RAG incorporando marcatori verificabili, ad esempio, “*hash*” crittografici di attribuzione, per garantire la tracciabilità.

2.1. RAG con marcatori verificabili.

Una RAG certificata utilizza marcatori verificabili – tipicamente *hash* crittografici o ancoraggi su registri distribuiti come la Blockchain¹⁰ – che rendono controllabili tutti i contributi.

Un *hash* crittografico è una funzione matematica unilaterale che trasforma un messaggio o una porzione di un documento di lunghezza arbitraria in una stringa di lunghezza fissa e apparentemente casuale, chiamata valore di *hash* o impronta digitale (in inglese, *digest*).

I *digest* hanno diverse proprietà rilevanti in ambito legale. Anzitutto, dato un documento A, è praticamente impossibile trovarne un altro B che produca lo stesso *digest*; allo stesso modo, dato il *digest* di A, è computazionalmente impossibile (o estremamente difficile) invertirlo per ricostruire il documento A originale. Questo garantisce l'unicità dell'impronta: se due documenti hanno lo stesso *hash*, con altissima probabilità sono identici *bit* per *bit*. D'altro canto, cambiare anche solo un *bit* del documento produce un valore di *hash* completamente diverso.

Consideriamo ora un esempio di applicazione di RAG certificata in ambito professionale. Un assistente medico basato su un LLM deve rispondere alla richiesta: “Qual è la terapia di prima linea per l'ipertensione resistente nel 2025?”. Le fasi di derivazione della risposta sono le seguenti:

1. Indicizzazione certificata (*offline*) Ogni porzione o “*chunk*” di un documento autorevole (ad esempio, le linee guida diagnostiche ESC/ESH 2024) viene associato a un *digest* di attribuzione:

LEWIS, W.T. YIH, T. ROCKTÄSCHEL, S. RIEDEL, D. KIELA, *Retrieval-augmented generation for knowledge-intensive NLP tasks*, cit.

¹⁰ L. MAURI, E. DAMIANI, S. CIMATO, *Be your neighbor's miner: Building trust in ledger content via reciprocally useful work*, in *2020 IEEE 13th International Conference on Cloud Computing (CLOUD)*, IEEE, 2020, 53-62.

$hi = \text{SHA256}(ci \parallel \text{URL} \parallel \text{timestamp} \parallel \text{versione})$

Il *digest hi* viene registrato su un registro blockchain con controllo dell'accesso (ad esempio, *Hyperledger Fabric*) o pubblico (ad esempio, *Ethereum Layer-2*).

2. Recupero dei dati con verifica crittografica (*runtime*). Il sistema RAG utilizza un algoritmo per la ricerca di frammenti rilevanti (densi semanticamente) nella base documentale. Supponiamo che la ricerca restituisca diversi *chunk* rilevanti, ciascuno con il proprio *digest* registrato:

verifica
 $\text{chunk}_1 \text{ --- } \rightarrow \quad h1 = \text{a3f1...c9e2}$

Se il valore del *digest* SHA-256 del *chunk* è diverso da $h1$, il *chunk* viene scartato perché non integro.

3. Invio del Prompt certificato al modello LLM. Si costruisce il contesto della domanda (*prompt*) da inviare al modello LLM, specificando il *prompt* propriamente detto, il contesto (*chunk* certificati provenienti dal RAG) e la raccomandazione proveniente dall'applicazione *client*

[Prompt]

Qual è la terapia di prima linea per l'ipertensione resistente nel 2025?

[Contesto certificato]

Fonte: ESC/ESH Guidelines 2024

Hash blockchain: a3f1...c9e2 (verificato 2025-11-25)

Contenuto: "In pazienti con ipertensione resistente, aggiungere spironolattone 25-50 mg/die..."

[Raccomandazione]

Rispondi citando esclusivamente le fonti certificate.

1. Risposta finale con attribuzione verificabile. Il modello LLM genera la risposta in base al *prompt*.

La terapia di prima linea per l'ipertensione resistente è l'aggiunta di spironolattone 25-50 mg/die (Hash blockchain: a3f1...c9e2).

Verifica l'integrità:

<https://blockexplorer.example/tx/a3f1c9e2>

Questo esempio illustra un meccanismo che correda una semplice citazione testuale di una prova crittografica di provenienza, soddisfacendo i requisiti di

trasparenza previsti dall'EU AI Act (Art. 13) e riducendo la responsabilità professionale nei confronti di terzi.

Strumenti più avanzati come RAGAS¹¹ sono in grado di valutare la fedeltà dell'attribuzione, misurando quanto le risposte fornite dal modello LLM all'utente siano allineate ai contesti recuperati dall'interrogazione RAG.

Le rassegne di letteratura¹² classificano i RAG nei paradigmi “*Naive*” (recupero di base), “*Advanced*” (indicizzazione ottimizzata) e “*Modular*” (composto da componenti, ognuno dei quali è aggiornabile separatamente rispetto alla provenienza attraverso codifiche multi-*hash*, come la tecnica del “pettine di *hash*” (*Hash-comb*) introdotta dal nostro gruppo di ricerca¹³.

Nell'ambito dell'iniziativa Human Hall for AI – HH4AI¹⁴ la ricerca sulla robustezza e resilienza delle tecniche RAG si propone di raggiungere i seguenti obiettivi:

- Fiducia e spiegabilità: gli utenti professionali hanno bisogno di citazioni verificabili per controllare le risposte dell'IA che utilizzano e preservare la fiducia dei loro clienti nella qualità del proprio lavoro.
- Conformità ed etica: la certezza della provenienza attenua la responsabilità dell'utente umano in caso di risultati parziali o errati.
- Efficienza e produttività: i sistemi certificati riducono i costi di documentazione attraverso l'aggiornamento dinamico delle basi di conoscenza.

Va notato che le risposte certificate includono collegamenti ipertestuali alla fonte, consentendo una verifica interattiva. Questa caratteristica è importante per l'istruzione professionale; la letteratura¹⁵ dimostra che il RAG certificato migliora la comprensione degli studenti attraverso l'indagine delle fonti.

¹¹ S. ES, J. JAMES, L. ESPINOSA-ANKE, S. SCHOCKAERT, *Ragas: Automated evaluation of retrieval augmented generation*, cit.

¹² Y. GAO, Y. XIONG, X. GAO, K. JIA, J. PAN, Y. BI, Y. DAI, J. SUN, Q. GUO, M. WANG *et al.*, *Retrieval-augmented generation for large language models: A survey*, cit.

¹³ A. ALMAHMOUD, E. DAMIANI, H. OTROK, Hash-comb: *A hierarchical distance-preserving multi-hash data representation for collaborative analytics*, in *IEEE Access*, vol. X, 2022, 34393-34403.

¹⁴ P. CERAVOLO, E. DAMIANI, M.E. D'AMICO, B.D.T. ERB, S. FAVARO, N. FIANO, P. GAMBATESA, S. LA PORTA, S. MAGHOOL, L. MAURI *et al.*, *HH4AI: A methodological framework for ai human rights impact assessment under the EUAI act*, 2025, disponibile su: <https://arxiv.org/abs/2503.18994>.

¹⁵ L. MAURI, E. DAMIANI, S. CIMATO, *Be your neighbor's miner: Building trust in ledger content via reciprocally useful work*, cit.

3. *Stato dell'arte tecnologico.*

La certificazione degli LLM sfrutta appieno la struttura tripartita dei sistemi RAG descritta in precedenza: recupero, aumento e generazione. In questa sezione esaminiamo i principali metodi di selezione delle fonti, concentrandoci sulle varianti certificate.

- **Recupero affidabile:** i sistemi RAG interattivi, come ColBERT¹⁶, utilizzano l'interazione successiva con il professionista umano per selezionare le fonti certificate tramite *hash*. Altri sistemi¹⁷ eseguono il recupero mediante una tecnica ibrida basata su semantica e parole chiave. ProvSEEK¹⁸ integra la tecnica RAG con un archivio sicuro che contiene le informazioni di provenienza, utilizzando agenti informatici per mitigare le minacce di sicurezza informatica all'archivio stesso.
- **Aumento della risposta:** la tecnica di certificazione detta "aumento della risposta" non crea un contesto separato, ma inserisce le informazioni recuperate dal RAG direttamente nella richiesta dell'utente; la certificazione non impiega quindi la tecnica dei *digest*, ma sfrutta tecniche di elaborazione del linguaggio naturale (NLP) per attribuire alla risposta un punteggio di fedeltà sulla base del vocabolario utilizzato¹⁹. Il sistema RAGE²⁰ utilizza il cosiddetto *pruning* ablativo per studiare l'impatto del RAG, rimuovendo le sezioni di contesto e osservando i cambiamenti della risposta fornita dal modello LLM.

¹⁶ K. SANTHANAM, O. KHATTAB, J. SAAD-FALCON, C. POTTS, M. ZAHARIA, *ColBERTv2: Effective and efficient retrieval via lightweight late interaction*, in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2022)*, Seattle, United States: Association for Computational Linguistics, luglio 2022, 3715-3734, disponibile su: <https://aclanthology.org/2022.naacl-main.272>.

¹⁷ Y. GAO, Y. XIONG, X. GAO, K. JIA, J. PAN, Y. BI, Y. DAI, J. SUN, Q. GUO, M. WANG *et al.*, *Retrieval-augmented generation for large language models: A survey*, cit.

¹⁸ K. MUKHERJEE, M. KANTARCIOGLU, *LLM-driven provenance forensics for threat investigation and detection*, 2025, disponibile su: <https://arxiv.org/abs/2508.21323>.

¹⁹ H. SANKARAMAN, M.N. YASIN, T. SORENSEN, A. DI BARI, A. STOLCKE, *Provenance: A light-weight fact-checker for retrieval augmented LLM generation output*, in F. DERNONCOURT, D. PREOTIUC-PIETRO, A. SHIMORINA (a cura di), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, Association for Computational Linguistics, 2024, 1305-1313 disponibile su: <https://aclanthology.org/2024.emnlp-industry.97>.

²⁰ Y. GAO, Y. XIONG, X. GAO, K. JIA, J. PAN, Y. BI, Y. DAI, J. SUN, M. WANG, H. WANG, *Retrieval-Augmented Generation with Explanations: A New Paradigm for Explainable AI*, cit.

- Generazione con output verificabili: un'interessante estensione rispetto alla certificazione che abbiamo descritto in precedenza ²¹ consiste nell'utilizzo di metriche calcolate da LLM "terzi" (*LLM-as-judge*) come G-Eval ²². Hoshi e altri hanno realizzato un sistema interattivo, RaLLe ²³, che confronta le informazioni recuperate via RAG con altre interrogazioni, usando metriche di provenienza per stimarne l'affidabilità.
- Database vettoriali certificati: lavori recenti come QA-RAG ²⁴ testano la robustezza delle risposte, utilizzando vettori (e database vettoriali) al posto dei *digest* ²⁵ per l'integrazione certificata dell'informazione esterna. I *database* vettoriali certificati rappresentano oggi l'avanguardia nella ricerca di dati ad alta dimensionalità per applicazioni di intelligenza artificiale, in grado di eseguire ricerche su miliardi di vettori in tempi dell'ordine dei millisecondi.

4. Questioni aperte e direzioni future di ricerca.

Nonostante i progressi fin qui descritti, persistono numerose questioni aperte ai contributi di ricerca interdisciplinari:

- Scalabilità: il recupero dei *chunk* rilevanti, estraendoli da corpora formati da centinaia di migliaia di documenti (corrispondenti a miliardi di miliardi di token) risulta impossibile con le tecniche di ricerca informatiche tradizionali; le tecniche ibride classico-quantistiche per la ricerca di passaggi

²¹ P. LEWIS, E. PEREZ, A. PIKTUS, F. PETRONI, V. KARPUKHIN, N. GOYAL, H. KUTTLER, M. LEWIS, W.T. YIH, T. ROCKTASCH, S. RIEDEL, D. KIELA, *Retrieval-augmented generation for knowledge-intensive NLP tasks*, in *Advances in Neural Information Processing Systems (NeurIPS)*, cit.

²² Y. LIU, D. ITER, Y. XU, S. WANG, R. XU, C. ZHU, *G-eval: NLG evaluation using GPT-4 with better human alignment*, in H. BOUAMOR, J. PINO, K. BALI (a cura di), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2023, 2511-2522, disponibile su: <https://aclanthology.org/2023.emnlp-main.153>.

²³ Y. HOSHI, D. MIYASHITA, Y. NG, K. TATSUNO, Y. MORIOKA, O. TORII, J. DEGUCHI, Ralle, *A framework for developing and evaluating retrieval-augmented large language models*, in Y. FENG, E. LEFEVER, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, 2023, 52-69, disponibile su: <https://aclanthology.org/2023.emnlp-demo.4>.

²⁴ A. MANSUROVA, A. MANSUROVA, A. NUGUMANOVA, *QA-RAG: Exploring LLM reliance on external knowledge*, in *Big Data and Cognitive Computing*, 9, 2024, 115, disponibile su: <https://www.mdpi.com/2504-2289/8/9/115>.

²⁵ C. ZHOU *et al.*, *Post-training Large Language Models for RAG evaluation*, in *Microsoft Research*, Tech. Rep., 2023.

rilevanti nei documenti sono promettenti e oggetto di ricerca anche in Italia²⁶.

- Multimodalità: le tecniche RAG descritte in questo contributo sono prevalentemente testuali; la certificazione della provenienza di immagini, video ed audio, di grande interesse per la scienza forense richiede una rappresentazione uniforme di testo e immagini²⁷.
- Controllo dei pregiudizi nella certificazione: gli LLM terzi utilizzati negli schemi “LLM-as-judge” possono introdurre o ereditare pregiudizi; l’uso di panel ibridi di valutatori con “human-in-the-loop” può mitigare questo effetto indesiderato.

Le direzioni future di ricerca su cui stiamo lavorando comprendono la messa a punto di RAG federati basati su database vettoriali per la tutela della privacy²⁸ e l’integrazione tra RAG e l’*Edge Computing* per la verifica della provenienza a bassa latenza²⁹. Il nostro gruppo di ricerca persegue attivamente anche la creazione di protocolli RAG industriali basati su blockchain³⁰.

5. Conclusioni.

Certificare la provenienza nelle risposte è essenziale per l’uso professionale dei modelli LLM; le tecniche RAG forniscono uno strumento prezioso per

²⁶ L. LEPORI, A. DUTTA, P. MAZZA, G. CARLEO, *Retrieving past quantum features with deep hybrid classical-quantum reservoir computing*, 2024, disponibile su: <https://arxiv.org/abs/2401.16961>.

²⁷ X. CHEN, C. WANG, Y. XUE, N. ZHANG, X. YANG, Q. LI, Y. SHEN, L. LIANG, J. GU, H. CHEN, *Unified hallucination detection for multimodal large language models*, in L.W. KU, A. MARTINS, V. SRIKUMAR (a cura di), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, 2024, 3235-3252, disponibile su: <https://aclanthology.org/2024.acl-long.178>.

²⁸ M. COLOMBO, R. ASAL, E. DAMIANI, L.M. ALQASSEM, A.A. ALMEMARI, Y. ALHAMMADI, *A quantization-based technique for privacy preserving distributed learning*, in *Future Generation Computer Systems*, 2025, disponibile su: <https://arxiv.org/abs/2406.19418>.

²⁹ C.A. ARDAGNA, R. ASAL, E. DAMIANI, N. EL IOINI, M. ELAHI, C. PAHL, *From trustworthy data to trustworthy IoT: A data collection methodology based on blockchain*, in *ACM Transactions on Cyber-Physical Systems*, 1, 2020, 421-425, disponibile su: <https://dl.acm.org/doi/10.1145/3418686>.

³⁰ D. HAWASHIN, M. NEMER, K. SALAH, R. JAYARAMAN, E. DAMIANI, D. SVETINOVIC, *Blockchain and NFT-based traceability and certification for UAV parts in manufacturing*, in *J. Industrial Information Integration*, vol. XXXV, 2024, 100549.

raggiungere questo scopo e costituiscono un terreno fertile per l'innovazione attraverso la collaborazione tra informatici e giuristi. Con la diffusione degli LLM nelle professioni forensi, la certificazione dei sistemi RAG legali garantirà che le risposte dei modelli non siano solo rapide, ma anche basate su fonti autorevoli. Il nostro gruppo di ricerca proseguirà nello sforzo di integrare il mondo della tecnologia nell'ambito regolatorio, con l'obiettivo di offrire un'intelligenza artificiale sicura e verificabile al servizio dei professionisti e degli utenti finali.

CAPITOLO V

I DIRITTI DELLE PERSONE CON DISABILITÀ AL
COSPETTO DEI SISTEMI DI INTELLIGENZA ARTIFICIALE.
UN'ANALISI DELL'AI ACT E DELLA LEGGE N. 132/2025

*Giuseppe Arconzo, Gaia Patarini**

SOMMARIO: 1. Disabilità e sistemi di intelligenza artificiale: opportunità e rischi. – 2. Le persone con disabilità come soggetti vulnerabili da tutelare secondo l'AI Act. – 3. (*Segue*): le altre condizioni richieste per l'applicazione del sistema di tutela nell'AI Act. – 4. Il rilievo, nell'AI Act, dell'accessibilità dei sistemi di IA. – 5. Persone con disabilità e intelligenza artificiale nella legge italiana: sfide e potenzialità di uno strumento in divenire ... ma a costo zero. – 6. Conclusioni.

1. Disabilità e sistemi di intelligenza artificiale: opportunità e rischi.

Gli studi giuridici, anche di diritto costituzionale, sui fenomeni di intelligenza artificiale (di seguito IA) sono ormai molto numerosi¹. Tra questi, al

* Il presente saggio è frutto delle riflessioni e del lavoro comune dei due Autori, anche se la redazione materiale dei §§ 1, 5 e 6 è ascrivibile a *Giuseppe Arconzo*, mentre quella dei §§ 2, 3 e 4 a *Gaia Patarini*.

¹ Si vedano, *ex multis*, M. D'AMICO, B. LIBERALI, *Intelligenza artificiale e discriminazioni di genere*, in *Manuale di diritto delle donne*, cap. IV, Milano, 2025, 465-486; C. NARDOCCI, *Algoritmi, eguaglianza, discriminazione. Le sfide dell'intelligenza artificiale*, Torino, 2025; O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, Milano, 2025; F. BASILE, M. BIASI, L. CAMALDO, G. CANESCHI, B. FRAGASSO, D. MILANI (a cura di), *Intelligenza artificiale. Diritto, giustizia, economia ed etica*, Torino, 2025; A. SIMONCINI, *Il linguaggio dell'intelligenza artificiale e la tutela costituzionale dei diritti*, in *Riv. AIC*, 2, 2023; A. PAJNO, F. DONATI, A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Diritti fondamentali, dati personali e regolazione*, vol. 1, Bologna, 2022; G.M. RICCIO, G. ZICCARDI, G. SCORZA (a cura di), *Intelligenza artificiale. Profili giuridici*, Padova,

momento, sono però rari quelli che si soffermano sul tema dei rapporti tra IA e diritti delle persone con disabilità². Eppure, sia il testo del regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 (di seguito AI Act), sia quello, più recente, della legge 23 settembre 2025, n. 132 (Disposizioni e deleghe al Governo in materia di intelligenza artificiale) dedicano alcune importanti previsioni a questo tema.

Del resto, è innegabile che i sistemi di IA rappresentino un'opportunità senza precedenti per la tutela dei diritti delle persone con disabilità, potendo potenzialmente contribuire allo sviluppo di tecnologie assistive in grado di garantire loro maggiore indipendenza, autonomia, partecipazione, nonché un migliore accesso a beni e servizi³.

2022; A. PAJNO, M. BASSINI, G. DE GREGORIO, M. MACCHIA, F.P. PATTI, O. POLLICINO, S. QUATTROCOLO, D. SIMEOLI, P. SIRENA, *Al: Profili giuridici. Intelligenza Artificiale: criticità emergenti e sfide per il giurista*, in *Riv. BioDiritto-BioLaw Journal*, 3, 2019, 205 ss.; A. D'ALOIA, *Ripensare il diritto nel tempo dell'intelligenza artificiale*, in G.C. FERRONI, C. FONTANA, E.C. RAFFIOTTA (a cura di), *AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Bologna, 2022, 99 ss.; G. SARTOR, *Artificial intelligence and human rights: between law and ethics*, in *Maastricht Journal of European and Comparative Law*, 27, 6, 2020, 705-719; ID., *L'intelligenza artificiale e il diritto*, Torino, 2022; F. BALAGUER CALLEJON, *La costituzione dell'algoritmo*, Milano, 2023.

² Si tratta soprattutto di dottrina internazionale: cfr. D. NEWMAN-GRIFFIS, J. SAGE RAUCHBERG, R. ALHARBI, L. HICKMAN, H. HOCHHEISER, *Definition drives design: disability models and mechanisms of bias in AI technologies*, in Arxiv.org, 2022, <https://doi.org/10.48550/arXiv.2206.08287>; T. KRUPYI, M. SCHEININ, *Disability discrimination in the digital realm: how the ICRPD applies to artificial intelligence decision-making processes and helps in determining the state of international human rights law*, in *Human Rights Law Review*, 23, 3, 2023; A. GUO, E. KAMAR, J. WORTMAN VAUGHAN, H. WALLACH, M. RINGEL MORRIS, *Towards fairness in AI for people with disabilities: a research roadmap*, in *SIG Access special interest group on accessible computing*, 125, 2019; A. ZARRA, S. FAVALLI, M. CERON, *Artificial intelligence and disability rights*, in A. QUINTAVALLA, J. TEMPERMAN (a cura di), *Artificial intelligence and human rights*, Oxford, 2023, 248 ss.; OECD (ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT), *Using AI to support people with disabilities in the labour market. Opportunities and challenges*, in *OECD Artificial Intelligence Papers*, 7, 2023. Nella dottrina italiana, V. PAGNANELLI, *Non discriminazione e diritto alla diversità: cosa c'è di nuovo per i disabili nell'era dell'AI?*, in *Riv. BioDiritto-BioLaw Journal*, 1, 2024, 331 ss.; F. GIRELLI, *L'inclusione artificiale. Le nuove tecnologie al servizio delle vulnerabilità*, in *Consultaonline*, 1, 2026.

³ Cfr., in generale, V. SINGH, *AI-Powered assistive technologies for people with disabilities: developing ai solutions that aid individuals with various disabilities in daily tasks*, in *Journal of Engineering Research and Reports*, 27, 2025, 292-309 e EUROPEAN DISABILITY FORUM, *Plug and pray? A Disability perspective on artificial intelligence, automated decision-making and emerging technologies*, 2018, <https://www.edf-feph.org/publications/plug-and-pray-2018/>.

Si pensi ad esempio ai progressi nel campo della guida autonoma, che potrebbe garantire una maggiore e più sicura mobilità alle persone con disabilità⁴; ai sistemi di *computer vision*, capaci di descrivere in modo sempre più performante le immagini e di ridurre le difficoltà di quanti convivono con disabilità visive⁵; ai sistemi di sottotitolazione automatica simultanea (*live captioning*) e agli assistenti vocali, che consentono alle persone con disabilità uditiva o intellettiva di comunicare più agevolmente; o, ancora, alle piattaforme di apprendimento adattivo, in grado di fornire un'esperienza didattica altamente personalizzata, attraverso l'analisi dei dati e delle esigenze dello studente.

Dall'altro lato, però, è innegabile che i sistemi di IA pongano alcune sfide da non sottovalutare, destinate ad amplificarsi proprio in relazione alla condizione delle persone con disabilità.

Nello specifico, le principali preoccupazioni ruotano attorno alla perpetrazione, attraverso l'IA, di *bias*, cioè di quelle distorsioni algoritmiche che contribuiscono ad amplificare le discriminazioni già presenti nella società, che possono essere causate dalla rappresentazione non adeguata – a causa di pregiudizi e stereotipi diffusi – della disabilità e della sua eterogeneità nella fase di addestramento dell'IA. In altri termini, le discriminazioni di cui sono e sono state oggetto le persone con disabilità nel passato e nel presente si riflettono nei dati che alimentano i sistemi di intelligenza artificiale, imprimendosi nella logica stessa dell'IA⁶.

In questo senso, ad esempio, alcune ricerche hanno dimostrato chiaramente come atteggiamenti sociali diffusi che definiscono in maniera negativa la disabilità vengano codificati all'interno di quei sistemi di IA che rilevano i discorsi d'odio: un modello di *machine learning* utilizzato per moderare le conversazioni ha in effetti classificato alcuni testi che menzionavano la disabilità come più

⁴ Si veda a tal proposito N. MINISCALCO, *L'intelligenza artificiale in movimento. L'impatto sui diritti costituzionali della smart mobility*, Padova, 2024, in particolare 79-102; C. GRIGIONI, F. CORRADINI, A. ANTONUCCI, J. GUZZI, F. FLAMMINI, *Safe road-crossing by autonomous wheelchairs: a novel dataset and its experimental evaluation* in <https://arxiv.org/abs/2403.08984>.

⁵ Si pensi anche all'applicazione di tali sistemi di intelligenza artificiale ai bastoni utilizzati dalle persone cieche: C. SUNKALP, S. UMANG, K. DEVESH, *A computer vision and depth sensor-powered smart cane for real-time obstacle detection and navigation assistance for the visually impaired*, in <https://arxiv.org/abs/2508.16698>.

⁶ M. WHITTAKER, M. ALPER, C.L. BENNETT, S. HENDREN, L. KAZIUNAS, M. MILLS, M. RINGEL MORRIS, J. RANKIN, E. ROGERS, M. SALAS, S. MYERS WEST, *Disability, Bias, and AI*, New York Center for Disability Studies, 2019, 8. Il report è scaricabile al seguente link <https://ainowinstitute.org/wp-content/uploads/2023/04/disabilitybiasai-2019.pdf>.

tossici e negativi⁷. Inoltre, l'acceso dibattito sui *bias* e sulla necessità di contrastarli attraverso il coinvolgimento delle persone con disabilità, sin dalla fase di progettazione di sistemi di IA inclusivi, chiama in causa anche ulteriori questioni, ugualmente critiche e sfidanti, quali il trattamento dei dati personali, la *privacy*, il consenso.

Le considerazioni sul rapporto tra IA e disabilità possono estendersi, con alcuni accorgimenti, anche ad altri storici fattori di discriminazione, come ad esempio il genere. Del resto, molte delle preoccupazioni, soprattutto sulla capacità fortemente selettiva e discriminatoria che possono avere gli algoritmi, riguardano anche e soprattutto le donne con disabilità, difficilmente coinvolte nella costruzione degli algoritmi. Anche le discriminazioni di cui sono vittime costoro si sedimentano dunque nei dati che alimentano i sistemi di intelligenza artificiale, finendo per influenzarne la logica stessa. Ciò comporta l'ingresso di stereotipi di genere nei dati che vengono utilizzati per addestrare l'IA, che si riproducono nel momento in cui vengono utilizzati⁸.

Indefinitiva, l'IA appare oggi come una straordinaria opportunità di miglioramento di vita, da una parte, e come fonte di ulteriori discriminazioni e potenziali pericoli, dall'altra: sono dunque questi i due opposti poli all'interno dei quali il giurista deve muoversi e i regolatori sono chiamati ad intervenire.

L'obiettivo di questo lavoro è allora proprio quello di verificare, nelle pagine che seguono, le traiettorie seguite dal legislatore europeo e da quello nazionale, al fine di analizzare le prospettive rispettivamente adottate, nonché per identificare i maggiori punti di contatto e le maggiori peculiarità di ciascuna delle due impostazioni.

⁷ Nel testo originale: «*a machine-learned model to moderate conversations classifies texts which mention disability as more "toxic"*». Così, B. HUTCHINSON, V. PRABHAKARAN, E. DENTON, K. WEBSTER, Y. ZHONG, S. DENUYL, *Unintended Machine Learning Biases as Social Barriers for Persons with Disabilities*, in *SIG Access Newsletter*, 125, 2019, 1, <http://sigaccess.org/newsletter/2019-10/hutchinson.html>.

⁸ M. D'AMICO, B. LIBERALI, *Intelligenza artificiale e discriminazioni di genere*, cit., 467-471, in cui si riportano anche alcuni casi: i pregiudizi alla base di algoritmi costruiti solo da uomini hanno ad esempio inciso sul reclutamento dei candidati per determinate posizioni di lavoro da parte di Amazon. L'algoritmo utilizzato favoriva sistematicamente candidati di sesso maschile nella selezione dei curricula, poiché i dati utilizzati per addestrarlo e crearlo derivavano da curricula in prevalenza maschili. Sulle discriminazioni di genere e l'intelligenza artificiale si vedano inoltre C. NARDOCCI, *Algoritmi, eguaglianza, discriminazione*, cit.; ID, *Intelligenza artificiale e discriminazioni*, in P. COSTANZO, P. MAGARÒ, L. TRUCCO (a cura di), *Il diritto costituzionale e le sfide dell'innovazione tecnologica*, Napoli, 2022, 241 ss.; G. CERRINA FERONI, *L'IA discrimina le donne, serve parità algoritmica*, in www.garanteprivacy.it, 16 luglio 2025.

2. *Le persone con disabilità come soggetti vulnerabili da tutelare secondo l'AI Act.*

Due sono le direttrici lungo le quali si muove l'AI Act nel disciplinare il rapporto tra IA e persone con disabilità.

Da una parte, come si evince dal considerando 80, la necessità di garantire la parità e la pienezza di accesso ai sistemi di IA. A tale scopo si ritiene essenziale che i produttori assicurino l'accessibilità degli strumenti di IA a tutte le persone, secondo i principi della progettazione universale. Si tratta di una condizione che si trae dal dovere – direttamente discendente dalla Convenzione ONU sui diritti delle persone con disabilità, di cui l'Unione Europea è firmataria⁹ – di promuovere l'eguaglianza di tali persone. Su questo specifico aspetto si tornerà con maggiore ampiezza nel par. 4.

La prospettiva principale dell'AI Act appare però quella volta a garantire forme di tutela alle persone con disabilità. In effetti, oltre a quanto si dirà immediatamente sull'art. 5, il rischio che i sistemi di IA producano discriminazioni nei confronti delle persone con disabilità è esplicitamente evidenziato nei considerando 32 (con particolare attenzione all'uso di sistemi di IA di identificazione biometrica remota), 56 (con riferimento ai sistemi di IA

⁹ La firma della Convenzione delle Nazioni Unite da parte dell'Unione Europea è avvenuta nel 2007. In seguito, il 26 novembre 2009, con la decisione 2010/48/CE, il Consiglio dell'Unione ha autorizzato l'approvazione della Convenzione, poi conclusasi il 23 dicembre del 2010. Con tale ratifica l'Unione Europea è divenuta per la prima volta parte contraente di un trattato internazionale sui diritti umani, partecipando inoltre attivamente ai negoziati che hanno portato alla redazione dello stesso. Pur essendo parte integrante del diritto dell'Unione – assumendo il rango di fonte intermedia, superiore al diritto derivato dell'Unione, ma inferiore rispetto ai Trattati – la Convenzione ONU non ha tuttavia efficacia diretta: ciò comporta la necessità dell'intervento degli Stati membri al fine di dare attuazione alle disposizioni della convenzione stessa. Ciò però non significa che la Convenzione non dispieghi comunque effetti di tipo interpretativo sul diritto secondario dell'Unione – regolamenti e direttive –, che deve essere interpretato conformemente alle disposizioni convenzionali, come mostra del resto lo stesso *IA Act*. Cfr. in dottrina D. FERRI, *The conclusion of the UN Convention on the Rights of Persons with Disabilities by the EC/EU: a constitutional perspective*, in L. WADDINGTON, G. QUINN (eds.), *European Yearbook of Disability Law*, Cambridge, 2010, 47-71; L. WADDINGTON, *The European Union and the United Nations Convention on the Rights of Persons with Disabilities: a story of exclusive and shared competences*, in *Maastricht Journal of European and Comparative Law*, 18, 4, 2011, 431-458; J. REISS, *Innovative governance in a federal Europe: implementing the Convention on the rights of persons with disabilities*, in *European Law Journal*, 20, 1, 2014, 107-125, nonché, volendo, G. ARCONZO, *I diritti delle persone con disabilità*, Milano, 2020, 78 ss.

nell'istruzione), 57 (con riferimento all'accesso nel mondo del lavoro e, più, in generale, all'occupazione), 58 (con riferimento all'accesso alle prestazioni di assistenza e sicurezza sociale, nonché per l'accesso ai sistemi essenziali e ai sistemi creditizi).

In questa prospettiva, occorre soffermarsi in particolare sull'art. 5 dell'AI Act, che individua le pratiche di IA espressamente vietate. Tra queste, secondo l'art. 5, par. 1, lett. b), «l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo in un modo che provochi o possa ragionevolmente provocare a tale persona o a un'altra persona un danno significativo».

In altri termini, rientrano nelle pratiche vietate dal regolamento europeo la diffusione e l'utilizzo di tutti quei sistemi di IA che perseguano l'obiettivo o abbiano l'effetto di provocare un danno significativo a quanti vivono una particolare condizione di vulnerabilità, che può essere legata a diversi e tipici fattori di discriminazione quali l'età, la disabilità o una specifica situazione sociale ed economica.

Il perno attorno al quale il legislatore europeo ha costruito questa previsione sembra dunque potersi individuare nel concetto di "vulnerabilità", da intendersi come la suscettibilità di subire dei danni (delle vere e proprie ferite, seguendo l'etimologia della parola *vulnera*) ad opera di un agente esterno.

Ci si trova, a dire il vero, di fronte ad una nozione dai confini incerti e su cui da tempo la dottrina delle scienze sociali, della filosofia, della politica e della bioetica si interroga, oscillando tra una visione universalistica e una particolaristica della stessa, per cui da un lato essa viene considerata caratteristica dell'intero genere umano e, dall'altro lato, esclusiva caratteristica di determinati gruppi e categorie di soggetti bisognosi di particolare protezione¹⁰.

Sebbene la visione universalistica abbia il merito di non stigmatizzare e isolare le minoranze, una simile visione rischia però di non prendere in debita

¹⁰ Sul concetto di vulnerabilità v. M.G. BERNARDINI, B. CASALINI, O. GIOLO (a cura di), *Vulnerabilità: etica, politica, diritto*, Roma, 2018; M.G. BERNARDINI, *Il soggetto vulnerabile. Status e prospettive di una categoria (giuridicamente) controversa*, in *Riv. fil. dir.*, 2, 2017, 365-384; M. LONGO, V. LORUBBIO, *Vulnerabilità, rischio e diritti umani tra riflessione sociologica e diritto internazionale*, in *Riv. trim. scienza amm.*, 3, 2021; B. PASTORE, *I molti volti della vulnerabilità*, in M.G. BERNARDINI, V. LORUBBIO (a cura di), *Diritti umani e condizioni di vulnerabilità*, Trento, 17-28; P. SCARLATTI, *I diritti delle persone vulnerabili*, Napoli, 2022.

considerazione la violenza, la discriminazione e lo sfruttamento strutturali che colpiscono alcuni particolari gruppi sociali, quali le persone con disabilità o le minoranze etniche¹¹.

L'AI Act non fornisce una chiara definizione della vulnerabilità a cui fa riferimento. Le scarse indicazioni contenute nelle linee guida della Commissione europea sulle pratiche di IA vietate dall'AI Act¹² sembrano accogliere una visione intermedia. Esse descrivono infatti il concetto di vulnerabilità come un concetto che fa riferimento a un ampio spettro di categorie: si tratta di forme di suscettibilità cognitive, emotive, fisiche e di altra natura, in grado di influenzare la capacità di un individuo o di un gruppo di persone di prendere decisioni consapevoli o, più in generale, in grado di condizionare il loro comportamento¹³.

In realtà, il testo dell'art. 5, par. 1, lett. b), non sembra proprio recepire questa interpretazione. Se è vero che esso fa dapprima riferimento alla vulnerabilità di qualsiasi persona o gruppo, lasciando così presagire un certo margine di elasticità nell'individuazione dei destinatari della norma, in realtà essa finisce poi per circoscrivere alquanto rigidamente le persone tutelate: non si parla in generale delle persone vulnerabili, ma si tratta esclusivamente di coloro la cui vulnerabilità è causata dall'età, dalla disabilità o da una – non meglio precisata – condizione sociale o economica. Altrimenti detto, l'AI Act individua queste, e soltanto queste, come categorie di soggetti che, per una serie di fattori esterni, sociali e culturali, potrebbero presentare una ridotta capacità di riconoscere pratiche manipolative o di sfruttamento da parte dell'IA, e che pertanto sono destinatarie di una tutela rafforzata¹⁴.

Del resto, non è la prima volta che nella normativa dell'Unione Europea viene utilizzato il concetto di vulnerabilità per individuare i destinatari di

¹¹ G. MALGIERI, J. NIKLAS, *Vulnerable data subjects*, in *Computer Law and Security Review*, 37, 2020, 3.

¹² EUROPEAN COMMISSION, *Guidelines on prohibited artificial intelligence (AI) practices, as defined by the AI Act*, 4 febbraio 2025, 32-40, <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act>.

¹³ In originale, nel testo (cfr. par. 102): «*That concept may be understood to encompass a broad spectrum of categories, including cognitive, emotional, physical, and other forms of susceptibility that can affect the ability of an individual or a group of persons to make informed decisions or otherwise influence their behaviour*».

¹⁴ Sulla vulnerabilità nell'AI Act si vedano G. MALGIERI, *Human vulnerability in the EU Artificial Intelligence Act*, in *blog.oup.com*, 27 maggio 2024 e F.P. LEVANTINO, I. NERONI REZENDE, *Rischio inaccettabile: usi proibiti*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., 167-170.

determinate misure¹⁵: basti qui menzionare la direttiva 2005/29/CE del Parlamento europeo e del Consiglio, che all'art. 5, par. 3, sancisce il divieto – da considerare complementare al divieto previsto dall'AI Act all'art. 5, par. 1, lett. b)¹⁶ – di pratiche commerciali sleali che comportino danni economici o finanziari ai consumatori particolarmente vulnerabili, a causa «della loro infermità mentale o fisica, della loro età o ingenuità».

Se si volge ora in particolare l'attenzione ai destinatari della tutela qui apprestata, si può notare che non creano particolari difficoltà interpretative, tra le cause individuate come fattori di vulnerabilità, né l'età, né la disabilità.

Per la prima deve ritenersi rilevante sia la condizione dei bambini e dei minori, sia quella degli anziani. Costoro, infatti, possono essere impossibilitati o più in difficoltà nel cogliere criticamente e prontamente le conseguenze dell'utilizzo degli strumenti di IA.

Per la disabilità, invece, il considerando 29 dell'AI Act rimanda esplicitamente alla definizione accolta nella direttiva 2019/882/UE (*Accessibility act*)¹⁷, secondo cui, in conformità con la Convenzione ONU sui diritti delle persone con disabilità, costoro sono definite come coloro che «hanno minorazioni fisiche, mentali, intellettuali o sensoriali a lungo termine che in interazione con

¹⁵ Si vedano, a titolo esemplificativo, la direttiva 2011/36/UE, concernente la prevenzione e la repressione del traffico degli esseri umani e la protezione delle vittime, in cui l'Unione Europea include tra i soggetti vulnerabili i minori e indica quali aspetti da tenere in considerazione nel valutare la vulnerabilità stessa della vittima il sesso, la gravidanza, lo stato di salute e la disabilità, il fatto di essere vittime di tortura, stupro o di altre forme di abuso sessuale, così come di violenza di genere. Nel medesimo senso si esprime la direttiva 2013/33/UE, recante norme relative all'accoglienza dei richiedenti protezione internazionale, che all'art. 21 prevede che, nelle misure nazionali di attuazione della stessa, gli Stati membri debbano tenere conto della specifica situazione di persone vulnerabili quali i minori, i minori non accompagnati, i disabili, gli anziani, le donne in stato di gravidanza, i genitori singoli con figli minori, le vittime della tratta degli esseri umani, le persone affette da gravi malattie o da disturbi mentali e le persone che hanno subito torture, stupri o altre forme gravi di violenza psicologica, fisica o sessuale, quali le vittime di mutilazioni genitali femminili.

¹⁶ Cfr. considerando 29, AI Act.

¹⁷ Per un'analisi dell'*European Accessibility Act* si rimanda a S. FAVALLI, *Servizi bancari e accessibilità digitale nel diritto internazionale, europeo e nazionale*, in C. RICCI, A. FACCHINETTI (a cura di), *Accessibilità digitale e inclusione delle persone con disabilità: regole e buone pratiche*, Torino, 2024, 49-57; A. BRODERICK, *The European Accessibility Act: a paradigm of inclusive digital equality for persons with disabilities?*, in C. RICCI (a cura di), *Building an inclusive digital society for persons with disabilities. New challenges and future potentials*, Pavia, 2019, 19 ss.; EUROPEAN DISABILITY FORUM, *Our analysis of the European Accessibility Act*, 2019, <https://www.edf-feph.org/newsroom-news-our-analysis-european-accessibility-act/>.

varie barriere possono impedire la loro piena ed effettiva partecipazione nella società su una base di eguaglianza con gli altri»¹⁸.

Viceversa, come già accennato, la nozione di «specificità situazione economica e sociale» appare più vaga. Sul punto, il citato considerando 29 riporta due esempi non particolarmente esaustivi, quali le persone che vivono in condizioni di povertà estrema e le minoranze etniche o religiose. Si deve tra l'altro osservare che non sono propriamente chiare le ragioni che determinerebbero la vulnerabilità di queste ultime con riferimento agli strumenti di IA.

Inoltre, secondo alcune interpretazioni, una lettura complessiva dell'AI Act consentirebbe di ampliare l'elenco di categorie vulnerabili includendo, ad

¹⁸ Sulla definizione di disabilità accolta nella Convenzione ONU cfr., *ex multis*, P. HARPUR, *Embracing the new disability rights paradigm: the importance of the Convention on the Rights of Persons with Disabilities*, in *Disability and Society*, 27, 1, 2012, 1-14; S. FAVALLI, *Disabilità, diritti umani e diritto internazionale*, Milano, 2021, 31 ss. e 101 ss.; G. ARCONZO, *I diritti delle persone con disabilità*, cit., 52-58; S. ROSSI, *Forme della vulnerabilità e attuazione del programma costituzionale*, in *Riv. AIC*, 2, 2017, 16 ss.; L. BUSATTA, *L'universo delle disabilità: per una definizione unitaria di un diritto diseguale*, in F. CORTESE, M. TOMASI (a cura di), *Le definizioni nel diritto. Atti delle giornate di studio 30-31 ottobre 2015*, Trento, 2016, 335 ss.

Quanto previsto dalla Convenzione ONU ha inoltre fortemente influenzato la definizione della disabilità nell'ambito dell'Unione Europea, in particolare ad opera della Corte di Giustizia; a partire dal 2005, stante l'assenza di una definizione di disabilità da parte della direttiva 2000/78/CE, i giudici di Lussemburgo sono stati chiamati in più occasioni a pronunciarsi in via interpretativa proprio sulla nozione di "handicap" contenuta nella direttiva, contribuendo, con una serie di sentenze, a plasmare la nozione di disabilità all'interno dell'ordinamento giuridico europeo. L'adozione della Convenzione ONU da parte dell'UE ha segnato così il passaggio da un'impostazione estremamente rigorosa – in cui la disabilità veniva intesa solo come limite di lungo periodo escludendo da quest'ambito le possibili conseguenze di una malattia – ad una definizione ispirata al modello bio-psicosociale, che certamente include anche le conseguenze di una malattia da cui derivino limitazioni che, in interazione con barriere di diversa natura, ostacolano di fatto la piena ed effettiva partecipazione della persona interessata. Non potendo in questa sede ripercorrere dettagliatamente la giurisprudenza della Corte di Giustizia in materia, si rinvia ad alcune delle sentenze più significative, Corte Giust. UE, sentenza 11 luglio 2006, Sonia Chacon Navas c. Eurest Colectividades SA, causa C-13/05; Corte Giust. UE, Grande sezione, sentenza 17 luglio 2008, Coleman c. Attridge Law, causa C-303/06; Corte Giust. UE, sentenza 11 aprile 2013, HK c. Denmark, cause riunite C-335/11 e C-337/11; Corte Giust. UE, sentenza 18 marzo 2014, C.D. c. S.T., causa C-167/12; Corte Giust. UE, sentenza 18 marzo 2014, Z. c. A Government Department the Board of Management of a Community School, causa C-363/12; Corte Giust. UE, sentenza 1° dicembre 2016, Daouidi c. Bootes Plus sl, causa C-395/15; Corte Giust. UE, sentenza 18 dicembre 2014, FOA c. Kommunernes Landsforening, causa C-354/13; nonché, volendo, alla ricostruzione contenuta in G. ARCONZO, *I diritti delle persone con disabilità*, cit., 89-98.

esempio, le persone in una condizione di basso reddito ovvero appartenenti a gruppi emarginati specifici, quali i migranti e la comunità LGBTQ+¹⁹. In effetti, questa visione sembra essere avallata anche dalle già menzionate e richiamate linee guida della Commissione, che forniscono, ancora una volta, una serie di delucidazioni in merito: la specifica situazione economica e sociale non sta ad indicare una caratteristica individuale unica, bensì uno *status* giuridico o l'appartenenza a uno specifico gruppo sociale o economico vulnerabile, le cui caratteristiche possono essere considerate relativamente stabili e a lungo termine, con la possibilità però che rilevino anche eventuali circostanze provvisorie (disoccupazione temporanea, sovraindebitamento, *status* migratorio, appunto)²⁰. Le linee guida chiedono inoltre, al fine di individuare le persone appartenenti a questo gruppo, di tenere in considerazione le caratteristiche legate a motivi di discriminazione tutelate dal diritto dell'Unione Europea in materia di uguaglianza²¹.

In ragione di ciò, dunque, anche alla luce del richiamo al diritto antidiscriminatorio dell'Unione²², non può effettivamente escludersi che l'elenco delle categorie vulnerabili tutelate dall'IA Act possa essere esteso anche ai soggetti poco sopra richiamati.

¹⁹ G. MALGIERI, *Human vulnerability*, cit.

²⁰ È interessante segnalare che il par. 109 delle linee guida esclude espressamente da questa categoria situazioni quali il lutto e la solitudine, in cui chiunque può trovarsi, ma che non sono tuttavia specifiche da un punto di vista socioeconomico. Viene comunque fatto presente che il loro eventuale sfruttamento può rientrare nell'ambito di applicazione dell'art. 5, par. 1, lett. a), dell'IA Act, che vieta «l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la loro capacità di prendere una decisione informata, inducendole pertanto a prendere una decisione che non avrebbero altrimenti preso, in un modo che provochi o possa ragionevolmente provocare a tale persona, a un'altra persona o a un gruppo di persone un danno significativo».

²¹ Cfr. linee guida, cit., par. 111.

²² Cfr., a titolo esemplificativo, la direttiva 2000/78/CE e la direttiva 2000/43/CE.

3. (Segue): *le altre condizioni richieste per l'applicazione del sistema di tutela nell'AI Act.*

Affinché un sistema di intelligenza artificiale sia vietato ai sensi della norma in commento, non è però sufficiente che questo abbia come obiettivo o effetto lo sfruttamento – inteso come «utilizzo oggettivo di tali vulnerabilità in modo dannoso per le persone o i gruppi di persone sfruttati o per altre persone»²³ – della vulnerabilità delle persone appartenenti alle categorie appena menzionate. Vi sono infatti due ulteriori e necessarie condizioni: in primo luogo, il sistema di IA deve determinare una distorsione materiale dei comportamenti di tali persone e, in secondo luogo, a tale distorsione deve conseguire – o può ragionevolmente conseguire – un «danno significativo» nei suoi confronti o eventualmente anche nei confronti di una persona diversa da quella il cui comportamento ha subito la distorsione evocata.

Anche in questo caso vengono in aiuto le linee guida della Commissione europea, che chiariscono una serie di concetti.

Anzitutto, un sistema di IA determina una «distorsione materiale del comportamento» di una persona o di un gruppo di persone quando, grazie all'utilizzo di tecniche subliminali volutamente manipolative o ingannevoli, è in grado di influenzare il comportamento delle persone a tal punto da pregiudicare in modo considerevole la loro capacità di prendere una decisione informata, inducendole così ad adottare un comportamento o a prendere una decisione che non avrebbero altrimenti assunto²⁴.

Sono chiare le similitudini con la disciplina europea in materia di tutela dei consumatori, in particolare con la direttiva 2005/29/CE – che le linee guida infatti richiamano – sulle pratiche commerciali sleali: essa vieta, al pari dell'*IA Act* con i sistemi di IA qui in esame, quelle pratiche commerciali ingannevoli e aggressive che possono indurre i consumatori a prendere decisioni di natura commerciale che altrimenti non avrebbero assunto. In questo contesto, le linee guida suggeriscono la possibilità di estendere la giurisprudenza della Corte di Giustizia dell'Unione Europea e gli orientamenti della Commissione sulle pratiche commerciali sleali anche ai sistemi di IA vietati, offrendo ulteriori strumenti a supporto dell'interprete. Di conseguenza, ad esempio, così come nel diritto dei consumatori è stato stabilito che non è necessario dimostrare che il comportamento economico del consumatore sia stato falsato, essendo sufficiente determinare che una pratica commerciale «è idonea a» (ossia può) influenzare la

²³ Cfr. linee guida, cit., par. 103.

²⁴ Cfr. linee guida, cit., parr. 77, 78, 79.

decisione di natura commerciale di un consumatore medio²⁵, lo stesso criterio troverà applicazione anche per un sistema di IA vietato²⁶.

Viene poi chiarita anche la nozione di «danno significativo», che sussiste quando sia altamente probabile che dallo sfruttamento derivino dei danni fisici, psicologici, finanziari ed economici che abbiano un rilevante impatto sulla vita di una persona, che per la sua condizione di particolare sensibilità allo sfruttamento, può risultare ancora più grave e complesso²⁷.

Ad una prima lettura della disposizione, e nonostante i chiarimenti appena condotti, non è però così agevole comprendere quali potrebbero essere in concreto i sistemi di IA vietati per i motivi sopracitati.

A questo proposito, e ancora una volta a supporto dell'interprete, le linee guida indicano per ciascuna categoria protetta degli esempi concreti delle misure che potrebbero rientrare nell'ambito del divieto. Così, ad esempio, non sarebbe consentito immettere sul mercato un gioco dotato di IA che – facendo leva sulla naturale curiosità dei bambini e sul loro desiderio di ottenere ricompense – li coinvolga, proponendo sfide di volta in volta più rischiose (arrampicarsi sui mobili, maneggiare oggetti appuntiti) in cambio di ricompense digitali, spingendoli verso comportamenti che possono causare gravi danni fisici.

Sarebbe poi vietato l'utilizzo di un *chatbot* con fini terapeutici di supporto psicologico per persone con disabilità psichiche. Un tale strumento potrebbe infatti sfruttare la loro condizione di vulnerabilità per convincerli ad acquistare prodotti medici particolarmente costosi o spingerli ad assumere comportamenti dannosi per sé stessi e per gli altri²⁸.

Per concludere sul punto, ai sensi dell'art. 5, par. 1, lett. b) dell'AI Act, un sistema di intelligenza artificiale è vietato se sussistono tre condizioni: la vulnerabilità di una persona o di un gruppo di persone; l'obiettivo o l'effetto (anche non intenzionale) di sfruttare tale vulnerabilità; la determinazione, ovvero la ragionevole possibilità, di determinare un danno significativo.

L'approccio del legislatore europeo, al di là dell'utilizzo di criteri piuttosto flessibili e non privi di incertezza, si inserisce nel più ampio disegno normativo dell'Unione Europea relativo alle sfide lanciate dal mondo delle nuove tecnologie:

²⁵ Corte Giust. UE, Sez. V, sentenza 26 ottobre 2016, Canal Digital Danmark A/S, causa C-611/14, punto 73.

²⁶ Cfr. linee guida, cit., par. 80.

²⁷ Cfr. linee guida, cit., parr. 85-97, 115.

²⁸ Gli esempi riportati sono numerosi, per questo si rimanda alle linee guida della Commissione europea, 33-46, <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act>.

esso sembra porsi in linea di continuità con altre esperienze regolatorie – come quella sulla protezione dalle pratiche commerciali sleali o quella in materia di protezione dei dati – concentrandosi soprattutto sui rischi e sui pericoli che il progresso tecnologico pone nei confronti delle categorie più vulnerabili²⁹. Un approccio, se così si può dire, improntato alla cautela, che rischia però di non guidare lo sviluppo delle potenzialità che tali strumenti possono avere nel miglioramento delle condizioni di vita delle persone con disabilità.

4. Il rilievo, nell'AI Act, dell'accessibilità dei sistemi di IA.

Come si accennava all'inizio del par. 2, l'AI Act, oltre alle norme di protezione, si sofferma anche su un'altra questione decisiva per le persone con disabilità: la completa accessibilità e fruibilità dei sistemi di IA. Si tratta di requisiti indispensabili affinché tale sistema possa considerarsi inclusivo e possa così rappresentare un'opportunità e non un ostacolo per le persone con disabilità.

Il tema dell'accessibilità delle tecnologie digitali è stato già affrontato nel recente passato dall'Unione Europea principalmente attraverso due strumenti: la direttiva sull'accessibilità dei siti *web* e il c.d. *European Accessibility Act*³⁰. Quest'ultimo, in conformità alla Convenzione ONU sui diritti delle persone con disabilità, promuove «la piena ed effettiva parità di partecipazione migliorando l'accesso ai prodotti e servizi generici che, grazie alla loro progettazione iniziale o al loro successivo adattamento, rispondono alle esigenze specifiche delle persone con disabilità»³¹, mentre il primo ricorda che, proprio in virtù

²⁹ Sull'approccio dei rischi adottato dal legislatore europeo si vedano, in termini critici, G. FINOCCHIARO, *I molteplici obiettivi perseguiti dal regolamento*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., 107-111; L. BALLOT JONES, J. THORNTON, D. DE SILVA, *Limitations of risk-based artificial intelligence regulation: a structuration theory approach*, in *Discover Artificial Intelligence*, 5, 14, 2025; N. RANGONE, L. MEGALE, *Risks without rights? The EU AI Act's approach to AI in Law and Rule-Making*, in *European Journal of Risk Regulation*, 3, 16, 2025, 1082-1097; V. D'ANTINO, *L'approccio basato sul rischio nell'AI Act*, in *federalismi.it*, 18, 2025; K. NOORI, *Una legge sull'intelligenza artificiale che includa la disabilità. Una guida per monitorare l'attuazione nel tuo paese*, in *EDF*, 2025, 68-69; con riguardo alle altre esperienze regolatorie dell'Unione P. DUNN, *The Artificial Intelligence Act: a title in the EU's digital risk-based approach*, in O. POLLICINO, F. DONATI, G. FINOCCHIARO, F. PAOLUCCI (a cura di), *La disciplina dell'intelligenza artificiale*, cit., 141-157.

³⁰ Direttive 2016/2102/UE (Relativa all'accessibilità dei siti web e delle applicazioni mobili degli enti pubblici) e 2019/882/UE (Sui requisiti di accessibilità dei prodotti e dei servizi).

³¹ Considerando 3, direttiva 2019/882/UE.

della ratifica della Convenzione ONU, «la maggior parte degli Stati membri e l'Unione si sono impegnati ad adottare misure adeguate per garantire alle persone con disabilità, in condizioni di parità con gli altri, l'accesso tra l'altro alle tecnologie e ai sistemi di informazione e comunicazione e a elaborare, adottare e monitorare l'attuazione di norme minime e linee guida per l'accessibilità alle strutture ed ai servizi aperti o forniti al pubblico nonché a promuovere l'accesso delle persone con disabilità ai nuovi sistemi e tecnologie di informazione e comunicazione, compreso internet»³².

Entrambe le direttive hanno però una portata applicativa limitata solo a determinati servizi e prodotti digitali, non esistendo a livello dell'Unione Europea un obbligo di accessibilità trasversale che si applichi a qualsiasi tipo di servizio digitale che transiti nelle competenze dell'Unione³³.

In tale scenario si inserisce l'AI Act che, sebbene presenti alcuni meriti nelle scelte attuate e rappresenti il primo vero tentativo di regolare l'intelligenza artificiale tenendo conto soprattutto dei rischi ad essa collegati, non è esente da criticità.

Vero che, attraverso il già citato considerando 80 del regolamento, il legislatore europeo – almeno in linea teorica – sembra riconoscere la fondamentale importanza dell'accessibilità digitale.

Nello specifico, il considerando 80 ricorda che l'Unione Europea e gli Stati membri, in qualità di firmatari della Convenzione ONU sui diritti delle persone con disabilità «sono tenuti, dal punto di vista giuridico, a proteggere le persone con disabilità dalla discriminazione e a promuoverne l'uguaglianza, a garantire che le persone con disabilità abbiano accesso, su un piano di parità con gli altri, alle tecnologie e ai sistemi di informazione e comunicazione e a garantire il rispetto della vita privata delle persone con disabilità». Nel ricordare ciò, il considerando richiama poi l'importanza dell'applicazione dei principi della progettazione universale a tutti i nuovi servizi e tecnologie, affermando che risulta essenziale che i fornitori si impegnino a garantire la piena conformità ai requisiti di accessibilità, nonché ai requisiti sanciti dalle direttive 2016/2102/UE e 2019/882/UE, sopracitate.

Tuttavia, tale requisito di accessibilità viene poi declinato in maniera diversa a seconda del tipo di sistema di intelligenza artificiale preso in considerazione: così, per i sistemi di IA c.d. ad alto rischio, gli Stati membri devono «garantire» l'accessibilità in conformità alle direttive di cui sopra (art. 16, lett. l); mentre

³² Considerando 12, direttiva 2016/2102/UE.

³³ Cfr. A. FACCHINETTI, *Accessibilità digitale tra attori pubblici e privati: quadro internazionale ed europeo*, in C. RICCI, A. FACCHINETTI (a cura di), *Accessibilità digitale e inclusione delle persone con disabilità*, cit., 25-41.

per la generalità dei sistemi di intelligenza artificiale non ad alto rischio, l'art. 95, par. 1, lett. e), «incoraggia e agevola» soltanto l'elaborazione di codici di condotta, volontari, finalizzati a promuovere l'applicazione di alcuni o di tutti i requisiti previsti per i sistemi ad alto rischio, tra cui la sostenibilità ambientale e l'accessibilità appunto ³⁴.

Tale lacuna era già stata messa in luce e criticata dal Forum europeo sulla disabilità in fase di negoziazione dell'AI Act; l'EDF aveva infatti invitato il Parlamento europeo a includere nel testo un requisito di accessibilità che fosse obbligatorio per tutti i sistemi di IA – requisito che risultava assente nella proposta della Commissione europea – sottolineando che tale assenza avrebbe potuto causare ulteriori barriere per le persone con disabilità ³⁵.

Malgrado ciò, il testo dell'AI Act è rimasto invariato, non prevedendo così un obbligo generale di accessibilità per tutti i sistemi di intelligenza artificiale da esso contemplati.

Non si tratta di un aspetto di poco conto: non prevedere un obbligo di accessibilità per tutti i sistemi di IA può contribuire ad amplificare il rischio per le persone con disabilità di subire discriminazioni nell'interazione con le tecnologie assistive che utilizzano tale tecnologia ³⁶.

Pur muovendo un passo significativo verso la regolazione dei rischi connessi all'utilizzo dei sistemi di intelligenza artificiale e pur riconoscendo – o meglio, dichiarando – l'esigenza di garantire i diritti delle persone con disabilità in conformità alla Convenzione ONU sui diritti delle persone con disabilità, l'AI Act si rivela dunque lacunoso proprio in relazione alla garanzia dell'accessibilità digitale ³⁷. Nell'ambito della regolamentazione dei sistemi di intelligenza

³⁴ Gli unici sistemi di intelligenza artificiale non necessariamente ad alto rischio per cui l'AI Act richiede l'applicazione dei requisiti di accessibilità, sono quelli individuati dall'art. 50, par. 1-4, che includono quelli di generazione o manipolazione di contenuti audio o video che costituiscono il c.d. *deep fake*. Cfr., inoltre, l'art. 71, par. 6, sull'accessibilità della banca dati UE per i sistemi di IA ad alto rischio.

³⁵ EUROPEAN DISABILITY FORUM, *AI Act: European Parliament must ensure accessibility of Artificial Intelligence systems*, 8 maggio 2023, disponibile al seguente link <https://www.edf-feph.org/ai-act-european-parliament-must-ensure-accessibility-of-artificial-intelligence-systems/>.

³⁶ V. A. ZARRA, S. FAVALLI, M. CERON, *Artificial Intelligence and Disability Rights*, cit., 248 ss.

³⁷ Come si vedrà nel paragrafo che segue, la legge n. 132/2025 prevede invece, all'art. 3, comma 7, che venga garantito «il pieno accesso ai sistemi di intelligenza artificiale e alle relative funzionalità o estensioni, su base di uguaglianza e senza alcuna forma di discriminazione e di pregiudizio, in conformità alle disposizioni della Convenzione delle Nazioni Unite sui diritti delle persone con disabilità [...]».

artificiale, il legislatore europeo ha infatti adottato un orientamento basato sulla valutazione del rischio, limitando l'obbligo di garantire l'accessibilità ai soli sistemi classificati come "alto rischio". Tale scelta, tuttavia, non assicura una tutela effettiva per le persone con disabilità. Per quest'ultime, infatti, la questione non si esaurisce nella mera necessità di prevenire fenomeni di manipolazione o di *bias* algoritmico, ma riguarda anche il pericolo – ben più ampio – di esclusione ed emarginazione derivante dall'inaccessibilità stessa delle tecnologie. Ne consegue, dunque, una disciplina parziale, frammentata, suscettibile di generare nuove barriere digitali e di svuotare di contenuto il complessivo sforzo dell'Unione Europea di contrastare i rischi e le discriminazioni connesse all'impiego dei sistemi di IA.

5. Persone con disabilità e intelligenza artificiale nella legge italiana: sfide e potenzialità di uno strumento in divenire ... ma a costo zero.

Rispetto alla prospettiva fatta propria dal legislatore europeo, la legge italiana n. 132/2025 (d'ora in poi legge n. 132) sembra concentrarsi maggiormente sulle potenzialità dell'IA per le persone con disabilità.

Nel tentativo di caratterizzarsi – in una dimensione antropocentrica – quale strumento di guida per una gestione più sicura e sostenibile dell'IA, il legislatore nazionale sembra aver seguito, almeno in parte, l'invito che lo stesso AI Act rivolge agli Stati membri «a sostenere e promuovere la ricerca e lo sviluppo di soluzioni di IA a sostegno di risultati vantaggiosi dal punto di vista sociale e ambientale, come le soluzioni basate sull'IA per aumentare l'accessibilità per le persone con disabilità, affrontare le disuguaglianze socioeconomiche o conseguire obiettivi in materia di ambiente, assegnando risorse sufficienti, compresi i finanziamenti pubblici e dell'Unione»³⁸.

In effetti, tra le disposizioni della legge n. 132, sono proprio quelle riferite ai diritti delle persone con disabilità che consentono di apprezzare le criticità e le potenzialità dell'approccio nazionale all'intelligenza artificiale; un approccio originale, che è stato definito «di sviluppo» alla stregua di una «terza via rispetto

³⁸ Continua «e, se del caso e a condizione che siano soddisfatti i criteri di ammissibilità e selezione, prendendo in considerazione soprattutto i progetti che perseguono tali obiettivi. Tali progetti dovrebbero basarsi sul principio della cooperazione interdisciplinare tra sviluppatori dell'IA, esperti in materia di disuguaglianza e non discriminazione, accessibilità e diritti ambientali, digitali e dei consumatori, nonché personalità accademiche» (considerando 142).

all'approccio Ue di prevenzione e difesa e a quella Usa che si presenta come una strategia di guerra»³⁹.

Come vedremo, le disposizioni della legge n. 132 esplicitamente dedicate ai diritti delle persone con disabilità, pur essendo soltanto tre e pur mantenendosi su un piano di enunciazioni di principio, costituiscono infatti un importante punto di partenza per uno sviluppo etico e responsabile dei sistemi di IA.

Innanzitutto, l'art. 3, che delinea i principi generali a cui la normativa si ispira, prescrive al comma 7 che debba essere garantito «alle persone con disabilità il pieno accesso ai sistemi di intelligenza artificiale e alle relative funzionalità o estensioni, su base di uguaglianza e senza alcuna forma di discriminazione e di pregiudizio, in conformità alle disposizioni della Convenzione delle Nazioni Unite sui diritti delle persone con disabilità». L'articolo non prevede ulteriori e più precise indicazioni, ma evidenzia l'interesse del legislatore ad evitare che le persone con disabilità possano essere escluse dall'accesso all'IA. Le differenze con l'approccio europeo sono alquanto evidenti: a tal proposito si possono dunque richiamare le osservazioni già svolte nel paragrafo 4.

Tuttavia, il più specifico riferimento alle persone con disabilità si trova nel comma 4 dell'art. 7 (*Uso dell'intelligenza artificiale in ambito sanitario e di disabilità*), in cui il legislatore italiano afferma che la legge «promuove lo sviluppo, lo studio e la diffusione di sistemi di intelligenza artificiale che migliorano le condizioni di vita delle persone con disabilità, agevolano l'accessibilità, la mobilità indipendente e l'autonomia, la sicurezza e i processi di inclusione sociale delle medesime persone anche ai fini dell'elaborazione del progetto di vita».

Si è di fronte ad una previsione molto ambiziosa, che riconosce il potenziale ruolo trasformatore che l'IA potrebbe ricoprire nella promozione dell'inclusione e dell'autonomia delle persone con disabilità, anche e soprattutto se si considera l'evoluzione che la materia sta vivendo grazie alla stagione di riforma attualmente in corso nel nostro ordinamento⁴⁰.

³⁹ Così M. D'AMICO, *Intelligenza artificiale: nella via italiana una strategia di sviluppo*, in *Il Sole 24 Ore*, 25 settembre 2025, 33.

⁴⁰ Con legge delega 22 dicembre 2021, n. 227 (delega al Governo in materia di disabilità), prevista nell'ambito del PNRR, è stato infatti inaugurato nel nostro ordinamento un ambizioso e complesso progetto di riforma di una parte significativa della normativa a tutela dei diritti delle persone con disabilità. A seguito dell'approvazione di tale legge, sono stati approvati, tra il dicembre 2023 e il maggio del 2024, tre decreti legislativi di attuazione: il d.lgs. n. 222/2023, che ha introdotto nel nostro ordinamento una definizione di accessibilità più innovativa e coerente con la Convenzione ONU del 2006; il d.lgs. n. 20/2024, che ha istituito l'Autorità Garante Nazionale dei diritti delle persone con disabilità con il compito di vigilare sul rispetto dei diritti delle persone con disabilità; infine, il d.lgs. n. 62/2024 – la parte centrale ma anche più

Tuttavia, e proprio per questo, stupiscono sia l'assenza di un articolo dedicato esclusivamente all'inclusione delle persone con disabilità, sia la collocazione stessa di questa previsione. Il tema della disabilità è infatti inserito nell'articolo dedicato all'uso dell'intelligenza artificiale in ambito sanitario, il che sembra purtroppo suggerire che la disabilità sia ancora percepita come una questione prettamente medica⁴¹. Anche se forse nell'impianto complessivo della legge n. 132 potrebbe sembrare una questione secondaria, è evidente che una diversa collocazione avrebbe consentito al legislatore di dare una maggiore e autonoma rilevanza alla tutela e all'inclusione dei diritti delle persone con

complessa della riforma – che ha introdotto una serie di novità rilevanti, tra cui un nuovo lessico in materia di disabilità, il concetto di accomodamento ragionevole, nuove procedure di accertamento della condizione di disabilità nonché una nuova configurazione del progetto individuale. La riforma, a ormai quattro anni dalla sua approvazione, sta faticando ad avviarsi, tra ritardi e rinvii nell'adozione dei provvedimenti. Si rimanda sul punto a G. ARCONZO, *L'incerto avvio delle riforme in tema di disabilità previste dalla legge n. 227 del 2021*, in *federalismi.it*, 16, 2025, 15-30.

⁴¹ I tre principali modelli che sono stati elaborati nella storia per classificare il concetto di disabilità sono il modello medico-assistenziale, il modello sociale ed infine il modello bio-psico-sociale.

Prima dell'approvazione della Convenzione ONU, il problema dell'inquadramento generale della disabilità veniva tendenzialmente risolto secondo due soluzioni contrapposte: vi era da un lato il c.d. modello medico, che estendeva il tradizionale metodo di classificazione delle malattie (*ICD International Classification of Diseases*) – poi sostituito dal modello ICIDH-1980 – anche alle conseguenze delle malattie, senza dare alcuna rilevanza al fattore ambientale; dall'altro lato vi era, sviluppatosi negli anni Settanta a seguito del *Disability Rights Movement* e scaturito dalla riflessione teorica dei *Disability Studies*, il c.d. modello sociale, che individuava la causa della disabilità solo ed esclusivamente nell'ambiente in cui la persona era inserita. Dalle critiche mosse a queste impostazioni – una troppo incentrata sul concetto di menomazione e l'altra all'opposto dimentica del ruolo che comunque la menomazione ricopre nella definizione della disabilità – scaturì l'ICF (*International Classification of Functioning, Disabilities and Health*), basato sul modello bio-psico-sociale, poi accolto anche a livello giuridico nella Convenzione ONU del 2006. Non essendo questa la sede opportuna per un adeguato approfondimento dell'evoluzione dei modelli classificatori di disabilità, non si può che rinviare alla dottrina in materia. Si vedano, *ex multis*, M.G. BERNARDINI, *Disabilità, giustizia, diritto. Itinerari tra filosofia del diritto e Disability Studies*, Torino, 2016, 1-72; F. MASCI, *Persona, scienza e linguaggio: l'inattuazione del programma costituzionale in materia di disabilità e dei relativi livelli essenziali delle prestazioni*, in *Corti supreme e salute*, 2, 2024, 33 ss.; B. ALTAMAN, *Disability definitions, models, classification schemes, and applications* in G. ALBRECHT, K.D. SEELMAN, M. BURY (a cura di), *Handbook of Disability Studies*, New York, 1996, 97-122; R. MEDEGHINI, E. VALTELLINA, *Quale disabilità? Culture, modelli e processi di inclusione*, Milano, 15-81; M. OLIVER, *The politics of disablement*, London, 1990; T. SHAKESPEARE, *Disability rights and wrongs*, London-New York, 2006.

disabilità, anche alla luce della nuova definizione di disabilità che ha fatto ingresso nel nostro ordinamento con il d.lgs. n. 62/2024⁴².

In modo non dissimile, l'art. 22 (*Misure di sostegno ai giovani e allo sport*), terza e ultima previsione che si riferisce esplicitamente alle persone con disabilità, assume una prospettiva legata alla tutela della salute, pur circoscritta all'attività sportiva. Il comma 3 dispone infatti che «[l]o Stato favorisce l'accessibilità ai sistemi di intelligenza artificiale per il miglioramento del benessere psicofisico attraverso l'attività sportiva, anche ai fini dello sviluppo di soluzioni innovative finalizzate a una maggiore inclusione in ambito sportivo delle persone con disabilità». La formulazione non è delle più chiare e anzi appare alquanto involuta. Essa sembra però lasciar intendere un chiaro *favor* per lo sviluppo di sistemi di IA che possano favorire la pratica sportiva delle persone con disabilità, in ragione degli effetti benefici che lo sport può comportare sulla salute⁴³.

Sebbene poi non esplicitamente riferite alle persone con disabilità, ci sono altre disposizioni destinate ad esplicare i loro effetti anche nei confronti di tali persone: si pensi alle previsioni in materia di trattamento dei dati personali (art. 9); a quelle sull'utilizzo dell'IA in materia di lavoro (art. 11), ovvero alla delega al Governo in materia di dati, algoritmi e metodi matematici per l'addestramento di sistemi di intelligenza artificiale (art. 16), dove in realtà colpisce l'omesso riferimento espresso alle persone con disabilità, spesso vittime – come si è segnalato in precedenza – delle alterazioni algoritmiche derivanti da dati densi di pregiudizi e stereotipi.

⁴² Si tratta di una definizione finalmente conforme a quella contenuta nella Convenzione ONU sui diritti delle persone con disabilità. A seguito della modifica apportata dal d.lgs. n. 62/2024, l'art. 3, comma 1, della legge n. 104/1992 definisce persona con disabilità «chi presenta durature compromissioni fisiche, mentali, intellettive o sensoriali che, in interazione con barriere di diversa natura, possono ostacolare la piena ed effettiva partecipazione nei diversi contesti di vita su base di uguaglianza con gli altri, accertate all'esito della valutazione di base».

⁴³ Sul ruolo dello sport, soprattutto in seguito alla sua introduzione in Costituzione, si vedano M. D'AMICO, *Lo sport come diritto della persona: analisi dei progetti di revisione costituzionale*, in *Gruppo di Pisa*, 1, 2022, 152-167; E. BATTELLI, *Il riconoscimento del valore educativo e sociale dell'attività sportiva in Costituzione*, in *Osservatorio AIC*, 2, 2024, 26-42; G. MARAZZITA, *Il riconoscimento del valore costituzionale dell'attività sportiva*, in *federalismi.it*, 1, 2024, 110-132; M. DI MASI, *Dall'etica alla costituzionalizzazione dello sport. Brevi note sulla riforma dell'articolo 33 della Costituzione*, in *federalismi.it*, 22, 2023, 124-134; G. DI ROSA, *Sport, diritto e inclusione sociale*, in *Nuova giur. civ. comm.*, 3, 2024, 725-729; G. AGRIFOGLIO, *Sport e soggetti fragili: tra ordinamento sportivo, tutele civilistiche e libertà costituzionali*, in *Riv. dir. economia, trasporti e ambiente*, 21, 2023, 1-20; C. SAGONE, *La costituzionalizzazione dell'attività sportiva: alla ricerca di nuovi traguardi (di tutela)*, in *Dir. reg.*, 2, 2023, 380-439.

Come anticipato, quelle descritte si rivelano perlopiù come previsioni di principio, la cui portata effettiva dipenderà soltanto dalla capacità degli attori coinvolti di tradurle in misure concrete; ciò nonostante, esse rivelano l'inedito e più positivo atteggiamento del legislatore italiano nei confronti dell'IA, diverso soprattutto rispetto a quello del legislatore europeo, focalizzato sui rischi piuttosto che sulle potenzialità e opportunità dei sistemi di intelligenza artificiale.

Il legislatore italiano sembra insistere volutamente su alcuni concetti a cui dedica maggiore attenzione ed enfasi: pieno accesso, sviluppo, studio, diffusione, agevolazione, soluzioni innovative; senza dimenticare però di fissare, in conformità all'AI Act, alcuni principi in materia di sicurezza, trasparenza dei dati, proporzionalità, riservatezza, accuratezza, non discriminazione, e sostenibilità.

Una strategia potenzialmente vincente, tradita però da una criticità non di poco conto e che, anzi, potrebbe risultare ostativa alle buone intenzioni scritte nella legge.

Malgrado sia innegabile che per la realizzazione dell'ambizioso quadro costruito dal legislatore con la legge n. 132 saranno necessarie ampie risorse economiche, l'art. 27 della legge pone una clausola di invarianza finanziaria: ciò significa che dall'attuazione delle previsioni legislative non devono derivare nuovi o maggiori oneri per la finanza pubblica. In particolare, è previsto che «le amministrazioni pubbliche interessate provvedono all'adempimento delle disposizioni della presente legge con le risorse umane, strumentali e finanziarie disponibili a legislazione vigente».

La totale assenza di risorse pubbliche con cui indirizzare il percorso di sviluppo e innovazione dell'IA nel campo dei diritti delle persone con disabilità, e non solo, solleva più di qualche dubbio circa la dimensione concreta che assumeranno le misure di attuazione della legge n. 132, che rischiano così di rimanere soltanto dei principi scritti con un inchiostro destinato a cancellarsi.

6. Conclusioni.

Il rapporto tra l'uso dell'IA e la tutela dei diritti delle persone con disabilità si colloca oggi al crocevia tra opportunità straordinarie e rischi altrettanto significativi. L'analisi svolta evidenzia come la disciplina europea e quella nazionale seguano impostazioni concettuali non completamente coincidenti, che riflettono priorità normative e culturali differenti.

Il modello europeo delineato nell'AI Act si caratterizza per una più marcata

impostazione *risk-based*, incentrata sulla prevenzione degli effetti distorsivi e discriminatori che l'impiego dei sistemi di IA può generare su gruppi sociali considerati vulnerabili. In tale quadro, le persone con disabilità vengono per lo più considerate come soggetti esposti a rischi specifici: a fronte di ciò, la risposta regolatoria si struttura attorno a meccanismi di interdizione e di mitigazione, con una scarsa attitudine proattiva. Il mancato riconoscimento di un obbligo generalizzato di accessibilità per tutti i sistemi di IA – previsto solo per quelli ad alto rischio – sembra inoltre rivelare una certa asimmetria tra la proclamata centralità dei principi della progettazione universale e la loro effettiva implementazione normativa.

La legge italiana n. 132/2025, viceversa, adotta un'impostazione più esplicitamente promozionale, attribuendo all'IA un ruolo potenzialmente abilitante nei processi di autonomia, partecipazione e inclusione delle persone con disabilità. Le previsioni dell'art. 7 e, in misura minore, degli artt. 3 e 22, pur muovendosi su un piano prevalentemente programmatico, configurano un orientamento normativo che valorizza l'IA come strumento di *empowerment*, in sintonia con i più recenti sviluppi del diritto interno e internazionale della disabilità. Tuttavia, tale prospettiva risulta significativamente indebolita dalla clausola di invarianza finanziaria di cui all'art. 27, che rischia di relegare la strategia italiana a un livello meramente dichiarativo, privo delle necessarie condizioni attuative.

Il confronto tra i due modelli mostra dunque come la regolazione europea privilegi la logica della protezione, mentre quella italiana propenda per una logica di sviluppo, senza però dotarsi degli strumenti finanziari indispensabili a sostenerla. Una disciplina realmente efficace dell'intelligenza artificiale per le persone con disabilità richiederebbe invece un'integrazione degli approcci: da un lato, un ampliamento delle garanzie di accessibilità e non discriminazione nel sistema europeo; dall'altro, l'allocazione di risorse adeguate che permettano all'ordinamento italiano di tradurre in azioni concrete le enunciate finalità promozionali. Solo attraverso un coordinamento tra misure di tutela e politiche di *empowerment* sarà infatti possibile evitare che l'IA riproduca disuguaglianze sistemiche e, al contrario, diventi uno strumento effettivo di inclusione, autodefinizione e partecipazione sociale.

CAPITOLO VI
MISOGINIA E *HATE SPEECH* ONLINE
NEL DIRITTO DELL'UE
Ilaria Anrò

SOMMARIO: 1. Introduzione: la lotta alla misoginia attraverso il contrasto all'*hate speech* online nel diritto dell'Unione europea. – 2. *Hate speech* online e strumenti di cooperazione giudiziaria penale. – 3. L'inserimento di *hate speech* e *hate crimes* tra i "reati europei". – 4. La risoluzione del Parlamento europeo del 18 gennaio 2024 a favore dell'estensione della lista *ex art. 83 TFUE*. – 5. Gli ostacoli tuttora esistenti e il cambio di strategia della Commissione. – 6. Alcune riflessioni conclusive.

1. *Introduzione: la lotta alla misoginia attraverso il contrasto all'hate speech online nel diritto dell'Unione europea.*

I dati della Mappa contro l'intolleranza, realizzata nell'ambito del progetto Vox e presentata il 13 marzo 2025, mostrano la misoginia tra le manifestazioni più frequenti di odio *online*¹. Secondo questi dati, l'*hate speech* prende di mira sempre più frequentemente il corpo delle donne, diversamente dal passato, quando a essere maggiormente colpita era la professionalità femminile. Un ulteriore dato che suscita preoccupazione riguarda l'origine delle manifestazioni di odio misogino *online*: il 20,81% dello *hate speech* risulta, infatti, proveniente dalle stesse donne (contro il 30,15% degli uomini). Da considerare anche il profilo dell'intersezionalità, essendo la discriminazione di genere sempre più connessa alla religione. A tal proposito, anche il *report* dell'Agenzia europea per i diritti fondamentali del 2025 sottolinea che «*discrimination, often combined with hate and racism, continues to plague today's society, affecting everyone who is seen as different. Hate speech and harassment, including*

¹ Cfr. Mappa dell'Intolleranza 8, <https://www.voxdiritti.it/ottava-edizione-mappa-intolleranza/>.

*hate crime and physical bias-motivated violent crime, are very real threats»*².

Nell'Unione europea la lotta alla discriminazione tra uomo e donna è sempre stata una costante dell'integrazione. Come noto, già nel trattato istitutivo della Comunità economica europea, pur nell'assenza di norme dirette a tutelare i diritti fondamentali, vi era l'art. 119 (oggi art. 157 TFUE) che prevedeva il principio della parità di retribuzione tra lavoratore e lavoratrice per lo stesso lavoro³. Sebbene tale norma fosse stata inserita, su proposta della Francia, per evitare fenomeni di *dumping* sociale, essa è stata considerata fin da subito come un baluardo della lotta alla discriminazione. Tale norma si rivolgeva agli Stati membri imponendo agli stessi di adottare le discipline a livello nazionale per garantire effettività al diritto alla parità retributiva per uno stesso lavoro. Come noto, la Corte di giustizia, con la sentenza *Defrenne II*⁴, ha successivamente rivelato tutto il potenziale di questa norma, sancendone l'effetto diretto e dunque l'invocabilità da parte dei privati anche in situazioni c.d. orizzontali: dunque, una norma del trattato poteva essere invocata direttamente dalla lavoratrice nell'ambito di una controversia con il proprio datore di lavoro, anche privato⁵.

Oggi la Carta dei diritti fondamentali dell'Unione europea prevede il principio della parità tra donne e uomini all'art. 23, come completamento del principio di non discriminazione di cui all'art. 21. La parità uomo-donna è anche un valore iscritto all'art. 2 TUE.

Senz'altro, la lotta alla misoginia affonda le sue radici nella lotta alla discriminazione, ma nell'attuale stadio dell'integrazione europea va oltre e si inserisce in una tendenza più recente di contrasto all'odio in tutte le sue forme. L'Unione europea, infatti, ha sviluppato una serie di norme e strumenti per il contrasto ai discorsi di odio, soprattutto per tenere contro dell'evoluzione delle nuove tecnologie che hanno enfatizzato il fenomeno, consentendone una propagazione al di là dei confini nazionali: il c.d. *hate speech online*⁶. Tali strumenti

² Fundamental Rights Agency, Fundamental Rights Report 2025, spec. 10.

³ Cfr. O. POLLICINO, *Discriminazione sulla base del sesso e trattamento preferenziale nel diritto comunitario*, Milano, 2005, spec. 29 ss.

⁴ Corte Giust. UE, 8 aprile 1976, causa 43/75, *Gabrielle Defrenne contro Société anonyme belge de navigation aérienne Sabena*, ECLI:EU:C:1976:56.

⁵ Sul punto v. F. COSTAMAGNA, *Diritti fondamentali e rapporti tra privati nell'ordinamento dell'Unione europea*, Torino, 2022.

⁶ Sull'emersione del fenomeno e il suo inquadramento giuridico v. G. ZICCARDI, *L'odio online. Violenza verbale e ossessioni in rete*, Milano, 2016; ID., *Online Political Hate Speech in Europe: the rise of New Extremism*, Oxford, 2020; M. SPATTI, *L'azione dell'Unione europea nella lotta all'odio online: la responsabilizzazione dei social media*, in AA.VV., *Temi e questioni di diritto*

hanno, tuttavia, evidenziato la ricerca di un delicato equilibrio tra i diversi valori e diritti in gioco, in particolare tra la tutela della libertà di espressione e della dignità umana e tra potere pubblico e autonomia privata⁷.

Come anticipato, l'*hate speech online* si presenta come un fenomeno fortemente transnazionale, tale da richiedere, per un contrasto efficace, una regolamentazione sovranazionale: un messaggio di odio in rete, infatti, può propagarsi velocemente e coinvolgere persone presenti in più Stati, nonché essere riproposto su diverse piattaforme. Esso risulta tuttora privo di una definizione univoca nell'ambito dell'Unione europea. Spesso si fa riferimento alla definizione di *hate speech* utilizzata dal documento *United Nations strategy and plan of action on hate speech* ai sensi della quale «*any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor*»⁸. Tale definizione, tuttavia, presenta dei profili di indeterminatezza che mal si conciliano con la libertà di espressione e manifestazione del pensiero.

L'ordinamento dell'Unione europea presenta diversi strumenti che concorrono al contrasto di tale fenomeno. In particolare, occorre osservare come l'azione dell'Unione europea per il contrasto all'*hate speech (online)* ma non solo) si svolga secondo tre direttrici: (i) l'adozione di norme dirette alla regolamentazione della condotta degli *internet service providers* (ISP), ovvero i fornitori di servizi nel modo digitale; (ii) gli strumenti di *soft law*; (iii) il ricorso a norme per l'incriminazione delle condotte di odio considerate penalmente rilevanti⁹.

dell'Unione europea. Scritti offerti a Claudia Morviducci, Bari, 2019, 909 ss.; M. D'AMICO, C. SICCARDI (a cura di), *La costituzione non odia. Conoscere, prevenire e contrastare l'hate speech online*, Torino, 2021; M. BASSINI, *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati". Spunti di comparazione*, in *Riv. it. inform. e dir.*, 2021, 43 ss.; N. FIANO, *Le radici del male. Antisemitismo e costituzione*, Torino, 2024, spec. 109 ss.; D. BARLETTA, A. COLLOLA, *Contrasting Hate Speech as a Practice of Digital Citizenship*, in R. TORINO (a cura di), *Digital Citizenship in the European Union Framework*, Roma, 2024, 151 ss.

⁷ Cfr. M. BASSINI, *Social Networks as New Public Forums?*, in *The Italian Review of International and Comparative Law*, 2022, 311 ss.; ID., *Libertà di espressione e social network, tra nuovi "spazi pubblici" e "poteri privati". Spunti di comparazione*, cit.; ID., *Internet e libertà di espressione. Prospettive costituzionali e sovranazionali*, Roma, 2019.

⁸ *United Nations strategy and plan of action on hate speech*, 18 giugno 2019.

⁹ Per una ricostruzione più ampia di tali linee direttrici si consenta di rinviare a I. ANRÒ, *Online Hate Speech: la prospettiva dell'Unione europea tra regolamentazione della condotta dei prestatori di servizi intermediari e ricorso al diritto penale*, in *Osservatorio sulle fonti*, 2023, 13 ss.

Con riferimento al punto (i), la direttiva 2000/31/CE del Parlamento europeo e del Consiglio dell'8 giugno 2000 relativa a taluni aspetti giuridici dei servizi della società dell'informazione¹⁰, che ha inserito le prime norme in tema di regolazione dei contenuti da parte dei prestatori di servizio *online*, ha rappresentato un primo strumento per la lotta all'*hate speech*. La direttiva si basava su un principio di assenza di un obbligo generale di sorveglianza (art. 15) e su una generale esenzione di responsabilità per i contenuti di terzi (artt. 12, 13, 14). L'esenzione di responsabilità operava a condizione che: 1) gli ISP non fossero effettivamente al corrente dell'illiceità dell'attività o dell'informazione o di fatti e circostanze da cui si potesse evincere la natura illecita delle attività o dei contenuti; 2) il fornitore di servizi, una volta ottenuta tale conoscenza o consapevolezza, agisse tempestivamente per rimuovere o disabilitare il contenuto. Gli stessi principi si ritrovano oggi nel *Digital Services Act* ("DSA")¹¹, il regolamento che ogni disciplina in maniera più ampia gli obblighi per gli ISP¹². Il DSA non ha modificato, nella sostanza, il regime di responsabilità e l'assenza di un obbligo generale di sorveglianza, continuando a prevedere l'esenzione purché in caso di notizia di informazioni o attività illecite il prestatore di servizi si attivi (il c.d. sistema di "*notice and take down*")¹³. Ha apportato, tuttavia, delle modifiche significative introducendo obblighi di mitigazione del rischio per un ecosistema digitale trasparente e sicuro, i quali variano a seconda della natura e delle dimensioni del servizio¹⁴. Il DSA, inoltre, introduce una nozione

¹⁰ Direttiva 2000/31/CE del Parlamento europeo e del Consiglio dell'8 giugno 2000 relativa a taluni aspetti giuridici dei servizi della società dell'informazione, in particolare il commercio elettronico, nel mercato interno (direttiva sul commercio elettronico), in *GUUE* L 178 del 17 luglio 2000, 1 ss.

¹¹ Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali), in *GUUE* L 277 del 27 ottobre 2022, 1 ss.

¹² Il DSA non ha abrogato la direttiva 2000/31, limitandosi a modificarla. Con specifico riferimento alle norme sulla responsabilità degli ISP per i contenuti di terzi l'art. 89 del DSA prevede che «1. Gli articoli da 12 a 15 della direttiva 2000/31/CE sono soppressi. 2. I riferimenti agli articoli da 12 a 15 della direttiva 2000/31/CE si intendono fatti rispettivamente agli articoli 4, 5, 6 e 8 del presente regolamento».

¹³ Cfr. G. ENES, I. NEVES, T. MORAIS ROCHA (eds.), *A Digital Europe for Citizens, Constitutional and Policymaking Challenge*, Cham, 2025. V. anche M. OROFINO, *Il Digital Service Act tra continuità (solo apparente) ed innovazione*, in F. PIZZETTI, S. CALZOLAIO, A. IANNUZZI, E. LONGO, M. OROFINO (a cura di), *La regolazione europea della società digitale*, Torino, 2024, 134 ss.

¹⁴ In tema di responsabilità degli internet services providers si veda: G. MORGESE, *Moderazione e rimozione dei contenuti illegali online nel diritto dell'Unione europea*, in *Federalismi*, 12 gennaio 2022, 80 ss.; M. BASSINI, *Libertà di espressione e social network, tra nuovi "spazi*

piuttosto ampia di «contenuto illegale», ritenendo tale «qualsiasi informazione che, di per sé o in relazione a un'attività, tra cui la vendita di prodotti o la prestazione di servizi, non è conforme al diritto dell'Unione o di qualunque Stato membro conforme con il diritto dell'Unione, indipendentemente dalla natura o dall'oggetto specifico di tale diritto» (art. 3, lett. h).

Con riferimento al punto *sub ii*) e quindi agli strumenti di *soft law*, basti ricordare il codice di condotta per lottare contro forme illegali di incitamento all'odio online promosso dalla Commissione europea il 31 maggio 2016 e successivamente adottato in maniera volontaria dalle più importanti piattaforme come Google (YouTube), Facebook, X (in precedenza, Twitter) and Microsoft, poi seguite Instagram, Dailymotion, Snap Inc., Jeuxvideo.com, TikTok, LinkedIn, Rakuten Viber e Twitch. Detto codice di condotta è stato rivisto inserendo maggiori oneri alla luce dell'entrata in vigore del DSA e il 20 gennaio 2025 è stato integrato nel quadro normativo della legge sui servizi digitali, a seguito di una valutazione positiva della Commissione e del comitato europeo per i servizi digitali ¹⁵.

Recentemente, inoltre, è stato istituito un panel europeo di cittadini sulla lotta contro l'odio nella società. Il *panel*, formato da 150 cittadini, selezionati in maniera casuale in tutti i 27 Stati membri, si riunisce per esaminare le cause profonde dell'odio e i modi per affrontarle. L'obiettivo di detto *panel* di cittadini è «formulare raccomandazioni su come costruire ponti tra gruppi e comunità frammentati» ¹⁶.

Alla luce del quadro sopra richiamato, il presente contributo si concentra sulla terza direttrice, ovvero il ricorso a norme incriminatrici comuni, ripercorrendo gli strumenti più recenti adottati con particolare riferimento alla tutela delle donne e al contrasto alla misoginia *online*.

2. Hate speech online e strumenti di cooperazione giudiziaria penale.

Qualche anno dopo la direttiva sul commercio elettronico, l'Unione europea è passata ad adottare norme in materia di cooperazione giudiziaria penale per una lotta più incisiva al discorso d'odio.

pubblici» e «poteri privati». Spunti di comparazione, cit.; M. SPATTI, L'azione dell'Unione europea nella lotta all'odio online: la responsabilizzazione dei social media, cit.

¹⁵ Cfr. il comunicato stampa della Commissione europea del 20 gennaio 2025, *Il codice di condotta per contrastare l'illecito incitamento all'odio online* +, reperibile online.

¹⁶ Cfr. il comunicato stampa della Commissione europea del 4 aprile 2024, *Panel europeo di cittadini per discutere della lotta contro l'odio nella società*, reperibile online.

Il primo importante strumento in questo senso è rappresentato dalla decisione quadro 2008/913/GAI che obbliga gli Stati membri a punire le condotte di odio, indipendentemente dal fatto che siano *online* o *offline*, ove essere siano forme ed espressioni di razzismo e xenofobia¹⁷. La decisione quadro precisa, al considerando 9, che la nozione di «odio» dovrebbe essere intesa come riferita «all'odio basato sulla razza, il colore, la religione, l'ascendenza o l'origine nazionale o etnica». Questa nozione copre sia le condotte di manifestazione di odio *online* che quelle *offline* e, dunque, appare piuttosto ampia sotto questo profilo. Al tempo stesso, tuttavia, tale definizione risulta limitata quanto al contenuto, essendo unicamente relativa all'odio di stampo etnico-razziale e religioso ed escludendo, ad esempio, l'odio per motivi diversi da quelli indicati, come la disabilità, il sesso, il genere, l'identità di genere.

La decisione quadro è diretta a garantire che gravi manifestazioni di razzismo e xenofobia siano punibili con sanzioni effettive, proporzionate e dissuasive nell'Unione europea, tramite l'armonizzazione della legislazione penale. In particolare, all'art. 1, par. 1, lett. a), essa sancisce il dovere per gli Stati di rendere punibili i comportamenti intenzionali consistenti nella «istigazione pubblica alla violenza o all'odio nei confronti di un gruppo di persone, o di un suo membro, definito in riferimento alla razza, al colore, alla religione, all'ascendenza o all'origine nazionale o etnica», mentre all'art. 4 viene affermato che «per i reati diversi da quelli di cui all'art. 1 e 2 [ovvero i reati di istigazione all'odio e apologia, negazione o minimizzazione dei crimini di genocidio e contro l'umanità] gli Stati membri adottano le misure necessarie affinché la motivazione razzista o xenofoba sia considerata una circostanza aggravante o, in alternativa, affinché tale motivazione possa essere presa in considerazione dal giudice all'atto della determinazione della pena».

Questa decisione quadro ha avuto una lunga gestazione, ben sette anni, dovuta principalmente alle differenze nazionali quanto al livello di tutela della libertà di espressione ed al bilanciamento con altri valori e diritti¹⁸. Anche la trasposizione negli Stati membri si è rivelata piuttosto difficoltosa, al pari della sua adozione. Nel 2014, la Commissione europea ha evidenziato che diversi Stati membri non avevano recepito del tutto o correttamente le disposizioni

¹⁷ Decisione quadro 2008/913/GAI del Consiglio del 28 novembre 2008 sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale, in *GUUE*, L 328 del 6 dicembre 2008.

¹⁸ Cfr. Relazione della Commissione al Parlamento e al Consiglio sull'attuazione della decisione quadro 2008/913/GAI del Consiglio sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale, 27 gennaio 2014, COM/2014/027 final.

della decisione quadro, in particolare quelle sui reati di negazione, apologia o minimizzazione grossolana di determinati crimini. Inoltre, secondo la Commissione, le disposizioni sull'istigazione all'odio, o alla violenza, di stampo razzista e xenofobo vigenti negli Stati membri non erano sempre del tutto conformi alla decisione quadro, sussistendo diverse lacune¹⁹. Per porre rimedio a tale situazione, la Commissione ha avviato una serie di colloqui bilaterali con gli Stati interessati. Lo Stato italiano provvedeva alla trasposizione solo nel 2017, mediante la legge europea di quell'anno, con notevole ritardo²⁰.

3. *L'inserimento di hate speech e hate crimes tra i "reati europei".*

La lotta contro il discorso d'odio attraverso il diritto penale è stata successivamente inserita tra le priorità della Commissione europea: nel settembre del 2020, infatti, la Presidente Ursula Von Der Leyen ha annunciato il proprio obiettivo di estendere la lista dei reati "europei" ex art. 83 TFUE per condannare «tutte le forme di crimini d'odio e incitamento all'odio»²¹, ovvero "hate crimes" e "hate speech" nella versione inglese dello stesso documento, sebbene ordinariamente la nozione di *hate speech* si riferisca già all'espressione di una condotta di odio più che ad un incitamento all'odio (già nel passaggio da una lingua all'altra si insinuano problemi definitivi).

Non è, inoltre, sempre lineare la distinzione tra la nozione di *hate speech* e *hate crimes*. Secondo la Guida dell'OSCE del 2014, gli *hate crimes* sono i crimini ispirati dall'odio, ovvero fatti penalmente rilevanti realizzati sulla base di pregiudizio e intolleranza: «il reato ispirato dall'odio consiste in un qualunque atto che sia: autonomamente tipizzato da una norma penale ('reato base' – base offence); e, in aggiunta, motivato dal pregiudizio basato su una specifica caratteristica della vittima ('la motivazione basata sul pregiudizio' – *bias motivation*)»²². Il reato base può essere diretto contro una o più persone o contro beni. La motivazione basata sul pregiudizio consiste nella discriminazione verso la vittima operata dall'autore dell'illecito, e centrata su una caratteristica che rappresenta un aspetto fondante ed essenziale di una comune identità di gruppo,

¹⁹ *Ibidem*.

²⁰ Legge 20 novembre 2017, n. 167, Disposizioni per l'adempimento degli obblighi derivanti dall'appartenenza dell'Italia all'Unione europea – legge europea 2017 (artt. 5, 6, 24).

²¹ Cfr. State of the Union 2020, Bruxelles, 16 settembre 2020.

²² Cfr. La guida dell'OSCE ODIHR, dal titolo "*Perseguire giudizialmente i crimini d'odio. Una guida pratica*", 2014, spec. 21.

come la razza, la lingua, la religione, l'etnia, la nazionalità, il genere o altra caratteristica.

Come noto, l'art. 83, par. 1, comma 1, TFUE prevede che Parlamento europeo e Consiglio possano adottare, mediante procedura legislativa ordinaria, «norme minime relative alla definizione dei reati e delle sanzioni in sfere di criminalità particolarmente grave che presentano una dimensione transnazionale derivante dal carattere o dalle implicazioni di tali reati o da una particolare necessità di combatterli su basi comuni». Tali reati sono attualmente individuati nella lista tassativa di cui all'art. 83, par. 1, comma 2, TFUE e sono: terrorismo, tratta degli esseri umani e sfruttamento sessuale delle donne e dei minori, traffico illecito di stupefacenti, traffico illecito di armi, riciclaggio di denaro, corruzione, contraffazione di mezzi di pagamento, criminalità informatica e criminalità organizzata. L'art. 83, par. 1, comma 3, prevede poi che «in funzione dell'evoluzione della criminalità» il Consiglio possa adottare una decisione che individui altre sfere di criminalità europea, ma a tal fine occorre una decisione all'unanimità, previa approvazione del Parlamento europeo.

Dunque, solo estendendo tale lista, con una modifica del Trattato, sarà possibile avere delle norme comuni in tema di *hate speech* e *hate crimes*, secondo il progetto della Commissione europea.

Successivamente, nel dicembre 2020, nella propria Comunicazione sul piano d'azione per la democrazia europea²³, la Commissione europea illustrava una vasta gamma di proposte e iniziative dirette a promuovere un più salubre ambiente digitale e informativo. In particolare, la Commissione europea, al fine di incoraggiare una più ampia ed eguale partecipazione democratica, prevedeva espressamente l'intenzione di compiere ulteriori sforzi nella lotta contro le fattispecie di *hate speech* digitale. Secondo il documento, la presenza in *internet* di forme di incitamento all'odio avrebbe, infatti, l'effetto di dissuadere le minoranze da esse colpite dall'esprimere le loro opinioni in rete e dal partecipare al discorso pubblico (par. 2.4). La Commissione prevedeva, pertanto, di adottare un'iniziativa per ampliare la sfera dei reati contenuta nell'art. 83 par. 1, TFUE per i quali è consentito a Parlamento e Consiglio stabilire norme minime di ravvicinamento delle legislazioni²⁴.

²³ Cfr. Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni sul piano d'azione per la democrazia europea, cit.

²⁴ Per un'analisi completa dell'art. 83 TFUE si v. C. AMALFITANO, *Art. 83 TFUE*, in A. TIZZANO (a cura di), *Trattati dell'Unione europea*, Milano, 2014, 896 ss.

Nel dicembre 2021 la Commissione europea ha adottato una nuova comunicazione e una proposta per avviare l'*iter* di inserimento di nuove fattispecie nell'elenco dei "reati europei" di cui all'art. 83, par. 1, TFUE²⁵. Come illustrato in detta comunicazione, il primo passo a tal fine, è l'adozione, da parte del Consiglio, all'unanimità, di una decisione che identifichi le condotte di hate speech e di hate crime come una sfera di criminalità tale da rendere necessaria una regolamentazione europea. Successivamente, la Commissione potrà adottare una proposta di direttiva diretta a stabilire regole minime per la definizione e l'incriminazione delle condotte di *hate speech* e *hate crime*.

La comunicazione illustra le motivazioni alla base di questa proposta. In prima battuta, la Commissione sottolinea come la lotta all'*hate speech* e agli *hate crimes* sia una delle priorità dell'Unione europea, richiamando le diverse iniziative già avviate: dall'adozione della decisione quadro 2008/913/GAI, al Codice di Condotta per gli ISP, dal regolamento circa i contenuti terroristici online all'EU internet forum e al DSA (par. 2).

Nonostante gli strumenti già in essere, la Commissione illustra la necessità dell'adozione di un atto per estendere la lista dei reati di cui all'art. 83 TFUE, in quanto *hate speech* e *hate crimes* si configurerebbero come un'area di criminalità particolarmente grave. Secondo la Commissione, infatti, questo tipo di fattispecie si porrebbe in spiccato contrasto con i valori ed i diritti fondamentali dell'Unione europea, racchiusi dall'art. 2 e dall'art. 6 TUE. La Commissione richiama anche la giurisprudenza della Corte EDU per rafforzare l'idea che si tratti di condotte particolarmente lesive dell'integrità fisica e morale di una persona (par. 3.2).

Hate speech e *hate crimes* violerebbero, soprattutto, il diritto fondamentale alla dignità e all'uguaglianza delle persone nonché il diritto alla libertà di espressione, con un forte impatto sulla persona e la sua comunità di appartenenza (par. 3.2.3).

Secondo la Commissione, inoltre, tali condotte avrebbero una forte dimensione sociale, costituendo una seria minaccia per i valori democratici, la stabilità e la pace della stessa società. Esse impatterebbero, inoltre, negativamente sulla libertà di espressione, poiché le vittime potrebbero essere dissuase dall'utilizzo

²⁵ Commissione europea, *Un'Europa più inclusiva e protettiva: estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio*, COM(2021)777 final. Si veda anche lo studio preparatorio: Commissione europea, Directorate-General for Justice and Consumers, Spark Legal Network and TIPIK, *Study to support the preparation of the European Commission's initiative to extend the list of EU crimes in Article 83 of the Treaty on the Functioning of the EU to hate speech and hate crime – Final report*, 2021.

dei social media in ragione dei contenuti di odio ivi riportati (par. 3.2.3). Nella propria comunicazione, la Commissione riporta esempi di *hate speech* e di *hate crime* riferiti, inter alia, a persone LGBTIQ, di origine asiatica, rom o africana, nonché legati a fenomeni di antisemitismo (par. 3.2.4).

La Comunicazione analizza poi gli altri elementi per giustificare la proposta, ovvero l'attuale normativa penale degli Stati membri e la dimensione transazionale del fenomeno (parr. 3.2.5 e 3.3). Nell'analizzare l'evoluzione della criminalità, la Commissione riferisce che nel 2019 il Parlamento europeo ha invitato gli Stati a condannare le condotte di *hate speech* e di *hate crimes* dei politici, poiché rafforzerebbero il clima di odio nella società. Inoltre, sottolinea come la pandemia da Covid-19 abbia provocato «uno tsunami di odio xenophobia» (par. 3.4).

La Commissione conclude, dunque, che l'iniziativa diretta ad ampliare la lista di reati *ex art.* 83 TFUE e quindi a consentire l'adozione di norme minime in materia penale da parte dell'Unione sia «la risposta più efficace, a livello dell'UE, alle sfide individuate» in quanto «solo un'iniziativa comune a livello dell'UE può proteggere efficacemente i valori comuni sanciti dall'art. 2 TUE che sono minacciati da tutte le forme di incitamento all'odio e di reati generati dall'odio» (par. 4.1). Tale conclusione non sarebbe rimessa in discussione dalla possibile sovrapposizione con i reati già previsti dall'art. 83 TFUE, come i reati informatici o il terrorismo ²⁶.

Recentemente, la Commissione è tornata a ricordare l'importanza di portare a termine l'estensione dei reati europei anche nella più recente comunicazione *Nessuno spazio per l'odio in un'Europa che, unita, lo ripudia* ²⁷. In tale documento, veniva a tal proposito evidenziato che «[i] recenti eventi sottolineano l'esigenza imperativa di una rapida adozione di tale decisione del Consiglio a difesa dei valori comuni dell'Unione, messi a repentaglio da tutte le forme di incitamento all'odio e dai reati generati dall'odio, indipendentemente dalle persone e dai gruppi colpiti» (par. 1).

²⁶ Sulle criticità di tale proposta e, in generale, dell'estensione dei crimini di cui all'art. 83 TFUE si consenta ancora un rinvio a I. ANRÒ, *Online Hate Speech: la prospettiva dell'Unione europea tra regolamentazione della condotta dei prestatori di servizi intermediari e ricorso al diritto penale*, cit.

²⁷ Commissione europea, Comunicazione congiunta al Parlamento europeo e al Consiglio, *Nessuno spazio per l'odio in un'Europa che, unita, lo ripudia*, Bruxelles, 6 dicembre 2023, JOIN(2023) 51 final.

4. *La risoluzione del Parlamento europeo del 18 gennaio 2024 a favore dell'estensione della lista ex art. 83 TFUE.*

Nel gennaio 2024, il Parlamento europeo adottava, con 397 voti favorevoli, 121 contrari e 26 astensioni, una risoluzione circa la proposta della Commissione di estendere l'elenco dei reati riconosciuti dall'Unione europea all'incitamento all'odio e ai reati generati dall'odio²⁸. In questa risoluzione, il Parlamento evidenziava innanzitutto che «tutte le forme e manifestazioni di odio e intolleranza, inclusi l'incitamento all'odio e i reati generati dall'odio, sono incompatibili con i valori dell'Unione della dignità umana, della libertà, della democrazia, dell'uguaglianza, dello Stato di diritto e del rispetto dei diritti umani» (lett. B) e che «l'incitamento all'odio e i reati generati dall'odio possono rappresentare reati particolarmente gravi» (E). Successivamente, il Parlamento ribadiva che «l'attuale quadro dell'UE riguarda solo l'incitamento all'odio e i reati generati dall'odio per motivi di razza, colore, religione, ascendenza e origine nazionale o etnica» e che «non esiste una definizione giuridica comune globale di discorso d'odio e di reati generati dall'odio a livello di UE», ma «vi è una chiara necessità di affrontare efficacemente il discorso d'odio e i reati generati dall'odio basati su altri motivi, quali il sesso, l'orientamento sessuale, il genere, l'identità di genere, l'espressione di genere, le caratteristiche sessuali, l'età, la disabilità e qualsiasi altra caratteristica fondamentale, in particolare quelle tutelate dall'articolo 21 della Carta, nonché qualsiasi combinazione di tali caratteristiche» (lett. N). Stante questo quadro, il Parlamento europeo evidenziava che la proposta della Commissione è in linea con i presupposti dell'art. 83 TFUE: esisterebbe, infatti, una chiara componente transfrontaliera dell'incitamento all'odio online che giustificerebbe una normativa europea (lett. O). Alla luce di tali considerazioni, il Parlamento esortava quindi il Consiglio ad adottare una decisione ai sensi dell'art. 83 TFUE per inserire l'incitamento all'odio e i reati generati dall'odio in detta lista, considerando la questione come prioritaria (punto 1), deplorando il fatto che fossero trascorsi quasi due anni dalla pubblicazione della comunicazione della Commissione senza che il Consiglio avesse compiuto progressi in merito (punto 3). Inoltre, il Parlamento raccomandava di modificare l'art. 83 TUE per renderlo soggetto ad una maggioranza qualificata rafforzata, anziché all'attuale unanimità e richiedeva nel caso di specie l'attivazione della clausola passerella (punto 7). Il Parlamento chiedeva

²⁸ Risoluzione del Parlamento europeo del 18 gennaio 2024, Estendere l'elenco dei reati riconosciuti dall'UE all'incitamento all'odio e ai reati generati dall'odio (2023/2068 (INI)), in *GUUE C*, 2024/5733.

poi alla Commissione di prendere in considerazione «un approccio aperto in base al quale l'elenco dei motivi di discriminazione non sia esaustivo» al fine di combattere efficacemente l'incitamento all'odio e i reati generali dall'odio (punto 11), enfatizzando che l'abuso di internet e dei modelli di business delle piattaforme di social media, basati sulla pubblicità micro-mirata, contribuisce ad amplificare i rischi di vittimizzazione ripetuta (punto 12). Particolare attenzione viene poi rivolta ai minori, per una maggiore protezione rispetto all'odio, al bullismo, al cyberbullismo, per ridurre al minimo l'impatto di tali fenomeni sulla loro salute mentale (punto 14).

La Commissione europea ha, successivamente, espresso la propria approvazione circa la risoluzione del Parlamento²⁹. La Commissione ha puntualizzato che, una volta adottata la decisione del Consiglio e creata una base giuridica per la legislazione secondaria, avvierà i lavori per definire i nuovi reati con riferimento a motivi specifici, oltre al razzismo e alla xenofobia, come stabilito nella decisione quadro sulla lotta contro talune forme ed espressioni di razzismo e xenofobia mediante il diritto penale. Le proposte legislative, secondo le rassicurazioni della Commissione, saranno formulate sulla base di ampie consultazioni, tenendo presente non solo le sfide attuali, ma anche la legislazione penale degli Stati. La Commissione ha poi ribadito il proprio impegno per un'attuazione rapida del DSA, nonché per la protezione dei minori, anche attraverso l'adozione di strumenti specifici. Infine, la Commissione ha ricordato i propri strumenti per rafforzare la capacità delle autorità nazionali di indagare e perseguire efficacemente i discorsi e i reati di incitamento all'odio sul campo. In particolare, ha ricordato che il gruppo di alto livello sulla lotta ai discorsi e ai reati di incitamento all'odio, istituito nel 2016, aiuta gli Stati membri a sviluppare la formazione e il rafforzamento delle capacità delle forze dell'ordine, a migliorare la registrazione dei reati di incitamento all'odio e la raccolta di dati, nonché a incoraggiare le vittime a denunciare tali reati.

5. Gli ostacoli tuttora esistenti e il cambio di strategia della Commissione.

Ad oggi, nonostante la risoluzione del Parlamento sopra ricordata, la proposta della Commissione europea di estendere la lista dei reati *ex art.* 83 TFUE non ha avuto seguito. Senz'altro, la necessità di raggiungere l'unanimità in una materia così delicata risulta particolarmente arduo: l'incriminazione di fattispecie

²⁹ Commissione europea, Reazione circa il testo adottato in plenaria, SP(2024)145 del 21 maggio 2024.

legate alle manifestazioni di odio implica un delicato bilanciamento tra la libertà di espressione e la dignità umana che ogni ordinamento effettua secondo i propri valori. Se è vero che il contrasto alle manifestazioni e ai reati di odio è un obiettivo e un valore condiviso a livello europeo, la scelta dello strumento penale non appare scontata, anche in presenza del ricco quadro normativo ricordato in apertura.

Nonostante la mancata prosecuzione dell'iniziativa di cui sopra, anche senza estendere la lista dei reati di cui all'art. 83 TFU, norme dirette all'inserimento di reati per combattere l'odio *online*, in particolare contro le donne, si ritrovano oggi nella direttiva (UE) 2024/1385 del Parlamento europeo e del Consiglio, del 14 maggio 2024, sulla lotta alla violenza contro le donne e alla violenza domestica³⁰.

Questa importante direttiva (su cui si soffermano altri contributi della presente colletanea) è stata adottata sulla base dell'attuale formulazione degli artt. 82, par. 2, e 83, par. 1, TFUE, per fornire un quadro giuridico generale per prevenire e combattere in maniera efficace la violenza contro le donne e la violenza domestica in tutta l'Unione europea. La direttiva stabilisce norme minime riguardanti la definizione dei reati e delle sanzioni in materia di sfruttamento sessuale femminile e minorile e di criminalità informatica, nonché circa i diritti delle vittime, la protezione, l'assistenza e la prevenzione. La direttiva inserisce norme affinché gli Stati provvedano a punire come reati le condotte intenzionali relative a: (i) mutilazioni genitali femminili (art. 3); (ii) matrimonio forzato (art. 4); (iii) condivisione non consensuale di materiale intimo o manipolato (art. 5); (iv) stalking *online* (art. 6); molestie *online* (art. 7); istigazione alla violenza o all'odio *online* (art. 8). Per quanto rileva ai fini del presente contributo occorre concentrarsi su quest'ultima fattispecie. A tal proposito, il considerando 25 chiarisce che «negli ultimi anni, l'aumento dell'uso di internet e dei social media ha portato a un'impennata dei casi di istigazione pubblica alla violenza e all'odio, anche basati sul genere. L'effetto disinibente di internet moltiplica la condivisione facile, rapida e vasta dei discorsi d'odio nel mondo digitale, in quanto il presunto anonimato sul web e il senso di impunità che ne deriva riducono il senso di inibizione che normalmente frenerebbe le persone». Il considerando sottolinea che l'odio online colpisce con più frequenza le donne: «[I]e donne sono spesso il bersaglio dell'odio sessista e misogino online, che può

³⁰ Direttiva (UE) 2024/1385 del Parlamento europeo e del Consiglio, del 14 maggio 2024, sulla lotta alla violenza contro le donne e alla violenza domestica in *GUUE* L, 2024/1385, del 24 maggio 2024. Per un primo commento v. M. FERRARI, *Violenza contro le donne: l'Unione europea adotta finalmente la direttiva (UE) 2024/1385*, in *Eurojus*, 17 giugno 2024.

degenerare in reati generati dall'odio nel mondo reale». Questo «[è] un fenomeno che va prevenuto e intercettato fin dalle prime fasi. Anche se il linguaggio usato in detto tipo di istigazione non rimanda sempre in maniera diretta al genere della persona presa di mira, il pregiudizio è facilmente deducibile dal contenuto o dal contesto generale del discorso». Il considerando successivo (n. 26) perimetra il reato di istigazione alla violenza o all'odio online precisando che questo debba implicare la diffusione al pubblico «da intendersi come il fatto di rendere accessibile a un numero potenzialmente illimitato di persone tramite TIC un dato materiale che si configuri come incitamento alla violenza o all'odio». A tal proposito, il considerando precisa altresì che «nel valutare se il materiale configuri istigazione all'odio o alla violenza, le autorità competenti dovrebbero tener conto del diritto fondamentale alla libertà di espressione sancito dall'art. 11 della Carta». Conseguentemente «gli Stati membri dovrebbero poter decidere di rendere punibili soltanto le condotte atte a turbare l'ordine pubblico o che sono minacciose, offensive o ingiuriose» (considerando 27).

L'art. 8 della direttiva, rubricato «istigazione alla violenza o all'odio online» dispone al primo paragrafo, coerentemente con i considerando ricordati, che «[g]li Stati membri provvedono affinché sia punita come reato la condotta intenzionale consistente nell'istigare alla violenza o all'odio nei confronti di un gruppo di persone o di un membro di detto gruppo definito con riferimento al genere, diffondendo al pubblico tramite TIC materiale contenente tale istigazione»; al secondo paragrafo, l'articolo dispone poi che «gli Stati membri possono decidere di configurare come reato soltanto le condotte atte a turbare l'ordine pubblico o che sono minacciose, offensive o ingiuriose».

Ciò che preme sottolineare ai fini del presente contributo è che, nonostante l'arresto dei lavori per l'ampliamento dei reati *ex art.* 83 TFUE al fine di includervi le condotte di odio e i reati generati dall'odio, la direttiva (UE) 2024/1385 raggiunge ugualmente l'obiettivo di introdurre una incriminazione della fattispecie di istigazione all'odio online (oltre che alla violenza). Si tratta di un inserimento circoscritto, in quanto limitato all'istigazione all'odio e con riferimento unicamente al genere: in ogni caso, aggiunge indubbiamente un reato assente nel panorama normativo, soprattutto nella direttiva quadro. Per tale norma, non si è reso necessario modificare il trattato perché l'art. 83 TFUE già prevede i reati contro lo sfruttamento sessuale e la criminalità informatica. Si tratta in ogni caso di norme minime, restando gli Stati liberi di adottare o mantenere norme di diritto penale più rigorose per quanto riguarda la definizione dei reati e delle sanzioni.

Con riferimento al coordinamento tra la direttiva (UE) 2024/1385 e la proposta circa l'estensione della lista dei reati *ex art.* 83 TFUE per l'inclusione di

hate speech e *hate crimes*, già nella proposta di direttiva la Commissione evidenziava che «[u]na volta che il Consiglio avrà adottato la decisione che aggiungerà l'incitamento all'odio e i reati generati dall'odio all'articolo 83, paragrafo 1, TFUE quale nuova base giuridica, la Commissione sarà in grado di proporre ulteriori norme miranti ad armonizzare la definizione di questi due tipi di reati aggiungendo motivazioni per le quali sia prevista una tutela»³¹. Pare ragionevole ritenere che, ove l'*iter* per l'inclusione di *hate speech* e *hate crimes* tra i reati europei venisse completato e la Commissione proponesse l'adozione di una nuova direttiva per introdurre norme comuni, occorrerebbe una modifica – se non un'abrogazione – dell'art. 8 della direttiva in esame per avere un regime coerente ed in un unico strumento circa la disciplina penale di tali fattispecie.

Nella recente comunicazione *Unione dell'uguaglianza: strategia per l'uguaglianza LGBTIQ+ 2026-2030*, la Commissione europea ha affermato che le molestie e le violenze motivate dall'odio nei confronti delle persone LGBTIQ+ continuano ad aumentare, ma al momento «il quadro giuridico attuale previsto dalla decisione quadro del Consiglio sulla lotta contro talune forme ed espressioni di razzismo e xenofobia e dalla direttiva sulla lotta alla violenza contro le donne e la violenza domestica riguarda rispettivamente soltanto i reati generati dall'odio basati su motivi razzisti e xenofobi e l'istigazione online alla violenza o all'odio basati sul genere. Non esiste ancora una legislazione che garantisca una protezione ai sensi del diritto penale contro i reati basati sull'orientamento sessuale e/o sull'identità di genere. Di conseguenza il quadro giuridico in tutti gli Stati membri è frammentato e incoerente»³². Pertanto, stante l'assenza di progressi per l'adozione della decisione del Consiglio ai fini dell'inclusione dei reati di *hate speech* e *hate crimes* nella lista *ex art.* 83 TFUE, «la Commissione sta valutando un'iniziativa legislativa basata sulle sfere di criminalità esistenti soggette all'applicazione dell'articolo 83, paragrafo 1, TFUE al fine di armonizzare la definizione dei reati generati dall'odio commessi online». In questo senso si è espresso anche il *panel* dei 150 cittadini.

La Commissione si sta quindi orientando nel senso di procedere sulla base del quadro giuridico esistente, senza attendere oltre l'adozione della decisione del Consiglio per la modifica dell'art. 83 TFUE e la conseguente creazione di una nuova e coerente base giuridica.

³¹ Cfr. Proposta di direttiva del Parlamento europeo e del Consiglio sulla lotta alla violenza contro le donne e alla violenza domestica, Proposta COM (2022) 105, 8 marzo 2022.

³² Commissione europea, 8 ottobre 2025, *Unione dell'uguaglianza: strategia per l'uguaglianza LGBTIQ+ 2026-2030*, COM(2025) 725 final, spec. par. 1.2.

6. Alcune riflessioni conclusive.

Se l'*iter* per l'inclusione di *hate speech* e *hate crimes* nella lista dei reati *ex art.* 83 TFUE pare essersi fermato, con l'entrata in vigore della direttiva (UE) 2024/1385 la Commissione ha, in ogni caso, raggiunto un obiettivo nella medesima direzione, ovvero l'obbligo per gli Stati membri di disporre norme penali per la repressione del reato di istigazione all'odio per motivi di genere. Dunque, il genere viene collocato accanto al razzismo e alla xenofobia quale motivo che rende l'istigazione alla violenza e all'odio *online* penalmente rilevante.

Il ricorso alle norme incriminatrici non pare tuttavia del tutto convincente. Innanzitutto, come sopra evidenziato, il ricorso al diritto penale dovrebbe essere limitato, in ragione del principio di proporzionalità, trattandosi di uno strumento particolarmente incisivo sui diritti della persona. In presenza di un quadro europeo sempre più attento alla regolamentazione dei contenuti offerti dagli ISP, sia attraverso norme di diritto derivato che attraverso strumenti di *soft law*, il ricorso al diritto penale pare non essere del tutto giustificato. In questo senso, la direttiva (UE) 2024/1385 pare segnare una compressione non del tutto proporzionata della libertà di espressione, diritto fondamentale consacrato dall'art. 11 CDFUE.

La direttiva (UE) 2024/1385 sembra, inoltre, aver raggiunto un obiettivo limitato (con riferimento ai profili considerati nel presente contributo), ovvero l'incriminazione di fattispecie di istigazione alla violenza e all'odio *online* per motivi di genere, senza tuttavia un chiaro coordinamento con la decisione quadro, anche ai fini della considerazione del fattore intersezionale, ove l'odio di genere si sommasse a quello xenofobo o religioso.

Come ricordato, la Commissione ha affermato di voler procedere sulla base dell'attuale formulazione dell'art. 83 TFUE ad adottare una proposta per armonizzare la definizione dei reati generati dall'odio commessi *online*. Se da un lato tale iniziativa si può comprendere, stante l'attuale frammentazione del quadro giuridico esistente, dall'altro lato essa pare in contraddizione con la comunicazione della stessa Commissione del 2021, ove si affermava la necessità di includere *hate speech* e *hate crimes* nella lista dell'art. 83 TFUE.

CAPITOLO VII

SISTEMI DI INTELLIGENZA ARTIFICIALE
E VIOLENZA DI GENERE: DUE ITINERARI
DI RIFLESSIONE TRA LORO CONVERGENTI

Irene Pellizzone

SOMMARIO: 1. Considerazioni introduttive sull'impatto dei sistemi di intelligenza artificiale sul fenomeno della violenza di genere: un itinerario a due sensi. – 2. I punti di forza dei Sistemi di intelligenza artificiale: strumenti di contrasto alla violenza di genere (richieste di aiuto ed esame dei fattori di rischio). – 3. Le donne come vittime predestinate dell'IA: *focus* sul *Deep Fake* non consensuale. – 4. Un importante passo avanti: l'intervento normativo attraverso la recente legge n. 132/2025. – 5. Prevenzione della violenza nei confronti delle donne e dell'intelligenza artificiale: un primo bilancio.

1. Considerazioni introduttive sull'impatto dei sistemi di intelligenza artificiale sul fenomeno della violenza di genere: un itinerario a due sensi.

Il tema dell'impatto dell'intelligenza artificiale sul fenomeno della violenza di genere è ancora poco studiato, benché meriti di essere analizzato sotto più profili.

Si ricordi al riguardo che l'indagine Istat del 2025, condotta ben 11 anni dopo l'ultima raccolta¹, mette chiaramente in evidenza che «sono circa 6 milioni e 400mila (il 31,9%) le donne che hanno subito almeno una violenza fisica o sessuale a partire dai 16 anni di età: il 18,8% (circa 3 milioni

¹ L'ISTAT ha già condotto l'Indagine sulla violenza contro le donne nel 2006 e nel 2014.

Oggi è previsto la legge n. 53/2022 sulle Disposizioni statistiche per misurare la violenza contro le donne (art. 2), la Convenzione di Istanbul e la direttiva (UE) 2024/1385 del Parlamento europeo e del Consiglio del 14 maggio 2024 sulla lotta alla violenza contro le donne e alla violenza domestica (art. 44) richiedono la raccolta di dati comparabili a livello europeo.

e 764 mila) ha subito violenze fisiche e il 23,4% violenze sessuali»².

Verso questa nuova esigenza di ricerca e studio spinge naturalmente la sempre maggiore attenzione data al fenomeno, grazie alla spinta proveniente in origine dal diritto internazionale e oggi maturata anche a livello interno, che ha portato sia le istituzioni, sia i media e l'opinione pubblica dedicare più energie all'analisi del problema e alla ricerca di possibili soluzioni³.

Dinanzi ad una casistica così ampia e ad un tema divenuto così centrale nel dibattito pubblico, potrebbe apparire scontato interrogarsi sui pro e i contra dell'intelligenza artificiale, ma in realtà occorre arrendersi dinanzi al dato di fatto che riflessioni strutturate in questo senso devono ancora essere sviluppate.

Non ci si riferisce ovviamente alle violenze semplicemente commesse attraverso internet, come ad esempio la diffusione sui social media di fotografie di parti intime della vittima per fini di "vendetta" o il discorso d'odio *online*⁴ – che pure, intersecandosi con alcuni nodi posti in queste pagine, saranno oggetto di analisi –, ma più specificamente al rapporto tra sistemi di intelligenza artificiale e violenza nei confronti di appartenenti al genere femminile.

Ciò premesso, il campo di indagine si presta ad essere approfondito in una duplice direzione.

In *primis*, l'intelligenza artificiale può essere oggetto di analisi in quanto strumento di ausilio al contrasto alla violenza di genere: seguendo l'impostazione delle c.d. "quattro P" della c.d. Convenzione di Istanbul⁵, essa può fungere da

² ISTAT, *La violenza contro le donne, dentro e fuori la famiglia primi risultati 2025*, 21 novembre 2025 (questa indagine non contiene dati relativi alle donne straniere, che saranno prodotti in un secondo momento, data la necessità di intervistarle di persona, per esigenze linguistiche e culturali).

³ V. ad esempio la sentenza n. 197/2023 della Corte costituzionale, al punto 5.4.1. del Considerato in diritto, ad esempio, dà atto di questo processo a livello normativo, mettendo in rilievo l'importanza della c.d. "Convenzione di Istanbul" e della legge n. 69/2019 (cosiddetto "Codice Rosso").

⁴ Lo studio in questi ambiti è già avanzato: v. ad es. F. BELOTTI, M. SCIANNAMBLO, P. PANARESE, S. PARISI, F. COMUNELLO, *Online gender-based violence as a socio-technical issue: the aware gaze of activists and practitioners*, in *About Gender – International Journal of Gender Studies*, 2025, 119 ss.

⁵ Convenzione del Consiglio d'Europa sulla prevenzione e la lotta contro la violenza nei confronti delle donne e la violenza domestica (cosiddetta "Convenzione di Istanbul"), ratificata con la legge 27 giugno 2013, n. 77 (Ratifica ed esecuzione della Convenzione del Consiglio d'Europa sulla prevenzione e la lotta contro la violenza nei confronti delle donne e la violenza domestica, fatta a Istanbul l'11 maggio 2011).

ausilio alla Prevenzione della violenza, da supporto alla Protezione della donna vittima di violenza, nonché da strumento per la Persecuzione penale dei presunti autori della violenza e, da ultimo, per l'elaborazione di Politiche sociali mirate⁶.

Queste potenzialità dei sistemi di intelligenza artificiale però, come già anticipato, devono ancora essere esplorate su un piano scientifico in modo davvero sistematico e strutturato, esistendo ad oggi, a parte alcuni studi offerti dalla letteratura internazionale⁷, singoli progetti di ricerca avviati a livello nazionale, molti dei quali in corso di svolgimento o appena terminati.

Si tratta di un primo punto su cui è bene focalizzare l'attenzione, perché occorre essere consapevoli che anche la scelta dell'investimento di risorse, per esplorare nuove opportunità offerte dall'intelligenza in un settore, piuttosto che in un altro, può derivare da una selezione di interessi orientata a criteri che trascurano i diritti e le posizioni delle vittime vulnerabili.

È infatti manifesto che queste nuove tecnologie hanno già rivoluzionato le nostre vite ed è dunque fondamentale verificare se possano essere ulteriormente valorizzate in ogni ambito, compreso quello dei diritti della donna violati dal partner o comunque dall'uomo violento⁸.

In secondo luogo, i sistemi di intelligenza artificiale possono essere strumentalizzati al fine di commettere violenza, trasformando questo fenomeno da personale a virtuale.

Possono infatti potenziare la diffusività, l'invasività e l'offensività della violenza e delle forme di controllo dell'uomo violento, amplificando la vergogna e

Sulla convenzione di Istanbul e sulla sua struttura, v. S. DE VIDO, M. FRULLI (a cura di), *Preventing and Combating Violence Against Women and Domestic Violence. A Commentary on the Istanbul Convention*, 2023, 85 ss.; F. POGGI, *Violenza di genere e Convenzione di Istanbul: un'analisi concettuale*, in *Diritti umani e diritto internazionale*, 2017, 65 ss.

⁶ Le "quattro p" – ovvero i principali capitoli in cui è strutturata la convenzione – sono infatti le seguenti: Prevenzione, Protezione, Persecuzione penale e Politiche integrate.

⁷ Cfr. ad es. E. GARCIA-VERGARA, N. ALMEDA, F. FERNÁNDEZ-NAVARRO, *Artificial intelligence extracts key insights from legal documents to predict intimate partner femicide*, in *Scientific Report*, 2023, 18212 ss.; P. NOVITZKY, J. JANSSEN, B. KOKKELER, *A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools*, in *Heliyon*, 2023; A. SUAYA, M. BUSTELO, K. MARTINEZ, *Innovations with Artificial Intelligence: An Exploratory Mapping in the Response to Violence Against Women*, 2025.

⁸ Per riflessioni in ambito di violenza di genere e utilizzo degli algoritmi nel processo penale, v. le prime riflessioni di T. ALESCI, *Processo algoritmico e contrasto alla violenza di genere. Punti di contatto*, in *Dir. pen. proc.*, 2025, 1330 ss., che ipotizza rischi di discriminazione di genere, laddove i dati elaborati dall'algoritmo potrebbero discriminare la vittima, minandone la credibilità.

il senso di colpa spingendo la donna a condotte di isolamento ancora più marcate. La lesione della libertà e della dignità della donna può essere per certi versi ancora maggiore se l'intelligenza artificiale è in campo.

I dati da questo punto di vista, come si vedrà a breve, sono impressionanti.

In questo scritto, si intende dunque dare avvio a riflessioni sul tema, facendo uso del metodo fatto proprio dagli studi di genere.

Si partirà dunque da un duplice presupposto: a) che la violenza di genere è costituita da quelle forme di violenza che attecchiscono in modo persistente nel nostro contesto sociale perché un genere, quello femminile, è collocato in posizione di subalternità: comportamenti violenti nei confronti del genere subalterno sono dunque o legittimati culturalmente⁹, o comunque tollerati (o non delegittimati¹⁰; b) che per trattare l'argomento della violenza di genere è necessario attrezzarsi con un bagaglio poroso agli approdi di altre discipline, come la psicologia e la sociologia.

Ogni volta che ci si accosta alla violenza di genere sfruttando le acquisizioni di campi di indagine extragiuridici (psicologia, sociologia), infatti, si amplia il raggio d'azione degli strumenti offerti dalle varie discipline, rendendolo più ampio e dinamico di quello che potrebbe apparire a prima vista.

Ad esempio, la consapevolezza dell'impatto psicologico della violenza domestica sulla vittima è essenziale per comprendere alcune dinamiche relazionali che rendono particolarmente complesso per la stessa donna spezzare il vincolo con l'uomo maltrattante. Ci si riferisce, in particolare, al c.d. ciclo della violenza, così chiamato perché l'uomo passa infatti da una strategia di tensione e conseguente attacco a rappacificazioni o "lune di miele"¹¹. Il ciclo poi riprende immutato, ma l'aggressione è sempre più potente: dinanzi a questa manifestazione intermittente della violenza, la donna è più comprensibilmente colpita da un'altalena di sentimenti diversi e, ogni volta che compie un ripensamento sulla rottura del legame, si trova irretita in una ragnatela sempre più difficile da strappare. Evidentemente, conoscendo questa modalità di esercizio della violenza su un piano psicologico, gli operatori sociali, le forze dell'ordine e i magistrati potranno davvero comprendere il significato e trattare correttamente i "ripensamenti" della donna stessa.

⁹ Cfr. M. D'AMICO, *Una parità ambigua. Costituzione e diritti delle donne*, Milano, 2011, 220 s.

¹⁰ V. B. PEZZINI, *L'Italia davanti alla Corte EDU per l'insufficiente protezione delle vittime della violenza di genere: una questione costituzionale*, in *Osservatorio AIC*, 2022, 280 ss.

¹¹ L.E. WALKER, *The battered Woman*, New York-Cambridge, 1980.

Solo un approccio *gender sensitive* (cioè sensibile alla specificità del fenomeno, connotato dalla subalternità sociale della vittima in ragione del suo genere e specularmente dalla indulgenza sociale verso l'autore della violenza)¹² può – per concludere su questo punto – consentire di comprendere le esigenze che potrebbero spingere la donna ad avvalersi dell'intelligenza artificiale nel chiedere aiuto o cogliere le reali manifestazioni dell'uso dell'intelligenza artificiale come strumento di violenza¹³.

2. I punti di forza dei Sistemi di intelligenza artificiale: strumenti di contrasto alla violenza di genere (richieste di aiuto ed esame dei fattori di rischio).

Quanto al primo versante, l'*intelligenza artificiale* è sinora stata sperimentata come strumento per facilitare la richiesta di aiuto della donna o per analizzare con maggiore velocità e precisione i fattori di rischio di *escalation* della violenza.

Pensiamo al ricorso alle *chat bot*, ovvero *software* progettati per simulare una conversazione umana, scritta o vocale, con un utente¹⁴.

¹² Particolarmente significative, per questo aspetto, sono le parole della concurring opinion di Pinto de Albuquerque nella sentenza *Valiuliene v. Lituania* (Corte eur. dir. uomo, Sez. II, sentenza 26 marzo 2013, *Valiuliene c. Lituania*). Nel caso, che vede agire come ricorrente una donna lituana vittima di ripetute violenze fisiche e psicologiche del coniuge, a seguito di denuncia, le istituzioni rimangono in un primo momento inerti ed il reato viene dichiarato prescritto, come in molti casi in Lituania, perché i termini di prescrizione sono molto brevi. Secondo il giudice della Corte europea dei diritti dell'uomo, l'effetto utile dell'applicazione in ambito di violenza domestica della Convenzione europea dei diritti dell'uomo «può essere raggiunto solo con una interpretazione e applicazione "gender-sensitive" delle sue previsioni, che tenga conto delle ineguaglianze di fatto tra donna e uomo e del loro impatto sulla vita delle donne. In questa luce, è auto-evidente che la natura profonda della violenza di genere riposa sulla umiliazione intima e sullo svilimento interiore della vittima, che sono esattamente gli scopi a cui l'aggressore mira».

¹³ Come sottolineano B. PEZZINI, *Il diritto e il genere della violenza: dal codice Rocco al codice rosso (passando per la convenzione di Istanbul)*, A. LORENZETTI, B. PEZZINI (a cura di), *Dal codice rosso al codice Rocco. Un itinerario di riflessione plurale attraverso la complessità del fenomeno*, Torino, 2020, 14 ss.; v. ancora B. PEZZINI, *La violenza di genere come questione costituzionale*, in M. D'AMICO, B. LIBERALI (a cura di), *I diritti delle donne. problematiche attuali e prospettive future*, Torino, 2024, 237; A. PITINO (a cura di), *Interventi di contrasto alla discriminazione e alla violenza sulle donne nella vita pubblica e privata: un'analisi multidisciplinare*, Torino, 2015.

¹⁴ Cfr. D. GIANCIPOLI, *Violenza sulle donne, chiedere aiuto è più facile con le app e gli strumenti digitali*, in *Il Sole 24 Ore*, 21 ottobre 2004, in cui si legge dell'iniziativa della Rete nazionale DIRE (Donne in Rete contro la Violenza) della chat chiamata "don't call me".

In questo caso, siamo dinanzi a sistemi di intelligenza artificiale generativi che permettono alle donne di chiedere in modo rapido e sicuro (senza nemmeno dover uscire di casa e in totale anonimato) informazioni sui servizi cui rivolgersi, ricevere un primo inquadramento di quanto sta accadendo e persino essere guidate in una eventuale fuga.

I vantaggi di questa nuova risorsa sono *prima facie* evidenti.

Da un lato, questo strumento è sempre reperibile e accessibile, è gratuito ed è meno rischioso di altri, dall'altro permette di ricevere risposte senza esporsi direttamente con una persona, potendo favorire la ricerca di un primo ascolto.

Infatti, spesso per una donna recarsi in un centro antiviolenza è già un passaggio che, per essere maturato, richiede di compiere un impegnativo percorso di autoconsapevolezza, che conduca la donna a superare il freno derivante dal timore di perdere l'anonimato, di essere scoperta mentre esce di casa per chiedere aiuto, il timore di essere giudicata, nonché il disagio e il senso di colpa, per avere sviluppato l'idea stessa di chiedere aiuto, paragonandola ad una denuncia ¹⁵.

Ovviamente, con questo non si intende sottovalutare l'importanza dell'elemento del contatto umano e dell'empatia che lo dovrebbe connotare: la donna che chiede aiuto si mette infatti in gioco e avvia una relazione con chi la supporterà, creando una rete di persone attorno a sé, che possano comprendere e condividere le sue emozioni, compresa la sua sofferenza e la sua paura, e calibrare in modo sempre recettivo al caso concreto le strategie di aiuto ¹⁶.

Tuttavia, è anche necessario ammettere che un primo ricorso ad un *chatbot* possa vantare elementi positivi, soprattutto se poi la donna trova quelle informazioni che le consentano di "rompere il ghiaccio" e la determinino a rivolgersi a strutture in cui è possibile instaurare un contatto in presenza con personale di supporto.

Durante il Covid, ad esempio, si sono riscontrati positivi elementi nella direzione di strumenti di supporto del genere ¹⁷.

¹⁵ V. P.M. INNOCENZI, *Ai e Violenza sulle donne: un prompt ci salverà?*, in *Agenda digitale*, 8 novembre 2004.

¹⁶ In tema v. B. SANZ, M. LOPEZ, BELLOSO, A.I. CHOPERENA, *Empathy, bias, and data responsibility: Evaluating AI chatbots for gender-based violence support*, in *Frontiers in Political Science*, 2025.

¹⁷ V. in tema G. BARBARA, A. VIERO, I. PELLIZZONE, L. BUGGIO, F. FACCHIN, C. CATTANEO, M.E. D'AMICO, P. VERCELLINI, A. KUSTERMAN, *Intimate Partner Violence in the COVID-19 Era: A Health, Psychological, Forensic and Legal Perspective*, in *International Journal of Environmental Research and Public Health*, 2022, 4973 ss.

Può essere interessante esaminare i dati ancor più articolati che due centri antiviolenza, uno milanese ed uno bergamasco, immediatamente attivatisi convertendo il loro supporto da

Lo sviluppo di sistemi sempre più raffinati e accessibili di questo tipo pare, dunque, che possa essere una importante priorità, il cui finanziamento potrebbe forse anche essere contenuto.

Tale approccio, tuttavia, non deve far trascurare l'importanza della assegnazione di risorse e strutture ai centri antiviolenza, alle case rifugio e alla rete di assistenza sociale coinvolti dopo la prima richiesta di aiuto al *chatbot*, che possano "agganciare" anche su un piano umano la vittima sulla spinta del ricorso al sistema di intelligenza artificiale.

Allo stesso tempo, i sistemi di intelligenza artificiale possono supportare l'esame degli indici di rischio, con cui viene valutata la probabilità di *escalation* della violenza da parte degli operatori e operatrici del settore.

A questo proposito, è bene ricordare che Anna Costanza Boldry ha elaborato nel 1995 in Canada il *Spousal Assault Risk Assessment*. Si tratta di uno strumento molto utile al fine di verificare l'esistenza di indici di rischio legati alle caratteristiche del reo, della vittima, della relazione vittima-omicida o del contesto in cui vivono (es. precedenti, tentati suicidi, possesso di armi, alcol dipendenza o tossicodipendenza, presenza di figli minori, disoccupazione), con cui classificare il rischio di una degenerazione della violenza e prevenirlo¹⁸.

In questo senso, l'esperienza spagnola è molto interessante e offre alcuni spunti di riflessione, ancorché non particolarmente promettenti.

In Spagna infatti è stato elaborato e reso operativo un sistema di valutazione del rischio, utilizzato dalle forze dell'ordine chiamate a intervenire nei casi di

modalità in presenza a modalità online, hanno condiviso con Regione Lombardia: nel relativo report, si sottolinea un incremento di telefonate da parte di donne che già erano in contatto con loro, avendo iniziato il percorso di aiuto, mentre le chiamate di soggetti nuovi ci sono state, ma sono in leggera decrescita (ISTAT, 2020). Questo dato è confermato anche dalla Commissione di inchiesta sul femminicidio del Senato (Relazione sui dati riguardanti la violenza di genere e domestica nel periodo di applicazione delle misure di contenimento per l'emergenza da Covid-19, 23). Si tratta di un aspetto preoccupante, perché fa ipotizzare che le limitazioni alla libertà di movimento possono avere reso più gravoso l'avvio di un percorso di uscita dal morso della violenza domestica da parte di donne che non hanno mai avuto modo di interagire con il centro antiviolenza e porta a supporre una maggiore difficoltà nell'allacciare un contatto con il centro di accoglienza via telefono o via chat. In questo senso, l'accessibilità a *chat bot* andrebbe resa effettiva con campagne di comunicazione adeguate, volte a rafforzare la fiducia delle donne nella accessibilità di aiuti veri, attraverso strumenti di sistemi artificiali o contatti umani.

¹⁸ A.C. BALDRY, *Dai maltrattamenti all'omicidio La valutazione del rischio di recidiva e dell'uxoricidio*, Milano, 2016; v. in tema anche V. BONINI, *Protezione della vittima e valutazione del rischio nei procedimenti per violenza domestica tra indicazioni sovranazionali e deficit interni*, in *Sistema penale*, 2023, 47 ss.

violenza domestica, chiamato VioGen (“Comprehensive Monitoring System in Cases of Gender Violence”) ¹⁹.

Tale sistema, previsto all’interno della legge organica del 2004 ²⁰, dedicata alla prevenzione e al contrasto alla violenza di genere, è partito nel 2007 ed è stato implementato e aggiornato nel 2019.

Il sistema di valutazione individua 35 fattori di rischio, raggruppati in cinque dimensioni o aree, così identificate: storia della violenza nel singolo caso; fattori di rischio del *sex offender*; fattori di vulnerabilità della vittima; circostanze aggravanti e indici di rischio personale della vittima.

Nel 2022, è stata condotta una indagine ²¹ sui risultati applicativi prodotti, che tuttavia ha messo in luce serie criticità, sia nella percezione delle vittime, sia nella valutazione del rischio, venendo le relative classificazioni svolte in modo automatizzato: si pensi che nel 95% dei casi le forze dell’ordine giungevano a modificare la valutazione del rischio, prevalendo la tendenza dell’uomo a sentirsi sostituito “dalla macchina” e ad appiattirsi su stereotipi e *bias* legati al genere, che conducevano ad una sottostima del pericolo, probabilmente anche in ragione dell’assenza di una sufficiente formazione specialistica e dell’assenza di una forte rete antiviolenza, che vedesse veramente collaborare e contaminarsi la lettura delle forze di polizia con quella di attori diversi, ma molto vicini anch’essi alla vittima ²².

Ebbene, un successivo studio ha messo in luce come, con un allenamento di sistemi di intelligenza artificiale di Machine Learning, la classificazione troppo bassa del rischio si sarebbe ridotta del 25% ²³.

¹⁹ V. in tema J.L. GONZÁLEZ-ÁLVAREZ, J.J. LÓPEZ-OSSORIO, C. URRUELA, M. RODRÍGUEZ-DÍAZ, *Integral Monitoring System in Cases of Gender Violence. VioGén System*, in *Behavior & Law Journal*, 2018, 29 ss.

²⁰ Legge organica 28 dicembre 2004, n. 1 (Ley orgánica de medidas de protección integral contra la violencia de género).

²¹ ÉTICAS FOUNDATION, *The External Audit of the VioGén System*, 2022.

²² Su cui v. la ricostruzione e le considerazioni di J.F. PEREIRA, *The Relevance of the Data Collection Process in the VioGén System from a Feminist Perspective*, in *Oñati Socio-Legal Series*, 2024, 1555 ss., in cui si sottolinea come l’assenza di contatto umano generi la diffidenza della vittima e conduca ad una sottovalutazione del rischio da parte della pubblica autorità, per mancanza di contatto diretto con la paura della vittima medesima, elemento fondamentale per comprendere la gravità della situazione; si mette altresì in luce il correlato rischio di *bias* basato sul genere e dunque di rivittimizzazione; cfr. inoltre le considerazioni critiche, in gran parte simili, di M.D. MORONDO TARAMUNDI, *Prevenzione versus predizione del rischio: l’uso del programma semi-automatizzato di decisione VioGén in Spagna*, in *Diritto e questioni pubbliche*, 2025, 69 ss.

²³ Così v. A. GONZÁLEZ-PRieto, A. BRÚ, J.C. NUÑO, J.L. GONZÁLEZ-ÁLVAREZ, *Hybrid*

L'esperienza spagnola immediatamente precedente insegna però che le tecniche di Machine Learning devono essere sempre accompagnate dall'investimento di risorse in formazione degli operatori di polizia e sensibilizzazione dei destinatari e delle destinatarie di tale politica della valutazione del rischio, nonché dalla condivisione trasparente dei dati, nel rispetto della riservatezza della vittima, e dalla cooperazione dei soggetti coinvolti nella rete antiviolenza.

In questo senso, l'ordinamento italiano si è di recente attrezzato, prevedendo, all'art. 5 della legge 24 novembre 2023, n. 168, Disposizioni per il contrasto della violenza sulle donne e della violenza domestica, che, «Al fine di favorire la specializzazione nella trattazione dei processi in materia di violenza contro le donne e di violenza domestica», uno o più procuratori aggiunti o uno o più magistrati «sono sempre specificamente individuati per la cura degli affari in materia di violenza contro le donne e domestica».

Si tratta di una disposizione importante, che è stata in effetti attuata da circolari emesse in varie Procure della repubblica, che non è ancora stata implementata, a quanto consta, attraverso modelli di Machine Learning o altri sistemi di Intelligenza artificiale. Nel caso in cui ciò avvenga, la lezione tratta dall'esperienza spagnola insegna che sarà fondamentale provvedere parallelamente all'adeguata formazione di tutto il personale coinvolto: pubblici ministeri, polizia giudiziaria, avvocati, ecc.

3. Le donne come vittime predestinate dell'IA: focus sul Deep Fake non consensuale.

Contemporaneamente, l'intelligenza artificiale viene in rilievo anche come possibile fonte di violenza verso le donne, giacché un cattivo uso dell'intelligenza artificiale può condurre a nuove forme di aggressione.

I dati sulla trasformazione e adattamento della violenza di genere agli strumenti messi a disposizione dall'intelligenza artificiale sono impressionanti²⁴.

A questo riguardo, occorre tornare a mettere a fuoco che la violenza di genere non costituisce solo la violazione di diritti costituzionali della persona

machine learning methods for risk assessment in gender-based crime, in *Knowledge-Based Systems*, 2023, 110130 ss.

²⁴ Alcuni dati sul modo in cui l'intelligenza artificiale ha influenzato le forme di manifestazione della violenza di genere sono reperibili nel rapporto UNESCO, R. CHOWDHURY, D. LAKSHMI, "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI, 2025.

appartenente al genere subalterno (vita, libertà personale e sessuale, dignità, salute, libertà di circolazione, diritto al lavoro), ma anche la punta estrema delle discriminazioni di genere.

Più nello specifico, particolarmente problematico è il fenomeno, prodotto dall'intelligenza artificiale generativa, della creazione di immagini e storie perfettamente credibili, in cui la vittima viene sottoposta a comportamenti violenti che minano la sua dignità. Si pensi a casi di *stalking* con richieste particolarmente credibili, prodotte da sistemi di intelligenza artificiale generativa (art. 612 *bis* c.p.).

La vera novità di questa tecnologica risiede però nella manipolazione delle immagini, che vengono sessualizzate, ad esempio attraverso la creazione di nudità.

Si tratta del fenomeno del *deep nude*, ovvero una specifica forma di *deep fake*.

Oggi tali fattispecie sono regolate dall'art. 612 *quater* (Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale) c.p., introdotto dall'art. 26 della legge 23 settembre 2025, n. 132 (Disposizioni e deleghe al Governo in materia di intelligenza artificiale), su cui si soffermerà nel prossimo paragrafo, che punisce, come si vedrà meglio nel prossimo paragrafo, genericamente tale fenomeno (a prescindere dalla sua connotazione sessuale).

Prima di esaminare la nuova disciplina, è importante offrire un inquadramento, da un lato, delle tecniche utilizzate e, dall'altro, come queste condotte incidano su dignità e diritti della vittima, nonché se si tratti di un reato ascrivibile alla violenza di genere.

Prima di tutto, è bene ricordare che il termine *deepfake* è composto da “*deep learning*” e “*fake*”. Solitamente, quando si parla di *deep fake*, ci si riferisce dunque a «foto, video e audio creati grazie a software di intelligenza artificiale (AI) 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 voce»²⁵.

Il *deep nude* è ovviamente una sottospecie del *deep fake*, che si caratterizza per la creazione di contenuti relativi alle parti intime della persona.

Ciò premesso, prima dell'intervento normativo che ha incriminato tali condotte, il fenomeno del *deep nude* aveva raggiunto proporzioni enormi e che era divenuto l'emblema del nuovo volto della violenza di genere, colpendo quasi esclusivamente destinatarie di genere femminile. I video pornografici generati con l'intelligenza artificiale e pubblicati nel 2023 sui quaranta principali siti

²⁵ Definizione reperibile nella scheda informativa del Garante della Privacy del dicembre 2020.

specializzati in immagini manipolate hanno accumulato 4,2 miliardi di visualizzazioni. Nel corso del 2023, il numero di video pornografici deepfake caricati online è cresciuto del 54% rispetto all'anno precedente. Tra il 90% e il 95% di questo materiale non consensuale ritrae donne²⁶.

Non ne erano vittime solo donne con alta visibilità (es. attrici, cantanti, politiche, giornaliste), ma anche donne prive di alcun risalto mediatico.

Un esempio significativo del primo tipo è costituito dalla vicenda che ha visto protagonista nel 2019 l'attrice Scarlett Johansson, vittima della creazione di vari video pornografici che la vedevano protagonista, andati incontro a 1,5 milioni di visualizzazioni²⁷.

Nel 2024, si raggiunge, con la cantante Taylor Swift, il numero di 47 milioni di visualizzazioni²⁸.

Non solo. Anche donne che ricoprono importanti cariche politiche ne sono vittima: Giorgia Meloni nel 2020 subisce infatti tale pratica²⁹.

Ma il fenomeno si è molto esteso, essendo stato per parecchio tempo relativamente semplice, "spogliare" virtualmente immagini di donne attraverso l'uso di applicazioni facilmente reperibili *online*.

Questa attività è purtroppo diffusa anche tra i giovani, con effetti drammatici rispetto alla loro salute psicologica³⁰.

Da ultimo, infatti, il 3 ottobre 2025, a seguito dell'esplosione in Italia di un caso riguardante decine di donne famose (attrici, politiche, giornaliste), il Garante per la protezione dei dati personali ha disposto, in via d'urgenza e con effetto immediato, la limitazione provvisoria del trattamento dei dati personali degli utenti italiani nei confronti di una società, con sede nelle Isole Vergini Britanniche, che gestisce l'app Clothoff, una delle tante con cui era particolarmente semplice ottenere la nudità³¹.

²⁶ Cfr. i dati citati da M. TORTORELLI, *Gli abusi sessuali tramite immagini. Limiti applicativi e prospettive di riforma dell'art. 612 ter c.p.*, in *Dir. pen. cont.*, 2024, 221.

²⁷ E. FORZINETTI, *AI. Video porno falsi, un business inarrestabile dietro i deepfakes*, in *Corriere della Sera*, 4 febbraio 2019.

²⁸ *Taylor Swift e AI: negli USA in arrivo la proposta di legge per punire i deepfake non consensuali*, in *Dir. & Giust.*, 2024.

²⁹ *Deepfake porno con Giorgia Meloni, ecco tutti i reati commessi*, in *Agenda Digitale*, 25 ottobre 2023: risulta che Meloni abbia intentato una causa per il risarcimento di 100.000 euro.

³⁰ G. CANÈ, *Studenti «spogliano» le compagne con l'intelligenza artificiale: sempre più deepfake nelle scuole (anche in Italia?)*, in *Corriere della Sera*, 7 novembre 2023.

³¹ Può valere la pena ricordare che questa app, utilizzata ovviamente anche all'estero, è stata oggetto di denuncia da parte di una ragazza diciassettenne americana: R. COSENTINO, *17enne*

Si tenga conto che le conseguenze di tale comportamento sono particolarmente difficile da estirpare, perché, anche se nell'ordinamento in cui si agisce è prevista una tutela legale, può risultare molto difficile cancellare l'immagine dal web e che l'effetto per la vittima può essere devastante, inducendola a nascondersi e arretrare dalla sfera pubblica: *internet never forgets*³².

Non potendosi addentrare oltre nella descrizione di una casistica, occorre a questo punto domandarsi se l'utilizzo di immagini creati dalla IA di corpi di una donna costituiscano violenza.

La risposta pare nettamente positiva, integrando il *deep nude*, da un lato, il furto di identità, per cui come si vedrà soccorrono le norme del diritto d'autore, e, dall'altro, una vera e propria violenza sessuale³³.

Infatti, con l'atto di spogliare e dunque esibire le parti intime del corpo femminile, l'autore della condotta s'impadronisce del corpo della donna e incide sull'immagine della sua sfera sessuale, interferendo con la sua libertà di autodeterminazione in tale ambito, oramai indubbiamente riconosciuta a livello costituzionale, negli artt. 2, 3 e 13 Cost. e dagli artt. 8 e 14 della Convenzione europea dei diritti dell'Uomo³⁴.

fa causa a ClothOff, l'app che «spoglia» le persone: un compagno di scuola la usava per creare sue foto di nudo, in *Corriere della Sera*, 17 ottobre 2025.

³² Come fa notare M. TORTORELLI, *Gli abusi sessuali tramite immagini. Limiti applicativi e prospettive di riforma dell'art. 612 ter c.p.*, cit., 210, cui si rinvia anche agli impressionanti dati che attestano l'impennata del fenomeno del *deep nude*.

³³ In tema v. G. FIORINELLI, *La violenza mediata dalla tecnologia. Dogmatica, profili politico-criminali e interpretazione della nozione di violenza nel diritto penale delle tecnologie digitali*, Torino, 2024.

³⁴ Cfr. da ultimo Corte cost., sentenza n. 141/2019, in cui si legge: «Che la libertà sessuale costituisca un diritto fondamentale della persona sarebbe pacifico, essendo stato affermato dalla Corte costituzionale con la sentenza n. 561 del 1987. Essa è garantita, altresì, da norme sovranazionali, quali gli artt. 8 e 14 della Convenzione per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali (CEDU), firmata a Roma il 4 novembre 1950, ratificata e resa esecutiva con legge 4 agosto 1955, n. 848» (punto n. 4.2 del Considerato in diritto).

In precedenza, con la sentenza n. 561/1987, la Corte ha affermato «la sessualità rappresenta uno degli essenziali modi di espressione della persona umana», con la conseguenza che «il diritto di disporre liberamente è senza dubbio un diritto soggettivo assoluto, che va ricompreso tra le posizioni soggettive direttamente tutelate dalla Costituzione ed inquadrato tra i diritti inviolabili della persona umana che l'art. 2 Cost. impone di garantire».

L'affermazione è stata resa «in rapporto a una fattispecie nella quale veniva in rilievo il profilo negativo di tale libertà, ossia il diritto ad opporsi a “intrusioni” altrui non volute nella propria sfera sessuale, e con riguardo alle pretese risarcitorie scaturenti dalla violazione di tale

Di più. Viene evidentemente leso il diritto all'identità personale e sessuale, violato dalla esibizione di parti intime create in modo artificiale, e della dignità nonché della reputazione individuale, calpestate dalla trasformazione dell'immagine della persona, da vestita a nuda.

Del resto, è oramai assodato nella giurisprudenza della Corte di Cassazione che la violenza sessuale possa consumarsi anche senza contatto³⁵.

Prima dell'intervento normativo sopra citato, che, nel 2025, ha introdotto nel codice penale l'art. 612 *quater*, ci si sarebbe potuti interrogare sulla riconducibilità di questa fattispecie nell'alveo del c.d. *revenge porn*, disciplinato dall'art. 612 *ter* c.p. (Diffusione illecita di immagini o video sessualmente espliciti), altra ipotesi di reato non di rado connotata dall'appartenenza alle figure penalmente rilevanti volte a reprimere forme di violenza di genere.

Si tratta infatti di una condotta ispirata a finalità di scherno, vendetta o ricatto, talvolta nate all'interno di una relazione di coppia, in cui la vittima è punita attraverso la diffusione di immagini captate in momenti intimi e col suo consenso, ma poi strumentalizzate dal partner, di genere maschile, che le ha scattate in precedenza.

Secondo questa norma, salvo che il fatto costituisca più grave reato, è punita con la reclusione da uno a sei anni e con la multa da euro 5.000 a euro 15.000, chiunque, dopo averli realizzati o sottratti, invia, consegna, cede, pubblica o diffonde immagini o video a contenuto sessualmente esplicito, destinati a rimanere privati, senza il consenso delle persone rappresentate.

La disposizione prevede altresì che la stessa pena si applichi a chi, avendo ricevuto o comunque acquisito le immagini o i video di cui al primo comma, li invia, consegna, cede, pubblica o diffonde senza il consenso delle persone rappresentate al fine di recare loro nocumento.

La differenza tra le due fattispecie è tuttavia evidente: nel primo caso, del c.d. *revenge porn*, l'autore dell'invasione nella sfera sessuale diffonde una foto intima, violando dunque prima di tutto il bene della riservatezza individuale³⁶.

diritto. Si lamentava, infatti, nell'occasione, che la disciplina sul trattamento pensionistico di guerra escludesse la risarcibilità dei danni non patrimoniali patiti dalle vittime di violenze carnali consumate in occasione di fatti bellici».

³⁵ Cfr. Cass., Sez. III, sentenza 19 novembre 2015 (dep. 5 maggio 2016), n. 18679, Pres. Amoresano, Rel. Aceto, con nota di M. CAPPALÀ, *La qualificazione delle "violenze" prive di un contatto corporeo corpori alla prova della recente giurisprudenza in tema di atti sessuali*, in *Dir. pen. cont.*, 2017, 43 ss.

³⁶ Come sottolinea M. TORTORELLI, *Gli abusi sessuali tramite immagini. Limiti applicativi e prospettive di riforma dell'art. 612 ter c.p.*, cit., 207 ss.

Può essere significativo, a questo proposito, soffermarsi un momento sul bene protetto dalla norma penale, messo a fuoco in giurisprudenza.

Particolarmente significativa al riguardo è una sentenza della Corte di Cassazione del 2023, in cui si afferma che «il reato, infatti, è configurabile come istantaneo e si consuma nel momento in cui avviene il primo invio dei contenuti sessualmente espliciti, non importa se diretto a familiari della vittima, che possano eventualmente avere interesse a non alimentare una successiva diffusione»³⁷.

È con tale primo invio che si verifica la diffusione penalmente rilevante, poiché la disposizione incriminatrice «non fa questione di reiterazione della condotta diffusiva né “quantifica” o qualifica in alcun modo la diffusione lesiva del bene protetto».

Il passaggio risulta particolarmente interessante anche perché consente alla Corte di soffermarsi sul bene giuridico tutelato dalla nuova incriminazione, enfatizzandone in modo originale la componente dell'autodeterminazione sessuale: «il reato è inserito tra quelli a tutela della libertà morale individuale e si rivolge alla sfera di intimità e della privacy, intesa quale diritto a controllare l'esposizione del proprio corpo e della propria sessualità, in un'ottica di autodeterminazione della sfera sessuale individuale».

Nel secondo caso, quello del *deep nude* l'autore del reato crea l'immagine lesiva e poi la diffonde; si incontra dunque un elemento peculiare, ovvero la *trasformazione* dell'immagine che ritrae parti del corpo della vittima, che vengono spogliate senza il suo consenso, ledendone – lo si ripete – la libertà di autodeterminazione in materia sessuale, l'identità e la dignità.

Per tale discrasia, la dottrina, negando che il *deep nude* potesse combaciare con l'area penalmente rilevante dell'art. 612 *ter*, era incline ad escludere che il *deep nude* potesse essere punito attraverso il reato del *revenge porn*³⁸, risultando dunque questa nuova forma di violenza di genere priva di incriminazione o comunque di altro trattamento sanzionatorio effettivamente “cucito” sulla sua specificità.

La disciplina applicabile era dunque limitata ad alcune norme sanzionatorie in tema di diritto d'autore: ovvero l'art. 10 e gli artt. 96 e 97, della legge 22

³⁷ Cass. pen., Sez. V, sentenza 22 febbraio 2023, n. 14927 (dep. 7 aprile 2023), Rel. Brancaccio.

³⁸ Così v. G.M. CALETTI, *Libertà e riservatezza sessuale all'epoca di internet. L'art. 612-ter c.p. e l'incriminazione della pornografia non consensuale*, in *Riv. it. dir. e proc. pen.*, 2019, 2044 ss. e M. TORTORELLI, *Gli abusi sessuali tramite immagini. Limiti applicativi e prospettive di riforma dell'art. 612 ter c.p.*, cit., 221 ss.

aprile 1941, n. 633, c.d. legge sul diritto d'autore³⁹, e all'art. 2059 c.c., circa il risarcimento del danno non patrimoniale.

4. *Un importante passo avanti: l'intervento normativo attraverso la recente legge n. 132/2025.*

La legge n. 132/2025, all'art. 26, è intervenuta dunque, anche a seguito dei gravi fatti di cronaca, citati nel precedente paragrafo, che avevano visto decine di donne di rilievo pubblico "spogliate" su siti internet, per far fronte a questo uso violento dei sistemi di intelligenza artificiale generativa.

Non solo. La norma può essere anche letta come forma di attuazione della Direttiva (UE) 2024/1385 del Parlamento europeo e del Consiglio del 14 maggio 2024 sulla lotta alla violenza contro le donne e alla violenza domestica⁴⁰, sebbene probabilmente aspiri ad un risultato più ampio, ricomprendendo nel suo raggio applicativo tutti i casi di *deepfake* non consensuali, compresi quelli realizzati a scopo di disinformazione e dunque a prescindere dalla loro connotazione sessuale⁴¹.

Dinanzi alla seria lacuna, che vedeva sfuggire a sanzione penale tale condotta, il legislatore, introducendo l'art. 612 *quater* c.p. (Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale), ha previsto quanto segue:

«Chiunque cagiona un danno ingiusto ad una persona, cedendo, pubblicando o altrimenti diffondendo, senza il suo consenso, immagini, video o voci falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuinità, è punito con la reclusione da uno a cinque anni.

³⁹ L'art. 96 stabilisce che l'immagine di una persona non può essere mostrata, riprodotta o commercializzata senza il suo permesso, salvo quanto previsto dall'articolo successivo. L'art. 97 chiarisce che il consenso della persona ritratta non è necessario quando l'utilizzo dell'immagine è giustificato dalla sua notorietà o dal ruolo pubblico ricoperto, da esigenze di giustizia o di polizia, oppure da finalità scientifiche, didattiche o culturali. Il consenso non è inoltre richiesto quando l'immagine è legata a fatti, eventi o cerimonie di interesse pubblico o svolti in luogo pubblico. Tuttavia, il ritratto non può essere esposto o venduto se l'esposizione o messa in commercio rechi pregiudizio all'onore, alla reputazione od anche al decoro nella persona ritrattata.

⁴⁰ In part., v. l'art. 5, par. 1, lett. b). In tema v. S. BRASCHI, *La nuova direttiva sulla lotta alla violenza contro le donne e alla violenza domestica e le sue ricadute nell'ordinamento nazionale*, in *Dir. pen. proc.*, 2024, 1367 ss.

⁴¹ Così v. B. FRAGASSO, *Intelligenza artificiale e responsabilità penale. Principi e categorie alla prova di una tecnologia "imprevedibile"*, Torino, 2025. 204 ss.

Il delitto è punibile a querela della persona offesa. Si procede tuttavia d'ufficio se il fatto è connesso con altro delitto per il quale si deve procedere d'ufficio ovvero se è commesso nei confronti di persona incapace, per età o per infermità, o di una pubblica autorità a causa delle funzioni esercitate».

Sul terreno del diritto costituzionale, è ovviamente fondamentale notare che, laddove invece sussista il consenso, la persona che ha prodotto le immagini falsificate o alterati mediante l'impiego di sistemi di intelligenza artificiale, è libera di diffonderle.

Occorre inoltre sottolineare che siamo dinanzi ad un reato di danno: infatti il reato non si integra quando le immagini, i video o le voci falsificati o alterati non sono «idonei a indurre in inganno sulla loro genuinità».

Non basta quindi che la persona si senta lesa da tale manipolazione della sua immagine, ma occorre che i destinatari della stessa possano travisare la assenza di veridicità della stessa.

Si tratta di una limitazione importante, funzionale ad evitare uno sconfinamento in forme limitative della libertà individuale di creare immagini rilevante ai fini dell'art. 21 o dell'art. 33 Cost., laddove funga da strumento di manifestazione del pensiero o da espressione di una forma artistica.

Il reato è perseguibile a querela (salvo ipotesi più gravi, come ad esempio quando la vittima è un minore o una persona incapace per età o per infermità o rappresenta una pubblica funzione e lo Stato), venendo lasciata alla persona la scelta sull'esistenza di un interesse a punire l'autore del reato.

Fermo restando che le conseguenze applicative di questa nuova fattispecie di reato andranno attentamente monitorate, è importante sottolineare che, in effetti, rispetto al *deep nude*, l'intrusione di queste condotte nella sfera sessuale è molto incisiva, ancorché si tratti di una intromissione figurata e mediata dall'immagine.

Ciò non toglie che simili attività siano manifestazione di una forma di violenza di genere commessa su larga scala e molto insidiosa su un piano psicologico ed eventualmente economico, dati i suoi aspetti reputazionali, che potrebbero danneggiare la posizione di chi ne è colpita.

5. Prevenzione della violenza nei confronti delle donne e dell'intelligenza artificiale: un primo bilancio.

In conclusione, è bene essere consapevoli che la violenza di genere è un fenomeno fortemente radicato. Lo dimostrano gli stereotipi di genere di cui ancora oggi è imbevuta la cultura e la società.

Proprio per questo motivo, è bene osservare la massima attenzione, dotandosi delle necessarie competenze interdisciplinari, quando ci si approccia ai sistemi di intelligenza artificiale generativa, che possono avere riflessi positivi o negativi per il fenomeno della violenza di genere.

Il bilancio a questo punto è estremamente provvisorio e apre uno squarcio su una realtà molto fragile.

Nel caso del primo itinerario (sistemi di intelligenza artificiale che possono supportare la vittima della violenza), la diffidenza a ampliare il novero degli deve essere certamente superata⁴², ma occorre sempre ricordare che l'intelligenza artificiale può celare insidie molto pericolose: la formazione dei *data set*, che possono nutrirsi di stereotipi che alimentano la indulgenza sociale verso la violenza di genere e dunque la vittimizzazione secondaria, va certamente curata al massimo; inoltre, il personale umano che ricorre alla intelligenza artificiale, per supportare le vittime di violenza, deve essere formato in modo altamente specialistico, per vigilare sul corretto funzionamento dello strumento dell'intelligenza artificiale e reattivo dinanzi alle peculiarità ed esigenze dei casi della vita.

La sfida più complessa, in questo senso, è costruire un sistema poroso rispetto ad uno sguardo *gender sensitive*, indispensabile per azioni che sappiano davvero identificare rischi, nonché punti di forza e di debolezza del caso, ed essendo richiesta la mutua partecipazione di dati rilevanti in modo diverso per le varie discipline coinvolte nell'ambito della violenza di genere.

Ciò richiederà un notevole sforzo di specializzazione, alta organizzazione che permetta la collaborazione di tutte le figure coinvolte, cura nel reperimento dei dati, risorse economiche.

D'altra parte, guardando al secondo itinerario (sistemi di intelligenza artificiale che possono essere usati per commettere violenza), l'intelligenza artificiale generativa può fare davvero male, portando la violenza a infiltrarsi in nuove

⁴² Come osserva Francesca Maur, intervistata da D. GIANCIPOLI, *Violenza sulle donne, chiedere aiuto è più facile con le app e gli strumenti digitali*, in *Il Sole 24 Ore*, 21 ottobre 2004: «Stiamo attivando nuovi servizi consapevoli del fatto che l'ampia diffusione dei media digitali in ogni ambito della vita ha determinato e continua a determinare nuovi processi di soggettivazione, sia online che offline, delle donne e delle ragazze. Questi media, infatti, da un lato condizionano gli immaginari e i modelli di genere, dall'altro sfidano le norme sui corpi, la sessualità e le relazioni esistenti. Sono quindi veicoli di una complessa ambiguità, in cui coesistono norme rigide sulle relazioni di genere insieme a processi di liberazione riguardo a queste stesse relazioni e ai comportamenti sessuali». È proprio su questa ambiguità che occorre concentrare l'attenzione, cercando di trasformare le potenzialità dell'intelligenza artificiale in strumenti per decostruire gli stereotipi e non per rafforzarli.

forme nella società e nella cultura: emblematico è l'esponenziale aumento del fenomeno del *deep nude*.

Vedremo se l'impatto dell'entrata in vigore dell'art. 614 *ter* c.p. porterà a risultati interessanti, agendo su un piano di prevenzione generale.

Anche se il rischio che la norma sia paralizzata dalla capacità del *web* di inghiottire le immagini lesive, una volta che ha iniziato a circolare, non può essere affatto sottovalutato.

Affinché tale rischio non si concretizzi, è ancora una volta dato da interventi di sensibilizzazione e formazione, che diffondano conoscenze circa gli effetti lesivi per la salute psicologica di simile prassi, specie tra le giovani generazioni.

Da questo punto di vista, dunque, i due itinerari di riflessione qui presentati sono convergenti.

Insomma, le potenzialità dell'intelligenza artificiale rischiano in maniera non certo lieve o limitata di trasformarsi in una nuova arma della cultura misogina, che ancora oggi è diffusa trasversalmente nella società.

Di per sé, l'intelligenza artificiale non costituisce naturalmente uno strumento di cui avere paura e da allontanare, necessariamente vietare e men che meno da demonizzare: la chiave di volta per raccogliere la sfida risiede probabilmente nella consapevolezza che, per limitarne le distorsioni e la forte capacità lesiva, per libertà e dignità della donna, dell'intelligenza artificiale, è ancora più importante diffondere una cultura della parità, in cui le radici della violenza nei confronti del genere femminile non attecchiscano.

Diversamente, non si assisterà ad altro se non all'affacciarsi sulla scena di nuove forme di manifestazione del fenomeno della violenza di genere.

In conclusione, la prospettiva del diritto costituzionale, da questo punto di vista, richiede di vigilare sull'evoluzione della realtà e cercare di contribuire ad un approccio attento ai sintomi di nuove forme di vittimizzazione del genere femminile senza rifuggire dalle opportunità che un buon uso di questo strumento potrà avere proprio in relazione all'evoluzione verso la parità di genere.

CAPITOLO VIII

SALUTE DELLE DONNE E REGOLAMENTAZIONE
DELL'IA: LA *GENDER-SENSITIVE GOVERNANCE* COME
PARADIGMA DI EGUAGLIANZA SOSTANZIALE

Marta Tomasi

SOMMARIO: 1. Introduzione. L'irrealizzato obiettivo dell'eguaglianza in salute. – 2. Salute delle donne e discriminazioni strutturali. – 3. L'automatizzazione delle diseguaglianze. – 4. La produzione del dato come spazio di diseguaglianza: la ricerca biomedica tra esclusione e *bias*. – 5. La pratica clinica come spazio di diseguaglianza: diagnosi, cura e dimensione strutturale della salute. – 6. La ricerca dell'eguaglianza di genere nello spazio della regolazione. – 7. Soluzioni normative a confronto: equità sostanziale e *gender-sensitive governance*.

1. Introduzione. L'irrealizzato obiettivo dell'eguaglianza in salute.

Il perseguimento di una piena eguaglianza sostanziale nel campo della salute richiede il superamento delle molteplici differenze che tuttora ne permeano le strutture e le pratiche; differenze che, di frequente, sono riconducibili a variabili quali il sesso, inteso nella sua dimensione biologica, e il genere, quale costruzione sociale di ruoli, aspettative e relazioni di potere¹. In questa prospettiva, la salute delle donne rappresenta una categoria di analisi prioritaria, non già per una pretesa di esclusività, ma in quanto punto di avvio necessario di un discorso più ampio, volto a interrogare criticamente concetti di salute e di medicina

¹ Si faccia riferimento al c.d. *sex-gender system*, descritto da G. RUBYN nel saggio *The Traffic in Women* (1975). Qui l'Autrice qualifica il *genere* come «l'insieme dei processi, adattamenti, modalità di comportamento e di rapporti, con i quali ogni società trasforma la sessualità biologica in prodotti dell'attività umana e organizza la divisione dei compiti tra gli uomini e le donne, differenziandoli l'uno dell'altro» (si v. C. SARACENO, S. PICCONE STELLA, *Genere: la costruzione sociale del femminile e del maschile*, Bologna, 1996, 7).

tradizionalmente costruiti su parametri maschili e solo in apparenza neutri ².

Naturalmente e di conseguenza, tale settore rappresenta anche un terreno elettivo per riflettere intorno alle potenzialità e, al tempo stesso, ai rischi connessi all'impiego di sistemi algoritmici in ambito sanitario. Poiché, infatti, l'intelligenza artificiale (IA) fonda il proprio funzionamento su processi di apprendimento automatico che si nutrono di dati storici e statistici, gli output che essa produce riflettono spesso un sapere medico costruito nel tempo su campioni prevalentemente maschili e su modelli androcentrici di normalità biologica ³. Ne deriva che i *bias* di genere, già radicati nella pratica clinica e nella ricerca scientifica, rischiano di essere replicati, amplificati e automatizzati dalle tecnologie digitali, assumendo così una forma meno visibile, ma più pervasiva. Si rende, dunque, sempre più necessaria una indagine circa la capacità dei sistemi normativi di individuare e correggere le asimmetrie che incidono sull'accesso alle cure, sulla ricerca scientifica e, più in generale, sulla definizione stessa dei bisogni sanitari.

2. Salute delle donne e discriminazioni strutturali.

La salute delle donne è storicamente attraversata da forme di discriminazione strutturale, spesso opache e difficili da intercettare perché radicate in pratiche, linguaggi e modelli epistemici apparentemente neutri. Tali disuguaglianze si sono tradotte, nel tempo, in marginalizzazione, sottorappresentazione e in una inadeguata considerazione delle specificità biologiche e sociali femminili. Come efficacemente sintetizzato da Silvio Garattini e Rita Banzi nel titolo del loro volume *La medicina che penalizza le donne* ⁴, la medicina contemporanea continua in larga misura ad assumere il corpo maschile quale modello di riferimento, relegando quello femminile a una sua mera variante. Questo paradigma ha teso a perimetrare le specificità della salute delle donne all'interno della sola

² L'elaborazione del concetto di "androcentrism", che si riferisce alla centralità e al privilegio delle prospettive maschili nella produzione del sapere, si deve a C.P. GILMAN, *The Man-Made World; Or, Our Androcentric Culture*, Gutenberg, 1911. Un concetto correlato è quello di "gender insensitivity", sviluppato da M.T. RUIZ-CANTERO *et al.*, *A framework to analyse gender bias in epidemiological research*, in *Journal of Epidemiology Community Health*, 61(suppl II), 2007, ii46-ii53.

³ S. ROSSER, *Re-visioning clinical research: gender and the ethics of experimental design*, in *Hypatia*, 2, 1989, 125-139. Si v. anche M. EICHLER, *Non sexist research methods: a practical guide*, New York, 2013, spec. 146.

⁴ S. GARATTINI, R. BANZI, *La medicina che penalizza le donne*, Roma-Bari, 2022.

dimensione riproduttiva, relegando altri ambiti di rilevante impatto clinico ed epidemiologico in un cono d'ombra. I dati più recenti, per esempio, mostrano che meno del 2% della ricerca e dello sviluppo globale in ambito sanitario è dedicato a patologie specificamente femminili, escluse quelle oncologiche. Le donne risultano, inoltre, sistematicamente sottorappresentate nelle sperimentazioni cliniche – sebbene negli ultimi anni si registri un parziale miglioramento – con la conseguenza che disponiamo di informazioni limitate sull'efficacia e, ancor più, sulla tossicità dei farmaci, spesso maggiore nelle donne a causa della calibrazione su popolazioni maschili. A ciò si aggiunge un ulteriore divario nella tempestività e accuratezza diagnostica: le donne attendono in media più degli uomini per ricevere una diagnosi corretta, soprattutto in presenza di patologie che si manifestano con sintomatologie differenziate per genere o che, se riferite da pazienti donne, sono state storicamente ricondotte, in ragione di persistenti stereotipi di genere, a cause psicosomatiche. Si pensi, ad esempio, alle malattie autoimmuni o alle diverse espressioni del dolore cronico ⁵.

In risposta a tali criticità, a partire dagli anni Novanta si è progressivamente affermato l'approccio della medicina di genere o medicina genere-specifica ⁶, volto a superare la visione tradizionalmente androcentrica della ricerca e della pratica medica. Tale prospettiva non si limita a considerare le differenze biologiche o sessuali tra uomini e donne, ma adotta un paradigma comprensivo e multidimensionale, che integra fattori sociali, culturali, economici e geografici,

⁵ *Ibidem*. Molti dati sono riportati nel WORLD ECONOMIC FORUM, *Closing the Women's Health Gap: A \$1 Trillion Opportunity to Improve Lives and Economies. Insight report*, 2024, online: WEF_Closing_the_Women's_Health_Gap_2024.pdf.

⁶ La letteratura giuridica è ormai piuttosto ampia sul tema. Si v. almeno gli scritti di F. RESCIGNO, *Medicina di genere e autodeterminazione femminile: un percorso giuridico accidentato*, in B. PEZZINI, A. LORENZETTI (a cura di), *70 anni dopo tra uguaglianza e differenza. Una riflessione sull'impatto di genere nella Costituzione e nel costituzionalismo*, Torino, 2019, 203 ss. e il lavoro monografico *Per un habeas corpus di genere*, Napoli, 2022. È essenziale ricordare che in Italia, il 31 gennaio 2018, è stata approvata la legge n. 3/2018 “*Delega al Governo in materia di sperimentazione clinica di medicinali nonché disposizioni per il riordino delle professioni sanitarie e per la dirigenza sanitaria del Ministero della Salute*”. L'art. 3 della legge, dedicato alla “Applicazione e la diffusione della medicina di genere nel Servizio sanitario nazionale”, richiedeva la predisposizione di «un Piano volto alla diffusione della medicina di genere mediante divulgazione, formazione e indicazione di pratiche sanitarie che nella ricerca, nella prevenzione, nella diagnosi e nella cura tengano conto delle differenze derivanti dal genere, al fine di garantire la qualità e l'appropriatezza delle prestazioni erogate dal Servizio sanitario nazionale in modo omogeneo sul territorio nazionale». Il piano, prodotto congiuntamente dal ministero della Salute e dal Centro di riferimento per la medicina di genere dell'Istituto superiore di sanità, è stato approvato il 13 giugno 2019.

con l'obiettivo di costruire percorsi diagnostici e terapeutici più accurati, equi e inclusivi ⁷.

La salute delle donne riveste dunque un ruolo paradigmatico, sebbene condiviso con altri gruppi marginalizzati, nell'analisi delle asimmetrie nell'ambito della salute; un ruolo che si conferma e si intensifica allorquando entri in relazione con le tecnologie, pervasive e trasformative, dell'IA ⁸.

3. *L'automatizzazione delle disuguaglianze.*

Sul descritto terreno di disuguaglianze strutturali e persistenti in ambito sanitario si innestano, oggi, le applicazioni dell'IA. Si tratta di sistemi che operano mediante l'elaborazione di dati storici e di categorie predeterminate, producendo risultati attraverso processi di inferenza statistica che possono generarsi in modo parzialmente autonomo e che, per questo, non sono sempre pienamente leggibili o verificabili ⁹. Proprio tale meccanismo di apprendimento automatico può condurre a decisioni discriminatorie in modo diretto o indiretto, negando l'accesso a un bene o a un'opportunità, oppure attribuendo esiti sfavorevoli sulla base di criteri irragionevoli, inappropriati o non trasparenti ¹⁰. Gli

⁷ MINISTERO DELLA SALUTE, *Il genere come determinante di salute. Lo sviluppo della medicina di genere per garantire equità e appropriatezza della cura*, in *Quaderni del Ministero della Salute*, 26, 2016, 3. Si veda anche K.W. SPRINGER, J. MAGER STELLMAN, R.M. JORDAN-YOUNG, *Beyond a catalogue of differences: A theoretical frame and good practice guidelines for researching sex/gender in human health*, in *Social Science & Medicine*, 74, 11, 2012, 1817-1824.

⁸ Su questi aspetti, si v. M. FASAN, *L'intelligenza artificiale alla prova dell'eguaglianza: l'evoluzione del diritto costituzionale tra scenari attuali e futuri*, in *Annali di studi religiosi*, 26, 2025, 539-558 e, più in generale, ID., *Intelligenza artificiale e costituzionalismo contemporaneo. Principi, diritti e modelli in prospettiva comparata*, Napoli, 2024.

⁹ Si v. gli scritti di una delle Autrici più attente al tema: C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, in *Riv. "Gruppo di Pisa"*, 3, 2021, 9-60; M. D'AMICO, C. NARDOCCI, *Intelligenza artificiale e discriminazioni di genere: rischi e possibili soluzioni*, in G. CERRINA FERONI, C. FONTANA, E.C. RAFFIOTTA (a cura di), *AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Bologna, 2022, 251-268; C. NARDOCCI, *Artificial Intelligence-based Discrimination: Theoretical and Normative Responses. Perspectives from Europe*, in *DPCE online*, 3, 2023, 2367-2393.

¹⁰ Secondo la nota e definizione di Batya Friedman e Helen Nissenbaum, i sistemi di IA sono biased quando «systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate» (B. FRIEDMAN, H. NISSENBAUM, *Bias in Computer Systems*, in *ACM Transactions on Information Systems*, 1996, 14, 330).

algoritmi rischiano quindi di replicare, quando non amplificare, le distorsioni già presenti nei sistemi sanitari e nei dati su cui essi si fondano, anche in assenza di un intento discriminatorio consapevole, rendendo più difficile individuarne la fonte e, conseguentemente, ricondurlo a responsabilità definite. In altri termini, l'IA – se non progettata, validata e implementata secondo specifici criteri – può agire come moltiplicatore dei *bias* esistenti, rendendoli più sistemici, pervasivi e opachi. L'effetto di “automazione della discriminazione”¹¹ che ne deriva pone dunque interrogativi rilevanti non solo sul piano tecnico, ma anche su quello giuridico, in quanto rischia di compromettere il principio di uguaglianza sostanziale e una concezione inclusiva del diritto alla salute¹².

È ormai ampia la produzione, tanto giuridica quanto sociologica, volta a individuare e categorizzare i punti di ingresso dei *bias* all'interno del ciclo di vita dei sistemi di IA. Una delle tassonomie più note in questo senso è quella proposta da Solon Barocas e Andrew Selbst, che distinguono cinque livelli nei quali la distorsione può insorgere e consolidarsi¹³. Di questi, tre risultano particolarmente rilevanti per l'ambito sanitario. Il primo concerne l'individuazione delle c.d. variabili target, ossia gli obiettivi che il sistema di IA è progettato per prevedere o ottimizzare. La scelta di tali variabili non è mai neutra: essa può infatti riflettere pregiudizi sociali o disuguaglianze storiche, determinando così l'inclusione, nel modello, di assunti impliciti che riproducono visioni parziali della realtà. Il secondo avviene nella fase di selezione dei dati di addestramento, nella quale la costruzione dei dataset può incorporare squilibri di rappresentazione, lacune nella raccolta o errori di classificazione, compromettendo la validità del modello sin dalla sua origine. Il terzo ha a che fare con la definizione dei criteri o modelli di inferenza, vale a dire la scelta delle variabili e dei pesi assegnati agli attributi che concorrono alla generazione del risultato. Scelte metodologiche scorrette o non trasparenti possono condurre a esiti distorti, anche in assenza di un intento discriminatorio consapevole. A tali dimensioni si aggiungono, secondo gli Autori, le variabili *proxy* – attributi non esplicitamente sensibili ma fortemente correlati con caratteristiche protette come il sesso,

¹¹ Sul tema della *discriminazione algoritmica* e della sua natura non intenzionale, v. S. WACHTER, B. MITTELSTADT, C. RUSSELL, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*, in *Computer Law & Security Review*, vol. 41, 2021, 105567. Secondo gli Autori, «*Compared to traditional forms of discrimination, automated discrimination is more abstract and unintuitive, subtle, intangible, and difficult to detect*».

¹² Su questi temi, in generale, L. BUSATTA, *La salute sostenibile. La complessa determinazione del diritto ad accedere alle prestazioni sanitarie*, Torino, 2018.

¹³ S. BAROCAS, A.D. SELBST, *Big Data's Disparate Impact*, in *California Law Review*, 104, 2016, 671-732.

l'etnia o lo status socioeconomico – e infine le ipotesi di discriminazione intenzionale, che si verificano quando il pregiudizio viene deliberatamente incorporato nel sistema, talvolta mascherato da criteri apparentemente neutrali.

Accanto ai punti di ingresso più strettamente tecnici, è importante richiamare un ulteriore livello, di natura strutturale e culturale, che Sulmicelli definisce “*bias* fondazionale”¹⁴. Esso si radica nella mancanza di diversità all'interno del settore dell'IA, nella scarsa rappresentatività di alcune categorie sociali nei processi di progettazione tecnologica e, più in generale, in una lettura parziale e culturalmente situata dei dati utilizzati. I *big data* non sono, infatti, entità neutre o oggettive: essi riflettono il contesto culturale, sociale e politico in cui vengono raccolti, selezionati e interpretati¹⁵. Di conseguenza, anche concetti apparentemente “oggettivi”, come il genere o la salute, risultano il prodotto di una costruzione culturale che incide in maniera significativa sul modo in cui tali attributi vengono registrati, codificati e, infine, trattati dai sistemi algoritmici. Questo livello fondazionale di distorsione richiede, pertanto, di interrogare i presupposti conoscitivi e le strutture di potere che informano la produzione dei dati e la definizione dei modelli predittivi.

È chiaro, dunque, in che termini l'uso dell'IA nei contesti sanitari imponga una riflessione rinnovata sui principi di uguaglianza sostanziale, non discriminazione e parità di accesso alle cure, quali corollari essenziali del diritto fondamentale alla salute. Invero, il rischio discriminatorio assume, nel caso dell'IA applicata alla salute, una natura capillare e trasversale, in ragione della pervasività della tecnologia lungo l'intera filiera sanitaria. Essa, infatti, è oggi in grado di incidere su ogni fase del sistema: dalla ricerca e sperimentazione clinica, alla diagnosi, alla prevenzione e alla cura, fino a influenzare gli aspetti organizzativi e gestionali delle strutture sanitarie. In questa estensione sistemica risiede, al tempo stesso, il potenziale trasformativo e la criticità dell'IA.

4. *La produzione del dato come spazio di disuguaglianza: la ricerca biomedica tra esclusione e bias.*

Il primo e più evidente punto di vulnerabilità emerge nella fase della ricerca, con riferimento alla rappresentatività delle popolazioni coinvolte negli studi

¹⁴ S. SULMICELLI, *Queer-responsive regulation for artificial intelligence in healthcare: a comparative study*, in *University of New South Wales Law Journal*, 48, 4, 2025.

¹⁵ Su questi temi si veda il fondamentale contributo di K. CRAWFORD, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, 2021.

clinici¹⁶. Si tratta di diseguaglianze scarsamente discusse, ma profondamente strutturanti, che determinano la qualità e l'affidabilità stessa dei dati scientifici su cui si fondano i modelli di IA.

Molti studi clinici, ancora oggi, vengono condotti su popolazioni demograficamente omogenee, composte in prevalenza da uomini giovani e caucasici¹⁷, con l'effetto di escludere una serie di variabili – tra cui il genere, oltre all'età, l'origine etnica e le condizioni socioeconomiche – che incidono in modo determinante sulla prevalenza delle patologie, sulla manifestazione sintomatologica, sugli esiti clinici, nonché sulla risposta ai farmaci (farmacodinamica, farmacocinetica e probabilità di eventi avversi)¹⁸.

Tali distorsioni a monte possono facilmente riflettersi nei dataset utilizzati per addestrare i sistemi di IA utilizzati in ambito sanitario, compromettendone l'affidabilità e generando una catena di errori sistemici che, nella fase di implementazione clinica, si potrebbero tradurre in decisioni mediche discriminatorie o non eque.

La sottorappresentazione femminile nella ricerca biomedica ha radici storiche profonde¹⁹. A partire dagli anni Settanta, le donne sono state sistematicamente escluse dagli studi clinici, in particolare da quelli di fase iniziale, a causa

¹⁶ Si v. M. FASAN, C.M. REALE, *Genere e sperimentazioni cliniche: il Regolamento (UE) n. 536/2014, un'occasione mancata?*, in *BioLaw Journal-Riv. BioDiritto*, 4, 2022, 251-276 e si consenta un riferimento anche a M. TOMASI, *Sperimentazioni cliniche e medicina di genere: la ricerca dell'eguaglianza attraverso la valorizzazione delle differenze*, in B. PEZZINI, A. LORENZETTI (a cura di), *70 anni dopo tra uguaglianza e differenza. Una riflessione sull'impatto di genere nella Costituzione e nel costituzionalismo*, Torino, 2019, 224 ss.

¹⁷ Alcuni dati sono riportati in N. DUMA *et al.*, *Representation of minorities and women in oncological clinical trials: review of the past 14 years*, in *Journal of Oncology Practice*, 14, 1, 2018, e1-e10; S. GELLER *et al.*, *Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: have we made progress?*, in *Journal of Women's Health*, 20, 3, 2011, 315-20; S.P. PHILLIPS, K. HAMBERG, *Doubly blind: a systematic review of gender in randomised controlled trials*, in *Global Health Action*, 9, 1, 2016 295-97.

¹⁸ A.L. SCHWARTZ *et al.*, *Why Diverse Clinical Trial Participation Matters*, in *The New England Journal of Medicine*, 388, 14, 2023, 1252-1254. Gli Autori individuano tre obiettivi fondamentali che guidano l'esigenza di diversificare le popolazioni coinvolte nelle sperimentazioni: «*earning and building trust, promoting fairness, and generating biomedical knowledge*». Sull'importanza di considerare le variabili di genere: L. SCHIEBINGER, *Scientific research must take gender into account*, in *Nature*, 7490, 2014, 9 ss. Il significativo impatto di sesso e genere sulle principali cause di mortalità e morbidità è analizzato in S. MAUVAIS-JARVIS *et al.*, *Sex and Gender: Modifiers of Health, Disease, and Medicine*, in *The Lancet*, 10250, 2020, 565-582.

¹⁹ Le vicende sono ripercorse da S. TUSINO, *Gender-specific medicine and women's involvement in research*, in *Italian Journal of Gender-Specific Medicine*, 2, 2021, 69-74.

di timori legati alle fluttuazioni ormonali e ai potenziali rischi per la gravidanza. Tale orientamento trova una delle sue principali espressioni nella tragica e nota vicenda della talidomide²⁰, un farmaco commercializzato alla fine degli anni Cinquanta come sedativo e antiemetico, che provocò gravi malformazioni congenite in oltre diecimila bambini e un aumento delle interruzioni di gravidanza. In reazione a tale evento, nel 1977, la Food and Drug Administration (FDA) emanò linee guida che sconsigliavano la partecipazione delle donne in età fertile agli studi clinici di fase iniziale²¹. Questa scelta, animata da un intento protettivo nei confronti della salute riproduttiva femminile, ha tuttavia avuto l'effetto di escludere le donne dalla produzione di conoscenza scientifica, contribuendo a consolidare una medicina fondata su un modello maschile di riferimento.

L'approccio iperprotettivo adottato in quel periodo, lungi dal garantire una maggiore sicurezza, ha generato un vuoto conoscitivo che ancora oggi condiziona la qualità dei dati biomedici: nonostante il cambio di rotta formale avvenuto all'inizio degli anni Novanta²² – con l'introduzione di politiche di inclusione delle donne negli studi clinici – la sottorappresentazione femminile nella ricerca biomedica continua a costituire un problema strutturale e persistente²³.

Anche limitando l'osservazione alla variabile biologica, studi recenti mostrano che le donne costituiscono meno del 40% dei partecipanti agli studi

²⁰ La vicenda è analizzata da W. REHMAN *et al.*, *The Rise, Fall and Subsequent Triumph of Thalidomide: Lessons Learned in Drug Development*, in *Therapeutic Advancement in Hematology*, 2, 5, 2011, 291-308. Si v. anche S. VARGESSON, *Thalidomide-Induced Teratogenesis: History and Mechanisms*, in *Birth Defects Research*, 8, 2015, 140-156.

²¹ U.S. FOOD AND DRUG ADMINISTRATION, *General Considerations for the Clinical Evaluation of Drugs*, 1977.

²² Nel 1993 l'FDA ha emanato le nuove linee-guida *Guideline for the Study and Evaluation of Gender Differences in the Clinical Evaluation of Drugs*. Il documento accompagnatorio spiegava che le linee guida del 1977 potevano aver scoraggiato la partecipazione delle donne agli studi sullo sviluppo dei farmaci e aver determinato una «*paucity of information about the effects of drugs in women*» adottando un approccio «*rigid and paternalistic, leaving virtually no room for the exercise of judgment by responsible female research subjects, physician investigators, and IRBs*». Si v. A.C. MASTROIANNI, R. FADEN, D. FEDERMAN, *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies: Volume I*, Washington, 1994 e C. LEVINE, *Changing view of justice after Belmont: Aids and the inclusion of "vulnerable" subjects*, in H.Y. VANDERPOOL (a cura di), *The ethics of research involving human subjects: facing the 21st century*, Hagerstown, 1996.

²³ Parlano di un obiettivo di parità ancora lontano dall'essere raggiunto, per esempio, A. NOWOGRODSKY, *Clinical research: Inequality in medicine*, in *Nature*, 550, 2017, S18-S19 e R.H. GOLDSTEIN, R.P. WALENSKY, *Where were the women? Gender parity in clinical trials*, in *New England Journal of Medicine*, 381, 26, 2019, 2491-2493.

clinici sulle patologie cardiovascolari e sull'ictus, pur essendo ampiamente colpite da tali condizioni; rappresentano soltanto il 42% dei soggetti coinvolti nelle sperimentazioni sui disturbi psichiatrici, sebbene costituiscano circa il 60% dei pazienti effettivi; sono solo il 41% nelle sperimentazioni oncologiche, nonostante rappresentino oltre la metà della popolazione affetta da tumori²⁴. Tale sottorappresentazione si manifesta non solo sul piano quantitativo, ma anche su quello temporale e qualitativo: le donne, infatti, vengono spesso incluse solo nelle fasi avanzate della sperimentazione clinica, quando gran parte dei dati è già stata raccolta su popolazioni maschili²⁵. Inoltre, nella maggior parte degli studi, il genere non viene considerato una variabile rilevante nell'analisi dei risultati, e anche nei casi in cui esso venga formalmente introdotto, non è sempre chiaro come venga definito – oltre la mera differenza biologica uomo-donna²⁶ – o reso operativo.

Peggiora il quadro il fatto che la scarsa considerazione della variabile sessuale o di genere non rappresenta un fenomeno isolato, ma si accompagna a un'analoga sottovalutazione di altri indici di vulnerabilità rilevanti. La ricerca biomedica contemporanea mostra, infatti, una netta predominanza di soggetti maschili e una forte concentrazione su profili cosiddetti WEIRD (Western, Educated, Industrialized, Rich, Democratic)²⁷. Una simile omogeneità dei campioni produce un impoverimento della diversità rappresentata nei dati, rendendo particolarmente difficile adottare una prospettiva intersezionale capace di cogliere le molteplici vulnerabilità che possono caratterizzare la condizione femminile in ambito sanitario²⁸.

²⁴ I dati sono riportati da A.Z. SOSINSKY *et al.*, *Enrollment of female participants in United States drug and device phase 1-3 clinical trials between 2016 and 2019*, in *Contemporary Clinical Trials*, 115, 2022, 106718.

²⁵ A. NOWOGRODSKY, *Clinical research: Inequality in medicine*, cit.

²⁶ Si rimanda qui agli insegnamenti di A. Fausto-Sterling che hanno dimostrato che le categorie binarie di sesso e genere sono una “decisione sociale” e non la rappresentazione di inevitabili realtà biologiche (A. FAUSTO-STERLING, *Sexing the body: gender politics and the construction of sexuality*, New York, 2000).

²⁷ H.J. HENRICH, S.J. HEINE, A. NORENZAYAN, *The weirdest people in the world?*, in *The Behavioral and Brain Sciences*, 33, 2010, 61-83.

²⁸ Il termine è stato coniato da Crenshaw nel 1989, con riferimento a una «*metaphor for understanding the ways that multiple forms of inequality or disadvantage sometimes compound themselves and create obstacles that often are not understood among conventional ways of thinking*» (K. CRENSHAW, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, in *University of Chicago Legal Forum*, 1989, 139-167), anche se il concetto è più risalente e rimane un quadro di riferimento imprescindibile per comprendere e affrontare le iniquità sociali.

Questa limitata diversità dei dataset potrebbe, peraltro, essere in parte condizionata dal quadro normativo vigente in materia di protezione dei dati personali. In ambito europeo, il principio di minimizzazione dei dati, sancito dall'art. 5, par. 1, lett. c) del Regolamento (UE) 2016/679 (GDPR), impone che i dati raccolti a fini di ricerca siano «adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati». Tale prescrizione, se da un lato tutela la riservatezza e la dignità della persona, dall'altro può determinare l'eliminazione di variabili sensibili – come quelle relative al sesso, al genere, all'orientamento sessuale, all'identità di genere, all'etnia o alle condizioni socio-economiche – che, pur considerate non essenziali ai fini della ricerca primaria, risultano cruciali per il riutilizzo secondario dei dati.

In un contesto scientifico sempre più orientato al *data reuse*, il rischio è che la sottorappresentazione possa risultare addirittura non intercettabile a causa dell'invisibilità di alcune variabili²⁹.

Se queste sono le caratteristiche dei dataset che alimentano i sistemi di IA, non appare improbabile che le distorsioni strutturali che caratterizzano la ricerca tradizionale siano destinate a essere riprodotte, amplificate e possibilmente “automatizzate”. Ne consegue che gli algoritmi impiegati in ambito sanitario rischiano di restituire risultati parziali, erronei o discriminatori, compromettendo l'affidabilità dei modelli predittivi e la stessa equità dell'accesso alla cura.

5. *La pratica clinica come spazio di diseguaglianza: diagnosi, cura e dimensione strutturale della salute.*

Nella fase clinica, che comprende attività diagnostiche e terapeutiche, i sistemi di IA rischiano di riflettere e amplificare le distorsioni accumulate nelle fasi precedenti della ricerca e della raccolta dati. Pur potendo offrire vantaggi significativi, dei quali si dirà brevemente, tali sistemi sono esposti a errori che possono produrre conseguenze dirette e immediate sulla salute dei pazienti. Di questi si è avuto conto, per esempio, quando alcune testate hanno riportato l'errore che avrebbe commesso Babylon Health, un fornitore di servizi sanitari digitali che ha sviluppato una piattaforma di IA (*GP at hand*) per l'erogazione

²⁹ Come è stato osservato: «The “data bodies” that digital medicine operates with often have no explicit gender attribution anymore; when digital epidemiology analyses entries in online search engines or the movement of people throughout a city via their phone's geolocation, these data are often used without any information on gender. The non- or underrepresentation of women and gender minorities is thus invisible» (M. POT, W. SPAHL, B. PRAINSACK, *The gender of biomedical data: challenges for personalised and precision medicine*, in *Somatechnics*, 9, 2-3, 2019, 170-187, 180).

di prestazioni primarie virtuali, in convenzione con il Servizio Sanitario Nazionale del Regno Unito. Secondo quanto riportato, durante alcuni test, il sistema ha diagnosticato erroneamente un attacco di ansia a una donna con sintomi compatibili con un attacco cardiaco, mentre un uomo con gli stessi sintomi e la medesima storia clinica ha ricevuto il corretto consiglio di rivolgersi immediatamente a un medico³⁰. Questo esempio, dai contorni specifici poco definiti, mette però bene in luce come i modelli algoritmici, se addestrati su dati parziali o sbilanciati, possano riprodurre discriminazioni preesistenti nella pratica clinica e non cogliere i bisogni di chi diverge dallo standard di una presunta “normalità”³¹.

Un ulteriore esempio particolarmente affascinante riguarda i cosiddetti *digital patient twins*³², strumenti innovativi della medicina personalizzata che integrano dati specifici del paziente in modelli predittivi di IA, con l'obiettivo di ottimizzare diagnosi e trattamenti. Pur rappresentando una promettente direzione della medicina, la maggior parte di questi modelli non è progettata per tenere conto delle differenze di genere, con il rischio che la personalizzazione rimanga parziale e incompleta³³. In altre parole, se da un lato l'analisi dei dati consente di caratterizzare i pazienti nelle loro dimensioni anatomo-genomiche, metabolomiche ed endocrinologiche, dall'altro può trascurare la dimensione sociale del genere, fondamentale per comprendere le diverse esposizioni al rischio, i determinanti sociali della salute e le disuguaglianze sistemiche.

Emergono qui, in tutta la loro evidenza, le potenzialità e i rischi dell'impiego dell'IA a sostegno di una delle direttrici più promettenti della medicina contemporanea: quella della personalizzazione³⁴.

³⁰ S. DAS, *It's Hysteria, Not a Heart Attack, GP Babylon App Tells Women*, in *The Sunday Times*, ottobre 2019 e N. CARDING, *Regulator Reveals “Concerns” Over Babylon's Chatbot*, in *Health Service Journal*, 4 March 2019, reperibile al link <https://www.hsj.co.uk/technology-and-innovation/regulator-reveals-concerns-over-babylons-chatbot/7029598.article>.

³¹ Aspetti, peraltro, che possono produrre significative conseguenze anche in punto di responsabilità dei medici. In tema, si v. B. BARTLETT, *The possibility of AI-induced medical manslaughter: Unexplainable decisions, epistemic vices, and a new dimension of moral luck*, in *Medical Law International*, 23, 3, 2023.

³² M. DE DOMENICO *et al.*, *Challenges and opportunities for digital twins in precision medicine from a complex systems perspective*, in *Nature Npj Digital Medicine*, 8, 37, 2025.

³³ N. WEINBERGER *et al.*, *Beyond the gender data gap: co-creating equitable digital patient twins*, in *Frontiers in digital health*, 30, 7, 2025, 1584415.

³⁴ Il Consiglio dell'UE (*Conclusioni del Consiglio su una medicina personalizzata per i pazienti*, 17 dicembre 2015, pt. 8), evidenziava la difficoltà di individuare una definizione comunemente accettata del termine “medicina personalizzata” e ne richiamava un significato più

Il potenziale dei c.d. Big Data e degli strumenti per la loro analisi è infatti straordinario, poiché essi consentono di raccogliere e correlare in modo sistematico informazioni estremamente dettagliate sulle caratteristiche individuali di ciascun paziente³⁵. In questa prospettiva, alcuni Autori hanno ipotizzato che, quando le persone saranno definite sulla base delle loro specificità anatomiche, genomiche, endocrinologiche, metabolomiche e di altre dimensioni biologicamente rilevanti, potremmo non avere più bisogno di “etichette” potenzialmente ambigue come il genere³⁶. Tale visione, tuttavia, rischia di tradursi in un abbaglio epistemologico: l’attenzione esclusiva alle dimensioni biologiche e molecolari, per quanto fine e dettagliata, potrebbe far dimenticare che una parte essenziale della soggettività risiede nelle coordinate sociali e culturali dell’identità. Come osserva Judith Butler³⁷, sesso e genere si intrecciano in un rapporto complesso e dialettico, nel quale la materialità del corpo non può essere disgiunta dai processi di significazione sociale che la plasmano.

specifico («si riferisce ad un modello medico che usa la caratterizzazione dei fenotipi e genotipi degli individui (ad esempio la definizione del profilo molecolare, l’imaging biomedico, i dati sullo stile di vita) per elaborare in modo mirato la strategia terapeutica giusta per la persona giusta al momento giusto, e/o per determinare la predisposizione alla malattia e/o praticare una prevenzione tempestiva e mirata») e uno più generico («La medicina personalizzata rimanda al concetto più ampio di assistenza orientata sul paziente, che tiene conto del fatto che, in generale, i sistemi di assistenza sanitaria devono essere più rispondenti ai bisogni del paziente»). Per un approfondimento sul tema della medicina personalizzata si veda K.R. BROTHERS, M.A. ROTHSTEIN, *Ethical, legal and social implications of incorporating personalized medicine into healthcare*, in *Personalized Medicine*, 1, 2015, 43-51 e si consenta un riferimento a M. TOMASI, *Genetica e Costituzione. Esercizi di eguaglianza, solidarietà e responsabilità*, Napoli, 2019, 200 ss.

³⁵ Le più promettenti applicazioni dell’IA nel perseguimento degli obiettivi della medicina personalizzata sono riportati nel *Libro bianco 2024* della Fondazione Onda *Dalla medicina di genere alla medicina personalizzata. Il ruolo dell’intelligenza artificiale*, online: <https://fondazioneonda.it/ondauploads/2024/11/Onda.pdf>.

³⁶ La citazione completa è la seguente: «*When people are characterised individually with their anatomical, genomic, endocrinological, metabolomic, and other relevant dimensions, we may no longer need generic labels such as gender. The newly founded groups could correspond with clusters along existing gender categories in some respects – such as around pregnancy and childbirth – but not in others, such as in endocrinological terms where ‘similar’ patients would be grouped together in ways that cut across traditional categories of sex and gender. Personalised and precision medicine could thus also be viewed as a way to overcome existing tensions between the desire for the abolishment of gender bias, on the one hand, and the important critique of the invisibility of women and gender minorities in biomedicine, on the other*» (M. POT, W. SPAHL, B. PRAINSACK, *The gender of biomedical data: challenges for personalised and precision medicine*, cit., 178).

³⁷ J. BUTLER, *The psychic life of power. Theories in subjection*, California, 1997.

In questo senso, è necessario evitare che la potenza analitica e computazionale dei sistemi di IA – capaci di “squadernare” l’identità biologica – finisca per oscurare o rendere opaca la dimensione macroscopica delle disuguaglianze sociali che continuano a incidere sulla salute e sull’accesso alle cure³⁸.

Questi profili problematici emergono con ancora maggiore evidenza allorché si considerino le applicazioni dei sistemi di IA in ambiti più ampi e sistemici rispetto alla cura individuale, quali la gestione della salute pubblica, l’epidemiologia e la pianificazione sanitaria. Lì, i modelli predittivi sono utilizzati sempre più di frequente, per esempio per prevedere l’insorgenza di malattie, ottimizzare l’allocazione delle risorse e orientare le strategie di intervento. Tuttavia, questi strumenti spesso trascurano le dimensioni di *genere* della salute, non tenendo conto di differenze rilevanti. Emblematico, in tal senso, è il caso della pandemia di Covid-19, durante la quale i primi modelli predittivi elaborati per la gestione della crisi non hanno integrato alcuni fattori (quali il carico sproporzionato di lavoro di cura non retribuito sostenuto dalle donne, i tassi più elevati di perdita dell’occupazione in alcuni settori e i maggiori rischi di esposizione nei ruoli sanitari di prima linea), producendo analisi e previsioni generalizzate che hanno finito per oscurare le specifiche vulnerabilità femminili. Analogamente, i modelli impiegati per il monitoraggio della salute materna, dei servizi di salute riproduttiva o delle campagne vaccinali tendono a trascurare fattori critici quali i determinanti sociali della salute, le barriere culturali e le disparità di accesso all’assistenza sanitaria. Tale omissione riduce l’efficacia complessiva delle strategie di salute pubblica e, soprattutto, rischia di compromettere il perseguimento dell’obiettivo di eguaglianza sostanziale nella tutela del diritto alla salute³⁹.

E tuttavia, proprio le caratteristiche di funzionamento delle nuove tecnologie di IA, se correttamente sfruttate, potrebbero portare dei complessivi benefici. In primo luogo, gli errori e le disfunzioni che tali sistemi possono (ri)produrre assumono una funzione di particolare valore: se adeguatamente intercettati e analizzati, magari attraverso i filoni di studio che hanno iniziato a occuparsi della cosiddetta *explainable AI* (xAI), che mira a rendere i processi decisionali degli algoritmi più trasparenti e comprensibili⁴⁰, essi possono contribuire a

³⁸ A. CARNEVALE *et. al.*, *Will Big Data and personalized medicine do the gender dimension justice?*, in *AI & Society*, 38, 2, 2021, 829-841.

³⁹ I due esempi sono descritti in H. CHUNG *et al.*, *Gender Bias in Artificial Intelligence: Severity Prediction at an Early Stage of COVID-19*, in *Frontiers in Physiology*, 29, 2021, 12, 778720.

⁴⁰ Sul concetto di XAI si v. A. SAJID, *Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence*, in *Information Fusion*, 99, 2023, 101805,

disvelare i meccanismi discriminatori occulti e le asimmetrie strutturali che continuano a permeare tanto la pratica clinica quanto l'organizzazione sanitaria.

In secondo luogo, la capacità dei sistemi di IA di integrare e correlare dati eterogenei – biologici, ambientali, sociali e comportamentali – li rende, almeno in linea teorica, strumenti privilegiati per promuovere un approccio realmente intersezionale e integrato alla salute, in grado di cogliere la complessità dei fattori che incidono sul benessere individuale e collettivo e di trasformare i paradigmi tradizionali della cura. Si tratta, dunque, di approcci con un grande potenziale, che potrebbero concorrere in modo determinante alla realizzazione di una medicina autenticamente personalizzata, capace di modulare diagnosi e trattamenti sulle specificità di ciascun individuo, tenendo conto delle variabili biologiche e molecolari e, al tempo stesso, abbracciando le coordinate sociali, culturali e sistemiche che plasmano l'esperienza concreta e individuale della salute e della malattia.

Perché ciò sia possibile, però, è necessario integrare la prospettiva di genere in tutto il ciclo di vita dell'IA, tanto nelle fasi di progettazione e addestramento, quanto in quella conseguente di applicazione dei modelli. Contrariamente, ogni promessa rischia di rovesciarsi in una eterogenesi dei fini: ciò che reca con sé il potenziale di uno strumento di equità e individualizzazione può, in concreto, tradursi in un meccanismo di riproduzione – per quanto tecnologicamente sofisticata – di disuguaglianze strutturali già radicate e forse anche di produrne di nuove.

6. *La ricerca dell'eguaglianza di genere nello spazio della regolazione.*

Quanto detto sino a qui evidenzia la sfida, epistemica, etica e istituzionale, volta a ripensare l'architettura della ricerca, della cura e della governance sanitaria alla luce di un modello di regolamentazione della tecnologia capace di orientare l'innovazione verso obiettivi di equità, inclusione e giustizia sostanziale. Resta ora da chiedersi quale sia il contributo che il diritto e le diverse opzioni normative possono fornire e quali siano le soluzioni sino a qui percorse, al fine di saggiarne l'efficacia.

online: <https://doi.org/10.1016/j.inffus.2023.101805> e A. ADADI, M. BERRADA, *Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)*, in *IEEE Access*, 6, 2018, 52138-52160. Per le applicazioni in campo sanitario si può fare riferimento a Z. SADEGHI, *A review of Explainable Artificial Intelligence in healthcare*, in *Computers and Electrical Engineering*, 118(A), 2024, 109370. Le criticità connesse a questo concetto sono esplorate da R.O. WEBER *et al.*, *XAI is in trouble*, in *AI Magazine*, 45, 3, 2024, 300-316.

Gli approcci tradizionali del diritto non discriminatorio costituiscono un punto di partenza evidentemente imprescindibile, ma faticano, al tempo stesso, a intercettare le forme più sottili e opache di discriminazione, spesso incorporate negli algoritmi e nei dataset su cui essi si basano. In tale contesto, caratterizzato dalla difficoltà di leggere gli indici di discriminazione, è stata da più parti segnalata l'esigenza di volgere verso approcci di *governance* maggiormente articolate consentano di intervenire sin dalle fasi di progettazione degli algoritmi e nella successiva interpretazione dei risultati prodotti, al fine di garantire una rappresentanza adeguata delle diverse identità e delle specificità sociali⁴¹. Proprio a tal fine, come si diceva, diventa essenziale adottare misure tecniche che assicurino la rappresentatività dei dataset, evitando sia la sovra- che la sottorappresentazione di categorie socialmente vulnerabili, comprese quelle storicamente marginalizzate. Ciò include, ad esempio, la stratificazione dei dati per genere, etnia, età, condizione socioeconomica e altre variabili rilevanti, così da consentire un'analisi più completa e intersezionale. Parallelamente, devono essere definiti stringenti requisiti di sicurezza, qualità e trasparenza dei dati e dei modelli, al fine di rendere verificabile il processo decisionale algoritmico e di permettere interventi correttivi nel caso in cui si rilevino distorsioni.

Di questi aspetti si trova una traccia nel c.d. AI Act europeo, il Regolamento UE 2024/1689, che, come noto, rappresenta un passaggio fondamentale e pionieristico nel delineare un modello di governance dell'IA basato su principi di proporzionalità del rischio, trasparenza e *accountability*. Pur con i limiti significativi di cui si dirà e in termini invero piuttosto generici, esso introduce obblighi per i sistemi classificati ad "alto rischio", tra i quali rientrano quelli destinati all'ambito sanitario, imponendo requisiti di qualità e rappresentatività dei dati, documentazione tecnica, tracciabilità e supervisione umana (artt. 8-15).

Più dettagliato il riferimento che si può rinvenire nell'art. 10 della *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* del Consiglio d'Europa (2024), dedicato all'uguaglianza e alla non discriminazione. La norma richiama esplicitamente anche l'uguaglianza di genere, un obiettivo il cui perseguimento impone agli Stati aderenti di adottare o mantenere misure volte a garantire l'uguaglianza, formale e sostanziale, in tutto il ciclo di vita dell'IA (art. 10, parr. 1 e 2). Significativamente, l'*Explanatory report*, nella parte di commento all'art. 10, sottolinea che l'avvento e la diffusione

⁴¹ Per uno sguardo articolato su questi temi, si v. il lavoro monografico di C. NARDOCCI, *Algoritmi, eguaglianza, discriminazione*, Torino, 2025. Si v. anche S. WACHTER, B. MITTELSTADT, C. RUSSELL, *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI*, cit.

dell'IA ha portato alla ribalta due ordini di problemi relativamente nuovi, quello dei *bias* tecnici e quello dei *bias* sociali. Questi ultimi, in particolare, sarebbero «*failures to properly account for historical or current inequalities in society in the activities within the lifecycle of artificial intelligence systems such as designing and training models*». Diseguaglianze che includono, per esempio, «*historical and structural barriers to gender equality and to fair and just treatment for persons belonging to groups that have been or are still partly underserved, discriminated against, or otherwise subject to persistent inequality*», che impongono di riconoscere che «*various individuals experience different impacts based on factors which are linked to their personal characteristics, circumstances or membership of a group*». A fronte di tali rinnovate sfide, l'approccio necessario non implica solo evitare che una persona sia trattata in modo sfavorevole «*without objective and reasonable justification*» sulla base di una o più caratteristiche protette, ma anche a «*overcoming structural and historical inequalities*», inevitabilmente, in un contesto normativo internazionale, «*to the extent permitted by [...] domestic and international human rights obligations*». Infine, i processi dovrebbero essere «*where appropriate, informed by the views of those impacted*» (parr. 76 e 77 Explanatory Report).

Altri ordinamenti si sono rivolti a strumenti più specifici e maggiormente innovativi, finalizzati a prevenire e mitigare i pregiudizi algoritmici di genere. Tra questi, il *Gender Impact Assessment Plus* (GBA Plus)⁴², adottato in Canada e richiamato anche nella *Canadian Directive on Automated Decision-Making*⁴³, rappresenta uno degli strumenti più avanzati di valutazione preventiva (*ex ante regulatory tool*) degli effetti discriminatori potenziali derivanti dall'uso di

⁴² Definito come «*an analytical tool used to assess how diverse groups of women, men and gender-diverse people may experience policies, programs and initiatives*» (Status of Women Canada (SWC), *What Is GBA+?*, 2017, online: <https://www.swc-cfc.gc.ca/gba-acis/index-en.html>).

⁴³ La *Directive on Automated Decision-Making* (2019) prevede che ogni sistema decisionale automatizzato impiegato per l'adozione di decisioni amministrative sia soggetto a una valutazione preventiva, imponendo ai responsabili dei programmi l'obbligo di effettuare una *Gender-based Analysis Plus* (GBA+) in fase di progettazione o revisione del sistema (Section 6.3.8). L'analisi: i) deve comprendere una valutazione degli impatti che il progetto di automazione può produrre sui diversi gruppi della popolazione, considerando sia gli effetti del sistema decisionale automatizzato e dei dati impiegati, sia le possibili conseguenze della decisione finale; ii) deve citare, laddove possibile, i dati utilizzati per l'analisi degli impatti, raccomandandosi che tali dati siano disaggregati per genere e includano altri fattori identitari intersezionali quali età, disabilità e origine etnica. Si prevede inoltre che, nel caso in cui i dati non siano disponibili, occorra individuare e segnalare le lacune informative rilevate. È infine necessario specificare le misure pianificate o già adottate per affrontare i rischi identificati tramite la *Gender-based Analysis Plus* o altre forme di valutazione (Appendix C – Impact Level Requirements).

tecnologie automatizzate. Esso mira ad analizzare in modo sistematico l'impatto delle politiche pubbliche e delle soluzioni tecnologiche sull'uguaglianza di genere, integrando una prospettiva intersezionale che considera anche variabili come età, etnia, disabilità e status socioeconomico. Di particolare interesse, per il contesto sanitario, è l'applicazione del GBA Plus alle sperimentazioni cliniche e alla ricerca biomedica, dove questo approccio consente di rilevare *ex ante* possibili disparità di rappresentanza, accesso e risposta ai trattamenti, favorendo così una progettazione più equa dei protocolli di studio. A partire dal 2011, il Canadian Institutes of Health Research (CIHR) ha previsto, nei propri bandi di finanziamento, l'obbligo per i ricercatori di esplicitare se le rispettive proposte integrino considerazioni relative al sesso e al genere, accompagnando tale dichiarazione con un'adeguata giustificazione⁴⁴. Le prime valutazioni sull'impatto di questa misura evidenziano un progressivo incremento dell'inclusione delle variabili di sesso e genere nella progettazione della ricerca biomedica; tuttavia, alcune analisi segnalano che tale evoluzione non ha ancora prodotto un cambiamento sostanziale né nelle modalità di conduzione degli studi, né nel modo in cui i risultati vengono riportati⁴⁵.

Alcuni Autori, inoltre, hanno messo in luce i limiti del modello GBA Plus, osservando come il suo approccio "gender-first" rischi di semplificare eccessivamente la complessità delle intersezioni tra genere e altri fattori di discriminazione. Pur dichiarandosi ispirato al paradigma intersezionale, che mira a comprendere le disuguaglianze come prodotto di interazioni multiple e simultanee tra categorie sociali, il GBA Plus fatica infatti a realizzarlo pienamente, poiché privilegia la dimensione di genere rispetto ad altre variabili rilevanti⁴⁶.

A questo tipo di preoccupazione sembra invece rispondere il disegno di legge *Dispõe sobre o uso da Inteligência Artificial* (PL 2338/2023), al quale si sta lavorando nell'ordinamento brasiliano⁴⁷. In questo testo si richiede agli operatori o fornitori di sistemi di IA classificati come ad alto rischio di adottare alcune

⁴⁴ J. JOHNSON *et al.*, *Does a change in health research funding policy related to the integration of sex and gender have an impact?*, in *PLoS One*, 9, 6, 2014 e Z. SHARMAN, J. JOHNSON, *Towards the inclusion of gender and sex in health research and funding: an institutional perspective*, in *Social Science & Medicine*, 74, 11, 2012, 1812-1816.

⁴⁵ V. WELCH *et al.*, *Reporting of sex and gender in randomized controlled trials in Canada: a cross-sectional methods study*, in *Research Integrity and Peer Review*, 2, 2017.

⁴⁶ La critica è avanzata da O. HANKIVSKY, L. MUSSEL, *Gender-Based Analysis Plus in Canada*, in *Canadian Public Policy/Analyse de Politiques*, 44, 4, 2018, 303-316, 309.

⁴⁷ Per un commento, si v. FRAZÃO, *Regulation of artificial intelligence in Brazil: examination of Draft Bill no. 2338/2023*, *UNIO-EU Law Journal*, 10, 1, 2024, 54-69.

misure di governance dei dati volte a prevenire effetti discriminatori (art. 20-IV). La parte più significativa della disposizione, tuttavia, è che, a tal fine, i dati devono essere valutati attraverso misure appropriate di controllo dei *bias* cognitivi umani (*vieses cognitivos humanos*) che possono influenzarne la raccolta e l'organizzazione. Di conseguenza, dovranno adottarsi misure correttive per prevenire l'incorporazione di *bias* sociali strutturali che possano essere perpetuati o amplificati dalla tecnologia. Un'altra previsione significativa, nel medesimo senso, impone la costituzione di un'équipe inclusiva per la progettazione e lo sviluppo del sistema, orientata alla ricerca della diversità ⁴⁸.

7. Soluzioni normative a confronto: equità sostanziale e gender-sensitive governance.

Se messe a confronto, le previsioni analizzate mostrano l'emergere di alcune linee che segnalano un'evoluzione nel modo in cui il diritto affronta i rischi di discriminazione algoritmica che possono rilevare nei contesti sanitari e di ricerca biomedica. In primo luogo, le esperienze canadese e brasiliana sottolineano l'opportunità, e forse la necessità, di considerare il genere insieme ad altri fattori identitari – quali età, disabilità, etnia, status socio-economico – pur divergendo nell'opzione, che è evidentemente politica e non priva di significative conseguenze, di attribuire al genere un ruolo prioritario o paritario rispetto ad altre dimensioni di vulnerabilità. In riferimento a questi aspetti, come si diceva, la capacità dei sistemi di IA di associare più dati potrebbe avere un grande potenziale, anche se i criteri di associazione non sono sempre trasparenti.

In secondo luogo, si osserva il superamento di un modello meramente "formale" di non discriminazione, basato sull'uguaglianza di trattamento, a favore di una *governance* sostanziale orientata all'equità, che agisce in modo proattivo nella fase di progettazione e implementazione delle tecnologie. Si tratta di un cambio di paradigma in linea con gli orientamenti più recenti del diritto antidiscriminatorio e con il dibattito sulla regolazione della tecnologia, che mira a integrare principi inclusivi già a livello di architettura dei sistemi ⁴⁹. In tal senso,

⁴⁸ Gli effetti di una mancata implementazione di un approccio intersezionale femminista nello sviluppo delle tecnologie digitali della salute è evidenziato da C.A. FIGUEROA *et al.*, *The need for feminist intersectionality in digital health*, in *The Lancet Digital Health*, 3, 8, 2021, e526-e533.

⁴⁹ Sono due, infatti, i livelli ai quali è possibile agire: «*We model a supervised machine learning algorithm as consisting of two components: a 'predictive algorithm,' which takes in training data consisting of outcomes and observed characteristics for a set of individuals and returns a prediction*

emerge l'importanza di interventi *ex ante*, che lavorino sul codice dell'algoritmo in chiave inclusiva⁵⁰. Giova anche segnalare che le stesse tecnologie di IA potrebbero contribuire in maniera significativa a migliorare la rappresentatività dei dati. In particolare, gli approcci algoritmici possono individuare e segnalare le popolazioni sottorappresentate e analizzare *bias* strutturali, quali divari di accesso digitale e disallineamenti linguistici. Inoltre, gli strumenti di *natural language processing* permettono di considerare i determinanti sociali della salute o fattori culturali che influenzano la partecipazione alle sperimentazioni, fornendo indicazioni utili per lo sviluppo di strategie di reclutamento più inclusive⁵¹.

Ciò che appare ancora carente, invece, è la capacità di alcuni strumenti normativi (soprattutto quelli improntati su un approccio di classificazione del rischio) di farsi realmente carico delle più profonde architetture discriminatorie, di operare non solo sui dati, ma anche sui processi cognitivi, percettivi e culturali che ne influenzano la raccolta, la classificazione e l'interpretazione, con la conseguenza che risultano non del tutto adeguati a fronteggiare direttamente i processi di esclusione che generano la discriminazione algoritmica in vari ambiti, incluso quello sanitario.

L'AI Act dell'Unione Europea rappresenta un esempio emblematico di questa impostazione⁵²: la classificazione dei sistemi di IA in base al livello di rischio (art. 6) tende a privilegiare, come da più parti segnalato, le preoccupazioni di natura economica o di sicurezza, più che quelle connesse alla tutela di altri diritti fondamentali. Così, ad esempio, la fase della ricerca è tendenzialmente esclusa dall'ambito di applicazione del regolamento⁵³ e le applicazioni di IA

function, and a 'decision rule,' which uses the constructed prediction function to make decisions. A policymaker therefore has two distinct tools available: influencing the design of the predictive algorithm and influencing how the decision rule uses the predictions», A. RAMBACHAN et al., An Economic Approach to Regulating Algorithms, 2021, National Bureau of Economic Research Working Paper No. 27111, online: <https://www.nber.org/papers/w27111>.

⁵⁰ Parlano della distinzione: O. BAR-GILL, C.R. SUNSTEIN, *Ex Post Policing and Ex Ante Regulation*, Oxford, 2025, spec. 108-113.

⁵¹ D.B. OLAWADE et al., *Artificial intelligence in clinical trials: A comprehensive review of opportunities, challenges, and future directions*, in *International Journal of Medical Informatics*, 206, 2026, 106141, online: <https://www.sciencedirect.com/science/article/pii/S1386505625003582>.

⁵² Le critiche sono espresse in maniera estesa in P.L. LAU, *Rewriting the Narrative of AI Bias: A Data Feminist Critique of Algorithmic Inequalities in Healthcare*, in *Law, Technology and Humans*, 7, 2, 2025.

⁵³ Il considerando 25 e l'art. 2(8) prevedono entrambi che l'esclusione della fase di ricerca non copra i test svolti in «*real world conditions*».

destinate alla salute femminile – note come FemTech⁵⁴ – non rientrano tra i sistemi ad “alto rischio”, nonostante la loro potenziale capacità di riprodurre o amplificare disuguaglianze di genere nella salute.

L’art. 10, inoltre, prevede meccanismi di controllo di eventuali *bias* che si basano però prevalentemente su metriche quantitative di equità, attraverso un riferimento al diritto antidiscriminatorio “standard”, che tendono a trattare il *bias* come un semplice errore statistico, piuttosto che come l’effetto di pratiche strutturali e sistemiche di esclusione. Inoltre, la scarsa attenzione alla trasparenza pubblica dei processi di controllo, che sono tendenzialmente interni, riduce la possibilità di un controllo democratico effettivo. Analogamente, gli obblighi di trasparenza previsti dall’art. 13 – pur significativi – non affrontano in modo sostanziale il fenomeno della *proxy discrimination*, ossia la riproduzione di esclusioni storiche sotto la veste della neutralità algoritmica, e non impongono interventi strutturali volti a correggere i *bias* sottostanti.

Il rischio più serio, connesso a quello che viene definito un “technocratic storytelling”, è quello di rappresentare i *bias* come un difetto correggibile, come una disfunzione tecnica da ottimizzare, piuttosto che come l’esito strutturale di pratiche di raccolta, classificazione e rappresentazione dei dati profondamente segnate da rapporti di potere e da disuguaglianze sistemiche⁵⁵.

Il regolamento europeo fa, probabilmente, ciò che una regolamentazione di prodotto deve e può fare: non contempla strategie per il superamento delle disparità strutturali, ma rimane ancorato a una logica di conformità formale più che di equità sostanziale, rinviando ai futuri atti di implementazione – standard tecnici e codici di condotta, in particolare – la possibile definizione di criteri concreti per l’inclusione di variabili di genere e intersezionali nei processi di progettazione e validazione algoritmica.

Nel contesto permeato dalle applicazioni dell’IA nel quale siamo immersi, però, restituire voce ai soggetti tradizionalmente marginalizzati nei processi di ricerca e nelle politiche sanitarie richiede interventi più strutturati, che consentano di intercettare la radice delle distorsioni derivanti da modelli di ricerca e sperimentazione storicamente centrati sul corpo maschile, promuovendo una progettazione realmente equa, inclusiva e attenta alle differenze di sesso, genere

⁵⁴ Sui rapporti fra Femtech e IA, si v. S. CORSO, *Sulla female technology o Femtech. Intelligenza artificiale, salute della donna e volontà della persona*, in *Biolaw Journal-Riv. BioDiritto*, 3, 2023, 81-112.

⁵⁵ Così, P.L. LAU, *Rewriting the Narrative of AI Bias: A Data Feminist Critique of Algorithmic Inequalities in Healthcare*, cit., 13.

e altri fattori intersezionali ⁵⁶. A tal fine è necessario elaborare strumenti normativi e tecnici “virtuosi” che, combinando rigore scientifico e consapevolezza sociale, contribuiscano a costruire sistemi di IA più trasparenti, inclusivi e responsabili. In Europa, come in altre parti del mondo, la sfida, in prospettiva, sarà quella di integrare in modo organico la dimensione della *gender-sensitive governance* in un ecosistema giuridico complesso ⁵⁷, così da renderlo capace di coniugare innovazione tecnologica, tutela dei diritti fondamentali e giustizia sociale.

⁵⁶ Il tema è sviluppato da C. CRIADO PEREZ, C. PALMIERI, *Invisibili. Come il nostro mondo ignora le donne in ogni campo: dati alla mano*, Torino 2022.

⁵⁷ Del quale fanno parte, fra gli altri atti, il regolamento generale sulla protezione dei dati (regolamento (UE) 2016/679), al regolamento sullo spazio europeo dei dati sanitari (regolamento (UE) 2025/327), al regolamento sui servizi digitali (regolamento (UE) 2022/2065) e al regolamento sui trials clinici (regolamento (UE) n. 536/2014). Per una panoramica su quasi tutti questi atti, si v. F. PIZZETTI (a cura di), *La regolazione della società digitale*, Torino, 2024.

CAPITOLO IX

I SISTEMI DI IA PER IL CONTROLLO
DELL'IMMIGRAZIONE: BENEFICI E RISCHI
PER I DIRITTI UMANI E LA PARITÀ DI GENERE

Cecilia Siccardi

SOMMARIO: 1. Premessa. – 2. *Smart borders* e *governance* digitale dell'immigrazione. – 3. L'impatto dei sistemi di IA sulla condizione delle donne migranti. – 4. L'AI Act e la disciplina dei sistemi per il controllo dell'immigrazione. – 5. Le garanzie previste per i sistemi "ad alto rischio" ed alcune eccezioni. – 6. Alcuni ostacoli all'applicabilità delle garanzie previste dall'AI Act. – 7. I sistemi vietati che vengono in rilievo nel contesto dell'immigrazione. – 8. Riflessioni conclusive.

1. *Premessa.*

Questo scritto intende indagare quale sia l'impatto dei sistemi di intelligenza artificiale nell'ambito del controllo delle frontiere e dell'immigrazione sui diritti umani delle persone migranti, tenendo in considerazione anche la specifica condizione delle donne migranti.

Da un lato, certamente i sistemi di intelligenza artificiale comportano enormi benefici, rendendo maggiormente efficienti diverse operazioni di monitoraggio dei flussi, come il controllo e il rilascio di visti e permessi di soggiorno, le procedure di riconoscimento della protezione internazionale, le verifiche sull'identità della persona.

Dall'altro lato, come evidenziato dal Comitato sulle migrazioni del Consiglio di Europa, «il rapido impiego dell'intelligenza artificiale nella gestione delle migrazioni presenta rischi significativi per i diritti umani e le libertà fondamentali»¹, come il diritto d'asilo (art. 10, comma 3, Cost., art. 18 CDFUE), il divieto di

¹ COUNCIL OF EUROPE, COMMITTEE ON MIGRATION, REFUGEES AND DISPLACED PERSONS, *Artificial Intelligence and Migration*, Report1 Rapporteur: Mr Petri Honkonen (Finland, ALDE), Parliamentary Assembly, Resolution, 25 June 2025, par. 12.

respingimento (art. 3 e art. 4, prot. 4, CEDU), la libertà personale (art. 13 Cost., art. 5 CEDU) e di circolazione (art. 16 Cost., art. 2 prot. 4, CEDU), il diritto di accesso alla giustizia (art. 24 Cost., art. 47 CDFUE, art. 13 CEDU). Inoltre, si deve tener conto che sistemi automatizzati di controllo delle frontiere e dell'immigrazione rischiano di non prendere in considerazione "caso per caso" le specificità individuali, non valutando ad esempio le peculiarità delle migrazioni femminili o di altri soggetti vulnerabili che si presentano alle frontiere.

Partendo da tale consapevolezza, lo scritto si propone, in primo luogo, di ricostruire il contesto attuale degli "smart borders" dell'Unione Europea (par. 2), che conosce da tempo l'utilizzo delle tecnologie per il controllo delle frontiere e nell'ambito del quale sono già state avviate le prime sperimentazioni di sistemi di intelligenza artificiale. In secondo luogo, si vuole comprendere l'impatto che tali sistemi hanno sulla specifica condizione delle donne migranti e sui loro diritti (par. 3).

In terzo luogo, lo scritto si concentrerà sul recente regolamento (UE) 2024/1689 sull'intelligenza artificiale (di seguito AI Act, par. 4 ss.), che disciplina l'impiego di sistemi di IA nei contesti dell'immigrazione, dell'asilo e del controllo delle frontiere, qualificandoli come sistemi ad "alto rischio" (all. III, par. 7).

2. Smart borders e governance digitale dell'immigrazione.

Tecnologia e intelligenza artificiale sono da tempo ampiamente utilizzate nel controllo delle frontiere europee, tanto che la complessa disciplina in materia è già stata efficacemente definita quale "governance digitale dell'immigrazione"² o degli "smart borders"³.

² Si vedano in particolare gli studi raccolti in F. BIONDI DAL MONTE, M. FORTI, L. RANIERI, *Migrazioni e governance digitale. Persone e dati alle frontiere dell'Europa*, Roma, 2024. Inoltre, si è occupato del tema, già prima dell'entrata in vigore dell'AI Act, S. PENASA, *Verso una governance smart delle migrazioni? L'utilizzo dell'Intelligenza artificiale nelle politiche sull'immigrazione in una prospettiva comparata*, in V. FAGGIANI, F.J. GARRIDO CARRILLO, *Desafíos y límites de la política migratoria en Europa y América*, Cizur Menor, 2022, 265 ss.; ID., *Smart borders o invisible walls? L'utilizzo di sistemi "intelligenti" al confine tra politiche migratorie e garanzie per gli stranieri*, in ADiMblog, dicembre 2022.

³ "Smart Borders" è un termine utilizzato dalla stessa Commissione europea per descrivere i sistemi del c.d. "Smart borders Package" composto da tecnologie volte alla raccolta, al raffronto e alla verifica dei dati delle persone straniere che attraversano i confini europei, come il sistema EES "Entry/Exit System" e il sistema ETIAS "European Travel Information and Authorisation System". Si veda in particolare il sito di Eu-Lisa "Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice".

Specialmente a partire dalla crisi migratoria del 2015⁴, l'UE si è progressivamente dotata di un apparato di norme volte ad introdurre i c.d. sistemi It di larga scala, digitalizzando e rendendo più efficienti diverse operazioni di controllo dei confini, come quelle volte all'identificazione, alle verifiche di sicurezza, alla presentazione delle domande di asilo⁵.

Tra tali sistemi è possibile menzionare: il Sistema d'Informazione Schengen (SIS)⁶, la banca di condivisione di informazioni più vasta e utilizzata per la gestione della sicurezza delle frontiere europee che consente alle autorità nazionali di effettuare e consultare segnalazioni su persone e oggetti; il Sistema di informazione sui visti (VIS)⁷, volto allo scambio di dati relativi ai visti d'ingresso nello Spazio Schengen; EURODAC⁸, la banca dati delle impronte digitali e dati biometrici volta a concorrere all'individuazione dello Stato membro competente nella gestione della domanda di asilo, al contrasto all'immigrazione irregolare e ai movimenti secondari.

A questi si aggiungono i sistemi del c.d. *Smart Borders Package*⁹: l'*Entry/Exit*¹⁰ ha sostituito il "vecchio" timbro sul passaporto con un archivio digitale che raccoglie i dati anagrafici e biometrici dei viaggiatori extra-UE, con lo scopo di facilitare controlli automatici e sistematici, l'individuazione di coloro che non rispettano la durata di soggiorno consentita, il contrasto a ingressi irregolari e frodi identitarie; il "nuovo" sistema ETIAS¹¹ che prevede un'autorizzazione

⁴ L'Agenda europea del 2015 prevedeva, tra i 4 pilastri per la gestione UE dell'immigrazione, anche quella della costruzione di "frontiere intelligenti" per "rendere più efficaci i valichi di frontiera" cfr. Comunicazione della commissione al parlamento europeo, al consiglio, al comitato economico e sociale europeo e al comitato delle regioni agenda europea sulla migrazione, Bruxelles, 13 maggio 2015 COM(2015) 240 final.

⁵ Tali sistemi sono menzionati all'allegato X dell'AI Act, regolamento (UE) 2024/1689.

⁶ Regolamento (UE) 2018/1860 relativo all'uso del sistema d'informazione Schengen per il rimpatrio di cittadini di paesi terzi; regolamento (UE) 2018/1862 sull'istituzione, l'esercizio e l'uso del sistema d'informazione Schengen (SIS).

⁷ Regolamento (UE) 2021/1133 per quanto riguarda la definizione delle condizioni di accesso agli altri sistemi di informazione dell'UE ai fini del sistema di informazione visti; regolamento (UE) 2021/1134 ai fini della riforma del sistema di informazione visti.

⁸ Regolamento (UE) 2024/1358 che istituisce l'"Eurodac" per il confronto dei dati biometrici.

⁹ La Commissione ha per la prima volta presentato la "*Smart borders proposal*" nel febbraio del 2013 cfr. *EU Commission, 'Smart borders': enhancing mobility and security*, Brussels, 28 February 2013.

¹⁰ Cfr. regolamento (UE) 2017/2226 che istituisce un sistema di ingressi/uscite.

¹¹ Regolamento (UE) 2018/1240 che istituisce un sistema europeo di informazione e autorizzazione ai viaggi (ETIAS).

per i viaggiatori esenti da visto riconosciuta a seguito di una valutazione algoritmica dei dati forniti dal richiedente e dei conseguenti indicatori di rischio per la sicurezza, salute pubblica o migrazione irregolare.

Questi sistemi sono stati ulteriormente rafforzati dal Patto UE sulla migrazione e sull'asilo del 2024¹², il quale «amplifica la sorveglianza ai confini delimitando un sistema di banche dati tra loro connesse e alimentate da un set sempre più ampio di dati biometrici»¹³. Nello specifico, il nuovo Patto, al fine di prevenire minacce per la sicurezza europea, ha inteso potenziare l'interoperabilità tra banche dati, imponendo verifiche congiunte tra i sistemi EURODAC¹⁴ e SIS¹⁵ anche con il sistema europeo di informazione sui casellari giudiziali riguardo ai cittadini di paesi terzi (ECRIS-TCN)¹⁶, nonché con quelli dell'archivio comune di dati di identità ("CIR")¹⁷.

In generale queste banche dati "digitali", progettate per gestire enormi quantità di dati biometrici e alfanumerici relativi a cittadini di Paesi terzi⁴, costituiscono un campo privilegiato per "testare" sistemi di intelligenza artificiale che sappiano rendere più fluidi ed efficaci i meccanismi di raffronto tra i dati e l'interoperabilità fra i predetti sistemi e le banche dati nazionali. La Corte di Giustizia dell'Unione Europea si è già espressa sulle caratteristiche che i sistemi "automatizzati" di gestione dei dati debbono presentare per non rischiare di ledere i diritti delle persone: essi debbono garantire un controllo giurisdizionale, la conoscibilità dei criteri automatizzati da parte del giudice, una "seconda fase" manuale per verificare ed eventualmente ridurre le distorsioni algoritmiche, l'informativa dell'interessato che deve essere messo in condizione di comprendere i criteri del trattamento automatizzato dei dati¹⁸.

¹² Per una ricognizione delle fonti del Patto UE, entrato in vigore il 14 maggio 2024 e che sarà applicabile a partire dal 2026 cfr. A. DI PASCALE, *Pubblicati gli atti che compongono il nuovo Patto sulla migrazione e l'asilo*, in *Eurojus*, 2024.

¹³ F. BIONDI DAL MONTE, *Confini digitali e dati dei migranti nel Patto sulla migrazione e l'asilo*, in F. BIONDI DAL MONTE, M. FORTI, L. RANIERI, *Migrazioni e governance digitale*, cit., 197.

¹⁴ Cfr. ancora regolamento UE 2024/1358 "Eurodac".

¹⁵ Cfr. in particolare regolamento (UE) 2024/1352 allo scopo di introdurre accertamenti nei confronti dei cittadini di paesi terzi alle frontiere esterne.

¹⁶ Istituito dal regolamento (UE) 2019/816.

¹⁷ Istituito dai regolamenti (UE) 2019/817 e (UE) 2019/818.

¹⁸ Cfr. in questo senso Corte Giust. UE, *Ligue des droits humains*, CG C-817/19, 21 giugno 2022. Commenta la decisione e ricostruisce le garanzie elencate S. PENASA, *Intelligenza artificiale e diritti: verso un diritto algoritmico dell'immigrazione?*, in F. BIONDI DAL MONTE, M. FORTI, L. RANIERI (a cura di), *Migrazioni e governance digitale*, cit., 28 ss.

Accanto a queste iniziative promosse a livello europeo, anche i singoli Stati membri hanno via via introdotto sistemi di intelligenza artificiale al fine di assolvere varie funzioni di controllo dei confini e dell'immigrazione¹⁹.

Diversi Stati membri hanno impiegato sistemi di IA per velocizzare la comprensione linguistica ai controlli di frontiera e durante le procedure di asilo, essendo tali sistemi al contempo utili ad individuare il Paese di origine o la zona di provenienza²⁰. In Italia, ad esempio, sistemi di traduzione automatica sono stati impiegati per trascrivere i colloqui con i richiedenti asilo.

Sempre in Italia un sistema algoritmico è stato testato dalla Polizia di Stato per facilitare le procedure di riconoscimento facciale²¹: il sistema SARI²² – i cui rischi per i diritti umani sono stati portati alla luce dal Garante della privacy come si dirà nel prossimo paragrafo – ha l'obiettivo di raffrontare immagini estrapolate da telecamere di sicurezza con le immagini contenute in altre banche dati per identificare soggetti nell'ambito di indagini investigative.

Infine, tra i progetti di IA più discussi vi sono quelli di rilevamento delle emozioni, messi in funzione a sostegno dell'autorità di frontiera, per scagionare eventuali inganni durante i controlli. Il più noto, tra tali progetti, è l'*iBorderCtrl*²³ che aveva l'obiettivo di sviluppare un "Avatar" di supporto decisionale a disposizione del personale di frontiera al fine di riconoscere le bugie, una sorta di

¹⁹ Per una mappatura dei sistemi di IA in uso per il controllo delle migrazioni si veda lo studio di C. DUMBRAVA, *Artificial intelligence at EU borders Overview of applications and key issues*, EPRS | European Parliamentary Research Service, July 2021; si veda inoltre P. MOLNAR, *The EU's AI Act and its Human Rights Impacts on People Crossing Borders. A project of the Migration Strategy Group on International Cooperation and Development*, in *Dot.Mig*, giugno 2022; D. OZKUL, *Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe. Algorithmic Fairness for Asylum Seekers and Refugees (AFAR) Project*, Oxford, 2023.

²⁰ È riportato da C. DUMBRAVA, *Artificial intelligence in asylum procedures in the EU*, EPRS | European Parliamentary Research Service, luglio 2025, 4.

²¹ In generale sul tema cfr. G. MOBILIO, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Napoli, 2021. Sulle tecniche di riconoscimento facciale con specifico riferimento al contesto delle migrazioni cfr. N. VAVOULA, *Artificial Intelligence (AI) at Schengen Borders: Automated Processing, Algorithmic Profiling and Facial Recognition in the Era of Techno-Solutionism*, in *European Journal of Migration and Law*, 2021.

²² Si veda la notizia sul sito del Ministero dell'Interno: <https://www.interno.gov.it/it/notizie/sistema-automatico-riconoscimento-immagini-futuro-diventa-realta>. Inoltre cfr. M. CARBONE, *SARI, il riconoscimento facciale nella pubblica sicurezza: servono regole e trasparenza*, in *agenda-digitale.eu*, 23 marzo 2021.

²³ Si veda il sito di tale progetto <https://www.iborderctrl.eu/>.

macchina della verità mediante l'osservazione delle micro-espressioni facciali²⁴. In merito a tale progetto, finanziato per 4,5 milioni di euro e testato a diversi valichi di frontiera europei, si è espressa anche la Corte di Giustizia UE²⁵ in una decisione che, pur riguardando profili specifici sulla trasparenza e sul diritto di accesso agli atti del progetto, riconosce il rischio che l'utilizzo di simili tecnologie possano comportare sui diritti fondamentali. In un significativo passaggio, la Corte di Giustizia infatti precisa che «la circostanza che i partecipanti al progetto *iBorderCtrl* siano tenuti a rispettare i diritti fondamentali e i principi riconosciuti, in particolare, dalla Carta, e che la Commissione sia tenuta a vigilare sul rispetto di detti diritti e di detti principi, non è in grado di far presumere l'assenza di una qualsiasi violazione di tali diritti e principi e di escludere l'esistenza di un interesse pubblico prevalente alla divulgazione dei documenti relativi a tale progetto a causa del possibile impatto delle tecniche utilizzate sulla protezione dei diritti fondamentali»²⁶.

3. *L'impatto dei sistemi di IA sulla condizione delle donne migranti.*

I sistemi di intelligenza artificiale in uso ai confini europei possono avere un significativo impatto sui diritti delle donne migranti, tanto in termini di “rischi”, tanto in termini di “benefici”.

I rischi sono prevalentemente connessi al fatto che i sistemi di intelligenza artificiale in uso per il controllo dell'immigrazione potrebbero non essere istruiti adeguatamente per prendere in considerazione la specifica condizione delle donne migranti, risultando discriminatori.

Tale situazione è dovuta alla falsa “neutralità” non solo dei sistemi di intelligenza artificiale²⁷, ma anche del diritto dell'immigrazione e dei rifugiati.

In dottrina è stato notato, infatti, come specialmente il sistema di protezione

²⁴ In letteratura si è discusso riguardo ai rischi per i diritti umani connessi a tale progetto cfr. J. SÁNCHEZ-MONEDERO, L. DENCİK, *The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl*, in *Information, Communication & Society*, 25, 3, 2022, 413 ss., i quali non presentano il progetto solo dal punto di vista tecnico ma riflettono sul suo impatto sui diritti fondamentali.

²⁵ Corte Giust. UE, 7 settembre 2023, XS C-135/22 P, Patrick Breyer c. European Research Executive Agency.

²⁶ *Ibidem*, par. 105.

²⁷ Si veda sul tema dal punto di vista costituzionalistico M. D'AMICO, *Parole che separano. Linguaggio, Costituzione*, Milano, 2023, 139 ss.

dei rifugiati sia stato costruito, a seguito delle atrocità avvenute durante i regimi totalitari e la Seconda Guerra mondiale, sull'idea del rifugiato uomo e bianco, in fuga per motivi politici²⁸. Così la definizione giuridica di rifugiato contenuta nella convenzione di Ginevra sullo status di rifugiato prescinde da qualsiasi riferimento al sesso. Ai sensi dell'art. 1 può essere considerato rifugiato colui che ha il fondato timore, a causa del rientro nel Paese di origine, di essere perseguitato «per motivi di razza, religione, nazionalità, appartenenza ad un determinato gruppo sociale o per le sue opinioni politiche».

Questa impostazione ha condizionato il sistema di protezione internazionale che, per lungo tempo, è stato privo di qualsivoglia approccio di genere essendo costruito sulla base dell'idea dell'asilante "uomo" in fuga per motivi politici²⁹. Passi avanti sono stati poi effettuati grazie all'entrata in vigore della Convenzione di Istanbul che ha assicurato il riconoscimento dello status di rifugiato alle donne vittime di violenza di genere, nonché grazie alla Giurisprudenza della Corte di Giustizia che ha dato rilievo anche al mancato rispetto del principio di parità³⁰.

Tuttavia, è evidente che, in questo campo, i sistemi di intelligenza artificiale rischiano di essere istruiti sulla base delle definizioni normative, che ad oggi non menzionano espressamente la protezione delle donne che hanno subito forme di persecuzione di genere. I sistemi di intelligenza artificiale in uso ai confini potrebbero non valutare attentamente la specificità della condizione di vulnerabilità delle donne migranti, non rilevando abusi, discriminazioni e persecuzioni nel Paese di origine o le violenze durante le rotte. In tal modo gli algoritmi costruiti sulla base dello stereotipo di rifugiato uomo non solo sono discriminatori, ma possono diventare un "filtro di ingresso" nel territorio, precludendo il diritto alla protezione di queste donne.

Il fatto che gli algoritmi di frontiera siano costruiti sulla base di un pacchetto di esperienze maschili è per ora confermato dai pochi dati che sinora si riescono a reperire in materia³¹. Uno studio afferma che l'algoritmo del

²⁸ Si veda su questi aspetti E. RIGO, *La straniera*, Roma, 2022, 43 ss.

²⁹ Sul sistema di protezione internazionale nella prospettiva di genere si consentito il rinvio a C. SICCARDI (a cura di), *Il diritto d'asilo delle donne. Principi, Norme e Procedure*, Torino, 2025, specialmente in contributi di M. BIANCHI, E. FIDENZI.

³⁰ Sulla giurisprudenza della Corte di Giustizia e sul quadro normativo sovranazionale cfr. S. GRECO, *Il sistema di protezione internazionale di fronte alla persecuzione in base al sesso: le convenzioni internazionali e il diritto dell'unione europea*, in C. SICCARDI (a cura di), *Il diritto d'asilo delle donne*, cit.

³¹ Si veda il contributo di R. FRADA, *Gendered Borders: Unveiling the Impact of EU's Technological Fortress on Women and Queer Refugees*, in *Gender and Geopolitics Institute blog*, 2023.

progetto *iBorderCtrl*, citato nel paragrafo precedente, è stato testato prevalentemente su uomini europei bianchi³². Il sistema, dunque, potrebbe non tenere in considerazione minimamente le divergenze “di genere”, nonché le diversità culturali ed etniche che ugualmente possono incidere sul linguaggio del corpo e sui micro movimenti facciali rilevati per decifrare le menzogne.

Un altro sistema che ha un alto potenziale di riprodurre pregiudizi di genere è quello del riconoscimento facciale. Si pensi che un’indagine del Mit del 2018 ha rilevato come un software di riconoscimento facciale abbia un tasso di errore del 34,7% per le donne dalla pelle scura, rispetto allo 0,8% di errore per i maschi dalla pelle chiara³³.

Al fine di individuare soluzioni per evitare che gli algoritmi impiegati per il controllo dell’immigrazione siano discriminatori nei confronti delle donne migranti, risultano particolarmente importanti alcune indicazioni contenute nella già citata risoluzione adottata dal Comitato sulle migrazioni del Consiglio di Europa del 24 giugno del 2025. In via generale, il Comitato sottolinea, in linea con la risoluzione n. 2343 (2020) “Prevenire la discriminazione causata dall’uso dell’intelligenza artificiale”, l’importanza di azioni specifiche per prevenire la discriminazione e l’impatto negativo sproporzionato su gruppi quali le donne, le minoranze e le persone più vulnerabili ed emarginate, compresi i migranti³⁴.

Nello specifico contesto migratorio il rischio di trattamento differenziato e di multipla discriminazione nei confronti delle donne migranti può essere limitato mediante il coinvolgimento “fin dalla fase di progettazione”, dei rifugiati (donne e uomini) e della società civile per aumentare la fiducia e l’affidabilità, nonché mediante «la sensibilizzazione e la formazione dei funzionari preposti all’asilo, delle ONG e degli sviluppatori di IA rafforzeranno l’implementazione etica dell’IA»³⁵.

Ancora, il report del Consiglio di Europa si concentra sulle tecniche di riconoscimento facciale che, come appena visto, presentano un potenziale discriminatorio molto elevato, affermando che esse devono essere sottoposte «a regolari controlli di imparzialità, verificandone l’accuratezza su diverse tonalità di pelle e generi»³⁶.

³² Cfr. A HADIB, *The Ongoing Digitisation of Europe’s Borders*, in *Digital Freedom Found blog*, 2021.

³³ Ne dà notizia L. HARDESTY, *Study finds gender and skin-type bias in commercial artificial-intelligence systems*, in *Mit News*, 2018.

³⁴ COUNCIL OF EUROPE, COMMITTEE ON MIGRATION, REFUGEES AND DISPLACED PERSONS, *Artificial Intelligence and Migration*, Report1 Rapporteur: Mr Petri Honkonen, cit., par. 27.

³⁵ *Ibidem*.

³⁶ *Ibidem*, par. 45.

Inoltre, la stessa risoluzione offre alcune indicazioni per garantire l'inclusività degli algoritmi anche nei confronti delle donne migranti. Ad esempio, strumenti di intelligenza artificiale progettati per ottimizzare il reinsediamento lavorativo in base alle competenze individuali e alla domanda di lavoro locale, devono tener conto anche delle specificità di genere e delle peculiarità del mercato del lavoro "femminile" al fine di evitare di rafforzare la segregazione lavorativa (ad esempio, indirizzando le donne in modo sproporzionato verso settori a bassa retribuzione)³⁷.

Infine, occorre segnalare che sempre dal punto di vista dell'inclusione, le donne migranti potrebbero beneficiare dei sistemi di IA di apprendimento e traduzione linguistica, poiché recenti studi dimostrano come la mancata conoscenza della lingua del Paese ospitante o di altra lingua veicolare costituiscono uno dei principali fattori di discriminazione per le donne, relegandole nell'ambiente domestico, anche una volta giunte in Europa³⁸.

Una regolamentazione degli algoritmi impiegati nel campo dell'immigrazione e dell'asilo risulta dunque fondamentale per evitare effetti lesivi dei diritti di donne e uomini, nonché per potenziare l'inclusione in alcuni settori cruciali: vedremo se le nuove norme dell'AI Act, analizzate nel prossimo paragrafo, si pongono in questa direzione.

4. *L'AI Act e la disciplina dei sistemi per il controllo dell'immigrazione.*

In tale contesto, tutt'altro che nuovo all'utilizzo di sistemi algoritmici, è intervenuto l'AI Act, il quale all'allegato III, par. 7 menziona espressamente i sistemi di intelligenza artificiale impiegati nel controllo delle frontiere, nell'asilo e nell'immigrazione qualificandoli come ad "alto rischio" (*high-risk AI systems*).

Nello specifico, il paragrafo 7 dell'allegato III individua quali sistemi ad alto rischio i sistemi di IA destinati a essere utilizzati dalle autorità pubbliche competenti, o per loro conto, o da istituzioni, organi e organismi dell'Unione, quali:

- a) poligrafi o strumenti analoghi di rilevamento delle emozioni;
- b) strumenti per valutare un rischio (compresi un rischio per la sicurezza, un rischio di migrazione irregolare o un rischio per la salute) posto da una persona fisica che intende entrare o è entrata nel territorio di uno Stato membro;

³⁷ *Ibidem*, par. 61.

³⁸ Sul tema cfr. F. MAURI, *Il quadro giuridico europeo in relazione all'apprendimento della lingua del paese ospitante come strumento di accoglienza ed inclusione degli stranieri*, in *Federalismi*, 12, 2025, 73 ss.

- c) strumenti che assistono le autorità pubbliche competenti nell'esame delle domande di asilo, di visto o di permesso di soggiorno e per i relativi reclami per quanto riguarda l'ammissibilità delle persone fisiche che richiedono tale status, comprese le valutazioni correlate dell'affidabilità degli elementi probatori;
- d) strumenti per il riconoscimento biometrico o la verifica dell'identità, ad eccezione della mera verifica dei documenti di viaggio.

La scelta di qualificare i sistemi appena elencati come ad alto rischio costituisce senz'altro una scelta opportuna.

Anzitutto il legislatore ha incluso tra i sistemi ad alto rischio quasi la totalità dei sistemi già impiegati per il controllo dei confini descritti nel paragrafo precedente e che avevano già mostrato rischi concreti per i diritti umani.

In primo luogo, si è già detto riguardo ai sistemi di rilevamento delle emozioni, indicati alla lett. a), il cui impatto sui diritti di uomini e donne migranti è stato messo in luce in letteratura³⁹, come anche dalla stessa Corte di Giustizia dell'Unione Europea in riferimento al progetto *iBorderCtrl*.

In secondo luogo, dall'uso di algoritmi nell'ambito di procedure di identificazione di cui alla lett. d), possono derivare non solo i rischi propri di processi automatizzati già evidenziati dalla Corte di Giustizia UE⁴⁰, ma anche veri e propri *bias*, idonei a produrre discriminazioni algoritmiche⁴¹. Nel contesto delle migrazioni dove le tecnologie si trovano a gestire dati concernenti nazionalità, culture ed etnie da ogni angolo del globo il rischio di stereotipizzazione e discriminazione è assai elevato. Tra le ragioni che rendono opportuna la classificazione dei sistemi nel settore della migrazione quali ad alto rischio vi è la consapevolezza del legislatore europeo, espressa al considerando 60 dell'AI Act che tali sistemi «hanno effetti su persone che si trovano spesso in una posizione particolarmente vulnerabile e il cui futuro dipende dall'esito delle azioni delle autorità pubbliche competenti»⁴².

In terzo luogo, direttamente connessi alle procedure di identificazione vi sono poi quegli strumenti, indicati alla lett. b), volti a valutare il “rischio” posto da una persona straniera per la sicurezza, per la migrazione irregolare o per la salute. Tra

³⁹ Si veda su questo aspetto F. CABITZA, A. CAMPAGNER, M. MATTIOLI, *The unbearable (technical) unreliability of automated facial emotion recognition*, in *Big Data & Society*, 9, 2, 2022.

⁴⁰ Cfr. in questo senso Corte Giust. UE, 21 giugno 2022, CG C-817/19, *Ligue des droits humains*, *supra* par. 2.

⁴¹ Su tale rischio cfr. C. DUMBRAVA, *Artificial intelligence in asylum procedures in the EU*, EPRS | European Parliamentary Research Service, luglio 2025, 9.

⁴² Considerando 60 dell'AI Act, regolamento (UE) 2024/1689.

questi potremmo menzionare, ad esempio, le già citate tecniche di riconoscimento facciale, utilizzate dalla Polizia di Stato italiana nell'ambito di attività investigative, le quali sono state definite dal Garante della privacy "intrusive" «per il diritto alla vita privata e alla dignità delle persone, unitamente al rischio di ripercussioni negative su altri diritti umani e sulle libertà fondamentali»⁴³.

Inoltre, merita di essere compiuta una riflessione a parte in relazione ai rischi dei sistemi, di cui alla lett. c), volti ad assistere le autorità pubbliche nelle procedure di asilo, di rilascio di visti e permessi di soggiorno o reclami in materia. Un eventuale errore algoritmico⁴⁴ nell'ambito delle predette procedure rischia di compromettere diritti e garanzie fondamentali del diritto dell'immigrazione, quali il divieto di respingimento ed espulsioni collettive, la libertà personale e la libertà di circolazione. Il considerando 60 dell'AI Act ricorda, a proposito, che i sistemi non debbono essere utilizzati «per violare in alcun modo il principio di non respingimento o per negare sicure ed efficaci vie legali di ingresso»⁴⁵. Ad esempio, se questi algoritmi non vengono istruiti in relazione alle specificità della condizione femminile potrebbero non valutare attentamente situazioni di violenza o discriminazione che dovrebbero costituire il presupposto per il riconoscimento del diritto alla protezione internazionale. In dottrina è stato segnalato come, da un lato, l'uso di algoritmi nell'ambito di procedure così delicate per i diritti della persona straniera rischia di non cogliere la complessità dei casi individuali⁴⁶; dall'altro lato, come l'opacità di tali processi decisionali potrebbe ostacolare anche l'effettività del diritto di accesso alla giustizia e la piena cognizione del giudice.

Ferma l'opportunità della classificazione quali sistemi ad alto rischio, in dottrina è stato evidenziato come l'elenco predisposto dall'allegato III non sia completo⁴⁷.

Secondo alcuni studiosi, la norma risulterebbe lacunosa non menzionando i sistemi predittivi, vale a dire quegli strumenti algoritmici destinati a fare delle

⁴³ GARANTE PRIVACY, *Parere sul sistema Sari Real Time*, 25 marzo 2021 [9575877].

⁴⁴ Riflette sulle conseguenze giuridiche degli errori algoritmici nel contesto delle migrazioni M. FORTI, *Errori algoritmici ai confini: questioni giuridiche e implicazioni politiche*, in F. BIONDI DAL MONTE, M. FORTI, L. RANIERI (a cura di), *Migrazioni e governance digitale*, cit., 109 ss.

⁴⁵ Considerando 60 dell'AI Act, regolamento (UE) 2024/1689.

⁴⁶ F. BIONDI DAL MONTE, *Confini digitali e dati dei migranti nel Patto sulla migrazione l'asilo*, in F. BIONDI DAL MONTE, M. FORTI, L. RANIERI (a cura di), *Migrazioni e governance digitale*, cit., 178.

⁴⁷ E. BROUWER, *EU's AI Act and Migration Control. Shortcomings in Safeguarding Fundamental Rights*, in *Verfassungsblog*, 12 dicembre 2024.

previsioni su comportamenti umani o fenomeni sociali. Per quanto riguarda questi specifici sistemi nel contesto delle migrazioni occorre effettuare una distinzione. Da una parte, il legislatore dell'UE ha incluso alla lett. b) tra i sistemi ad alto rischio i sistemi predittivi che potremmo definire "individuali", vale a dire quegli strumenti algoritmici destinati a stimare la probabilità che una persona commetta reati, fuga o violi le condizioni del soggiorno. Dall'altra parte, il legislatore non ha invece incluso tra i sistemi ad alto rischio gli strumenti predittivi "generalisti", utilizzati al fine di conoscere con anticipo l'entità e le caratteristiche dei flussi in entrata. Pur trattandosi di strumenti apparentemente innocui che rendono maggiormente efficiente l'operato dell'autorità di frontiera, mettendola nelle condizioni di conoscere le caratteristiche dei flussi e organizzare con anticipo l'accoglienza, la gestione delle domande di asilo e i rimpatri⁴⁸, vi è chi in letteratura ne ha evidenziato le criticità: essi potrebbero infatti costituire un modo per limitare gli ingressi da alcuni Paesi, aggirando così il divieto di espulsioni collettive e aumentando il rischio di "profilazione" e discriminazione automatica in base all'origine nazionale⁴⁹.

5. Le garanzie previste per i sistemi "ad alto rischio" ed alcune eccezioni.

Ai sistemi ad alto rischio impiegati nel contesto migratorio si applicano sia le garanzie previste, in via generale dal regolamento (UE) 2024/1689 per i sistemi ad alto rischio⁵⁰, sia alcune norme che valgono per lo specifico contesto migratorio.

Per quanto riguarda le garanzie generali, l'AI Act impone una serie di obblighi di conformità ai sistemi qualificati come ad alto rischio, rivolti tanto ai fornitori (*providers*) quanto agli utilizzatori (*deployers*) pubblici. Tra questi obblighi, che sono oggetto di singoli commenti in questo volume, vale la pena ricordare: l'adozione di un sistema di gestione del rischio, la valutazione dell'impatto sui diritti fondamentali (*Fundamental Rights Impact Assessment*), la

⁴⁸ Su questi benefici cfr. A. BEDUSCHI, *International migration management in the age of artificial intelligence*, in *Migration Studies*, 9, 3, 2021, 576 ss.

⁴⁹ *Ibidem*.

⁵⁰ Per un commento alle norme dell'AI Act che vengono in rilievo in materia di migrazione cfr. E. BROUWER, *EU's AI Act and Migration Control. Shortcomings in Safeguarding Fundamental Rights*, in *Verfassungsblog*, 12 dicembre 2024; L.S. STEWART, *The regulation of AI-based migration technologies under the EU AI Act: (Still) operating in the shadows?*, in *European Law Journal*, 30, 1-2, 122 ss.; Z. VASILIKOV, *Artificial intelligence in the service of border control in the EU*, in *ADiM Blog, Analysis & Opinions*, dicembre 2024.

documentazione tecnica e istruzioni d'uso chiare, l'accuratezza, sicurezza e robustezza tecnica, gli obblighi di sorveglianza.

Di particolare importanza per il settore in esame è la *Fundamental Rights Impact Assessment*⁵¹, prevista all'art. 27 dell'AI Act ai sensi del quale i *deployers* devono effettuare una valutazione di impatto sui diritti fondamentali, prima dell'impiego del sistema di IA. Tale garanzia dovrà essere calibrata sui rischi e sui diritti che vengono in rilievo nello specifico contesto migratorio, come il diritto d'asilo, il divieto di respingimento, la libertà personale, la libertà di circolazione, nonché sulla specifica condizione delle donne migranti. Inoltre, grazie alle previsioni di cui alla lett. c) e d) dell'art. 27 – che impongono di tenere in considerazione «le categorie di persone fisiche e gruppi interessati» dall'uso del sistema di IA, nonché «i rischi specifici di danno che possono incidere sulle categorie di persone fisiche o sui gruppi di persone individuati»⁵² – i *deployers* sono obbligati a valutare attentamente le diverse vulnerabilità che si presentano nel contesto migratorio, evitando ingiuste discriminazioni. Come visto, ad esempio, per evitare il rischio di multiple discriminazioni nei confronti delle donne migranti è fondamentale coinvolgere già nella fase di progettazione rifugiati e rifugiate, ONG e esperti in materia.

La disciplina dei sistemi impiegati nel contesto migratorio prevede però alcune eccezioni rispetto alle garanzie accordate a tutti gli altri sistemi. Tali eccezioni riguardano gli obblighi di trasparenza, registrazione e sorveglianza umana.

Per quanto concerne gli obblighi di trasparenza, l'AI Act prevede diverse eccezioni che trovano ragione negli interessi pubblici sottesi al controllo dei confini, quali l'ordine pubblico e la sicurezza. Non a caso l'art. 79 in materia di “Riservatezza” prevede un paragrafo dedicato ai sistemi in oggetto, imponendo che «le informazioni scambiate in via riservata tra le autorità nazionali competenti, o tra le autorità nazionali competenti e la Commissione, non siano divulgate senza previa consultazione dell'autorità nazionale competente e del *deployer* che hanno prodotto tali informazioni, qualora tale divulgazione rischi di compromettere gli interessi pubblici e di sicurezza nazionale»⁵³.

Di conseguenza anche la documentazione che riguarda i sistemi di IA ad alto

⁵¹ In generale sulla valutazione di impatto sui diritti umani si veda M. D'AMICO, *I diritti umani nella rivoluzione dell'intelligenza artificiale*, in *4C Portal AI*, 7 maggio 2025; dal punto di vista tecnico e interdisciplinare (giuridico e informatico) si veda M. TAMBORINI *et al.*, *HH4AI: A methodological Framework for AI Human Rights impact assessment under the EUAI ACT*, in *Computer Science*, 23 marzo 2025.

⁵² Art. 27 dell'AI Act, regolamento (UE) 2024/1689.

⁵³ Art. 79 dell'AI Act, regolamento (UE) 2024/1689.

rischio in materia di controllo delle frontiere e dell'immigrazione non è soggetta agli stessi obblighi di trasparenza e pubblicità applicabili a tutti gli altri sistemi.

Ad esempio, l'art. 59 impone ai fini del trattamento dei dati nello spazio di sperimentazione dell'IA, la pubblicazione sul sito web delle autorità competenti della «sintesi del progetto di IA, dei suoi obiettivi e dei risultati attesi»⁵⁴. Tale obbligo non si applica, per espressa previsione normativa, alle attività delle autorità competenti in materia di migrazione. Analogamente, la documentazione tecnica prodotta dai fornitori deve «rimanere nei locali delle autorità» e non può essere resa pubblica⁵⁵. Solo l'autorità di vigilanza del mercato può richiederne una copia.

Come è stato evidenziato nella risoluzione sul tema del Consiglio di Europa, «la mancanza di trasparenza, la comprensione limitata da parte del pubblico, specialmente tra le popolazioni migranti vulnerabili, compromettono la capacità di mettere in discussione o ricorrere contro le decisioni basate sull'intelligenza artificiale»⁵⁶.

Direttamente connessa agli obblighi di trasparenza, vi è la disciplina concernente la registrazione nella Banca Dati UE per i sistemi ad alto rischio, istituita appositamente dall'art. 71 dell'AI Act con lo scopo di «agevolare il lavoro della Commissione e degli Stati membri nel settore dell'IA e di aumentare la trasparenza nei confronti del pubblico»⁵⁷.

Diversamente, per i sistemi di IA ad alto rischio nei settori delle attività di contrasto, della migrazione, dell'asilo e della gestione del controllo delle frontiere, gli obblighi di registrazione devono essere adempiuti in una sezione non pubblica della banca dati dell'UE⁵⁸. L'accesso a tale sezione non pubblica sicura è strettamente limitato alla Commissione e alle autorità di vigilanza del mercato.

Un'altra differenza di rilievo riguarda gli obblighi di sorveglianza umana prescritti per tutti i sistemi ad alto rischio al fine di «prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere»⁵⁹. Nello specifico, tale sorveglianza è svolta per la maggioranza dei sistemi separatamente da almeno due persone fisiche dotate della necessaria competenza, formazione e autorità. Tuttavia, il requisito della sorveglianza operata separatamente da almeno due persone non è richiesto proprio per i sistemi

⁵⁴ Art. 59 dell'AI Act, regolamento (UE) 2024/1689.

⁵⁵ Art. 78, par. 3 dell'AI Act, regolamento (UE) 2024/1689.

⁵⁶ COUNCIL OF EUROPE, COMMITTEE ON MIGRATION, REFUGEES AND DISPLACED PERSONS, *Artificial Intelligence and Migration* Report1 Rapporteur: Mr Petri Honkonen, cit., par. 12.

⁵⁷ Art. 71 dell'AI Act, regolamento (UE) 2024/1689.

⁵⁸ Considerando 131 dell'AI Act, regolamento (UE) 2024/1689.

⁵⁹ Art. 14 dell'AI Act, regolamento (UE) 2024/1689.

ad alto rischio impiegati nel controllo delle frontiere e dell'immigrazione.

Oltre alle norme specifiche dell'AI Act, occorre considerare che i sistemi di intelligenza artificiale devono ovviamente sottostare agli obblighi e alle garanzie previste dal diritto dell'immigrazione e dal sistema comune europeo di asilo.

È lo stesso AI Act infatti a richiamare i principi fondamentali in materia, precisando che i sistemi impiegati nel settore in esame «non dovrebbero in alcun caso essere utilizzati dagli Stati membri o dalle istituzioni, dagli organi o dagli organismi dell'Unione come mezzo per eludere gli obblighi internazionali a essi derivanti a titolo della convenzione delle Nazioni Unite relativa allo status dei rifugiati firmata a Ginevra il 28 luglio 1951, modificata dal protocollo del 31 gennaio 1967»⁶⁰.

6. Alcuni ostacoli all'applicabilità delle garanzie previste dall'AI Act.

Dopo aver analizzato le garanzie previste dall'AI Act, occorre mettere in luce come alcune scelte del legislatore UE possano comportare un vuoto di tutela, non assicurando l'applicabilità delle norme appena descritte ad alcuni sistemi di IA.

Anzitutto, è opportuno ragionare sulla clausola prevista all'art. 2, par. 3 dell'AI Act che precisa come il regolamento non debba pregiudicare «le competenze degli Stati membri in materia di sicurezza nazionale, indipendentemente dal tipo di entità incaricata dagli Stati membri di svolgere compiti in relazione a tali competenze»⁶¹.

La norma pone un aspetto critico non di poco conto: l'elaborazione di sistemi di IA per il controllo dell'immigrazione e dei confini potrebbe essere inclusa nelle competenze statali in materia di sicurezza nazionale escludendo così l'applicabilità del regolamento UE?

Vi è in effetti un'ambiguità di fondo che caratterizza la materia dell'immigrazione in generale. Nonostante la competenza esclusiva statale in materia di sicurezza nazionale sia sancita anche dai Trattati⁶², il richiamo alla competenza nazionale nell'ambito dell'immigrazione è assai raro. Anzi in tale settore, da anni di competenza UE, gli Stati sono maggiormente portati ad invocare una condivisione di responsabilità. Quanto appena affermato sembra confermato dalle previsioni della legge italiana sull'IA di attuazione del regolamento (UE)

⁶⁰ Considerando 60 dell'AI Act, regolamento (UE) 2024/1689.

⁶¹ Art. 2, par. 3 dell'AI Act, regolamento (UE) 2024/1689.

⁶² Cfr. in particolare art. 72 TFUE ai sensi del quale «Il presente titolo non osta all'esercizio delle responsabilità incombenti agli Stati membri per il mantenimento dell'ordine pubblico e la salvaguardia della sicurezza interna».

2024/1689⁶³, la quale esclude espressamente dal proprio ambito di applicazione le attività in materia di sicurezza nazionale⁶⁴ e non menziona invece in alcun modo le attività di controllo dell'immigrazione, evidentemente rimesse alla disciplina europea.

Un altro possibile ostacolo all'applicabilità delle garanzie previste dall'AI Act potrebbe essere dovuto alla natura necessariamente transnazionale dei sistemi di controllo dell'immigrazione. Non è escluso che i fornitori e *deployers*, pur sempre con lo scopo di controllare le frontiere UE, agiscano in Paesi terzi di origine dei migranti, tentando così di aggirare gli obblighi imposti dall'AI Act. In realtà il legislatore UE ha cercato di limitare il più possibile tale rischio, precisando all'art. 2 che l'AI Act si applica «ai fornitori che immettono sul mercato o mettono in servizio sistemi di IA, indipendentemente dal fatto che siano stabiliti o ubicati nell'Unione o in un paese terzo», nonché «ai fornitori e ai *deployer* di sistemi di IA che hanno il loro luogo di stabilimento o sono situati in un paese terzo, laddove l'output prodotto dal sistema di IA sia utilizzato nell'Unione»⁶⁵.

Ancora, occorre tenere in considerazione che l'allegato III – pur facendo riferimento alla maggior parte dei sistemi già in uso ai confini europei e descritti nel paragrafo 2 di questo scritto, come i sistemi di rilevamento dell'inganno, i sistemi volti a facilitare l'identificazione e l'incrocio di dati biometrici – non si applica ai sistemi già in uso. Infatti, il considerando 177 dell'AI Act precisa che il regolamento «si applica ai sistemi di IA ad alto rischio che sono stati immessi sul mercato o messi in servizio prima della data generale di applicazione dello stesso, solo se, a decorrere da tale data, tali sistemi sono soggetti a modifiche significative della loro progettazione o finalità prevista»⁶⁶.

Inoltre, per quanto riguarda i sistemi di IA impiegati sui sistemi It su larga scala elencati all'allegato X, l'art. 111 specifica che gli operatori «dovrebbero adottare le misure necessarie per conformarsi ai requisiti del presente regolamento, rispettivamente, entro la fine del 2030»⁶⁷.

Simili previsioni comportano un vuoto di tutela escludendo l'applicabilità delle garanzie previste dal regolamento per i sistemi ad alto rischio ai sistemi di IA già in uso o posticipandone l'applicazione nell'ambito dei sistemi It di larga

⁶³ Legge 23 settembre 2025, n. 132, Disposizioni e deleghe al Governo in materia di intelligenza artificiale (25G00143). Sulla legge si veda M. D'AMICO, *Intelligenza artificiale: nella via italiana una strategia di sviluppo*, in *Il Sole 24 Ore*, 25 settembre 2025.

⁶⁴ Cfr. art. 6, legge 23 settembre 2025, n. 132.

⁶⁵ Art. 2 dell'AI Act, regolamento (UE) 2024/1689, che definisce l'ambito di applicazione.

⁶⁶ Considerando 177 dell'AI Act, regolamento (UE) 2024/1689.

⁶⁷ Art. 111 dell'AI Act, regolamento (UE) 2024/1689.

scala, che hanno un ruolo fondamentale nella governance digitale delle migrazioni e degli *smart borders*⁶⁸.

7. I sistemi vietati che vengono in rilievo nel contesto dell'immigrazione.

Tra i sistemi vietati dall'AI Act, elencati all'art. 5, ve ne sono alcuni che possono venire in rilievo nel contesto delle migrazioni e sui quali è opportuno porre l'attenzione.

Fra gli strumenti vietati, l'art. 5 indica quelli che sfruttano «la vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo in un modo che provochi o possa ragionevolmente provocare a tale persona o a un'altra persona un danno significativo»⁶⁹. Non si può negare come nel contesto migratorio vi sia il concreto rischio che i sistemi di IA possano «sfruttare» la situazione di vulnerabilità in cui si trovano le persone migranti dopo lunghi e pericolosi viaggi, al fine, ad esempio, di captare informazioni e selezionare i migranti da respingere. Tuttavia, non si può non notare una criticità della norma in commento, la quale non menziona tra i fattori di vulnerabilità il sesso. Eppure, nel contesto delle migrazioni è molto frequente che siano proprio le donne a rischiare di essere «sfruttate» divenendo vittime di tratta o di sfruttamento sessuale e lavorativo.

Ancora è opportuno evidenziare che l'art. 5 vieta sistemi volti a valutare o prevedere il rischio che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità. Di conseguenza, l'uso di tecnologie per la previsione del comportamento dei migranti, o per la selezione automatizzata tra richiedenti asilo e migranti economici, potrebbe rientrare tra i divieti, in quanto basato su tecniche predittive potenzialmente manipolatorie e discriminatorie.

La norma, inoltre, vieta il raffronto di dati biometrici utilizzati «per trarre deduzioni o inferenze in merito a razza, opinioni politiche, appartenenza sindacale, convinzioni religiose o filosofiche, vita sessuale o orientamento sessuale»⁷⁰; rischio quest'ultimo che potrebbe non essere raro nel contesto migratorio.

⁶⁸ Su tali aspetti critici cfr. E. BROUWER, *EU's AI Act and Migration Control. Shortcomings in Safeguarding Fundamental Rights*, in *Verfassungsblog*, 12 dicembre 2024.

⁶⁹ Art. 5, par. 1, lett. b) dell'AI Act, regolamento (UE) 2024/1689.

⁷⁰ Art. 5, par. 1, lett. g) dell'AI Act, regolamento (UE) 2024/1689.

Infine, particolarmente importante il divieto di identificazione remota in spazi pubblici⁷¹ che impone di evitare il ricorso a tali tecnologie nei centri di detenzione o durante i controlli alle frontiere.

8. *Riflessioni conclusive.*

Dallo scritto emerge come gli “smart borders” pongano diverse sfide che non attengono solo al rapporto tra tecnologia e diritti, ma coinvolgono il difficile bilanciamento tra l’esigenza costituzionale di controllo dei confini e i diritti delle persone migranti, senza dimenticare la specificità delle migrazioni femminili. Si è visto, infatti, come in questo ambito la “falsa neutralità” non si cela solo dietro gli algoritmi, ma dietro la stessa costruzione giuridica del sistema di asilo e del diritto degli stranieri, impostato sull’idea del migrante straniero/uomo.

Per tale ragione, ancor più che in altri settori, gli algoritmi utilizzati per il controllo dell’immigrazione e dell’asilo potrebbero produrre ingiuste discriminazioni, anche intersezionali, specialmente nei confronti delle donne migranti, la cui situazione peculiare non viene presa in considerazione dalla tecnologia.

Certamente l’AI Act rappresenta un passo significativo verso una “giusta” regolamentazione dell’intelligenza artificiale nell’Unione Europea anche nello specifico contesto del controllo dell’immigrazione e dei confini.

Da un lato, il riconoscimento dei sistemi di IA impiegati per il controllo delle frontiere come “ad alto rischio” è senz’altro un segnale positivo che impone l’applicabilità di garanzie fondamentali per la tutela dei diritti e per il contrasto alla discriminazione algoritmica.

Dall’altro lato occorrerà prestare particolare attenzione a quelle aree in cui l’effettività degli obblighi e delle garanzie previsti dal regolamento rischia di essere compromessa e ritardata, come la deroga temporale fino al 2030 per i sistemi It elencati nell’allegato X, che costituiscono come visto il cuore della governance dell’immigrazione e del controllo delle frontiere.

In altri termini, il settore dell’immigrazione potrà costituire senz’altro uno dei banchi di prova più critici per verificare la concreta applicabilità delle garanzie previste dal regolamento e della loro efficacia per la tutela dei diritti della persona, donne e uomini.

⁷¹ Art. 5, par. 1, lett. h) dell’AI Act, regolamento (UE) 2024/1689.

CAPITOLO X

IL POTERE PERFORMATIVO DEL LINGUAGGIO
NELL'ERA DELL'IA
Nannerel Fiano

SOMMARIO: 1. L'intelligenza artificiale e il suo linguaggio *vis à vis* con la Costituzione. – 2. Il funzionamento dei *LLM*. – 3. L'*hate speech*, questo (s)conosciuto: dalla filosofia del linguaggio alla tutela costituzionale. – 4. La regolamentazione europea dell'intelligenza artificiale: quale linguaggio? – 5. La legge n. 132/2025 e i profili linguistici. – 6. Conclusioni: per un linguaggio umano e tecnologico non discriminatorio.

1. *L'intelligenza artificiale e il suo linguaggio vis à vis con la Costituzione.*

Il presente contributo intende analizzare il potere del linguaggio nel contesto delle tecnologie di intelligenza artificiale, con particolare riferimento agli effetti che il linguaggio algoritmico può produrre sui diritti fondamentali.

In particolare, l'analisi, incardinata in una prospettiva costituzionale, si concentra sul tema dei *bias* derivanti dai processi di apprendimento automatico, esaminando la normativa europea, il regolamento (UE) 1689/2024 (c.d. AI Act) e nazionale, rispondente alla legge n. 132/2025, rubricata "*Disposizioni e deleghe al Governo in materia di intelligenza artificiale*".

La riflessione muove dall'interrelazione tra l'art. 21 Cost., che sancisce la libertà di manifestazione del pensiero, e l'art. 3 Cost., che sancisce il principio di non discriminazione¹.

Nella nostra democrazia costituzionale, fondata sul principio egualitario e non discriminatorio², infatti, il linguaggio non dovrebbe «differenziare in

¹ M. D'AMICO, *Parole che separano. Linguaggio, Costituzione, diritti*, Milano, 2023, 30 ss.

² V. M. DELLA MORTE, *La dis-eguaglianza nello Stato costituzionale, Atti del Convegno di Campobasso 19-20 giugno 2015, 2016*, Napoli, 4.

modo irragionevole», poiché esso esprime la qualità di uno Stato e di una democrazia: quanto appena scritto è autorevolmente confermato da Agamben, secondo il quale «[è] dunque solo nella misura in cui riusciremo a mettere in questione questo assunto fondamentale, che una vera trasformazione della nostra cultura diventerà possibile. Dobbiamo provare a ripensare da capo che cosa facciamo quando parliamo, calarci in quella zona opaca e interrogarci non sulla grammatica e il lessico, ma sull'uso che facciamo del nostro corpo e della nostra voce mentre le parole sembrano uscire quasi da sole dalle nostre labbra. Vedremmo allora che in quest'esperienza ne va dell'apertura di un mondo e delle nostre relazioni con i nostri simili e che, pertanto, l'esperienza del linguaggio è, in questo senso, la più radicale esperienza politica»³.

Pertanto, a parere di chi scrive, lo sviluppo di un linguaggio aderente al principio di uguaglianza, escludendo qualsivoglia intervento dall'alto, onde evitare il rischio dell'imposizione di un linguaggio antidiscriminatorio – di per sé di difficile armonizzazione con il nostro Stato di diritto⁴ – è un'esigenza connaturata al consolidamento della struttura democratica.

Tale considerazione ha radici molto antiche: Aristotele chiamava l'uomo animale che parla e animale politico: in altri termini, «[il] linguaggio si lega anche alla sovranità e alla costituzione dello Stato e quindi il mondo dei nomi è il nostro mondo»⁵.

Il rapporto tra linguaggio e democrazia plurale è presto spiegato⁶: linguaggio e azione sono intrinsecamente legati l'uno con l'altra, (anche) poiché la lingua consta (o, meglio, può constare) di una parte performativa⁷.

E ancora, nell'indagare il tema della libertà di manifestazione del pensiero che si esprime mediante il linguaggio nella dimensione algoritmica o, se si vuole, nell'era della digitalizzazione, è possibile appuntare l'attenzione su due profili: quello dell'influenza esercitata dagli algoritmi mediante linguaggio sulla volontà e sull'autodeterminazione del singolo e quello sul rapporto tra la libera

³ G. AGAMBEN, *L'esperienza del linguaggio è un'esperienza politica*, in <https://www.quodlibet.it/giorgio-agamben-l-u2019esperienza-del-linguaggio-n-u2019esperien>.

⁴ M. D'AMICO, *Odio online: limiti costituzionali e sovranazionali*, in M. D'AMICO, C. SICCARDI (a cura di), *La Costituzione non odia. Conoscere, prevenire e contrastare l'hate speech online*, Torino, 2021, 29-30.

⁵ P. VALORE, *Linguaggio e realtà. Prospettive storiche e teoretiche*, in *Interlingvistikaj Kajeroj*, 2:1, 2011, 96.

⁶ M. D'AMICO, *Parole che separano*, cit., 139 ss.

⁷ Si veda, diffusamente, J. BUTLER, *Parole che provocano. Per una politica del performativo*, Milano, 2010.

(e lecita) manifestazione del pensiero e l'(illecita) propaganda o istigazione all'odio⁸.

In buona sostanza, l'intelligenza artificiale e il suo linguaggio fanno pieno ingresso nell'orizzonte del costituzionalismo, modificando, forse, il modo in cui sono applicabili i principi costituzionali.

Non solo, si è evidenziato che l'IA potrebbe avere modificato le stesse categorie del diritto antidiscriminatorio, insinuandosi nel fenomeno discriminatorio, per così dire, classico, e in quello intersezionale⁹.

La Costituzione, dunque, si trova per la prima volta a trovare concreta applicazione in un mondo simmetrico a quello "in carne ed ossa" che conosce insidie: si pensi, ad esempio, al clima di post-verità¹⁰ e al pericolo che vi sia una riduzione della partecipazione libera e consapevole del dialogo nella dimensione pubblica¹¹.

2. Il funzionamento dei LLM.

L'intelligenza artificiale, e in particolare i sistemi di elaborazione del linguaggio naturale (*Natural Language Processing, NLP*)¹², forniscono *outputs* che non solo riflettono, ma riproducono e amplificano i pregiudizi contenuti nei dati da cui apprendono: la conseguenza immediata è che il linguaggio digitale può discriminare, escludere o rendere invisibili soggetti e categorie, con effetti tangibili sui diritti fondamentali¹³.

⁸ Punite all'art. 604 *bis*. c.p.

⁹ C. NARDOCCI, *Algoritmi, eguaglianza, discriminazione. Le sfide dell'intelligenza artificiale*, Torino, 2025, 91-499.

¹⁰ Per un approfondimento, si veda G. FIORIGLIO, *Post verità, paura e controllo dell'informazione: quale ruolo per il diritto?*, in *Governare la paura*, 2019.

¹¹ Si v., sul punto, A. CREMONE, *Libertà di espressione e nuove tecnologie: a complicated relationship*, in *Riv. it. inform. e dir.*, 1, 2025; si rimanda anche alle autorevoli considerazioni di F. BALAGUER CALLEJÓN, *Freedom of Speech and Personality Rights*, in L. MEZZETTI (a cura di), *Science, Technology and Law. Mutual Impact and Current Challenges*, Bologna, 2014, 71-76.

¹² Esempi ne sono Chat GPT e Gemini.

¹³ Sul pericolo di perdere il controllo su tali meccanismi, v. C. ARUZZOLI, *Intelligenza artificiale e nudging: riflessioni intorno al principio di non esclusività della decisione algoritmica*, in F. FABRIZZI, L. DURST (a cura di), *Controllo e predittività. Le nuove frontiere del costituzionalismo nell'era dell'algoritmo*, Napoli, 2024, 71-73.

Infatti, i dati non sono neutri, ma sono situati, vale a dire rispecchiano un modo di pensare e, quindi, un modo di progettare¹⁴.

La problematica, poi, si complica laddove l'algoritmo apprenda da sé soluzioni per fornire nuovi *outputs*; infatti, lo stereotipo radicato nella società viene rispecchiato dal linguaggio della macchina, proprio perché l'IA riproduce, nel suo flusso di apprendimento – in modo particolarmente persuasivo¹⁵ – le credenze e i tropi umani seguendo uno schema che potremmo definire oggettivo e non relazionale, poiché completamente avulso da un possibile contesto di riferimento.

Gli stereotipi che sono insiti nell'esperienza di allenamento dei *LLM*, poi, riproducono il modello culturale cui fanno riferimento i dati di cui si nutrono (es. immagini con alcuni riferimenti alla cultura statunitense)¹⁶.

Non è un caso, quindi, che in dottrina si sia fatto riferimento ai riflessi collettivi derivanti dall'utilizzo dell'intelligenza artificiale¹⁷, dal momento che essa, tendenzialmente, riposa su uno specifico punto di vista culturale e sociale.

Va da sé, quindi, che l'intelligenza artificiale plasma «la realtà fisica e intellettuale dell'uomo, incidendo sul modo in cui questi, come organismo informazionale (*inforge*), si relaziona con gli altri e con se stesso in un ambiente *onlife* del tutto nuovo e in costante evoluzione»¹⁸, pur non avendo l'IA alcun elemento di tipo volontaristico¹⁹: e questo, a ben vedere, è l'aspetto che più rileva se consideriamo gli effetti dell'utilizzo degli strumenti di intelligenza artificiale sui consociati e sulla relazione tra Stato e singoli²⁰.

¹⁴ Sul punto, v. D. COLUMBRO, *Quando i dati discriminano. Bias e pregiudizi in grafici, statistiche e algoritmi*, Trento, 2024.

¹⁵ G. FASANO, *Le "informazioni sintetizzate" generate dai large language models e le esigenze di tutela del diritto all'informazione: valori costituzionale e nuove regole*, in *Diritti fondamentali*, 1, 2024, 114.

¹⁶ A. FERRARA, *Le macchine del linguaggio. L'uomo allo specchio dell'intelligenza artificiale*, Torino, 2025, 368 ss.

¹⁷ V. C. NARDOCCI, *Algoritmi, eguaglianza, discriminazione. Le sfide dell'intelligenza artificiale*, cit., 155 ss.

¹⁸ G. PISTORIO, *Chat GPT e la sfida della regolamentazione normativa*, in *Costituzione e intelligenza artificiale*, in *Riv. AIC*, 5, 2023.

¹⁹ Si v. rispetto all'aspetto "volitivo" dell'AI, G. DE MINICO, *Le fonti del diritto: un argine all'intelligenza artificiale?*, in *Riv. AIC*, 3, 2025, 91-96.

²⁰ Sulla necessità che i diritti fondamentali governino l'utilizzo delle nuove tecnologie, si vedano le preziose riflessioni di R. ARNOLD, *Technological Progress and Constitutional Values*, in *Science, Technology and Law. Mutual Impact and Current Challenges*, cit., 40-41.

A questo proposito, mi piace riportare le parole di Zagrebelsky, il quale, pur non trattando del tema delle nuove tecnologie dell'informazione e della comunicazione, ha affermato che «[s]ono dittature ideologiche i regimi che manipolano i fatti, li travisano o addirittura li creano o li ricreano ad hoc, attraverso Ministeri della verità come descritti da George Orwell in 1984»²¹.

3. *L'hate speech, questo (s)conosciuto: dalla filosofia del linguaggio alla tutela costituzionale.*

L'*hate speech*, secondo una recente prospettiva, può essere interpretato come un atto linguistico di subordinazione: secondo la tesi di Bianchi, l'enunciato *I neri non possono votare*, «può essere concepito come un atto illocutorio di subordinazione: se proferito in circostanze appropriate pone in essere un fatto nuovo, e fa sì che i neri siano privati del diritto di voto»²². Il discorso d'odio non si limita a offendere: produce esclusione, limitazione della libertà di manifestazione del pensiero nelle vittime – con ciò incidendo sull'art. 21 Cost.²³ –, paura e marginalizzazione, così restringendo il campo di applicazione di cui all'art. 3 Cost.

La dimensione dei discorsi d'odio è a geometria variabile, a seconda della natura e del contesto, oltre che a seconda della natura umana o algoritmica, considerato che le formule algoritmiche presenti sui *social network* estremizzano i *bias* e «tendono a suggerire o restituire agli user contenuti sovente faziosi»²⁴.

Sul piano sovranazionale, oltre alla nota *General Policy Recommendation No. 15 on Combating Hate Speech* dell'ECRI rileva la CM/Rec(2022)16 – *Recommendation of the Committee of Ministers to member States on combating hate speech (Adopted by the Committee of Ministers on 20 May 2022 at the 132nd Session of the Committee of Ministers)*, che ha ribadito l'esigenza di contrastare il linguaggio d'odio con strumenti normativi diversificati sul piano penale, civile e amministrativo.

²¹ G. ZAGREBELSKY, *L'onestà delle parole*, in www.accademiadellacrusca.it, 13 dicembre 2006.

²² Si v. C. BIANCHI, *Linguaggio d'odio, autorità e ingiustizia discorsiva*, in *Riv. di estetica*, 64, 2017, 18.

²³ Sul tema, molto interessante, si rimanda alla lettura del volume di A. NICITA, *Nell'età dell'odio*, Bologna, 2025.

²⁴ Così P. DUNN, *Piattaforme digitali e moderazione dei contenuti d'odio: nodi giuridici e pratici*, in *Media Laws*, 3 maggio 2021.

Secondo la Raccomandazione, ai sensi del par. 1.2 dell'appendice «si intende per discorso d'odio qualsiasi forma di espressione mirante a stimolare, promuovere, diffondere o giustificare la violenza, l'odio o la discriminazione nei confronti di una persona o un gruppo di persone, o a denigrare una persona o un gruppo di persone per motivi legati alle loro caratteristiche o situazioni personali, reali o presunte, quali la “razza”, il colore della pelle, la lingua, la religione, la nazionalità o l'origine nazionale o etnica, l'età, la disabilità, il sesso, l'identità di genere e l'orientamento sessuale»²⁵.

Di rilievo è l'approccio olistico di cui si fa portatrice la Raccomandazione: il diritto non può sradicare, da solo, l'«emozione ostile»²⁶, ma necessita di un intervento – coordinato – anche sul piano della prevenzione, ovvero su quello dell'educazione e della formazione.

Decisamente interessante e lungimirante è, poi, il punto ove è specificato, al par. 3.33 che «[g]li intermediari della rete internet dovrebbero compiere tutti gli sforzi necessari per garantire che l'utilizzo di strumenti di automazione o di intelligenza artificiale sia verificato dalla moderazione di contenuti “umana” e che la moderazione dei contenuti prenda in considerazione le specificità dei contesti giuridici, locali, culturali, socio-politici e storici pertinenti. Nell'ambito degli sforzi per tenere conto di tali specificità, dovrebbero esaminare la possibilità di decentralizzare la moderazione dei contenuti».

In sostanza, si individua il principio dello *human in the loop*²⁷: nel contrasto all'*hate speech* l'intelligenza artificiale può fungere da supporto, non certamente da strumento dotato di una vita propria, privo di verifiche e di controlli umani.

La finalità esplicitata dalla raccomandazione oggetto di analisi, e cioè il contrasto all'*hate speech*, è certamente lodevole; è pur vero, però, che il raggiungimento di una definizione univoca di discorso d'odio, e soprattutto di quella definizione che prelude alla necessaria attivazione degli strumenti di diritto penale, rischia di fungere da detonatore “antidemocratico” ove vi siano democrazie che non conoscono una piena separazione tra poteri, e laddove il rapporto tra Stato e singoli non sia strutturato in modo tale da prevenire rovesciamenti pericolosi per l'esercizio delle libertà fondamentali e per il godimento dei diritti, siano essi civili e/o sociali.

La considerazione di cui sopra è stata fatta propria, con molta probabilità,

²⁵ Cfr. allegato alla raccomandazione CM/REC(2022)16, 1, comma 2.

²⁶ Si tratta di un termine preso a prestito da M. SANTERINI, *La mente ostile. Forme dell'odio contemporaneo*, Milano, 2021, 31-64.

²⁷ Si v. D. MARTIRE, *Human in the loop. L'essere umano come fattore condizionante della – o condizionato dalla – intelligenza artificiale*, in *Riv. it. inform. e dir.*, 2, 2024, 468 ss.

da parte del legislatore europeo: il *Digital Services Act*, al cons. 12, infatti, non definisce che cosa sia il discorso d'odio, ma si limita a specificare che: «[p]er conseguire l'obiettivo di garantire un ambiente online sicuro, prevedibile e affidabile, ai fini del presente regolamento il concetto di "contenuto illegale" dovrebbe rispecchiare ampiamente le norme vigenti nell'ambiente offline. In particolare, il concetto di "contenuto illegale" dovrebbe essere definito in senso lato per coprire anche le informazioni riguardanti i contenuti, i prodotti, i servizi e le attività illegali. Tale concetto dovrebbe, in particolare, intendersi riferito alle informazioni, indipendentemente dalla loro forma, che ai sensi del diritto applicabile sono di per sé illegali, quali l'illecito incitamento all'odio o i contenuti terroristici illegali e i contenuti discriminatori illegali, o che le norme applicabili rendono illegali in considerazione del fatto che riguardano attività illegali».

Per questa ragione, come suggerisce la raccomandazione adottata dal Consiglio di Ministri del Consiglio d'Europa, è opportuno individuare strumenti normativi di tutela di natura diversa, non necessariamente rientranti, se non nei casi più gravi, nella dimensione penalistica.

4. *La regolamentazione europea dell'intelligenza artificiale: quale linguaggio?*

Il regolamento (UE) 2024/1689, noto come AI Act, rappresenta il primo tentativo organico di disciplinare l'intelligenza artificiale in un'ottica di tutela dei diritti fondamentali²⁸.

L'approccio è quello del *risk-based model*: più elevato è il rischio di impatto sui diritti, più stringenti sono gli obblighi di conformità.

L'AI Act distingue tra sistemi ad alto rischio e sistemi a rischio limitato, specificando, peraltro, che i sistemi che incidono materialmente sul processo decisionale umano rientrano nella categoria più delicata.

Tuttavia, proprio in questa distinzione si annida un possibile equivoco: anche un sistema apparentemente "ausiliario" – ad esempio un correttore linguistico o un classificatore di testi – può influenzare in modo significativo il discorso pubblico e le rappresentazioni sociali.

²⁸ Si v. G. ZICCARDI, *Diritti digitali. Informatica giuridica per le nuove professioni*, Milano, 2022, 249-254.

Vi è chi ha parlato di una vera e propria «prova di resistenza per la legalità algoritmica», G. LO SAPIO, *L'Artificial Intelligence Act e la prova di resistenza per la legalità algoritmica*, in *Federalismi*, 16, 2024, 265 ss.

Il linguaggio, infatti, anche quando non decide, orienta.

L'AI Act fa riferimento al linguaggio nel punto 53 del "considerando" a proposito dell'ingresso dell'IA nel processo decisionale: nello specifico, è possibile leggere: «[t]enuto conto di tali caratteristiche, il sistema di IA fornisce solo un livello aggiuntivo a un'attività umana con conseguente riduzione del rischio. Tale condizione si applicherebbe, ad esempio, ai sistemi di IA destinati a migliorare il linguaggio utilizzato in documenti redatti in precedenza, ad esempio in relazione al tono professionale, allo stile accademico del linguaggio o allineando il testo a una determinata comunicazione di marchio».

Il regolamento sull'IA non disciplina, quindi, le modalità di contrasto di un linguaggio capace di incidere sul principio di uguaglianza o sulla dignità dei singoli, ma si limita all'aspetto di cui sopra.

È pur vero, però, che il regolamento muove intorno a sette principi: intervento e sorveglianza umani, robustezza tecnica e sicurezza, vita privata e *governance* dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale e ambientale e responsabilità, che devono informare l'elaborazione stessa dell'IA, che, per sua natura, secondo il legislatore europeo, deve muoversi lungo la direttiva dell'antropocentrismo e del rispetto della democrazia e dello Stato sociale.

5. La legge n. 132/2025 e i profili linguistici.

La legge italiana di regolamentazione dell'utilizzo dell'intelligenza artificiale, da interpretare e applicarsi conformemente al regolamento sull'IA²⁹, ha adattato alcune disposizioni al contesto nazionale³⁰.

Tra gli obiettivi dichiarati vi sono la protezione dei diritti fondamentali, la democrazia, lo Stato di diritto e la sostenibilità.

Il testo normativo fa riferimento al linguaggio all'art. 4, comma 3, con riferimento alla necessità che le informazioni e comunicazioni relative al trattamento dei dati che siano correlate all'utilizzo di strumenti di intelligenza artificiale siano «rese con linguaggio chiaro e semplice» al fine di rendere conoscibili all'utente i rischi correlati; allo stesso modo, anche l'art. 13, comma 2, fa riferimento alla necessità che le informazioni «relative ai sistemi di intelligenza artificiale utilizzati dal professionista sono comunicate al soggetto

²⁹ Art. 1, comma 2.

³⁰ V. E. GUARNACCIA, F. TRAINA, *Algoritmi e diritti: il nuovo disegno di legge italiano sull'intelligenza artificiale*, in *Lavoro, Diritti, Europa*, 7 luglio 2025.

destinatario della prestazione intellettuale con linguaggio chiaro, semplice ed esaustivo».

Eppure, anche in questo caso, l'art. 3, comma 1, non può passare inosservato agli occhi del/della costituzionalista, secondo cui «[l]a ricerca, la sperimentazione, lo sviluppo, l'adozione, l'applicazione e l'utilizzo di sistemi e di modelli di intelligenza artificiale per finalità generali avvengono nel rispetto dei diritti fondamentali e delle libertà previste dalla Costituzione, del diritto dell'Unione Europea e dei principi di trasparenza, proporzionalità, sicurezza, protezione dei dati personali, riservatezza, accuratezza, non discriminazione, parità dei sessi e sostenibilità».

Di particolare interesse, poi, sempre nell'ottica del rapporto tra linguaggio e democrazia è l'art. 1, comma 4, secondo cui «[l]'utilizzo di sistemi di intelligenza artificiale non deve pregiudicare lo svolgimento con metodo democratico della vita istituzionale e politica e l'esercizio delle competenze e funzioni delle istituzioni territoriali sulla base dei principi di autonomia e sussidiarietà e non deve altresì pregiudicare la libertà del dibattito democratico da interferenze illecite, da chiunque provocate, tutelando gli interessi della sovranità dello Stato nonché i diritti fondamentali di ogni cittadino riconosciuti dagli ordinamenti nazionale ed europeo».

6. Conclusioni: per un linguaggio umano e tecnologico non discriminatorio.

L'analisi condotta mostra come il linguaggio – umano o artificiale – sia uno strumento di potere performativo capace di incidere sulla realtà giuridica e sociale.

L'intelligenza artificiale, nel replicare il linguaggio umano, che «rivela il mondo interiore del parlante»³¹, ne assorbe anche le distorsioni, rischiando di amplificare disuguaglianze preesistenti.

Un diritto dell'intelligenza artificiale che ignori la dimensione linguistica rischia di essere monco.

Occorre, dunque, integrare nella regolazione del settore una consapevolezza linguistica esplicita, prevedendo obblighi di inclusività, trasparenza semantica e verificabilità dei processi linguistici automatici, così promuovendo, auspicabilmente, un «diritto costituzionale ad un linguaggio tecnologico comprensibile»³².

³¹ G. ZAGBERELSKY, *Il dubbio e il dialogo. Il labirinto di Norberto Bobbio*, Torino, 2024, 7.

³² Così A. SIMONCINI, *Il linguaggio dell'intelligenza artificiale e la tutela costituzionale dei diritti*, in *Riv. AIC*, 2, 2023, 34-39.

L'obiettivo non è soltanto prevenire la discriminazione e la compressione di principi costituzionali fondamentali, ma promuovere una cultura giuridica e tecnologica capace di valorizzare la diversità³³.

L'“intelligenza naturale” – ben si capisce, fondata sul principio di non discriminazione – deve guidare quella artificiale: solo così il linguaggio algoritmico può rispettare le fondamenta della nostra Carta costituzionale³⁴.

³³ È fondamentale, in questo senso, uno sforzo da parte di legislatori, giudici e società civile: cfr. G. FORMICI, *La censura ai tempi di Internet: dal blocco della rete al deplatforming*, in A. OSTI, M. DELLA MALVA (a cura di), *Libertà di espressione e forme moderne di censura*, Torino, 2024, 85.

³⁴ Interessante, in un'ottica extragiuridica, il testo di D. ANDLER, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, Torino, 2024.

CAPITOLO XI

LA PROTEZIONE DEI DATI PERSONALI AL TEMPO
DELL'INTELLIGENZA ARTIFICIALE:
IL COORDINAMENTO TRA AI ACT E GDPR

Pietro Villaschi

SOMMARIO: 1. Introduzione: i dati come linfa vitale della società digitale. – 2. La perdurante attualità di una regolazione che limiti un trattamento “selvaggio” dei dati personali degli utenti della società digitale. – 3. AI ACT e GDPR: punti di contatto e complementarità delle normative europee. – 4. AI ACT e GDPR: possibili problemi di coordinamento. – 5. Tra proposte di integrazione e rischi di sovra-regolazione.

1. Introduzione: i dati come linfa vitale della società digitale.

Nel contesto della società digitale contemporanea, i dati rivestono un'importanza cruciale sia per la gestione delle attività quotidiane sia per il corretto funzionamento degli algoritmi di intelligenza artificiale.

Innanzitutto, i dati rappresentano il materiale grezzo dalla quale le organizzazioni – pubbliche e private – traggono informazioni cruciali, orientano decisioni strategiche e sviluppano nuovi prodotti e servizi.

Questo è particolarmente evidente quando si considera che la nostra interazione all'interno dell'universo digitale lascia tracce che, opportunamente raccolte e analizzate, permettono di comprendere meglio comportamenti, tendenze e processi degli abitanti dell'info-sfera ¹.

In secondo luogo, quando consideriamo la dimensione degli algoritmi di intelligenza artificiale – in particolare quelli basati su un apprendimento automatico e sulle reti neurali – diventa evidente che senza dati non c'è apprendimento.

¹ Per riprendere la lungimirante definizione di L. FLORIDI, *La quarta rivoluzione. Come l'infosfera sta cambiando il mondo*, Milano, 2017, 1-312.

Gli algoritmi hanno bisogno quindi dei dati personali, mentre non è vero il contrario: la società dei dati, infatti, è esistita prima dell'avvento dell'intelligenza artificiale.

I dati, infatti, fungono da strumento per l'addestramento, la validazione e i test dei sistemi di intelligenza artificiale: attraverso dataset etichettati e non etichettati gli algoritmi imparano a riconoscere pattern, prevedere esiti e adattarsi, evolvendo e auto-apprendendo. In questo senso, i sistemi di intelligenza artificiale generativa ne sono una dimostrazione plastica.

Inoltre, non è solo la quantità ma, soprattutto, la qualità del dato che impatta direttamente sull'efficacia del sistema: dati inaccurati, incompleti o distorti possono portare a errori o a *bias*, che si amplificano nel contesto della società digitale. La regolazione, quindi, del trattamento dei dati personali è condizione necessaria per un efficace contrasto alle discriminazioni prodotte dagli algoritmi.

Infine, l'importanza dei dati travalica l'ambito strettamente tecnico e si estende a quello sociale, culturale e politico. I dati alimentano una forma di *governance* digitale: consentono ai decisori pubblici di monitorare fenomeni complessi (ad esempio, salute, mobilità, ambiente, sicurezza) e permettono ai cittadini di partecipare in modo informato alle decisioni che interessano la società.

In questo senso, la corretta gestione dei dati – dal punto di vista dell'accessibilità, della trasparenza, della qualità e della tutela della *privacy* – diventa condizione fondamentale per una società digitale equa². In definitiva, possiamo affermare che la società digitale non “funziona” senza dati e che gli algoritmi di intelligenza artificiale non possono esistere né operare in modo affidabile senza una base dati solida, ben gestita e coerente con principi costituzionali e sovranazionali: è la capacità di assicurare questa sinergia che plasmerà il futuro del rapporto uomo-macchina nel mondo interconnesso.

2. *La perdurante attualità di una regolazione che limiti un trattamento “selvaggio” dei dati personali degli utenti della società digitale.*

La regolazione dei dati personali, incarnata in particolare dal regolamento (UE) 2016/679 – GDPR, è stata un pilastro fondamentale su cui la società digitale europea è stata edificata³.

² S. ZUBOFF, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, 2023, 20.

³ Cfr. F. PIZZETTI, *Il governo dei dati personali. Dalla disciplina sulla privacy al regolamento europeo (ue) 2016/679*, Torino, 2018, 50.

In un'epoca in cui ogni click, transazione, sensore o interazione online può generare informazioni personali – e queste informazioni alimentano modelli algoritmici, sistemi predittivi e decisioni automatizzate – la protezione offerta dal GDPR assume un significato strategico per costruire un'intelligenza artificiale realmente antropocentrica ⁴.

Sotto questo punto di vista, può a ragione affermarsi che il GDPR, pur essendo stato scritto in un periodo storico in cui ancora non era diffusa l'intelligenza artificiale, è un regolamento con un'impostazione lungimirante, capace di adattarsi all'evoluzione che la società digitale ha intrapreso negli ultimi anni.

Basti pensare alla latitudine applicativa della nozione di dato personale che, ai sensi del regolamento, comprende qualsiasi informazione attinente a una persona fisica identificata o identificabile, anche indirettamente.

Inoltre, il GDPR stabilisce principi chiave – come liceità, correttezza e trasparenza; limitazione delle finalità; minimizzazione dei dati; accuratezza; conservazione per il tempo necessario; integrità e riservatezza – che dimostrano la loro centralità nel contesto della nuova rivoluzione digitale dell'intelligenza artificiale, pur dovendo essere calati in una realtà diversa da quella per la quale erano stati pensati.

Infine, il GDPR conserva la propria centralità anche dal punto di vista economico e sociale: garantendo trasparenza e rafforzando la fiducia degli utenti nei servizi digitali, consente alle imprese di operare in un contesto più stabile e sicuro ⁵.

La regolazione dei dati personali non dovrebbe quindi essere percepita come un freno all'innovazione: essa è, al contrario, condizione perché l'innovazione digitale sia realmente antropocentrica ⁶.

⁴ Su cui v. B. BODO, K. IRION, H. JANSSEN, A. GIANNOPOULOU, *Personal data ordering in context: the interaction of meso-level data governance regimes with macro frameworks*, in *Internet Policy Review*, 10, 3, 2021, 8 ss.

⁵ Cfr. R. D'ORAZIO, G. FINOCCHIARO, O. POLLICINO, G. RESTA, *Codice della privacy e data protection*, Milano, 2021, XXXVII-1795.

⁶ *Ex plurimis*, A. QUINTAVALLA, J. TEMPERMAN (a cura di), *Artificial Intelligence and Human Rights*, Oxford, 2023, 70 ss.; A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal-Riv. BioDiritto*, 1, 2019, 63 ss.; A. PAJNO, M. BASSINI, G. DE GREGORIO, M. MACCHIA, F.P. PATTI, O. POLLICINO, S. QUATTROCOLO, D. SIMEOLI, P. SIRENA, *AI: Profili giuridici intelligenza artificiale: Criticità emergenti e sfide per il giurista*, in *Riv. BioDiritto-BioLaw Journal*, 3, 2019, 2-35; F. DONATI, *La protezione dei diritti fondamentali nel regolamento sull'intelligenza artificiale*, in *Riv. AIC*, 1, 2025, 1-20.

3. *AI ACT e GDPR: punti di contatto e complementarità delle normative europee.*

La sinergia tra GDPR ed AI ACT si intuisce già solo per il fatto che i sistemi di intelligenza artificiale si nutrono di un'imponente mole di dati personali, che sono il petrolio su cui si alimenta tutta l'info-sfera digitale⁷.

Occorre, quindi, comprendere come i due regolamenti operino e si integrino tra loro e quali obblighi sviluppatori e utilizzatori sono tenuti a seguire⁸.

Una delle criticità segnalata da più parti è, infatti, il rischio che l'approvazione, in un breve lasso temporale, di molti e complessi regolamenti possa pregiudicare la certezza del diritto e minare la competitività del mercato dell'Unione.

Di qui, l'importanza di previsioni normative chiare, comprensibili a tutti e coordinate tra loro.

Da questo punto di vista, l'AI ACT lascia espressamente impregiudicato il GDPR, che rimane quindi in questo campo la normativa di riferimento (art. 2, comma 7, AI ACT).

Restano, quindi, salvi gli obblighi dei fornitori e dei *deployers* nel loro ruolo di titolari o responsabili del trattamento di dati ai sensi del GDPR. In tal senso, il GDPR si configura come una disciplina complementare e integrativa tutte le volte in cui a rilevare sono diritti legati alla tutela del "dato personale", secondo la definizione, molto ampia, data dal GDPR stesso.

I due regolamenti dovrebbero, in quest'ottica, operare in sinergia per giungere all'obiettivo del legislatore europeo di assicurare lo sviluppo di un'intelligenza artificiale antropocentrica.

Diverse sono le ipotesi di applicazione convergente dei due regolamenti. Ad esempio, uno sviluppatore di un sistema di intelligenza artificiale deve svolgere una valutazione di conformità prima di immettere il prodotto nel mercato, cosa che si accompagna ad alcuni adempimenti previsti dal GDPR, tra cui la valutazione di impatto sulla protezione dei dati (DPIA). Valutazione che richiama un'altra valutazione, quella di impatto sui diritti fondamentali (FRIA) richiesta ai *deployers* dall'AI ACT. Ed è proprio l'AI ACT, correttamente, a stabilire che,

⁷ Cfr. P. FALLETTA, A. MARSANO, *Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR*, in *Riv. it. inform. e dir.*, 1, 2024, 1-19; M.C. POLLICINO, *Gli effetti della "sommatoria" tra il GDPR e il nuovo Regolamento sulle intelligenze artificiali nell'ambito dell'attività amministrativa*, in *Riv. it. inform. e dir.*, 1, 2025, 232-256.

⁸ Su cui R. TARCHI, A. GATTI, *Intelligenza artificiale e protezione dei dati personali: problemi di metodo e di procedura*, in *DPCE*, 2, 2024, 1175-1194.

onde evitare inutili duplicazioni, se un obbligo è già assolto nella DPIA non va riproposto e adempiuto con la FRIA.

Altro punto di contatto tra i due regolamenti risiede nel fatto che la medesima autorità deputata a livello nazionale all'osservanza della disciplina in tema di protezione dei dati personali può essere anche designata come autorità per la vigilanza del mercato ai sensi dell'AI ACT. Si evita, quindi, una proliferazione di autorità regolatorie molteplici.

Simili sono anche le valutazioni di rischio, in termini di impatto sui diritti fondamentali, che i due regolamenti prescrivono, sebbene la prima con riferimento al solo trattamento dei dati personali. Un approccio, quindi, fondato sul rischio caratterizza tutta l'architettura europea in materia⁹.

Comune anche l'esigenza di una supervisione umana, onde evitare di incorrere nei casi di decisioni completamente automatizzate (cosa che, tuttavia, può verificarsi in molti sistemi di IA, specie ad alto rischio, se non adeguatamente regolati). In questo senso dimostra la sua lungimiranza il GDPR, che, già molto tempo prima dell'avvento dei sistemi di IA, prescriveva, all'art. 22, che «l'interessato ha il diritto di 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».

Entrambe le normative, infine, richiedono misure come la pseudonimizzazione, la crittografia, l'audit e, non meno importante, la formazione del personale.

4. AI ACT e GDPR: possibili problemi di coordinamento.

Nonostante, quindi, sotto molti aspetti, le due normative risultino tra loro coordinate per espressa volontà del legislatore europeo, non mancano i profili di contrasto tra i due regolamenti.

L'IA spesso richiede grandi quantità di dati (ciò è in conflitto con i principi di minimizzazione, limitazione della finalità e conservazione propri del GDPR). Alcuni diritti sanciti dal GDPR (ad esempio, il diritto all'oblio, la portabilità, l'accesso) sono difficili da garantire tecnicamente in certi modelli di IA. Più in generale, la materia su cui gravitano i due regolamenti è differente (i dati personali, l'uno, i sistemi di IA, specie ad alto rischio, l'altro), risentendo anche dei

⁹ Cfr. B. HOHMANN, G. KOLLAR, *Reflections on the data protection compliance of AI systems under the EU AI Act*, in *Cogent Social Sciences*, 11, 2025, 10 ss.

diversi contesti storici (e tecnologici) in cui le due normative sono state scritte.

A testimonianza di ciò, è significativo quanto rilevato dal Servizio di Ricerca del Parlamento europeo (*European Parliamentary Research Service* – “EPRS”), il quale ha evidenziato in particolare una criticità con riferimento al trattamento dei dati sottoposti a protezione speciale di cui all’art. 9 del GDPR, quando si tratti di sistemi di IA regolati dall’AI ACT ¹⁰.

Uno degli obiettivi principali dell’AI ACT è, infatti, come noto, quello di mitigare i *bias* nello sviluppo, nell’implementazione e nell’uso dei sistemi di IA, specie ad alto rischio. A tal fine, l’art. 10, par. 5, dell’AI ACT consente il trattamento di categorie particolari di dati personali, purché siano adottate misure adeguate a garantire la tutela della protezione dei dati, al fine di individuare e prevenire eventuali discriminazioni nell’uso di tali tecnologie.

Questo però rischia di entrare in conflitto con la disciplina stringente del GDPR in materia.

Affinché le disposizioni dell’AI ACT sulla rilevazione dei *bias* siano compatibili con il GDPR, occorre allora che: il trattamento di dati “sensibili” avvenga nel rispetto dei principi del GDPR, come la minimizzazione dei dati, la limitazione della finalità e della conservazione, l’integrità e la riservatezza, la *privacy by design* e *by default*; siano adottate misure di sicurezza tecniche e organizzative robuste, al fine di prevenire violazioni dei dati personali; il trattamento sia “strettamente necessario” per proteggere altri diritti fondamentali, come il diritto alla non discriminazione; soprattutto, sussista una base giuridica idonea ai sensi dell’art. 9 GDPR, eventualmente riconducibile ai motivi di interesse pubblico rilevante *ex* art. 9, par. 2, lett. f) e g) del GDPR, dove l’AI ACT rappresenterebbe il fondamento normativo dell’Unione e la lotta alle discriminazioni costituirebbe l’interesse pubblico rilevante ¹¹.

Peraltro, la discriminazione potrebbe derivare anche dal trattamento di dati non rientrante nelle “categorie particolari” dell’art. 9 del GDPR, come ad esempio la disabilità, l’età o, ancora, il genere. In questi casi, la base giuridica applicabile andrebbe individuata piuttosto nell’art. 6 del GDPR, che offre basi giuridiche più ampie rispetto all’art. 9 GDPR, come, ad esempio, l’interesse legittimo, rendendo più agevole giustificare il trattamento dei dati per ridurre i *bias*.

¹⁰ EUROPEAN PARLIAMENTARY RESEARCH SERVICE, *Algorithmic discrimination under the AI Act and the GDPR*, 2025, 1-2.

¹¹ Sull’art. 9 GDPR e sulle relative problematiche applicative, cfr. P. QUINN, G. MALGIERI, *The Difficulty of Defining Sensitive Data – The Concept of Sensitive Data in the EU Data Protection Framework*, Cambridge, 19 gennaio 2022; A. BUSACCA, *Le “categorie particolari di dati” ex art. 9 GDPR: divieti, eccezioni e limiti alle attività di trattamento*, in *Ordine internazionale e diritti umani*, 2018, 36-54.

Vi è, quindi, un'incertezza diffusa su come interpretare la disposizione dell'AI ACT relativa al trattamento di categorie particolari di dati, al fine di evitare forme di discriminazione algoritmica.

Il GDPR, imponendo limiti stringenti al trattamento di tali dati, potrebbe, inoltre, rivelarsi un ostacolo nell'attuale contesto economico e digitale, dove l'IA viene impiegata in molti settori e implica il trattamento massivo di dati personali e non personali, costituendo, quindi, un freno alla competitività del mercato dell'Unione.

5. Tra proposte di integrazione e rischi di sovra-regolazione.

Per affrontare le problematiche e le criticità evidenziate nelle pagine precedenti, l'EPRS suggerisce due possibili linee di intervento.

Da un lato, si avanza la proposta di una riforma del GDPR, che chiarisca la sua interazione con l'AI ACT e che adegui il regolamento alle nuove frontiere tracciate dall'intelligenza artificiale.

In particolare, si propone di agire con modifiche mirate, chiarendo, ad esempio, le basi giuridiche per il trattamento delle categorie speciali di dati, rafforzando il principio di proporzionalità, rivitalizzando i codici di condotta settoriali, rivedendo la DPIA (eventualmente ragionando di un'unica valutazione d'impatto comune ad entrambi i regolamenti), semplificando i sistemi di segnalazione di illeciti, migliorando i meccanismi di collaborazione tra le autorità regolatorie europee e nazionali.

Progetto sicuramente ambizioso, ma di non facile percorribilità, già solo se si considera che sarebbe necessario rimettere mano al regolamento.

Dall'altro, si suggerisce una linea di intervento più morbida, consistente nello stilare specifiche linee guida per fornire maggiore certezza giuridica ai soggetti coinvolti nel trattamento dei dati, specie per finalità connesse al contrasto delle discriminazioni. Ciò nella consapevolezza che i sistemi di intelligenza artificiale si nutrono di dati e hanno bisogno di essi proprio per prevenire e contrastare i rischi derivanti dalla loro stessa applicazione. Tali linee guida dovrebbero favorire una coerenza applicativa tra i due strumenti normativi, riducendo incertezze per operatori e autorità, senza dover necessariamente intervenire sul testo del GDPR.

In definitiva, il GDPR e l'AI ACT, pur nascendo in contesti completamente differenti, hanno l'obiettivo (ambizioso) di completarsi, operando in sinergia tra loro, con lo scopo, da un lato, di tutelare i diritti individuali che ruotano attorno ai dati personali, dall'altro, di disciplinare l'affidabilità e la sicurezza dei sistemi di IA.

Una *compliance* efficace richiede, infatti, l'integrazione di entrambi, con valutazioni d'impatto coordinate, una supervisione costante e adeguate misure tecniche e organizzative.

Rimane sullo sfondo il tema di come rendere tutti coloro che operano nel campo del digitale, dalle imprese ai singoli utenti, consapevoli di questa "selva" di regole e di come scongiurare il rischio di una sovra-regolazione normativa, che mini la competitività del mercato dell'Unione, specie se paragonato ad altre realtà più lontane (si pensi agli Stati Uniti).

Non è questa la sede per rispondere a questi difficili interrogativi.

Può, però, affermarsi che un approccio realmente antropocentrico nello sviluppo e nel governo dei sistemi di intelligenza artificiale passa attraverso un aggiornamento e una riscoperta delle regole in tema di protezione dei dati personali, primo fondamentale architrave su cui l'intero edificio della società digitale europea è stato costruito.

CAPITOLO XII

AI ACT E DISCRIMINAZIONI DI GENERE: TUTELA EXTRA-GIUDIZIALE E GIUDIZIALE A CONFRONTO *

Paolo Gambatesa

SOMMARIO: 1. Introduzione: l'algoritmo in una prospettiva costituzionale quale chiave di lettura per il contrasto delle discriminazioni di genere. – 2. Il contrasto alle discriminazioni di genere nei sistemi vietati e ad alto rischio. Una lettura in filigrana dell'AI Act. – 3. (*Segue*): le previsioni più esplicite della legge n. 132/ 2025. – 4. Luci e ombre dello strumento extra-giudiziale della valutazione di impatto ai sensi dell'art. 27 dell'AI Act. – 5. Le possibili prospettive di tutela giudiziale in un'ottica di integrazione dei livelli di protezione dei diritti. – 6. Considerazioni conclusive.

1. Introduzione: l'algoritmo in una prospettiva costituzionale quale chiave di lettura per il contrasto delle discriminazioni di genere.

Nel corso degli ultimi anni, nel dibattito politico e scientifico si è consolidata la consapevolezza della necessità di una regolamentazione giuridica capace di governare l'espansione dell'intelligenza artificiale. Il rapido sviluppo dei sistemi di IA ha sollevato interrogativi che investono la tutela dei diritti fondamentali, la sicurezza dei sistemi e la necessità di un controllo democratico sull'utilizzo degli algoritmi. Le implicazioni dell'IA si estendono a settori nevralgici della società – dalla giustizia al mercato del lavoro, dalla sanità all'istruzione – rendendo imprescindibile l'adozione di strumenti normativi in grado di bilanciare innovazione e protezione dei principi costituzionali.

* Il presente contributo, pur con significative modifiche e opportuni cambiamenti, riprende le tesi già espresse nel contributo *“La costituzionalizzazione dell'algoritmo tra implicazioni normative e livelli di tutela dei diritti fondamentali”*, pubblicato nella raccolta *La Constitución del algoritmo. Recensiones, presentaciones y entrevistas*, a cura di F. BALAGUER CALLEJÓN, Fundación Manuel Giménez Abad, Zaragoza, 2025.

Nel contesto attuale, si riscontra una crescente evidenza di fenomeni discriminatori e di pregiudizi sistemici che, lungi dall'essere superati, trovano nelle tecnologie di intelligenza artificiale nuovi canali di amplificazione, incidendo in modo particolare sulla condizione delle donne¹. Lo studio *Gender Shades*² è stato tra i primi sul tema e ha rivelato come alcuni sistemi di riconoscimento facciale presentino performance significativamente peggiori nei confronti dei volti femminili, segnalando una forma di trattamento algoritmico iniquo. Analogamente, i sistemi automatizzati di selezione del personale³ – basati su dati storici – hanno manifestato la tendenza a penalizzare i curricula provenienti da candidate donne. Inoltre, l'UNESCO ha recentemente evidenziato che i modelli linguistici di grandi dimensioni (LLM) nell'ambito dell'IA generativa mostrano un tasso di *bias* di genere particolarmente elevato, che finisce per incidere negativamente sulle prestazioni, oltre che sulla funzionalità stessa, di tali applicativi di IA⁴.

¹ I riferimenti in letteratura sull'argomento sono molteplici, per tutti, si v. M. D'AMICO, *Una parità ambigua: costituzione e diritti delle donne*, Milano, 2020, 313 ss.; F. BALAGUER CAL-LÉJON, *La trasformazione dei diritti nella società digitale e il suo impatto sulla parità*, in M. D'AMICO, B. LIBERALI (a cura di), *I diritti delle donne. Problematiche attuali e prospettive future*, Torino, 2024, 201 ss. e, nella medesima raccolta si v. anche E.C. RAFFIOTTA, *Intelligenza artificiale e tutela dell'uguaglianza di genere*, 169 ss.; inoltre, v. E. STRADELLA, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in *Liber amicorum per Paquale Costanzo*, 2020, in [Consultaonline.org](https://consultaonline.org); C. NARDOCCI, *Intelligenza artificiale e discriminazioni*, in *Riv. "Gruppo di Pisa"*, 3, 2021, 9 ss. (spec. 35 ss.).

² J. BUOLAMWINI, T. GEBRU, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in *Proceedings of Machine Learning Research*, 81, 2018, 1 ss. Lo studio, in particolare, ha rivelato disparità nei sistemi di riconoscimento facciale, attraverso l'analisi delle tecniche di IA adoperate da tre aziende leader per la classificazione di genere. È stato stimato un tasso di errore nel riconoscimento di volti maschili bianchi inferiore all'1%, mentre per le donne nere superava il 34,7%. Similmente, un altro studio ha evidenziato gli effetti discriminatori di sistemi di riconoscimento facciale sulla base dell'età, specificatamente, a danno delle generazioni più anziane che più difficilmente vengono riconosciute correttamente dai sistemi di IA (cfr. J. SUNG PARK *et al.*, *Understanding the Representation and Representativeness of Age in AI Data Sets*, in *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021, 834 ss.).

³ Cfr. G. ELISABETH BIRKELUND *et al.*, *Gender Discrimination in Hiring: Evidence from a Cross-National Harmonized Field Experiment*, in *European Sociological Review*, 38, 2022, 337 ss. In aggiunta, sulle *policies* da implementare negli Stati europei in questo ambito, v. EIGE, *Artificial intelligence, platform work and gender equality*, report pubblicato nel dicembre 2021 e disponibile al seguente indirizzo https://eige.europa.eu/publications-resources/publications/artificial-intelligence-platform-work-and-gender-equality?language_content_entity=en.

⁴ UNESCO, *Challenging systematic prejudices: an investigation into bias against women and*

Le cause profonde di tali distorsioni possono essere ricondotte alla cronica sottorappresentazione delle donne nei settori tecnologici. Sebbene i progressi scientifici siano stati notevoli, l'ecosistema dell'intelligenza artificiale resta prevalentemente governato da figure maschili, con una partecipazione femminile limitata nei processi decisionali, nella ricerca e nelle fasi di sviluppo ⁵. Tale squilibrio non solo perpetua disparità consolidate, ma riduce anche la pluralità di prospettive necessaria alla costruzione di sistemi di IA equi, affidabili e socialmente responsabili.

Occorre, pertanto, tenere traccia di questi importanti fattori di squilibrio del sistema e immaginare soluzioni che rendano immuni le tecnologie di IA da questi annosi e radicati problemi legati alle diseguaglianze di genere.

A suffragio di una simile esigenza interviene, in una prospettiva costituzionale ed europea, il principio di uguaglianza che costituisce perno dei sistemi ordinamentali, tanto del nostro Stato costituzionale (art. 3 Cost.) quanto di quello unionale (artt. 2 TUE e 21 CdUE).

Abbracciando questa prospettiva, le riflessioni che seguono non possono che prendere le mosse dal concetto, teorizzato da autorevole dottrina, di costituzionalizzazione degli algoritmi. L'espressione sta ad indicare anzitutto la necessità di «preservare il contributo del costituzionalismo ai progressi della civiltà, a livello statale, sovranazionale e globale, ricorrendo a fonti di legittimazione che sono al centro delle sue storiche conquiste: i diritti fondamentali, il controllo del potere e la canalizzazione dei conflitti attraverso il diritto» ⁶.

In quest'ottica, infatti, l'esigenza di “reagire” allo sviluppo tecnologico con norme di diritto positivo produce l'effetto, da un lato, di consolidare la prospettiva di controllare tale sviluppo e, dall'altro, di impedire che per mezzo di

girls in large language models, 2024, disponibile al seguente link: <https://unesdoc.unesco.org/ark:/48223/pf0000388971>. In argomento, si v. altresì il precedente studio del 2019, *I'd blush if I could: closing gender divides in digital skills through education*, disponibile al seguente link: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>. A commento v. Y. LI *et al.*, *Integrating AI in Education: Navigating UNESCO Global Guidelines, Emerging Trends, and Its Intersection with Sustainable Development Goals*, in *ChemRxiv*, 2025.

⁵ Cfr. M. D'AMICO, *Una parità ambigua: costituzione e diritti delle donne*, cit., 314-315.

⁶ F. BALAGUER CALLEJÓN, *La Costituzione dell'Algoritmo*, Milano, 2023, 106. Con la ulteriore precisazione per cui, secondo l'A., «costituzionalizzare l'algoritmo non significa semplicemente regolamentare la realtà digitale aggiungendo alla costituzione “analogica” le clausole corrispondenti per incorporare questo aspetto specifico. Al contrario, è necessario partire dalla premessa che la costituzione deve riordinare i suoi contenuti per incorporare il mondo ibrido in tutta la sua dimensione, il che significa adottare una nuova mentalità quando si tratta di riforme costituzionali» (*ibidem*).

quest'ultimo si pongano nel nulla le "storiche conquiste" del costituzionalismo contemporaneo. In questo senso, se è l'obiettivo delle costituzioni è (e auspicabilmente sempre sarà) quello di organizzare la realtà sociale, allora «la costituzione dell'algoritmo deve tenere conto dei limiti derivanti dalla configurazione specifica della nuova realtà digitale, in particolare del carattere globale dei fattori reali di potere che stanno contribuendo in modo decisivo a plasmarla»⁷.

A queste prime osservazioni, occorre aggiungere che la considerazione per cui «[s]olo l'Unione Europea ha la dimensione necessaria per far fronte alla globalizzazione e agli effetti negativi dello sviluppo tecnologico, allargando il cerchio del processo decisionale politico (compreso quello degli Stati membri e delle loro costituzioni interne)»⁸. Questa considerazione porta, quindi, a collocare la stessa costituzionalizzazione dell'algoritmo nella dimensione costituzionale europea e, conseguentemente, a dare rilievo ai diversi livelli che integrano quest'ultima. Ovverosia, quella dimensione in cui il crocevia tra costituzioni e norme UE (tra quest'ultime anche quelle della Carta dei diritti fondamentali) ci consente di esplorare con maggiore precisione «l'integrazione della realtà costituzionale nazionale degli Stati membri e della loro interazione con l'Unione europea»⁹.

In questi termini, si individuano le coordinate di fondo per comprendere e approfondire il fenomeno dello sviluppo tecnologico e dell'espansione dell'intelligenza artificiale e del loro impatto sui diritti fondamentali e sul contrasto alle discriminazioni di genere, e cioè impostando il ragionamento giuridico su più livelli di tutela che si integrano reciprocamente¹⁰.

Precisato il contesto entro cui si intende collocare il presente contributo, le riflessioni che seguono si perseguono un duplice obiettivo.

Il primo è rintracciare quale ruolo viene riservato al contrasto alle discriminazioni di genere nel diritto positivo, e in quest'ottica l'analisi si biforcherà, da un lato, tra una lettura in "filigrana" del regolamento UE sull'IA che consentirà, pur in assenza di esplicite previsioni sul tema, di rintracciare una tutela nelle

⁷ *Ibidem*.

⁸ F. BALAGUER CALLEJÓN, *La Costituzione dell'Algoritmo*, cit., 102.

⁹ F. BALAGUER CALLEJÓN, *Profili metodologici del Diritto costituzionale europeo*, in *La Cittadinanza Europea*, 1, 2015, 40.

¹⁰ Sul tema diversi livelli di tutela dei diritti fondamentali tra l'ordinamento nazionale e quello europeo, si v. F. BALAGUER CALLEJÓN, *Niveles y técnicas internacionales e internas de realización de los derechos en Europa. Una perspectiva constitucional*, in *Revista de derecho constitucional europeo*, 1, 2004, 25 ss., e F. BALAGUER CALLEJÓN, *La Costituzione dell'Algoritmo*, cit., 102 ss.

previsioni regolamentari (*infra* par. 2), dall'altro, il fuoco dell'analisi si incentrerà sulla legge italiana n. 132/2025, che invece introduce più espliciti riferimenti in materia (*infra* par. 3).

Il secondo obiettivo verterà, invece, sui possibili rimedi, extragiudiziali (*infra* par. 4) e giudiziali (*infra* par. 5), che possono condurre alla rimozione di ipotesi di discriminazione contro le donne realizzate, più o meno consapevolmente, da applicativi di IA.

2. *Il contrasto alle discriminazioni di genere nei sistemi vietati e ad alto rischio. Una lettura in filigrana dell'AI Act.*

L'art. 1 del *corpus* normativo europeo sull'IA, nel definire il suo oggetto precisa che esso mira sia a «migliorare il funzionamento del mercato interno», sia a «promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile», attraverso la garanzia di un «livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente [...]»¹¹.

L'espresso richiamo ai diritti della Carta di Nizza consente, in via implicita, di richiamare gli artt. 21 e 23 di quest'ultima, che rispettivamente impongono, da un lato, il divieto di forme di discriminazioni fondate sul sesso e, dall'altro, la parità di trattamento tra uomini e donne «in tutti i campi, compreso in materia di occupazione, di lavoro e di retribuzione». Ciò sino al ritenere, nel secondo periodo del par. 1 dell'art. 23, del tutto legittime azioni positive volte ad attribuire «vantaggi specifici a favore del sesso sottorappresentato»¹².

L'assonanza giuridica più vicina di queste disposizioni della Carta la si rinviene nel concetto di vulnerabilità¹³, che emerge in relazione alla necessità di

¹¹ Sulle implicazioni, i possibili rischi e le prospettive derivanti da un approccio di regolamentazione dell'IA basato sulla tutela dei diritti, si v. M. ALMADA, N. PETIT, *The EU AI act: a medley of product safety and fundamental rights?*, in *EUI, RSC, Working Paper*, 59, 2023, disponibile al seguente indirizzo <https://hdl.handle.net/1814/75982>.

¹² Al fine di concretizzare l'attuazione dei principi di non discriminazione e di parità, la Commissione europea ha adottato nel marzo 2020, la strategia «Un'Unione dell'uguaglianza: la strategia per la parità di genere 2020-2025» (COM(2020) 152 final), ove si precisa la necessità di un intervento da parte dei maggiori attori istituzionali, nazionali e sovranazionali, per affrontare efficacemente l'incidenza negativa dell'IA sulle donne.

¹³ Sul concetto di «vulnerabilità» e le sue ricadute nell'Era digitale si rinvia alle più ampie riflessioni di G. MALGIERI, *Vulnerability and Data Protection Law*, Oxford, 2023.

tenere indenni persone singole o gruppi ¹⁴ da effetti distorsivi dei sistemi di IA.

Il regolamento, però, non definisce compiutamente il concetto di vulnerabilità e il suo impiego presenta diverse sfumature a seconda del sistema di IA cui la disciplina rivolga attenzione.

L'art. 5, in materia di sistemi di IA vietati, al par. 1, lett. b) dispone che ad essere vietato è tanto l'immissione sul mercato quanto la messa in servizio e l'uso di sistemi che deliberatamente *sfruttino* la vulnerabilità di una persona fisica o di uno specifico gruppo, in relazione «all'età, alla disabilità o a una specifica situazione sociale o economica».

Da una prima lettura della disposizione pare si possa desumere che sistemi di IA che sortiscano effetti discriminatori su determinati soggetti a motivo del loro genere non siano da considerarsi vietati. Tuttavia, si potrebbe dare una lettura estensiva dell'art. 5, par. 1, lett. b), al fine di ricomprendere anche il genere nel novero dei fattori che concretizzano una situazione di vulnerabilità.

A suffragio di una simile interpretazione possono considerarsi due argomenti.

Il primo è legato alle parole «specifica situazione sociale ed economica», che seguono i concetti di più facile determinazione, quale “disabilità” ed “età” e che pare rinviino a precise situazioni la cui definizione spetta all'interprete. In quest'ottica, potrebbe rilevarsi come le discriminazioni di genere radicandosi in una società patriarcale, ove proliferano stereotipi contro le donne, possano ritenersi fonte di una *specifica situazione sociale* ¹⁵.

In altri termini, in una società ancora fortemente permeata da una cultura insensibile alla parità devono ritenersi necessarie tutte quelle misure volte a rimuovere gli ostacoli che sollecitano le disuguaglianze di genere.

¹⁴ Come osserva, G. MALGIERI, *Vulnerability and Data Protection Law*, cit., 49-51, intorno alla vulnerabilità sono stati sviluppati due distinti approcci: il primo enfatizza la connotazione particolaristica della vulnerabilità, dal momento che essa caratterizzerebbe determinate persone o gruppi sulla base di specifiche situazioni socio-economiche; la seconda, invece, ritiene la vulnerabilità quale condizione universale che accomuna tutti gli esseri umani, pur potendo essa variare a seconda del tempo e del luogo in cui emerge. Un tentativo di composizione delle due teorie è quello di F. Luna (spec. nt. 22, 50) che propone una concezione “stratificata” della vulnerabilità, secondo cui «*layers of vulnerability are not fixed attributes of specific individuals or groups but are features constructed by an individual's status, time, and location. In this sense, the concept of layering provides an opening to a more intersectional approach and stresses its cumulative and transitory potential*» (50-51).

¹⁵ In questi termini, *mutatis mutandis*, è possibile leggere il fenomeno della violenza contro le donne. Per un approfondimento del tema in questi termini si v. M. D'AMICO, C. NARDOCCI, S. BISSARO (a cura di), *Le violenze contro la donna. Origini, forme, strumenti di prevenzione e repressione della violenza di genere*, Milano, 2023, e i saggi ivi contenuti.

Il secondo argomento si potrebbe rintracciare nella stessa sovraordinazione nel sistema delle fonti dell'Unione delle norme della Carta di Nizza. L'idea di limitare la tutela non discriminatoria solo a fattori come la "disabilità" e l'"età" genererebbe una ingiustificata (se non irragionevole) differenziazione rispetto agli altri elementi¹⁶ che a norma dell'art. 21 della Carta posso fondare un trattamento discriminatorio. Proseguendo in questo solco, in un futuro rinvio pregiudiziale potrebbe dubitarsi della stessa validità dell'art. 5, par. 1, lett. b) in relazione all'art. 21 della Carta dei diritti fondamentali dell'UE.

La nozione di vulnerabilità emerge, poi, anche in relazione ai sistemi ad alto rischio, che costituiscono il cuore della regolamentazione europea. Ciò in quanto tali sistemi risultano caratterizzati da una intrinseca propensione a ledere i diritti fondamentali e per tale ragione possono essere utilizzati solo nel caso in cui essi rispettino le condizioni previste dal regolamento (specificamente le norme contenute nel Capo III).

Nella determinazione dei sistemi ad alto rischio l'art. 6, parr. 1 e 2, opera un rinvio agli allegati I e III. Quest'ultimo, in particolare, individua le otto macro materie in cui può aversi un sistema ad alto rischio.

Vi è modo di ritenere che la maggior parte dei sistemi di IA che potenzialmente potranno generare una discriminazione di genere saranno collocati proprio in questo novero.

Si pensi, ad esempio, agli algoritmi di selezione del personale che, basandosi su dati storici, possono perpetuare *bias* di genere nell'output, escludendo candidate che non rientrino nei modelli precedentemente assunti come standard¹⁷.

Inoltre, le applicazioni di riconoscimento facciale, come già detto, hanno dimostrato tassi di errore più elevati per le donne, in particolare per le donne di colore, rispetto agli uomini bianchi, mettendo in evidenza la necessità di affrontare le disuguaglianze intrinseche nei dati di addestramento¹⁸.

¹⁶ In particolare, gli ulteriori fattori di discriminazione individuati dall'art. 21 della Carta dei diritti fondamentali sono: il sesso, la razza, il colore della pelle o l'origine etnica o sociale, le caratteristiche genetiche, la lingua, la religione o le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita e l'orientamento sessuale.

¹⁷ Cfr. allegato III, par. 4, lett. a), («i sistemi di IA destinati a essere utilizzati per l'assunzione o la selezione di persone fisiche, in particolare per pubblicare annunci di lavoro mirati, analizzare o filtrare le candidature e valutare i candidati»).

¹⁸ A norma dell'allegato III, par. 2 rientrerebbero sistemi che fanno leva sui dati biometrici, ed in particolare:

«a) i sistemi di identificazione biometrica remota. Non vi rientrano i sistemi di IA destinati

L'art. 7 descrive la procedura per emendare l'allegato III e, a tal fine, viene conferito alla Commissione il potere di adottare atti delegati, attraverso i quali si possono apportare modifiche o aggiunte in relazione ai sistemi di IA individuati nell'allegato in virtù del loro impatto negativo sui diritti fondamentali (art. 7, par. 1, lett. b). In aggiunta, la medesima disposizione, al par. 2, elenca una serie di criteri che devono orientare la scelta emendativa. Tra questi, alla lett. h), viene dato rilievo all'esistenza di «uno squilibrio di potere» o alla circostanza per cui «le persone che potrebbero subire il danno o l'impatto negativo si trovino in una situazione vulnerabile rispetto al deployer di un sistema di IA, in particolare a causa della condizione, dell'autorità, della conoscenza, della situazione economica o sociale o dell'età».

Il riferimento alla *situazione vulnerabile* darebbe credito all'idea di una vulnerabilità potenzialmente transitoria, perlopiù legata al contesto in cui emerge piuttosto che ad elementi fissi predeterminabili *ex ante*¹⁹. Ad essere determinante è, infatti, la specifica posizione di squilibrio che emerge tra il deployer e «le persone».

Se, da un lato, questa specifica circostanza rende più agevole considerare le discriminazioni di genere nel novero dei sistemi di IA non ancora ricompresi in quelli ad alto rischio, dall'altro, però, il ruolo dell'interprete risulta sempre più discrezionale e per questo suscettibile di interpretazioni eterogenee in spregio ad una più certa tutela dei diritti su tutto il suolo europeo.

La vulnerabilità, così considerata, si lega all'eventualità che emerga l'impatto negativo o il danno, ora non più considerato come «significativo», nei confronti delle persone.

Un ulteriore profilo di interesse per la nostra analisi è costituito dall'articolato sistema di *compliance* in relazione ai sistemi di IA ad alto rischio.

In particolare, l'art. 9 prescrive un sistema di gestione dei rischi continuo, ovvero «come un processo iterativo continuo pianificato ed eseguito nel corso dell'intero ciclo di vita di un sistema di IA ad alto rischio [...]» (par. 2, primo

a essere utilizzati per la verifica biometrica la cui unica finalità è confermare che una determinata persona fisica è la persona che dice di essere;

b) i sistemi di IA destinati a essere utilizzati per la categorizzazione biometrica in base ad attributi o caratteristiche sensibili protetti basati sulla deduzione di tali attributi o caratteristiche;

c) i sistemi di IA destinati a essere utilizzati per il riconoscimento delle emozioni».

Fanno eccezione quei sistemi di identificazione biometrica remota in tempo reale che, invece, sono annoverati tra quelli vietati (art. 5, par. 1, lett. h) ss.).

¹⁹ In questi termini, G. MALGIERI, *Human vulnerability in the EU Artificial Intelligence Act*, in *Oxford University Press Blog*, 27 maggio 2024.

periodo)²⁰. E l'analisi del rischio si sviluppa sul binomio: uso improprio ragionevolmente prevedibile, da un lato, e accettabilità del rischio residuo, dall'altro.

Nell'effettuare tale analisi, precisa il par. 9, «i fornitori prestano attenzione [...] all'eventualità che il sistema di IA ad alto rischio possa avere un impatto negativo sulle persone di età inferiore a 18 anni o, a seconda dei casi, su altri gruppi vulnerabili».

Pertanto, il contrasto a precise situazioni discriminatorie assume importanza anche nell'ambito della *compliance*, cementificando così gli obblighi di controllo da parte degli attori principali che intervengono nel processo sia di emissione sia di utilizzazione dei sistemi ad alto rischio.

3. (Segue): le previsioni più esplicite della legge n. 132/2025.

Nell'ambito della regolamentazione dell'intelligenza artificiale, la legge italiana n. 132/2025 esplicita una maggiore attenzione ai rischi di discriminazione, con particolare riguardo a quella di genere, da parte di applicativi di IA.

In questa prospettiva, la legge n. 132/2025 può essere letta come una risposta normativa alla trasformazione tecnologica in atto, volta non solo a regolare lo sviluppo dell'intelligenza artificiale, integrando le previsioni del regolamento europeo e, allo stesso tempo, preservando i principi fondamentali del costituzionalismo contemporaneo.

Ancora una volta emerge come l'esigenza di introdurre norme di diritto positivo in materia di IA mira, da un lato, a governare il processo tecnologico mediante strumenti di controllo giuridico e, dall'altro, a scongiurare il rischio che l'autonomia operativa degli algoritmi comprometta conquiste giuridiche e sociali ormai radicate, tra cui la tutela della dignità, la parità di trattamento e il divieto di discriminazione, con particolare riferimento al genere.

Ne è prova l'art. 3 della legge, che inserisce esplicitamente i principi di "non discriminazione" e di "parità dei sessi" tra i criteri fondamentali che devono orientare ogni fase della progettazione, sperimentazione e applicazione dell'intelligenza artificiale.

La positivizzazione di tali principi assume il valore di una barriera costituzionale che impedisce all'innovazione tecnologica di divenire veicolo di regressione sociale.

Da questo punto di vista, la disciplina dell'intelligenza artificiale appare

²⁰ Sui punti di forza e sulle problematiche che potranno sorgere in ambito applicativo in relazione al sistema basato sul rischio, si v. di C. NOVELLI, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in *federalismi.it*, 2, 2024, 95 ss.

come un banco di prova della capacità dell'ordinamento giuridico di adattarsi a una realtà digitale in rapido mutamento, mantenendo al contempo saldo l'orizzonte assiologico della Carta costituzionale.

La previsione contenuta nell'art. 11 della legge, secondo cui l'intelligenza artificiale non può essere impiegata nel lavoro in modo contrario alla dignità umana, né con modalità discriminatorie fondate sul sesso, sull'età o sulle condizioni personali e sociali, mostra come il legislatore abbia inteso affermare che la tutela della persona prevale sull'efficienza algoritmica.

A ciò si aggiunga, l'art. 24 della legge che, nel prevedere percorsi di formazione e alfabetizzazione anche nelle discipline STEM, mira a incidere sulle radici materiali delle discriminazioni di genere, contrastando la storica sotto-rappresentazione femminile nei settori scientifici e tecnologici e aprendo la strada a un'interpretazione dinamica del principio di uguaglianza.

La legge, in quest'ottica, contribuisce a delineare un modello di protezione multilivello: costituzionale, legislativo e sovranazionale. Tale stratificazione della tutela giuridica permette di affrontare il fenomeno dell'IA non come mera questione tecnica, ma come nodo cruciale del costituzionalismo contemporaneo, nel quale la difesa dei diritti fondamentali, e tra essi la parità tra uomo e donna, costituisce il parametro essenziale per misurare la legittimità dell'innovazione tecnologica. Infatti, è proprio attraverso l'intreccio tra costituzione nazionale, norme europee e vincoli sovranazionali che si delineano le coordinate giuridiche necessarie per comprendere l'impatto dell'intelligenza artificiale sui diritti e per scongiurare l'emergere di nuove forme di discriminazione. La legge n. 132/2025, pur nella sua natura nazionale, indica chiaramente che il governo dell'IA, specie con riguardo alle discriminazioni di genere, non può che fondarsi su un paradigma di tutela integrato e multilivello, capace di rendere effettivi i principi democratici nello spazio digitale.

4. Luci e ombre dello strumento extra-giudiziale della valutazione di impatto ai sensi dell'art. 27 dell'AI Act.

Come già anticipato, la tutela dei diritti assume un ruolo centrale nell'impianto generale dell'approccio europeo. A tal proposito, l'art. 27 dell'AI Act introduce un meccanismo specifico di tutela preventiva volto a mitigare i rischi derivanti dall'impiego di sistemi di intelligenza artificiale ad alto rischio: la valutazione di impatto sui diritti fondamentali²¹. Specificamente, si tratta

²¹ Per ulteriori approfondimenti, sia consentito il rinvio a M. D'AMICO, P. GAMBATESA, *La*

di uno strumento extragiudiziale, pensato per prevenire possibili violazioni prima che queste si traducano in danni concreti e, dunque, prima che l'intervento della magistratura si renda necessario. L'obiettivo è quello di spostare il baricentro della protezione giuridica dalla fase contenziosa a quella di monitoraggio e controllo preventivo, evitando che le conseguenze negative dell'IA si manifestino e richiedano, *ex post*, un intervento riparatorio spesso tardivo o inefficace.

La valutazione di impatto impone ai *deployer* – ovvero ai soggetti che utilizzano sistemi di IA ad alto rischio²² – di condurre un'analisi dettagliata sui potenziali effetti negativi della tecnologia impiegata, con particolare attenzione ai soggetti vulnerabili²³. Il procedimento si sviluppa in più fasi: inizialmente, il *deployer* deve descrivere in appositi formulari il sistema di IA e specificare le finalità d'uso, il contesto applicativo e le categorie di individui coinvolti. Successivamente, deve procedere a un'analisi dei rischi, individuando i possibili effetti distorsivi sulle libertà individuali, sulla non discriminazione e sul diritto alla protezione dei dati personali. Una volta identificati i rischi, il soggetto responsabile è chiamato a individuare misure di mitigazione, che possono consistere nell'introduzione di controlli umani rafforzati, nella riduzione dell'opacità algoritmica o nella limitazione dell'uso dell'IA in determinati contesti sensibili. Una volta completata la valutazione, il *deployer* ha l'obbligo di notificare il documento all'autorità di vigilanza competente, la quale ha il potere di richiedere chiarimenti, imporre modifiche o, nei casi più critici, vietare l'impiego del sistema.

Trattandosi di un meccanismo di prevenzione extragiudiziale, sarà fondamentale verificare in che modo le autorità nazionali coordineranno la loro

valutazione d'impatto sui diritti fondamentali per i sistemi di IA ad alto rischio. Commento all'art. 27, in F. BALAGUER CALLÉJON, M. D'AMICO, A. AGUILAR CALAHORRO, P. GAMBATESA, S. DI GIOVANNI (a cura di), Commentario tematico del Regolamento (UE) 2024/1689 e della legge italiana 23 settembre 2025, n. 132, Torino, 2026, in corso di pubblicazione.

²² Esprime perplessità sull'ambito soggettivo delineato all'art. 27 dell'*AI Act*, F. DONATI, *La protezione dei diritti fondamentali nel Regolamento sull'intelligenza artificiale*, in *Rivista AIC*, 1, 2025, 1 ss. L'A. (17) osserva che “[s]e il sistema di verifica ex ante previsto dall'*AI Act* è ritenuto non sufficiente ad offrire una adeguata tutela dei diritti fondamentali contro i rischi derivanti dall'impiego di sistemi di IA ad alto rischio, infatti, l'obbligo di procedere alla FRIA andrebbe imposto a tutti gli utilizzatori di tali sistemi, non invece ad una limitata categoria di essi”.

²³ Sul tema sia consentito il rinvio a P. GAMBATESA, *Alla ricerca degli “anticorpi” contro le discriminazioni di genere nell'*AI Act**, in M. TOMASI, L. BUSATTA, M. FASAN, C. NARDOCCI, S. PENASA, S. SULMICELLI (a cura di), *Vulnerabilità e Intelligenza Artificiale*, in *BioLaw Journal*, fasc. monografico 1, 2024, 253 ss.

azione di supervisione. Il regolamento lascia un ampio margine di discrezionalità agli Stati membri nell'individuare le autorità competenti incaricate di ricevere, valutare e monitorare le valutazioni di impatto²⁴. Tuttavia, proprio questa discrezionalità potrebbe tradursi in un'applicazione disomogenea, con il rischio che gli *standard* di tutela varino significativamente da uno Stato all'altro. In particolare, la creazione di un sistema efficace di coordinamento tra le diverse autorità nazionali sarà cruciale per evitare disparità applicative e garantire che la protezione dei diritti fondamentali sia effettiva su tutto il territorio dell'Unione Europea. L'assenza di un chiaro meccanismo di raccordo potrebbe portare a differenze interpretative, con alcuni Stati che potrebbero adottare un approccio più restrittivo e altri più permissivo, minando così l'uniformità del quadro regolatorio.

Un ulteriore profilo critico riguarda la capacità delle autorità di vigilanza di esercitare un controllo efficace e tempestivo. La valutazione di impatto è concepita come un documento dinamico, che deve essere aggiornato periodicamente qualora emergano nuove evidenze o modifiche nel funzionamento del sistema di IA. Tuttavia, il regolamento non chiarisce in che misura le autorità nazionali saranno dotate delle risorse necessarie per esaminare con continuità queste valutazioni, verificandone l'effettività e l'adeguatezza. Senza un'adeguata dotazione di mezzi e personale specializzato, il rischio è che il meccanismo di valutazione di impatto resti un mero adempimento formale, privo di reali conseguenze operative²⁵.

Nel complesso, la valutazione di impatto dell'AI Act rappresenta un primo tentativo di costruire un sistema di *governance* preventiva sull'intelligenza artificiale, riducendo il ricorso al contenzioso giudiziario ed eventualmente

²⁴ In Italia, ad esempio, il ruolo di autorità nazionale di notifica sarà ricoperto dall'Agenzia per l'Italia Digitale (AgID) e dall'Agenzia per la Cybersicurezza Nazionale (ACN).

²⁵ Tale strumento può essere confrontato con la valutazione di impatto sulla protezione dei dati (DPIA) prevista dall'art. 35 del GDPR (regolamento UE 2016/679). In primo luogo, la finalità delle due valutazioni non coincide: mentre la DPIA è finalizzata esclusivamente a garantire la tutela dei dati personali, la valutazione di impatto dell'*AI Act* ha un ambito di applicazione più ampio, mirato a individuare qualsiasi rischio per i diritti fondamentali, anche al di là della protezione dei dati. Inoltre, l'obbligo di notifica rappresenta un altro elemento di distinzione: nel GDPR, la DPIA non deve essere obbligatoriamente trasmessa alle autorità di controllo, salvo nei casi in cui il rischio residuo sia particolarmente elevato. Al contrario, nell'*AI Act*, la valutazione di impatto deve essere sempre notificata, rendendo il controllo pubblico un elemento strutturale della regolamentazione. Infine, la normativa sull'IA prevede poteri di intervento più incisivi per le autorità di vigilanza, che possono arrivare a vietare direttamente l'uso di un sistema qualora la valutazione di impatto evidenzii profili di rischio inaccettabili.

affidando il controllo a meccanismi di vigilanza amministrativa. Tuttavia, la sua effettiva efficacia dipenderà dalla capacità degli Stati membri di garantire un'applicazione uniforme e un coordinamento efficace tra le diverse autorità nazionali. Se questo non dovesse avvenire, il rischio è che la protezione dei diritti fondamentali dipenda dal grado di efficienza del singolo Stato, compromettendo l'obiettivo di creare un sistema di regolamentazione equo e uniforme a livello europeo. In questa prospettiva, il coordinamento tra gli organi nazionali e le istituzioni europee sarà una delle principali sfide per garantire che l'IA non diventi un fattore di disuguaglianza e di arbitrio algoritmico, ma uno strumento capace di integrarsi nel tessuto giuridico senza ledere i principi di uguaglianza, trasparenza e tutela effettiva dei diritti fondamentali.

5. Le possibili prospettive di tutela giudiziale in un'ottica di integrazione dei livelli di protezione dei diritti.

Se il regolamento europeo, da un lato, si preoccupa di prevedere uno specifico strumento extragiudiziale per prevenire e sanzionare violazioni dei diritti fondamentali, dall'altro, si può affermare che esso sia del tutto silente sui possibili risvolti giudiziali delle suddette violazioni.

Occorre, pertanto, provare ad impostare il ragionamento in una prospettiva *de jure condendo* sulla necessità per cui, tanto le norme europee sull'IA, quanto le norme nazionali che rechino ulteriori dettagli alla disciplina europea, dovranno essere interpretate alla luce di *standard* integrati – costituzionali ed europei – di tutela dei diritti fondamentali.

In quest'ottica, ad avviso di chi scrive, un ruolo cruciale potrà essere svolto dalla Corte costituzionale italiana, che investita nei c.d. casi di doppia pregiudizialità potrà, se del caso previo rinvio alla Corte costituzionale, fornire le coordinate costituzionali ed eurounitarie circa l'applicazione dei principi e diritti fondamentali.

In altri termini, tale approccio, collocandosi nell'alveo del diritto costituzionale europeo, mira a stabilire una forte e stabile connessione tra le norme costituzionali espressive di diritti fondamentali e quelle europee, specificamente quelle della Carta dei diritti fondamentali dell'Unione, attraverso il giudizio di costituzionalità sollecitato in via incidentale dai giudici comuni.

Vi è, quindi, modo di credere che questa giurisprudenza sul *come* determinare standard integrati possa fornire elementi utili per le future controversie nazionali, nel cui ambito si lamenterà la lesione di diritti fondamentali a danno di applicativi di IA.

Ma andiamo per gradi e tentiamo qui di ricostruire, sia pur in sintesi, le coordinate di tale filone della giurisprudenza costituzionale italiana.

La sentenza n. 269/2017 della Corte costituzionale italiana segna un momento cruciale nell'evoluzione del sistema di tutela dei diritti fondamentali, inaugurando una nuova fase giurisprudenziale che tiene conto del complesso dialogo tra l'ordinamento nazionale e quello dell'Unione Europea. Il cuore della riflessione riguarda la coesistenza e l'integrazione tra le garanzie previste dalla Costituzione italiana e quelle sancite dalla Carta dei diritti fondamentali dell'Unione Europea²⁶, resa giuridicamente vincolante dall'art. 6 del Trattato sull'Unione Europea nella versione *post-Lisbona*²⁷.

La Corte costituzionale riconosce che, con gli effetti giuridici vincolanti della Carta, si è assistito a un'espansione del sistema di protezione dei diritti fondamentali, che ora opera su un doppio binario: quello nazionale e quello europeo. Di conseguenza, i giudici ordinari si trovano, oggi, nella posizione di poter scegliere tra tre strumenti giuridici concorrenti: la rimessione alla Corte costituzionale o alla Corte di Giustizia oppure la disapplicazione diretta della norma nazionale in contrasto con una disposizione europea dotata di effetto diretto. Questo potere di scelta, tuttavia, comporta un rischio: che la disapplicazione possa trasformarsi in una forma impropria di controllo di costituzionalità "difuso", in contrasto con il principio di accentramento del giudizio costituzionale previsto dal sistema italiano.

Per prevenire tale rischio, la Corte ha introdotto un'*eccezione*²⁸ al tradizionale

²⁶ Per un inquadramento dogmatico in chiave costituzionale, v. A. AGUILAR CALAHORRO, *Naturaleza y eficacia de la Carta de Derechos Fundamentales de la Union Europea*, Madrid, 2021.

²⁷ Sulla c.d. trattatizzazione del Carta si v. M. D'AMICO, *Trattato di Lisbona: principi, diritti e "tono costituzionale"*, in M. D'AMICO, P. BILANCIA (a cura di), *La nuova Europa dopo il Trattato di Lisbona*, Milano, 2009, 70.

²⁸ Tra le numerose tesi sviluppate dalla dottrina a commento dell'*obiter dictum* della sentenza n. 269/2017, per tutti, si v. M. D'AMICO, *L'Europa dei diritti: tra "aperture" e "chiusure"*, in A. PÉREZ MIRAS, G.M. TERUEL LOZANO, E. RAFFIOTTA, M.P. IADICICCO (a cura di), *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Madrid, 2020, 403 ss., l'A. ne evidenzia l'"effetto dirompente", segnalando anche che «la nostra Corte costituzionale non pare mossa dalla volontà di "chiudere" il nostro ordinamento ai diritti protetti a livello sovranazionale, ma solo dalla volontà di non essere estromessa dai relativi giudizi, e allo stesso tempo dalla volontà di valorizzare la Costituzione italiana come fonte di tutela dei diritti fondamentali»; collocano la "svolta", operata dalla pronuncia n. 269 nell'ambito più generale "declino dell'incidentalità", R. ROMBOLI, *Il sistema di costituzionalità tra "momenti di accentramento" e "di diffusione"*, in *Riv. "Gruppo di Pisa"*, 2, 2020, 1 ss.; G. SCACCIA, *L'inversione della "doppia pregiudiziale" nella sentenza della Corte costituzionale n. 269 del 2017: presupposti teorici*

schema interpretativo derivante dalle sentenze *Simmenthal* e *Granital*, che legittimava il giudice ordinario a disapplicare norme interne in contrasto con il diritto europeo direttamente efficace. Con la nuova giurisprudenza viene affermata la possibilità – e in taluni casi la preferenza – per il giudice comune di sollevare la questione di legittimità costituzionale anche quando si tratta di norme nazionali che potrebbero essere disapplicate sulla base del diritto europeo. In questo modo, la Corte costituzionale rivendica un ruolo primario nel garantire l'interpretazione e l'applicazione coordinata dei diritti fondamentali, contribuendo alla costruzione di un dialogo costruttivo e leale con la Corte di Giustizia dell'Unione Europea.

Questo meccanismo mira a evitare decisioni frammentarie e disomogenee da parte dei giudici ordinari, assicurando l'uniformità e l'efficacia *erga omnes* delle decisioni in materia di diritti fondamentali²⁹. La Corte, però, precisa che tale percorso non esclude in alcun modo la possibilità, per il giudice ordinario, di effettuare un rinvio pregiudiziale alla Corte di Giustizia, anzi, ribadisce il valore aggiunto di questo strumento, che arricchisce il sistema di tutele e ne amplia l'efficacia, senza restringere i rimedi disponibili.

La Corte sottolinea, inoltre, che l'esame di costituzionalità può e deve avvenire tenendo conto anche dei parametri europei, in particolare degli artt. 11 e 117 Cost., i quali vincolano l'ordinamento italiano al rispetto degli obblighi comunitari e internazionali. Ciò consente una lettura sistemica e coerente dei diritti, volta a garantire che le disposizioni della Carta siano interpretate in armonia con le tradizioni costituzionali comuni agli Stati membri, come previsto anche dall'art. 53 della stessa.

Un aspetto rilevante emerso nelle pronunce successive, come le sentenze nn. 20 e 63/2019, è la riaffermazione del potere del giudice comune di procedere alla disapplicazione anche dopo un eventuale rigetto della questione da parte della Corte costituzionale. Questo conferma che, pur promuovendo un ruolo attivo del Giudice delle leggi, l'ordinamento mantiene intatta la flessibilità dei

e problemi applicativi, in *Forum Quad. cost.*, 25 gennaio 2018, 4 ss. Tra le tesi di segno contrario sono state sviluppate, per tutti, si v. L.S. ROSSI, *La sentenza 26972017 della Corte costituzionale italiana: obiter "creativi" (o distruttivi?) sul ruolo dei giudici italiani di fronte al diritto dell'Unione europea*, in *federalismi.it*, 3, 2018, 2 ss.; R. MASTROIANNI, *Da Taricco a Bolognesi, passando per la ceramica Sant'Agostino: il difficile cammino verso una nuova sistemazione del rapporto tra Carte e Corti*, in *Osservatorio sulle fonti*, 1, 2018, 2 ss.; D. GALLO, *Efficacia diretta del diritto UE, procedimento pregiudiziale e Corte costituzionale: una lettura congiunta delle sentenze n. 269/2017 e 115/2018*, in *Riv. AIC*, 1, 2019, 159 ss.

²⁹ Corte cost., sentenza 12 febbraio 2024, n. 15.

rimedi giudiziali, tutelando in modo effettivo i diritti fondamentali in un sistema multilivello.

Da ultimo, con la sentenza n. 181/2024³⁰, la Corte ha ulteriormente esteso le maglie del suo giudizio. In particolare, la c.d. regola 269 subisce una notevole trasformazione, generata dall'introduzione del "tono costituzionale" della questione, che fa leva potenzialmente su tutte le disposizioni UE che esibiscano effetti diretti. Così la cognizione delle questioni di legittimità costituzionale si amplia e con essa anche il novero dei casi in cui il giudice potrà sollevare questione incidentale dinnanzi alla Corte costituzionale, in luogo di procedere immediatamente alla disapplicazione.

In sintesi, questo nuovo filone rappresenta una tappa fondamentale per rafforzare il dialogo tra i sistemi di giustizia costituzionale ed eurolunitaria, promuovendo una tutela integrata, efficace e coordinata dei diritti fondamentali. E vi è modo di credere che tale terreno potrà essere fertile per la risoluzione delle c.d. discriminazioni algoritmiche. Infatti, nei futuri contenziosi che si instaureranno dinnanzi ai giudici comuni, l'eventuale incidente di costituzionalità sarà fondamentale per stabilire (eventualmente previo rinvio alla Corte di Giustizia) lo standard comune di tutela dei diritti fondamentali cui gli applicativi di IA dovranno conformarsi.

6. *Considerazioni conclusive.*

Le trasformazioni imposte dall'intelligenza artificiale ai sistemi costituzionali non possono essere considerate un semplice problema di regolazione tecnica o di governance tecnologica; esse interrogano, più radicalmente, la capacità degli ordinamenti democratici di preservare l'effettività dei principi su cui si fondano. La rivoluzione digitale in corso, se non adeguatamente governata, rischia di produrre effetti distorsivi e strutturali in grado di compromettere la protezione dei diritti fondamentali, in particolare laddove emergano fenomeni di discriminazione algoritmica nei confronti delle donne e dei soggetti più vulnerabili. La recente legge italiana n. 132/2025 si colloca proprio in tale orizzonte,

³⁰ Cfr. N. ZANON, *Il ruolo della Corte costituzionale nella difesa del principio di uguaglianza e degli obblighi europei ex art. 117, primo comma, Cost.*, in *federalismi.it*, n. 30, 2024, p. IV ss.; S. SCIARRA, *Antonio La Pergola e l'Europa del diritto: a partire dalla sentenza Granital*, in *Giurisprudenza costituzionale*, n. 1, 2025, 213 ss.; nella medesima rivista v. anche E. Malfatti, *Per un tentativo di lettura "conforme" degli sviluppi giurisprudenziali recenti ai paradigmi della sentenza La Pergola*, n. 1, 2025, 213 ss.; e S. LEONE, *Dalla certezza del diritto, e di altri beni di rilievo costituzionale, da Granital ad oggi*, 2024, n. 6, 2024, 2739 ss.

mettendo in evidenza – in modo più netto rispetto ad altri ordinamenti europei – l’urgenza di contrastare le discriminazioni di genere generate o amplificate dai sistemi di IA, sancendo tra i principi generali quelli di “non discriminazione” e di “parità dei sessi” (art. 3), nonché vietando esplicitamente ogni impiego dell’intelligenza artificiale che possa produrre effetti discriminatori nel mondo del lavoro (art. 11).

L’IA può certamente diventare strumento di equità sostanziale, efficienza e partecipazione democratica; tuttavia, ciò richiede una chiara volontà normativa e un controllo costante sugli algoritmi che ne definiscono l’operatività. La legge n. 132/2025 mostra che il diritto non può restare spettatore passivo di fronte a un potere tecnologico capace di incidere sulle dinamiche sociali e decisionali: esso deve assumere una funzione attiva di indirizzo e vigilanza, ponendo limiti chiari alla possibilità che l’innovazione digitale riproduca – sotto nuove forme – i modelli storici di esclusione, in particolare nei confronti delle donne³¹. In tale prospettiva, non basta enunciare un generico principio di neutralità tecnologica; occorre invece garantire, mediante controlli *ex ante* e verifiche sull’impatto algoritmico, che l’IA non consolidi forme di discriminazione strutturale, già presenti nei dati e negli schemi decisionali da cui essa trae alimento.

In questo contesto, emerge con forza l’esigenza di costruire un costituzionalismo digitale capace di riconoscere che la parità di genere non è solo una questione di tutela formale, ma rappresenta un parametro sostanziale di legittimazione dell’innovazione tecnologica. La legge n. 132/2025 offre un modello normativo che, pur ancora perfettibile, tenta di integrare i principi del diritto costituzionale nell’architettura dell’IA, promuovendo una visione orientata all’inclusione e alla giustizia sociale. Tale direzione deve essere ulteriormente rafforzata attraverso strumenti di formazione, alfabetizzazione digitale e controllo democratico, affinché i cittadini – e in particolare le donne – non diventino meri destinatari passivi di decisioni automatizzate, ma possano partecipare consapevolmente alla definizione del loro ambiente tecnologico.

In definitiva, è necessario che oggi nel riaffermare la capacità del diritto di orientare l’innovazione, si preservino le conquiste sociali e si ponga al centro il tema del contrasto alle nuove forme di discriminazione, *in primis* quelle di genere. La sfida non si esaurisce nella predisposizione di un sistema di regole tecniche, ma richiede un rinnovato patto sociale in cui la dignità, l’uguaglianza e la parità tra uomo e donna costituiscano i fondamenti irrinunciabili della

³¹ M. D’AMICO, *Le nuove frontiere delle discriminazioni di genere: l’impatto dell’intelligenza artificiale*, in questo volume.

società digitale che si va costruendo. Solo un ecosistema normativo capace di integrare tali valori potrà garantire che l'intelligenza artificiale sia realmente uno strumento di emancipazione e non una nuova forma di subordinazione.

CAPITOLO XIII

L'UTILIZZO DI SISTEMI DI IDENTIFICAZIONE
BIOMETRICA REMOTA NEL REGOLAMENTO
SULL'INTELLIGENZA ARTIFICIALE
TRA MECCANISMO DI AUTORIZZAZIONE
E MECCANISMO DI SORVEGLIANZA

Giulia Agrati

SOMMARIO: 1. Premessa. – 2. Introduzione: il divieto di utilizzo di sistemi di identificazione biometrica remota in “tempo reale” in spazi accessibili al pubblico e a fini di attività di contrasto alla criminalità e le rispettive eccezioni. – 3. Il meccanismo di autorizzazione e il ruolo dell'autorità giudiziaria o autorità amministrativa indipendente. – 4. Il meccanismo di controllo e il ruolo dell'autorità di vigilanza del mercato e dell'autorità nazionale per la protezione dei dati. – 5. Conclusioni.

1. *Premessa.*

La regolamentazione dell'intelligenza artificiale è percepita oggi come una delle sfide politiche e normative più rilevanti nel contesto europeo (e non solo).

Dopo anni di discussioni, il regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 ha introdotto, a livello comunitario¹, una cornice giuridica comune per lo sviluppo, la diffusione e l'uso dei sistemi di intelligenza artificiale.

Il legislatore comunitario, come ben noto, ha inteso garantire, attraverso un

¹ Nella consapevolezza che, a seguito del Trattato di Lisbona, l'Unione Europea abbia sostituito e sia succeduta ufficialmente alla Comunità europea (cfr. art. 1 TUE), nel presente contributo si è scelto di fare ancora uso del termine “comunitario”. Quest'ultimo, infatti, non è scomparso dal linguaggio comune e, questo, per la sua forte valenza “identitaria” che lo rende difficilmente sostituibile.

approccio basato sul rischio, un equilibrio tra promozione dell'innovazione tecnologica e tutela dei diritti fondamentali, delineando un quadro regolatorio direttamente applicabile negli ordinamenti nazionali.

Emblematica dell'approccio del legislatore comunitario alla materia è la regolamentazione dell'utilizzo di sistemi di identificazione biometrica remota in "tempo reale" in spazi accessibili al pubblico e a fini di attività di contrasto alla criminalità².

Secondo il regolamento, tale pratica è generalmente vietata, in considerazione del rischio inaccettabile di lesione dei diritti fondamentali che da essa deriverebbe.

Il divieto si basa, infatti, sul rischio di sorveglianza di massa che un libero utilizzo di tali sistemi potrebbe comportare, dal momento che questi ultimi consentirebbero alle autorità di polizia di identificare individui a distanza, in tempo reale e senza il loro consenso, confrontando immagini "dal vivo" con dati biometrici archiviati.

Ad ogni modo, proprio al fine di contemperare gli interessi sottesi alla regolamentazione, il legislatore comunitario ha previsto specifiche eccezioni a tale divieto, tassativamente individuate all'art. 5, par. 1, lett. h) del regolamento. In tali ipotesi, l'uso (consentito) dei sistemi di identificazione biometrica remota è strettamente necessario al perseguimento di un rilevante interesse pubblico, in particolare quello alla "sicurezza pubblica".

Ebbene, il presente contributo intende analizzare il quadro normativo e procedimentale sottostante all'utilizzo dei sistemi di identificazione biometrica remota, proprio quale espressione dell'approccio basato sul rischio adottato dal legislatore comunitario nella regolamentazione dell'intelligenza artificiale.

Tale quadro si fonda su di un duplice meccanismo: (i) un sistema di autorizzazione preventiva, rilasciata, caso per caso, da parte di un'autorità giudiziaria o amministrativa indipendente; (ii) un sistema di controllo e vigilanza generale, affidato all'autorità di vigilanza del mercato, all'autorità nazionale per la protezione dei dati e alla Commissione europea.

² Per consentire una maggior agilità nella lettura del commento, nel prosieguo ci si riferirà ai «sistemi di identificazione biometrica remota in "tempo reale" in spazi accessibili al pubblico e a fini di attività di contrasto alla criminalità», utilizzando soltanto l'abbreviazione "sistemi di identificazione biometrica remota".

2. Introduzione: il divieto di utilizzo di sistemi di identificazione biometrica remota in “tempo reale” in spazi accessibili al pubblico e a fini di attività di contrasto alla criminalità e le rispettive eccezioni.

Come anticipato, l'uso di sistemi di identificazione biometrica remota in “tempo reale” in spazi accessibili al pubblico e a fini di attività di contrasto alla criminalità costituisce una pratica di intelligenza artificiale generalmente vietata, in quanto connotata da un rischio inaccettabile di lesione dei diritti fondamentali dei soggetti interessati³. Confrontando immagini “dal vivo” (assunte, ad esempio, per mezzo di telecamere) e dati biometrici archiviati, le autorità di polizia sarebbero infatti in grado di identificare individui senza il loro consenso, a distanza e in tempo reale, potendo così attuare forme inaccettabili di sorveglianza di massa⁴.

L'uso di tali sistemi è, pertanto, generalmente vietato, salvo alcune ipotesi specifiche «elencate in modo esaustivo e definite rigorosamente, nelle quali l'uso è strettamente necessario per perseguire un interesse pubblico rilevante, la cui importanza prevale sui rischi»⁵.

In queste ipotesi, in cui ne è ammesso l'utilizzo, i sistemi di identificazione biometrica remota sono qualificati come “ad alto rischio” (cfr. n. 1, lett. a) dell'allegato III al regolamento europeo) e, pertanto, soggetti all'espletamento della valutazione d'impatto di cui all'art. 27 del regolamento europeo⁶.

Le eccezioni al divieto sono elencate al par. 1, lett. h), dell'art. 5 del regolamento europeo e riguardano casi in cui l'uso di questi sistemi è strettamente connesso con il perseguimento di un interesse pubblico. In particolare, l'uso di sistemi di identificazione biometrica remota è consentito esclusivamente per finalità che attengono alla “sicurezza pubblica”, ossia: (i) la ricerca di persone

³ Cfr. art. 5, par. 1, lett. h) del regolamento europeo.

⁴ Tale preoccupazione è rilevata dallo stesso regolamento europeo, secondo cui l'uso di sistemi di identificazione biometrica remota «è particolarmente invasivo dei diritti e delle libertà delle persone interessate, nella misura in cui potrebbe avere ripercussioni sulla vita privata di un'ampia fetta della popolazione, farla sentire costantemente sotto sorveglianza e scoraggiare in maniera indiretta l'esercizio della libertà di riunione e di altri diritti fondamentali» (considerando 32).

⁵ Considerando 33 del regolamento europeo.

⁶ A mente dell'art. 27 del regolamento europeo, di fronte a un sistema di intelligenza artificiale ad alto rischio è obbligo dei *deployer*, che sono organismi di diritto pubblico o enti privati che forniscono servizi pubblici, effettuare una valutazione dell'impatto sui diritti fondamentali che l'uso di tale sistema potrebbe avere. Ciò allo scopo di adottare eventuali misure di mitigazione per ridurre il rischio di lesione dei diritti prima che il sistema venga effettivamente utilizzato.

scomparsa o vittime di reati gravi; (ii) la prevenzione di minacce imminenti alla vita; e (iii) l'identificazione di sospetti o condannati per reati gravi.

Posta tale macro-cornice di deroghe a livello comunitario, è però lasciata a ciascuno Stato membro la decisione se consentire o meno l'utilizzo di siffatti sistemi all'interno dei rispettivi ordinamenti nazionali, potendo sempre gli Stati membri introdurre "disposizioni più restrittive" rispetto a quelle contemplate dal regolamento⁷.

Come già osservato in dottrina⁸, il regolamento europeo affronta il tema dell'uso dei sistemi di identificazione biometrica remota introducendo una proibizione che si traduce, essenzialmente, «nella formulazione di una base giuridica che ne consente l'uso», lasciando però agli Stati membri l'onere di "implementare" o meno la deroga a livello nazionale.

Ciò non senza criticità, delle quali si tratterà appresso (cfr. *infra* par. 3).

Considerata la delicatezza della materia, l'utilizzo di tali sistemi – che devono essere stati oggetto della valutazione d'impatto di cui all'art. 27 del regolamento europeo – non è soltanto limitato a ipotesi specifiche, ma è anche contenuto all'interno di una cornice procedimentale ben precisa.

I parr. 3-8 dell'art. 5 del regolamento prevedono che l'utilizzo di siffatti sistemi sia soggetto a un meccanismo di autorizzazione (da parte di un'autorità giudiziaria o di un'autorità amministrativa indipendente) (cfr. *infra* par. 3) e, parallelamente, a un sistema di controllo (da parte dell'autorità di vigilanza del mercato e dell'autorità nazionale per la protezione dei dati) (cfr. *infra* par. 4). Nella fase di vigilanza il regolamento europeo prevede anche un ruolo della Commissione, unica autorità, tra quelle coinvolte, in grado di avere uno sguardo complessivo, a livello comunitario, sull'uso dei sistemi di identificazione biometrica remota.

⁷ Art. 5, par. 5, del regolamento europeo. Cfr. anche considerando 37, secondo cui l'uso di sistemi di identificazione biometrica remota nel territorio di uno Stato membro è possibile «solo nel caso e nella misura in cui lo Stato membro interessato abbia deciso di prevedere espressamente la possibilità di autorizzare tale uso nelle regole dettagliate del proprio diritto nazionale». Sempre secondo il considerando in esame, «gli Stati membri restano [...] liberi [...] di non prevedere affatto tale possibilità o di prevederla soltanto per alcuni degli obiettivi idonei a giustificare l'uso autorizzato di cui al presente regolamento».

⁸ F.P. LEVANTINO, I. NERONI REZENDE, *Rischio inaccettabile: usi proibiti*, in F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, O. POLLICINO (a cura di), *La disciplina dell'intelligenza artificiale*, Milano, 2025, 196.

3. *Il meccanismo di autorizzazione e il ruolo dell'autorità giudiziaria o autorità amministrativa indipendente.*

Al fine di scongiurare qualsiasi automatismo che potrebbe ledere diritti e libertà fondamentali, il regolamento subordina l'utilizzo dei sistemi di identificazione biometrica remota a «un'autorizzazione preventiva rilasciata [a livello nazionale] da un'autorità giudiziaria o da un'autorità amministrativa indipendente».

Non è sufficiente, pertanto, che un dato utilizzo del sistema di cui trattasi sia consentito, in termini generali, a livello normativo, ma è necessario che tale utilizzo venga autorizzato in concreto, caso per caso, tramite decisione giudiziale o provvedimento esecutivo.

L'art. 5, par. 3, del regolamento europeo fornisce alcune indicazioni di principio circa il meccanismo di autorizzazione:

- l'autorizzazione deve essere preventiva⁹;
- competente al rilascio deve essere un'autorità giudiziaria o un'autorità amministrativa indipendente dello Stato membro in cui deve avvenire l'uso dei sistemi di identificazione biometrica (in proposito, cfr. anche considerando 35);
- il rilascio deve avvenire previa domanda di parte e sulla base di prove oggettive o indicazioni chiare fornite dall'istante;
- il rilascio deve avvenire con decisione vincolante;
- il rilascio deve essere subordinato all'accertamento dei requisiti di necessità e proporzionalità dell'utilizzo dei sistemi in parola rispetto al conseguimento di uno degli obiettivi di sicurezza pubblica di cui alle ipotesi derogatorie dell'art. 5, par. 1, lett. h);
- il rilascio deve essere limitato a quanto strettamente necessario in termini di tempo, spazio e persona per il conseguimento di uno degli obiettivi menzionati.

Un ultimo principio sancito a livello comunitario è quello per cui «nessuna decisione che produca effetti giuridici negativi su una persona può essere presa

⁹ In via di eccezione, l'art. 5, par. 3, del regolamento europeo prevede che «in una situazione di urgenza debitamente giustificata» sia possibile «iniziare a utilizzare il sistema senza autorizzazione, a condizione che tale autorizzazione sia richiesta senza indebito ritardo, al più tardi entro 24 ore». Se la richiesta di autorizzazione è respinta, «l'uso è interrotto con effetto immediato e tutti i dati nonché i risultati e gli output di tale uso sono immediatamente eliminati e cancellati».

unicamente sulla base dell'*output* del sistema di identificazione biometrica remota "in tempo reale"»¹⁰, declinazione nel caso di specie del trasversale approccio antropocentrico ai sistemi di intelligenza artificiale del regolamento europeo.

Poste tali indicazioni, la disciplina dettagliata del procedimento di autorizzazione (inerente a richiesta, rilascio ed esercizio dell'autorizzazione) è demandata, a mente del par. 5, ai singoli Stati membri.

Emerge fin da subito, pertanto, come gli Stati membri abbiano un ampio margine di discrezionalità nella regolamentazione dei casi e delle modalità di utilizzo dei sistemi di identificazione biometrica remota.

E ciò sebbene i sistemi in questione trattino una delle tecnologie più delicate e problematiche dal punto di vista del bilanciamento tra esigenze di sicurezza pubblica e tutela dei diritti e delle libertà fondamentali dell'individuo.

In tale ambito, come meglio si dirà, l'approccio del legislatore comunitario, particolarmente rispettoso delle sovranità nazionali, pare perfino allontanarsi dalla finalità di unificazione delle discipline dei singoli Stati membri propria del "regolamento", per avvicinarsi, invece, a un'ottica di armonizzazione propria della "direttiva".

Le previsioni contenute nei paragrafi in esame, pur trattando un tema cruciale nell'ambito della regolamentazione dell'intelligenza artificiale, non paiono infatti immediatamente applicabili, così come l'intervento richiesto allo Stato non sembra soltanto di carattere esecutivo. Anzi, agli ordinamenti nazionali pare essere richiesta una vera e propria attività di normazione, nel momento in cui è lasciata loro la scelta se consentire, in quali specifiche ipotesi e con quali modalità, l'utilizzo dei sistemi di identificazione biometrica remota.

In tale prospettiva, essenziale sarà il ruolo di regolatore attribuito alla Commissione, la quale, tramite l'elaborazione di orientamenti potrà fornire indicazioni omogenee per l'implementazione del regolamento europeo, al fine di evitare un'eccessiva (e non voluta) frammentazione della disciplina a livello nazionale.

L'attuale articolazione regolamentare non è, infatti, senza conseguenze sul piano della tutela offerta a diritti e libertà fondamentali da parte dei singoli ordinamenti nazionali: la discrezionalità conferita agli Stati membri si traduce infatti in un'inevitabile disparità di trattamento rispetto a situazioni tendenzialmente analoghe all'interno dello stesso territorio europeo, con ripercussioni circa lo *standard* di protezione dei diritti fondamentali associati all'utilizzo dei sistemi di identificazione biometrica remota.

La questione del margine di discrezionalità lasciato agli Stati membri e delle

¹⁰ Cfr. art. 5, par. 3, del regolamento europeo.

connesse ripercussioni sullo *standard* di tutela offerto ai cittadini comunitari è particolarmente evidente con riferimento alla possibilità di scegliere, a livello nazionale, se attribuire la competenza al rilascio dell'autorizzazione a un'autorità giudiziaria o a un'autorità amministrativa indipendente¹¹.

In effetti, quella che a un primo sguardo potrebbe sembrare una scelta meramente procedurale fra due autorità interscambiabili per i fini dell'attività di autorizzazione di cui trattasi, è in realtà una scelta sostanziale, che porta con sé importanti conseguenze circa il bilanciamento tra esigenze di sicurezza pubblica e tutela dei diritti fondamentali¹².

Il modello dell'autorità amministrativa indipendente si presta certamente meglio ad affrontare i tecnicismi propri del settore dell'intelligenza artificiale; quindi, nel caso di specie, a comprendere la complessità e le criticità tecniche dei sistemi di identificazione biometrica remota¹³. Ciò, anzitutto, in ragione della composizione di questo genere di autorità, che è espressione delle (migliori) professionalità tecniche nel settore di riferimento.

Inoltre, considerate le peculiarità del settore in esame, tale modello pare più idoneo a fornire risposte tempestive alle richieste di autorizzazione, adattandosi meglio alle esigenze operative dell'attività di contrasto alla criminalità.

D'altro canto, l'interpretazione e l'applicazione dei diritti fondamentali rientrano nella funzione propria della magistratura: di conseguenza, il modello dell'autorità giurisdizionale si rivela più adatto al delicato compito di bilanciamento di diritti ed interessi particolarmente sensibili, come quelli coinvolti nel funzionamento dei sistemi di identificazione biometrica remota. E, d'altronde, la delicatezza della materia emerge chiaramente anche dall'impostazione del regolamento europeo, che dedica particolare attenzione alla tutela dei diritti fondamentali di fronte ai diversi sistemi di intelligenza artificiale.

¹¹ Come anticipato, ai sensi dell'art. 5, par. 3, del regolamento europeo «ogni uso di un sistema di identificazione biometrica remota “in tempo reale” in spazi accessibili al pubblico a fini di attività di contrasto è subordinato a un'autorizzazione preventiva rilasciata da un'autorità giudiziaria o da un'autorità amministrativa indipendente».

¹² In merito alla scelta tra autorità cfr. F. PAOLUCCI, *The enforcement of the Artificial Intelligence Act: Looking Forward to a Commission Implementing Decisions for Protecting Fundamental Rights*, in F. DONATI, G. FINOCCHIARO, F. PAOLUCCI, O. POLLICINO (a cura di), *La disciplina dell'intelligenza artificiale*, cit., 356-364; F. PAOLUCCI, *Shortcoming of the AI Act: Evaluating the New Standards to Ensure the Effective Protection of Fundamental Rights*, in *VerfBlog*, 14 marzo 2024.

¹³ Sul modello dell'autorità amministrativa indipendente cfr. *inter alia* G. GIRAUDI, M.S. RIGHETTINI, *Le autorità amministrative indipendenti. Dalla democrazia della rappresentanza alla democrazia dell'efficienza*, Roma, 2001; AA.VV., *Regolazione e garanzia del pluralismo. Le autorità amministrative indipendenti*, Milano, 1997.

Si noti che, sotto il profilo del bilanciamento dei diritti fondamentali, la principale differenza tra le due tipologie di autorità risiede nell'assunto di base per cui l'una è espressione del potere esecutivo ¹⁴, mentre l'altra di quello giurisdizionale.

Sebbene il modello dell'autorità amministrativa indipendente si caratterizzi per l'assenza di subordinazione al controllo politico, tale autorità è infatti soggetta a requisiti di indipendenza meno stringenti rispetto a quelli a cui è tenuto un organo giurisdizionale: l'autorità amministrativa non è, infatti, obbligata esclusivamente all'applicazione della legge, in un'ottica di tutela dei diritti del singolo, ma può operare delle scelte anche in una prospettiva, più discrezionale, di cura dell'interesse collettivo. Da questo punto di vista, il modello più flessibile dell'autorità amministrativa indipendente parrebbe adattarsi poco alle peculiarità della materia: allorché si tratti di ingerenze in diritti fondamentali, le ipotesi derogatorie devono operare nei limiti dello stretto necessario, con assoluta restrizione del margine di valutazione consentito all'autorità di contrasto e, dunque, all'autorità competente al rilascio dell'autorizzazione ¹⁵.

Inoltre, se l'indipendenza dell'organo giurisdizionale è sancita dalla Costituzione, in particolare, all'art. 101 che dispone che «i giudici sono soggetti soltanto alla legge», quella dell'autorità amministrativa indipendente è prevista esclusivamente a livello normativo, negli atti istitutivi delle singole autorità amministrative, senza un ancoraggio a garanzie costituzionali.

Aggiungasi che, generalmente, l'autorità amministrativa indipendente non svolge soltanto una funzione di controllo a carattere para-giurisdizionale, ma anche funzioni di vigilanza e regolamentazione. Ebbene, tale commistione di funzioni tende ad attenuare ulteriormente il grado di indipendenza di questo genere di autorità: l'attività di applicazione della norma, che avviene sia a livello esecutivo che a livello giurisdizionale, in questo caso è infatti svolta dal medesimo soggetto, chiamato sia a elaborare gli atti esecutivi, sia a vigilare sul loro rispetto.

Nell'esercizio dell'attività di controllo, le autorità amministrative indipendenti possono quindi trovarsi ad applicare al caso concreto le stesse disposizioni

¹⁴ Riconoscere a un determinato soggetto la caratteristica dell'indipendenza non significa necessariamente fare del medesimo un "giudice speciale". L'appartenenza delle autorità amministrative indipendenti all'"amministrazione statale", quindi al potere esecutivo, è stata più volte affermata dal giudice nazionale (cfr., da ultimo, TAR Lazio, 28 febbraio 2018, n. 2252) e dal giudice dell'Unione (cfr. Corte Giust. UE, 16 settembre 2020, causa C-462/19; Corte Giust. UE, 31 maggio 2005, causa C-53/03).

¹⁵ European Data Protection Board (EDPB), *Guidelines on the use of facial recognition technology in the area of law enforcement*, 26 aprile 2023, 21-22.

che hanno contribuito a elaborare in sede esecutiva, con potenziali ricadute sul principio di separazione dei poteri.

Nel caso di specie, la verifica dei presupposti di necessità e proporzionalità per il rilascio dell'autorizzazione potrebbe quindi essere affidata allo stesso soggetto che ha elaborato la disciplina del relativo procedimento di autorizzazione e che, più in generale, svolge un'attività di regolamentazione in materia.

Aggiungasi a tali considerazioni che, anche a livello procedimentale, i due modelli di autorità offrono un differente *standard* di garanzia dei diritti fondamentali: allo stato dell'arte, soltanto organi propriamente giurisdizionali sono infatti legittimati a rimettere una questione di costituzionalità alla Corte costituzionale¹⁶ e a sollevare un rinvio pregiudiziale dinanzi alla Corte di Giustizia dell'Unione Europea¹⁷.

In conclusione, è evidente che il margine di discrezionalità che il regolamento europeo lascia al singolo Stato membro circa l'uno o l'altro modello di autorità a cui conferire la competenza al rilascio dell'autorizzazione in questione

¹⁶Con sentenza n. 13/2019, la Corte costituzionale pare aver chiuso ogni prospettiva di apertura alla legittimazione incidentale delle autorità amministrative indipendenti, dichiarando l'inammissibilità della questione di costituzionalità ad essa rimessa dell'AGCM (con ordinanza del 3 maggio 2018). In proposito, *inter alia*, O. SPATARO, *Autorità amministrative indipendenti e controllo di costituzionalità: considerazioni a partire dalla sentenza n. 13 del 2019 della Corte costituzionale*, in *federalismi.it*, 30 dicembre 2019; M.C. MARZOCCA, *La Corte fa il punto sulla (non) legittimazione delle Autorità indipendenti come giudice a quo*, in *Riv. AIC*, 4 febbraio 2020; G. MAZZANTINI, M.C. MARZOCCA, *Legittimazione dell'Autorità quale giudice a quo di fronte alla Corte costituzionale e alla Corte di giustizia europea*, in *Osservatorio costituzionale*, 26 giugno 2018.

¹⁷Nel settore della concorrenza, dopo una prima apertura nei confronti della ricevibilità di un rinvio pregiudiziale proveniente dall'autorità spagnola garante della concorrenza e del mercato (Corte Giust., 16 luglio 1992, causa C-67/91), il giudice dell'Unione ha subito "invertito la rotta", respingendo, proprio escludendo la natura giurisdizionale dell'organo *a quo*, prima un rinvio pregiudiziale proposto dall'autorità antitrust greca (Corte Giust., 31 maggio 2005, causa C-53/03), poi un rinvio proveniente dall'autorità antitrust spagnola (Corte Giust., 16 settembre 2020, causa C-462/19). Nel settore dei trasporti, un'unica altra ipotesi di riconoscimento di un'autorità amministrativa indipendente quale "organo giurisdizionale" ai fini del rinvio pregiudiziale è quella operata in Corte Giust., 22 novembre 2012, causa C-136/11, in cui il giudice dell'Unione ha ritenuto ricevibile un rinvio presentato dall'autorità di regolazione dei trasporti austriaca. Tale approccio è stato però abbandonato molto in fretta dal giudice dell'Unione che è, da ultimo, ha dichiarato irricevibile il rinvio proposto dall'autorità di regolazione dei trasporti ceca (Corte Giust., 3 maggio 2022, causa C-453/20). In proposito, A. CIRCOLO, *"This is Out of my Jurisdiction". The Court of Justice Confirms that the Exercise of Administrative Functions does not Get Along with Preliminary References: the CityRail Case (C-453/20)*, in *EU Law Live*, 4 giugno 2022; V. CAPUANO, *Il rinvio pregiudiziale e la nozione di giurisdizione in cerca di nuovi confini: la sentenza CityRail*, in *BlogDUE*, 29 maggio 2022.

non è senza conseguenze, anzi contribuisce a delineare l'intensità della tutela che ogni ordinamento nazionale reputa di apprestare ai diritti fondamentali dinanzi ai sistemi di identificazione biometrica remota.

Come detto, quello descritto, è solo uno degli esempi da cui emerge l'ap-proccio del regolamento europeo alla disciplina dell'intelligenza artificiale, spesso caratterizzato dalla necessità di una spiccata attività di implementazione, attribuita agli Stati membri.

In questo contesto, cruciale è dunque il ruolo della Commissione, unica autorità che, tramite la propria attività di regolamentazione, potrebbe essere in grado di prevenire e risolvere i conflitti che dovessero insorgere tra le differenti discipline nazionali. In particolare, come osservato rispetto all'uso dei sistemi di identificazione biometrica remota, l'intervento della Commissione si preannuncia fondamentale nel fornire orientamenti alle autorità nazionali rispetto alla delicata attività di bilanciamento tra esigenze di sicurezza pubblica e tutela dei diritti fondamentali dell'individuo, allo scopo di allineare i singoli Stati membri al più ampio quadro giuridico e valoriale che fa parte dell'*aquis* comunitario.

4. Il meccanismo di controllo e il ruolo dell'autorità di vigilanza del mercato e dell'autorità nazionale per la protezione dei dati.

Accanto al sistema descritto al paragrafo precedente, di esame e rilascio, caso per caso, dell'autorizzazione all'uso dei sistemi di identificazione biometrica remota, il regolamento prevede anche un meccanismo di controllo generale sull'utilizzo di tali sistemi, affidato all'autorità nazionale di vigilanza del mercato, all'autorità nazionale per la protezione dei dati e alla Commissione.

L'art. 5, par. 4, del regolamento prevede, anzitutto, che ogni uso dei sistemi in questione debba essere notificato, secondo le relative procedure nazionali, alla pertinente autorità di vigilanza del mercato e all'autorità nazionale per la protezione dei dati. Anche in questo caso la disciplina comunitaria si premura soltanto di fornire alcune indicazioni di principio circa la notifica, che deve esplicitare l'esito della richiesta e non includere "dati operativi sensibili", lasciando invece agli Stati membri l'implementazione delle «necessarie regole dettagliate [...] per le attività di controllo e comunicazione ad esse relative»¹⁸.

La notifica all'autorità di vigilanza del mercato è coerente con l'apparato di supervisione predisposto dal regolamento europeo in materia di intelligenza

¹⁸ Art. 5, par. 5, del regolamento europeo.

artificiale, che affida a tale autorità il controllo dei sistemi ad alto rischio che vengono immessi sul mercato (categoria in cui ricadono, qualora ne è autorizzato l'uso, i sistemi di identificazione biometrica remota).

La notifica all'autorità nazionale per la protezione dei dati vede, invece, la sua *ratio* nella particolare forma di "trattamento" dei dati in cui si sostanzia l'utilizzo dei richiamati sistemi: evidente è, infatti, che ogniqualevolta sussista un "trattamento" di dati personali debba esser coinvolta la rispettiva autorità nazionale, a maggior ragione se, come nel caso di specie, il trattamento riguarda dati biometrici (quindi particolarmente sensibili), senza consenso dell'interessato.

In proposito, si noti che il regolamento sull'intelligenza artificiale e, in particolare, l'art. 5, si pone quale *lex specialis* rispetto alla disciplina comunitaria circa il "trattamento di categorie particolari di dati personali" di cui all'art. 10 della direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio del 27 aprile 2016, dovendo prevalere su di essa ogniqualevolta si dovesse verificare un contrasto tra norme ¹⁹.

Ricevute tali notifiche, l'art. 5, par. 6, del regolamento richiede alle autorità nazionali di vigilanza del mercato e di protezione dei dati di presentare alla Commissione relazioni annuali sull'uso dei sistemi di identificazione biometrica remota nei singoli ordinamenti nazionali, da cui si evinca il numero di richieste e il loro esito.

Infine, sulla base di tali dati aggregati, è onere della Commissione pubblicare ogni anno una relazione sull'uso di siffatti sistemi a livello comunitario.

In questa fase di controllo e vigilanza pare assumere rilievo decisivo il ruolo della Commissione europea, quale unica istituzione in grado di disporre di una visione d'insieme sull'impiego dei sistemi di identificazione biometrica remota nei diversi ordinamenti nazionali.

Più in generale, come già anticipato, a fronte di un sistema fondato su di un meccanismo di autorizzazione caso per caso, da rilasciarsi a livello nazionale, e su di un meccanismo di vigilanza attuato sempre a livello nazionale, pare essenziale la funzione svolta dalla Commissione di analisi dei dati aggregati provenienti dalle autorità dei singoli Stati membri. Tale attività consente infatti di individuare tendenze comuni, ma anche criticità ricorrenti e divergenze applicative, nell'approccio ai sistemi di identificazione biometrica remota dei diversi ordinamenti nazionali, allo scopo di offrire una base conoscitiva trasversale circa l'utilizzo di tali sistemi.

Questa base conoscitiva diviene particolarmente strategica se si considera che la Commissione è anche l'istituzione competente ad adottare linee guida

¹⁹ In proposito considerando 38 del regolamento europeo.

per l'attuazione uniforme del regolamento europeo, potendo quindi – in presenza di spiccate divergenze – orientare l'evoluzione delle discipline nazionali verso modelli più omogenei e coerenti con i principi dell'ordinamento comunitario.

Anche in questo caso, in ultima analisi, l'effettività e la coerenza dell'ordinamento comunitario potrebbero dipendere, dunque, dalla capacità della Commissione di esercitare il suo ruolo di indirizzo e coordinamento che, pur nel rispetto delle competenze nazionali, assicuri un livello uniforme di tutela dei diritti fondamentali all'interno dell'Unione Europea.

5. Conclusioni.

L'analisi del quadro normativo e procedimentale relativo all'uso dei sistemi di identificazione biometrica remota nel regolamento sull'intelligenza artificiale rivela la complessità dell'approccio comunitario alla materia, volto a bilanciare (o, quantomeno, provare a bilanciare) la promozione dell'innovazione con la tutela dei diritti fondamentali.

In tale ottica, nelle ipotesi in cui l'utilizzo dei sistemi in parola è consentito, il legislatore comunitario ha istituito un duplice meccanismo di garanzia, allo scopo di far fronte ai rischi che esso comporta: un sistema di autorizzazione preventiva caso per caso, che deve essere rilasciata da un'autorità giudiziaria o amministrativa indipendente, e un sistema di controllo generale affidato alle autorità nazionali e alla Commissione europea.

Tuttavia, il margine di discrezionalità lasciato agli Stati membri nell'implementazione delle deroghe e nella scelta dell'autorità competente comporta alcune criticità, che rischiano di indebolire il bilanciamento degli interessi. Tale discrezionalità rischia, infatti, di tradursi in una frammentazione della disciplina e in una disparità nello *standard* di protezione dei diritti fondamentali all'interno dell'Unione Europea.

In questo contesto, assume un rilievo decisivo il ruolo della Commissione europea, unica istituzione in grado di disporre di una visione d'insieme. Come osservato, attraverso l'analisi dei dati aggregati provenienti dalle relazioni annuali degli Stati membri e l'elaborazione di orientamenti, la Commissione dovrà esercitare con efficacia la sua funzione di indirizzo e coordinamento, cruciale per orientare l'evoluzione delle discipline nazionali verso modelli più omogenei, assicurare un livello uniforme di tutela dei diritti nell'Unione Europea e un idoneo bilanciamento tra tale tutela e la promozione dell'innovazione tecnologica.

CAPITOLO XIV

LE DISCRIMINAZIONI DI GENERE NELL'ACCESSO AL
LAVORO GENERATE O AMPLIFICATE DAI SISTEMI DI
INTELLIGENZA ARTIFICIALE: PROFILI COSTITUZIONALI

Niccolò Panigada

SOMMARIO: 1. I sistemi di IA e la discriminazione di genere nell'accesso al lavoro: origine del problema a partire da alcuni casi concreti. – 2. L'ancoraggio costituzionale e sovranazionale. – 3. La legge italiana sull'intelligenza artificiale in relazione al mondo del lavoro. – 4. Rischi e criticità giuridiche ancora presenti. – 5. I rimedi previsti dall'ordinamento. – 6. Conclusioni.

1. I sistemi di IA e la discriminazione di genere nell'accesso al lavoro: origine del problema a partire da alcuni casi concreti.

L'impatto dell'intelligenza artificiale sul genere, specie nel mondo del lavoro, è stato un tema emerso fin dai primi impieghi degli algoritmi in questo settore. Nello specifico, la preoccupazione riguardava lo scarso impiego delle donne nei settori tecnologici e nella di scrittura dei codici algoritmici¹, problematica a sua volta dovuta dal numero esiguo di laureate nelle materie STEM; si è parlato in questo senso di «segregazione orizzontale»². Il mondo tecnologico, infatti, è stato – ed è tuttora – dominato dagli uomini a causa di uno stereotipo, protrattosi a lungo, secondo il quale le donne non sarebbero state portate per le materie scientifiche³. Negli ultimi anni si può assistere ad un cambio di approccio che, tuttavia, deve far fronte all'incontrollabile velocità dello sviluppo tecnologico.

¹ Sul punto si veda M. D'AMICO, *Una parità ambigua*, Milano, 2020, 313.

² Ne parla anche M. D'AMICO, *Le nuove frontiere delle discriminazioni di genere: l'impatto dell'intelligenza artificiale*, in questo volume, spec. 5 ss.

³ Questo tema viene affrontato in M. D'AMICO, M. P. ABBRACCHIO, *Donne nella scienza. La lunga strada della parità*, Milano, 2023.

In altri termini, per colmare il divario tra uomini e donne in ambito tecnico-scientifico si richiede che il superamento di questo tipo di discriminazioni nei confronti del genere femminile vada ad un passo di gran lunga superiore rispetto a quello della rivoluzione digitale a cui stiamo assistendo; è chiaro come ciò costituisca un ostacolo non indifferente alla realizzazione di una parità effettiva tra uomini e donne.

In generale, una delle frontiere più delicate del rapporto tra tecnologia e tutela della parità di genere è rappresentata dall'impiego di sistemi di intelligenza artificiale nei processi di selezione e gestione delle risorse umane. Ci sono diversi casi noti che dimostrano lo stretto legame tra questi due aspetti; emblematica è stata la vicenda che ha coinvolto Amazon nel 2018. Il caso, relativo all'abbandono di un algoritmo di *recruitment* che penalizzava sistematicamente e "involontariamente" i *curricula* femminili, ha evidenziato in modo paradigmatico come il rischio discriminatorio non derivi necessariamente da una volontà dolosa o consapevole del programmatore, bensì dal modo in cui i dati di addestramento riflettono strutturalmente *bias* preesistenti nella realtà sociale. In quella circostanza, l'algoritmo era stato istruito a partire da dati storici aziendali che riproducevano un contesto di predominanza maschile nelle assunzioni; di conseguenza, la presenza di elementi associati al genere femminile (come la frequenza di college femminili o l'uso di pronomi femminili nei *curricula*) veniva interpretata come un indicatore negativo di qualità professionale.

Casi simili si sono verificati con Facebook e Google Ads dove alle donne venivano presentate opportunità lavorative meno retribuite.

Alla luce di queste situazioni da un lato risulta evidente come il *bias* diventi «lo stereotipo pronto a trasformarsi in scelta discriminatoria»⁴, dall'altro, gli elementi fondamentali diventano i dati che vengono utilizzati per "allenare" gli algoritmi e i sistemi di intelligenza artificiale⁵. Tali dati, infatti, oltre a poter essere direttamente portatori di *bias*, possono anche essere rielaborati dai sistemi di intelligenza artificiale, portando a risultati discriminanti nei confronti di determinati soggetti, senza che ciò sia prevedibile in fase di programmazione.

Il rischio che sistemi di IA amplifichino disparità di genere o altre forme di disuguaglianza si iscrive, dunque, nella più ampia problematica della discriminazione algoritmica, in cui l'apparente neutralità del codice si traduce in una riproduzione, o persino in un rafforzamento, delle disuguaglianze di fatto. Tale

⁴ E. STRADELLA, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in *Consulta online. Liber amicorum per Pasquale Costanzo*, 30 marzo 2020.

⁵ Sull'importanza dei dati si sofferma G. CERRINA FERONI, *L'IA discrimina le donne, serve parità algoritmica*, in www.garanteprivacy.it, 16 luglio 2025.

fenomeno assume una particolare rilevanza nel contesto lavorativo, dove l'accesso all'impiego rappresenta non solo una dimensione della libertà economica ⁶, ma anche un presidio della dignità e della parità sostanziale costituzionalmente garantiti.

Non a caso, il regolamento europeo sull'intelligenza artificiale (c.d. AI Act) colloca i sistemi di IA utilizzati per il reclutamento e la selezione del personale nella categoria dei "sistemi ad alto rischio" (art. 6 che rimanda all'allegato III, par. 4 e il considerando 57). Tale qualificazione implica l'assoggettamento a rigorosi obblighi di trasparenza, controllo e gestione del rischio, volti a prevenire effetti discriminatori, garantendo che la progettazione e l'utilizzo degli algoritmi avvengano nel rispetto dei diritti fondamentali. Tuttavia, l'effettività di tali garanzie resta condizionata alla possibilità di comprendere e rendere verificabili i criteri decisionali adottati dai sistemi di IA: un obiettivo ancora lontano, se si considera la complessità tecnica e l'opacità che spesso caratterizzano i modelli di *machine learning*.

In questo senso, il nodo del rischio discriminatorio nei sistemi di IA non è solo tecnico, ma anche e soprattutto giuridico-costituzionale, poiché interroga la capacità dell'ordinamento di garantire un controllo effettivo sui processi decisionali automatizzati che incidono su diritti fondamentali come l'uguaglianza, la dignità e il lavoro.

2. *L'ancoraggio costituzionale e sovranazionale.*

Il rischio di discriminazione di genere prodotto o amplificato dai sistemi di intelligenza artificiale si colloca al crocevia tra il principio di uguaglianza e le nuove sfide poste dalla digitalizzazione dei processi decisionali. Sotto il profilo costituzionale i principi incisi sono molteplici. In questa sede si intende esclusivamente impostare il quadro di riferimento delineando i confini entro cui la materia si sviluppa.

L'art. 3 Cost. rappresenta il fulcro del discorso: da un lato, come risaputo, afferma l'eguaglianza formale, vietando ogni discriminazione fondata sul sesso e su altre condizioni personali; dall'altro, impone alla Repubblica il compito di rimuovere gli ostacoli che, di fatto, limitano la libertà e l'uguaglianza dei cittadini ⁷.

⁶ Così T.H. MARSHALL, *Cittadinanza e classe sociale*, a cura di S. MEZZADRA, Roma-Bari, 2002, 18.

⁷ Sull'art. 3 della Costituzione la dottrina è vastissima. In questa sede si possono ricordare *ex multis* i contributi di L. PALADIN, *Il principio costituzionale di eguaglianza*, Milano, 1965; P.

È proprio questa seconda dimensione – quella dell’uguaglianza sostanziale – a rendere maggiormente evidente come il diritto costituzionale non possa restare indifferente rispetto agli effetti indiretti e strutturali dei sistemi di IA in relazione al fattore di discriminazione del sesso. Anche un algoritmo apparentemente neutro può, infatti, consolidare disuguaglianze preesistenti, riproducendo nel linguaggio della tecnica ciò che la Costituzione impone allo Stato di correggere attraverso politiche di eguaglianza effettiva.

A ciò si aggiunge l’art. 37 Cost., che riconosce il diritto delle donne ad una parità di retribuzione a parità di lavoro rispetto agli uomini⁸. Tale disposizione, letta oggi alla luce delle trasformazioni tecnologiche, non si limita a tutelare l’accesso e le condizioni lavorative in senso tradizionale, ma si estende a proteggere da nuove forme di esclusione mediate da strumenti digitali e automatizzati. In questa prospettiva, l’uso discriminatorio dell’intelligenza artificiale rappresenta una violazione non solo del principio di eguaglianza, ma anche della dignità e della libertà professionale della persona.

Sul piano sovranazionale, la Carta dei diritti fondamentali dell’Unione Europea (artt. 21 e 23)⁹ rafforza la tutela contro le discriminazioni fondate sul sesso e consacra l’obiettivo di una effettiva parità tra uomini e donne in tutti i settori, compreso quello lavorativo. L’art. 21 sancisce il divieto generale di discriminazione, mentre l’art. 23 stabilisce che «la parità tra uomini e donne deve essere assicurata in tutti i campi, compreso l’impiego, il lavoro e la retribuzione». Tali principi, letti in combinato disposto con l’art. 8 della Convenzione europea dei diritti dell’uomo, impongono agli Stati membri di predisporre strumenti di prevenzione e rimedi effettivi contro le discriminazioni anche quando esse derivano da processi automatizzati.

Il dialogo tra ordinamento costituzionale interno e ordinamento europeo impone, dunque, una rilettura sistematica del principio di uguaglianza alla luce della “governance algoritmica”. L’esigenza non è solo quella di reagire a posteriori alle discriminazioni algoritmiche, ma di costruire *ex ante* un paradigma di

BISCARETTI DI RUFFIA, *Uguaglianza (principio di)*, in *Noviss. Dig. it.*, Torino, 1973; A. CERRI, CERRI, *Uguaglianza (principio costituzionale di)*, in *Enc. giur.*, vol. XXXII, Roma, 1994.

⁸ A. BARBERA, *Discriminazioni ed eguaglianza nel rapporto di lavoro*, Milano, 1991; T. TREU, *Art. 37*, in G. BRANCA (a cura di), *Commentario della Costituzione*, Bologna, 1979, 153 ss.; C. SALAZAR, *Art. 37*, in R. BIFULCO, A. CELOTTO, M. OLIVETTI (a cura di), *Costituzione commentata*, Torino, 2007.

⁹ Questi articoli sono da intendersi come «canoni interpretativi dell’azione dell’Unione europea» alla luce dell’art. 8 TFUE: così E.C. RAFFIOTTA, *Intelligenza artificiale e tutela dell’uguaglianza di genere*, in M. D’AMICO, B. LIBERALI (a cura di), *I diritti delle donne problematiche attuali e prospettive future*, Torino, 2024, 190.

regolazione inclusiva, capace di assicurare che la progettazione, l'addestramento e l'utilizzo dei sistemi di intelligenza artificiale siano conformi ai valori costituzionali e sovranazionali di parità e non discriminazione.

Ciò risulta ben evidente, per il tema qui affrontato, nel Considerando 36 dell'AI Act dove si evidenzia che l'intelligenza artificiale in ambito lavorativo rischia di «perpetuare modelli storici di discriminazione, ad esempio nei confronti delle donne».

3. La legge italiana sull'intelligenza artificiale in relazione al mondo del lavoro.

La legge n. 132/2025 sull'intelligenza artificiale rappresenta il primo tentativo organico dell'ordinamento italiano di delineare una cornice normativa generale per lo sviluppo e l'impiego dei sistemi di IA in coerenza con i valori costituzionali e con il quadro europeo definito dall'AI Act¹⁰. Il testo assume particolare rilievo in quanto introduce, tra i principi generali (art. 3), quello della parità dei sessi, espressamente richiamato quale parametro di orientamento per la progettazione, l'adozione e l'utilizzo dell'intelligenza artificiale nei diversi ambiti di applicazione.

Anche l'art. 11, comma 3, della legge è, sotto questo profilo, di particolare interesse: esso stabilisce che «l'intelligenza artificiale nell'organizzazione e nella gestione del rapporto di lavoro garantisce l'osservanza dei diritti inviolabili del lavoratore senza discriminazioni in funzione del sesso». Tale formulazione, che riprende quasi tutti i fattori di discriminazione già enunciati dall'art. 3 Cost., sancisce un principio di neutralità non solo formale, ma sostanziale dell'IA nel contesto lavorativo, imponendo che gli strumenti algoritmici rispettino il divieto di discriminazione diretta e indiretta, indipendentemente dall'intenzionalità soggettiva del datore di lavoro o del fornitore del software.

Il successivo art. 12 istituisce un Osservatorio sull'adozione dei sistemi di intelligenza artificiale nel mondo del lavoro, con la finalità di monitorare l'impatto dell'IA sull'occupazione, individuare potenziali rischi discriminatori e promuovere buone pratiche di progettazione inclusiva. Tale previsione, se adeguatamente implementata, potrebbe costituire uno strumento di governance proattiva, capace di coniugare sviluppo tecnologico e tutela dei diritti fondamentali, in linea con il principio di eguaglianza sostanziale di cui all'art. 3,

¹⁰ G. ZICCARDI, *L'intelligenza artificiale e la regolamentazione giuridica: una relazione complessa*, in M. BIASI (a cura di), *Diritto del lavoro e intelligenza artificiale*, Milano, 2024, 13 ss.

comma 2, Cost. Il problema principale è che tale Osservatorio dovrebbe contare su «risorse umane, strumentali e finanziarie disponibili a legislazione vigente», privandolo in questo modo di nuovi mezzi che avrebbero potuto incentivarne l'attività ¹¹.

Pertanto, la portata innovativa della legge va valutata anche in termini critici. La disposizione non chiarisce, infatti, quali meccanismi di controllo e sanzione siano previsti in caso di violazione del principio di parità né definisce i criteri tecnici con cui verificare la conformità dei sistemi di IA agli standard di non discriminazione. La mancata previsione di un obbligo di *audit* indipendenti o di trasparenza algoritmica rischia di tradurre i principi in enunciazioni di carattere programmatico, prive di effettività, i cui unici strumenti di opposizione rimarrebbero le ordinarie azioni antidiscriminatorie e, in via stragiudiziale, la valutazione d'impatto sui diritti fondamentali per i sistemi ad alto rischio così come prevista dall'art. 27 dell'AI Act.

In questo senso, la legge si colloca in una fase di transizione: da un lato, recepisce la sensibilità costituzionale ed europea verso l'uguaglianza di genere e la dignità del lavoratore e della lavoratrice; dall'altro, mostra la necessità di un'evoluzione normativa che renda vincolante l'integrazione dei diritti fondamentali nella progettazione *ex ante* dei sistemi di IA, secondo una logica di "*constitutional by design*" ¹². Solo così sarà possibile evitare che la tecnologia divenga un fattore di regressione rispetto ai principi di uguaglianza e inclusione su cui si fonda l'ordinamento democratico.

4. Rischi e criticità giuridiche ancora presenti.

L'introduzione dei sistemi di intelligenza artificiale nei processi di selezione, valutazione e gestione del personale comporta una serie di rischi e criticità giuridiche che incidono direttamente sui principi costituzionali di uguaglianza e di tutela della dignità del lavoratore e della lavoratrice ¹³. Il primo e più evidente

¹¹ La questione viene attenzionata da M. BIASI, *Il lavoro nel disegno di legge governativo in materia di intelligenza artificiale: principi, regole, parole, silenzi*, in *Dir. rel. ind.*, 3, 2024, 657.

¹² Interessanti considerazioni sul concetto di "*constitutional by design*" sono svolte da G. DE MINICO, *Towards an "Algorithm Constitutional by Design"*, in *BioLaw Journal-Riv. BioDiritto*, 1, 2021.

¹³ Il problema viene ampiamente esaminato da A. RICCOBONO, *Intelligenza artificiale e limiti al social media profiling nella selezione del personale*, in M. BIASI (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., 253 ss.

profilo di criticità è rappresentato dalla possibilità di una discriminazione indiretta: anche in assenza di un intento discriminatorio da parte del datore di lavoro o del programmatore, l'algoritmo può generare effetti di esclusione o penalizzazione sistematica nei confronti di gruppi protetti, in particolare delle donne (il fenomeno è conosciuto anche come *indirect proxy discrimination*¹⁴). In tali casi, la violazione della parità sostanziale discende dal modo in cui i modelli di *machine learning* apprendono regolarità dai dati storici, che riflettono disuguaglianze già presenti nella società o nel mercato del lavoro.

Questa forma di discriminazione algoritmica, difficilmente riconducibile alle categorie tradizionali della responsabilità soggettiva, mette in crisi l'impianto classico del diritto antidiscriminatorio, fondato sull'individuazione di un autore, di un comportamento e di un nesso causale diretto. La domanda diventa allora: chi risponde dell'errore dell'algoritmo? Il datore di lavoro che lo utilizza? Il programmatore che lo ha sviluppato? Il fornitore del software? In assenza di una chiara attribuzione di responsabilità, si rischia di generare una "zona grigia" di irresponsabilità diffusa, in cui la lesione del diritto all'uguaglianza rimane priva di un effettivo rimedio giuridico.

A ciò si aggiunge il problema dell'opacità algoritmica ("*black box problem*")¹⁵: la complessità tecnica dei modelli di IA, specie quelli basati su reti neurali profonde, rende spesso impossibile comprendere le logiche interne di funzionamento e, di conseguenza, contestare in giudizio la correttezza o la legittimità di una decisione automatizzata. Tale opacità non solo ostacola l'esercizio del diritto di difesa, ma mina la stessa effettività del controllo giurisdizionale, in quanto impedisce di verificare se il trattamento differenziato sia fondato su criteri obiettivi e ragionevoli, come richiesto dall'art. 3 Cost. e dalla giurisprudenza costituzionale¹⁶.

Un'ulteriore criticità riguarda l'assenza di strumenti di *audit* e di *accountability* realmente vincolanti. La semplice previsione di obblighi di trasparenza o di principi generali, se non accompagnata da meccanismi di verifica indipendenti e da sanzioni effettive, rischia di produrre un regime di mera autoregolazione,

¹⁴ Il tema, anche con riferimento alla *direct proxy discrimination*, è affrontato da C. NARDOCCI, *Algoritmi, eguaglianza, discriminazione. Le sfide dell'intelligenza artificiale*, Torino, 2025, 114 ss.

¹⁵ Anche questo problema viene ripreso nel contributo di M. D'AMICO, *Le nuove frontiere delle discriminazioni di genere*, cit., 8.

¹⁶ Fin dalla sentenza n. 15/1960 della Corte costituzionale, infatti, è stato stabilito che la valutazione sulla violazione del principio di uguaglianza deve tenere in considerazione anche il canone della ragionevolezza nel trattamento differenziato.

insufficiente a prevenire e correggere gli effetti discriminatori. Da questo punto di vista, la prospettiva di un controllo *ex post* – affidato al giudice o alle autorità di garanzia – rischia di rivelarsi inadeguata a fronte della rapidità e complessità dei processi automatizzati: servono, invece, strumenti di prevenzione *ex ante*, capaci di assicurare che la progettazione stessa degli algoritmi sia conforme ai principi costituzionali e sovranazionali di uguaglianza e non discriminazione.

In definitiva, la vera sfida giuridica non consiste soltanto nell'adattare le categorie tradizionali del diritto del lavoro o del diritto antidiscriminatorio all'era digitale, ma nel ripensare l'intero paradigma della responsabilità alla luce dell'autonomia funzionale dei sistemi di IA. Si tratta di un compito eminentemente costituzionale: riaffermare che anche l'innovazione tecnologica, per quanto sofisticata, deve restare soggetta al controllo dei principi supremi dell'ordinamento e non può tradursi in una zona franca di non diritto.

5. I rimedi previsti dall'ordinamento.

Dinanzi ai rischi di discriminazione derivanti dall'impiego di sistemi di intelligenza artificiale nei processi di accesso e gestione del lavoro, l'ordinamento italiano dispone già di un articolato quadro normativo volto a prevenire e reprimere le condotte discriminatorie, in particolare attraverso le disposizioni contenute nel Codice delle pari opportunità (d.lgs. n. 198/2006). Tuttavia, l'applicazione di tali strumenti alla dimensione algoritmica richiede un adattamento interpretativo che consenta di colmare le lacune derivanti dalla novità tecnologica dei fenomeni in esame.

L'art. 27 del Codice vieta in modo espresso «qualsiasi discriminazione per quanto riguarda l'accesso al lavoro [...], compresi i criteri di selezione e le condizioni di assunzione», indipendentemente dalle modalità di assunzione e dal settore di attività. Tale previsione, letta in combinato disposto con l'art. 3 Cost. e con l'art. 21 della Carta dei diritti fondamentali dell'Unione Europea, deve oggi essere intesa come comprensiva anche delle decisioni automatizzate o assistite da sistemi di IA, le quali, seppur formalmente neutrali, possono produrre effetti discriminatori di fatto. In questa prospettiva, la discriminazione algoritmica si configura come una nuova tipologia di discriminazione indiretta, in cui la lesione del diritto all'uguaglianza deriva dall'uso non consapevole o non controllato di dati e modelli che riflettono pregiudizi sistemici.

Un ruolo particolarmente significativo è svolto dall'art. 40 del Codice, che prevede un onere probatorio attenuato: qualora il ricorrente fornisca elementi

di fatto idonei a fondare, anche solo in termini presuntivi, l'esistenza di un comportamento discriminatorio, spetta al convenuto dimostrare l'insussistenza della discriminazione. Questa inversione dell'onere della prova assume una valenza decisiva nel contesto algoritmico, dove la parte lesa difficilmente dispone degli strumenti tecnici necessari per comprendere e dimostrare il funzionamento dell'algoritmo. In tali casi, l'obbligo del datore di lavoro di provare la neutralità del sistema di IA si traduce in un potente incentivo alla progettazione trasparente e documentata dei modelli decisionali, favorendo un approccio di "responsabilità preventiva" coerente con la logica della parità sostanziale.

L'art. 38 del Codice, infine, prevede la possibilità per il giudice di disporre non solo il risarcimento del danno ma anche la cessazione del comportamento illegittimo e la rimozione degli effetti discriminatori. Tale previsione apre la strada a una tutela potenzialmente "trasformativa": se interpretata in senso costituzionalmente orientato, essa può estendersi fino a includere l'obbligo di correggere o sostituire il sistema algoritmico che ha generato la discriminazione, imponendo così una modifica strutturale del processo decisionale.

Nonostante ciò, permane una significativa asimmetria tra la forza normativa degli strumenti antidiscriminatori e la complessità tecnica delle decisioni automatizzate. La possibilità di individuare concretamente l'origine della discriminazione, di accedere ai dati di addestramento o di ottenere una piena *disclosure* dell'algoritmo rappresenta, di fatto, una condizione spesso irrealizzabile. Da qui l'esigenza, sempre più avvertita, di affiancare ai rimedi giurisdizionali *ex post* meccanismi di *compliance ex ante*, fondati, ad esempio, su valutazioni d'impatto sui diritti fondamentali, *audit* indipendenti e obblighi di trasparenza sostanziale.

In sintesi, la disciplina antidiscriminatoria italiana contiene già i principi idonei a contrastare le disuguaglianze generate o amplificate dall'IA, ma necessita di un'evoluzione interpretativa e applicativa che consenta di tradurre tali principi in strumenti effettivi di giustizia algoritmica. Solo così sarà possibile assicurare che l'innovazione tecnologica, lungi dal perpetuare stereotipi o asimmetrie di potere, diventi un veicolo di realizzazione concreta del principio di uguaglianza sancito dall'art. 3 Cost.

6. Conclusioni.

L'analisi dei rischi discriminatori connessi all'impiego dei sistemi di intelligenza artificiale nel mondo del lavoro mostra con chiarezza come le tecnologie emergenti non pongano soltanto questioni di efficienza o innovazione, ma incidano direttamente sulla tenuta assiologica dell'ordinamento costituzionale.

L'uguaglianza, la dignità e la libertà del lavoratore e della lavoratrice – principi cardine della nostra Costituzione e del diritto europeo – rischiano di essere compromessi non tanto da scelte consapevolmente discriminatorie, quanto da meccanismi algoritmici che riproducono, in forma automatizzata, disuguaglianze sociali e culturali preesistenti.

Gli strumenti giuridici oggi disponibili – dal Codice delle pari opportunità alla legge n. 132/2025, fino ai principi sovranazionali dell'AI Act – delineano un quadro normativo in evoluzione, che riconosce la necessità di governare l'innovazione tecnologica entro i confini dei diritti fondamentali. Tuttavia, tali strumenti, pur potenzialmente efficaci, operano ancora prevalentemente in una logica reattiva, volta a sanzionare le discriminazioni una volta verificatesi, più che a prevenirle. In un contesto in cui le decisioni algoritmiche sono per loro natura opache, automatizzate e pervasive, questa impostazione rischia di rivelarsi insufficiente. Tra gli strumenti da valorizzare, pertanto, c'è sicuramente quello della valutazione d'impatto sui diritti fondamentali per i sistemi di intelligenza artificiale ad alto rischio previsto dall'AI Act, costituendo essa un mezzo potenzialmente efficace per anticipare effetti discriminatori.

Diventa indispensabile promuovere un approccio costituzionale e preventivo alla regolazione dell'intelligenza artificiale, fondato sul principio di “non discriminazione *by design*”. Ciò implica che la tutela dei diritti non possa essere una fase successiva o correttiva, ma debba essere incorporata sin dall'origine nei processi di progettazione, addestramento e implementazione dei sistemi di IA. Tale prospettiva, già presente nel dibattito europeo sul “*trustworthy AI*”, deve trovare un radicamento esplicito nel diritto interno, trasformando il principio di uguaglianza da limite esterno a criterio interno di costruzione della tecnologia.

In questa direzione, la previsione dell'onere probatorio attenuato nel contenzioso antidiscriminatorio assume un valore strategico, in quanto incentiva le imprese e i datori di lavoro a sviluppare sistemi di IA trasparenti e verificabili. Il rischio reputazionale e giuridico di una discriminazione automatizzata si traduce così in un potente meccanismo di autodisciplina virtuosa, che spinge il mercato verso soluzioni tecnologiche più eque e responsabili.

In definitiva, la sfida che si pone al giurista e al legislatore non è solo quella di aggiornare il diritto al passo con la tecnologia, ma di riaffermare, attraverso il diritto, la centralità della persona in un ecosistema digitale che tende a spostare il baricentro decisionale dalla volontà umana al calcolo algoritmico. La Costituzione, in questa prospettiva, non rappresenta un vincolo allo sviluppo, bensì il fondamento normativo di un'innovazione inclusiva, capace di orientare l'intelligenza artificiale verso la piena realizzazione del principio di uguaglianza sostanziale.

CAPITOLO XV

IA GENERATIVA E DIFFUSIONE DEGLI STEREOTIPI
ONLINE. LA CONCETTUALIZZAZIONE DI UN AGENTE
CONTRO IL LINGUAGGIO DISCRIMINATORIO

Marta Annamaria Tamborini

SOMMARIO: 1. Premessa. – 2. Le nuove frontiere della discriminazione algoritmica. – 3. *Bias* nel linguaggio, *bias* nell'algoritmo e diffusione di odio. – 4. L'IA generativa come alleato e non come minaccia. – 5. Il progetto dell'identificazione degli stereotipi e del linguaggio d'odio attraverso l'*Agent for Anti Discriminatory Language* (AAL). – 6. Conclusioni.

1. *Premessa.*

Il fenomeno dell'intelligenza artificiale e, in particolare, dell'IA generativa, sta lentamente trovando un inquadramento nei vari ordinamenti giuridici mondiali. “Lentamente” non è un'espressione casuale dal momento che ci sono voluti quasi 5 anni per arrivare ad un inquadramento sul piano europeo e la diffusione dei sistemi di intelligenza artificiale generativa segna un passaggio di fase rispetto alle tradizionali forme di automazione¹. Se, per i sistemi di scoring

¹ Per un primo inquadramento del quadro normativo, sul versante europeo v. il regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce norme armonizzate sull'intelligenza artificiale (AI Act), in *GUUE* L 206 del 12 luglio 2024, primo tentativo di disciplina “onnicomprensiva” dei sistemi di IA a livello mondiale. Quanto al recepimento interno, v. ora la legge 23 settembre 2025, n. 132, Disposizioni e deleghe al Governo in materia di intelligenza artificiale, in *GU* n. 223 del 25 settembre 2025, che affida a una pluralità di decreti delegati la definizione dell'assetto interno di governance dell'IA. Sul versante statunitense, in assenza (allo stato) di una legge federale organica in materia di IA, rilievo assumono, da un lato, gli atti di indirizzo della Casa Bianca – in particolare l'Executive Order 14179, Removing Barriers to American Leadership in Artificial Intelligence, 23 gennaio 2025, e il successivo documento strategico Winning the AI Race: America's AI Action Plan,

e di decisione automatizzata, il tema principale riguardava l'impatto su singoli procedimenti amministrativi, giudiziari o contrattuali, la generativa incide direttamente sulla produzione di contenuti, linguaggi e rappresentazioni che circolano nello spazio pubblico. Il problema non è più soltanto «che cosa decidono gli algoritmi», ma «come parlano» e «di chi parlano»: quali gruppi vengono resi visibili o invisibili, in quali ruoli sono collocati, quali stereotipi vengono riprodotti o corretti. In questo quadro, il costituzionalista è chiamato a confrontarsi con una forma di discriminazione che agisce sul terreno simbolico prima ancora che su quello materiale. Le categorie degli artt. 2, 3 e 21 Cost., lette alla luce della CEDU e della Carta dei diritti fondamentali dell'Unione Europea, offrono la griglia per interrogare la capacità della generativa di normalizzare linguaggi d'odio e rappresentazioni diseguali, ma anche, in senso potenzialmente opposto, di divenire strumento di emersione e decostruzione degli stereotipi². A livello *nazionale*, il quadro è oggi completato dalla legge 23 settembre 2025, n. 132, "*Disposizioni e deleghe al Governo in materia di intelligenza artificiale*"³, che, in coerenza con l'AI Act, afferma un uso antropocentrico e responsabile dell'IA e attribuisce espressamente rilievo alla vigilanza sui rischi economici e sociali e sull'impatto sui diritti fondamentali. Il presente contributo si colloca in questo contesto, prendendo le mosse dalle nuove frontiere della discriminazione algoritmica e il nesso tra *bias* linguistici e *bias* nei modelli, si propone di valutare se e a quali condizioni l'IA generativa possa essere utilizzata come alleato nella tutela dell'eguaglianza, anche attraverso l'esempio di un agente dedicato all'identificazione degli stereotipi (AAL)⁴.

reso pubblico il 23 luglio 2025, espressamente orientati a "vincere la corsa" globale all'IA; dall'altro, le prime leggi statali di carattere generale, quali lo Utah Artificial Intelligence Policy Act (SB 149/2024), il Colorado Artificial Intelligence Act (SB 24-205/2024, con applicazione differita al 30 giugno 2026) e la più recente Transparency in Frontier Artificial Intelligence Act della California (SB 53/2025), che compongono un quadro frammentato ma fortemente pro-innovazione.

² C. NARDOCCI, *Algoritmi, eguaglianza, discriminazione. Le sfide dell'intelligenza artificiale*, Torino, 2025.

³ M. D'AMICO, *Intelligenza artificiale: nella via italiana una strategia di sviluppo*, in *Il Sole 24 Ore*, inserto Innovazione, 25 settembre 2025.

⁴ M.A. TAMBORINI, F. MOHAMMADI, P. CERAVOLO, S. MAGHOOL, M.E. D'AMICO, C. SICCARDI, *Agent for Anti-Discriminatory Language (AAL): A Human-Centered AI Approach to Discriminatory Speech in Italian*, in *Proceedings of the IEEE Cyber Humanities Conference (IEEE CH 2025)*, Firenze, 2025.

2. *Le nuove frontiere della discriminazione algoritmica.*

Nella tradizione costituzionale italiana la discriminazione è stata a lungo ricondotta al comportamento di un soggetto umano, pubblico o privato, che pone in essere trattamenti differenziati ingiustificati in violazione dell'art. 3, comma 1, Cost. La centralità del profilo soggettivo, cioè l'intento di escludere o svantaggiare un determinato gruppo, ha orientato sia la costruzione dogmatica, sia le tecniche di tutela⁵. L'impiego crescente di algoritmi nei processi decisionali, dall'accesso a servizi essenziali alla selezione del personale, impone tuttavia di spostare lo sguardo dall'autore umano alla struttura dei meccanismi automatizzati. La discriminazione algoritmica non deriva necessariamente da un intento lesivo, ma può configurarsi come effetto di correlazioni statistiche, scelte di modellizzazione e criteri di trattamento dei dati che riproducono e, spesso, amplificano, gerarchie preesistenti. L'eguaglianza, nelle sue dimensioni formale e sostanziale, è così chiamata ad operare non solo come parametro *ex post* di sindacato sulle decisioni, ma come criterio di progettazione, addestramento e controllo dei sistemi di intelligenza artificiale. Infatti, una caratteristica peculiare della discriminazione algoritmica è l'opacità dei sistemi, accentuata nei modelli di apprendimento automatico e profondo. A tale opacità tecnica si sommano profili giuridici ed economici, segreti industriali, clausole di non divulgazione, asimmetrie informative tra fornitori, amministrazioni e utenti che rendono difficile individuare dove, nel ciclo di vita dell'IA, si generi la disparità. In questo contesto emergono fenomeni di *proxy discrimination* che avviene quando il modello non utilizza direttamente categorie sensibili, ma fa leva su variabili fortemente correlate (residenza, storia creditizia, percorsi di studio, abitudini di consumo) che producono esiti sistematicamente sfavorevoli per determinati gruppi⁶. La "discriminazione per correlazione" sfugge agli schemi tradizionali, perché non si fonda su un criterio esplicito, ma su combinazioni di dati apparentemente neutri. Se si parte da un'idea di eguaglianza sostanziale, il

⁵ Per una panoramica completa e aggiornata dello stato attuale del diritto antidiscriminatorio in Italia, in particolare sotto il profilo della discriminazione di genere, cfr. M. D'AMICO, B. LIBERALI, *Manuale di diritto delle donne*, Milano, 2025.

⁶ Corte Giust. UE, Sez. I, 27 febbraio 2025, causa C-203/22, CK / Magistrat der Stadt Wien, con intervento di Dun & Bradstreet Austria GmbH, ECLI:EU:C:2025:117. La sentenza, resa nel caso CK su rinvio del Verwaltungsgericht Wien, chiarisce che nel credit scoring automatizzato la persona interessata ha diritto ad accedere a «informazioni significative sulla logica» del trattamento ai sensi degli artt. 15, par. 1, lett. h), e 22 GDPR, e impone al giudice e all'autorità di controllo di bilanciare tale esigenza di trasparenza con la tutela del segreto commerciale e dei dati personali di terzi.

controllo costituzionale non può arrestarsi all'*output*, ma deve investire l'intero *design* dell'infrastruttura algoritmica, ossia la raccolta e selezione dei dati, la preparazione e la definizione delle funzioni obiettivo, nonché delle metriche di performance.

Con la diffusione della IA generativa, le frontiere della discriminazione algoritmica si estendono però alla dimensione comunicativa e simbolica. I modelli producono testi, immagini, audio e video che entrano nello spazio pubblico e contribuiscono a plasmare l'immaginario sociale. Pertanto, gli stereotipi incorporati nei dati di addestramento e nelle scelte di ottimizzazione si traducono in descrizioni e narrazioni che reiterano rappresentazioni diseguali. La serializzazione degli *output* rende, inoltre, più probabile la normalizzazione di associazioni stereotipiche (professioni e genere, devianza e origine etnica) e moltiplica la portata dell'odio quando i contenuti generati si innestano su dinamiche di polarizzazione già proprie delle piattaforme digitali⁷. Il legislatore europeo ha colto solo in parte questo mutamento. L'AI Act, pur costruito attorno alla categoria dei sistemi "ad alto rischio", riconosce che l'IA può incidere sull'uguaglianza e sui diritti fondamentali e introduce obblighi specifici anche per determinati sistemi generativi, specie quando idonei a manipolare il comportamento o a ingannare gli utenti. Sul versante delle piattaforme, il Digital Services Act impone doveri di valutazione e mitigazione dei rischi sistemici, tra cui la diffusione di contenuti d'odio e la minaccia ai diritti fondamentali, indipendentemente dal fatto che tali contenuti siano prodotti dall'utente o generati e amplificati dall'IA⁸. La combinazione fra questi due strumenti è stata letta come primo tassello di un "costituzionalismo digitale", nel quale dignità, eguaglianza e libertà di espressione vengono tradotti in obblighi di progettazione, trasparenza e responsabilità tecnologica⁹.

Sul piano *nazionale*, la legge 23 settembre 2025, n. 132 rappresenta il primo intervento organico del legislatore italiano in materia di intelligenza artificiale da cui risulta un quadro multilivello nel quale le "nuove frontiere" della discriminazione algoritmica non si collocano soltanto nelle scelte di scoring o nelle

⁷ Per una riflessione sulle "radici del male" e sul ruolo della Costituzione come argine contro l'antisemitismo e le derive discriminatorie, N. FIANO, *Le radici del male. Antisemitismo e costituzione*, Torino, 2024.

⁸ Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio, del 19 ottobre 2022, relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (Digital Services Act), in *GUL* 277 del 27 ottobre 2022, 1-102.

⁹ F. BALAGUER CALLEJON (a cura di), *La constitución del algoritmo. Recensiones, presentaciones y entrevistas*, Fundación Manuel Giménez Abad, Zaragoza, 2025.

decisioni amministrative automatizzate, ma sempre più nel governo dei contenuti generati con l'IA, là dove il confine tra libera manifestazione del pensiero, discorso d'odio e discriminazione strutturale si fa più incerto.

3. Bias *nel linguaggio*, bias *nell'algoritmo* e *diffusione di odio*.

Il linguaggio è spesso il primo strumento di diffusione della discriminazione. Che le parole non si limitino a descrivere la realtà, ma contribuiscano a costruirla è un principio ormai acquisito, anche in diritto costituzionale¹⁰. Le parole tracciano confini tra un "noi" legittimo e un "loro" stigmatizzato, incidono sulla dignità delle persone e sulla loro effettiva partecipazione alla vita democratica. Gli stereotipi non sono meri errori di rappresentazione, bensì schemi cognitivi e discorsivi che normalizzano la disegualianza, presentando tratti di gruppo relativi a genere, origine etnica, religione, orientamento sessuale o condizione socio-economica come base naturale di trattamenti peggiorativi. Il linguaggio d'odio rappresenta il punto più evidente di un continuum che va da espressioni sottilmente svalutative a forme apertamente violente¹¹. L'IA generativa innesta questo continuum in un nuovo ambiente, nel quale sistemi non umani producono essi stessi linguaggio a partire da corpora che riflettono la stratificazione storica dei pregiudizi. La questione non riguarda solo l'insulto esplicito, ma la capacità dei modelli di replicare e moltiplicare narrazioni che collocano stabilmente alcuni gruppi nella sfera dell'inferiorità, della pericolosità o della non piena appartenenza alla comunità politica e questo perché i dati di addestramento dei modelli non rappresentano la società nella sua capillare eterogeneità ma una società fortemente connotata. È stato, infatti, argomentato che interrogare modelli di fondazione come Chat GPT equivale ad interrogare un maschio americano, bianco, etero e sulla cinquantina¹².

Il nesso tra *bias* linguistici e *bias* algoritmici, quindi, può essere letto distinguendo, in via analitica, tre livelli: dati, modello, *output*. Il modello può attenuare o amplificare questi pregiudizi in base alle scelte di architettura, parametrizzazione e alle funzioni obiettivo, ad esempio se si privilegia la "plausibilità

¹⁰ M. D'AMICO, B. LIBERALI, *Manuale di diritto delle donne*, cit., 363 ss.

¹¹ N. FIANO, *Il linguaggio dell'odio in Germania: tra Wehrhafte Demokratie e Netzwerkdurchsetzungsgesetz*, in M. D'AMICO, C. SICCARDI (a cura di), *La Costituzione non odia. Conoscere, prevenire e contrastare l'hate speech on line*, Torino, 2021, 155-165.

¹² A. FERRARA, *Le macchine del linguaggio. L'uomo allo specchio dell'intelligenza artificiale*, Torino, 2025, 82 ss.

statistica” rispetto al pluralismo dei punti di vista, il risultato sarà la riproduzione delle narrazioni maggioritarie, con compressione delle voci minoritarie. Negli *output*, infine, il *bias* si traduce in risposte che, a parità di input, associano sistematicamente determinate categorie a ruoli subordinati, ridicoli, patologizzati o criminalizzati e l’algoritmo rischia, così, di divenire lo specchio deformante di una realtà già ingiusta, cristallizzandola in forma automatizzata.

La diffusione di odio attraverso contenuti generati con l’IA si colloca all’incrocio fra produzione, moderazione e raccomandazione dei contenuti. Restano possibili *output* che si avvicinano al discorso d’odio in senso stretto (insulti, delegittimazione radicale di gruppi protetti, forme di disumanizzazione) specie mediante l’utilizzo di prompt malevoli o contesti borderline, nonostante i meccanismi di filtraggio implementati nei principali modelli. Ancora più frequente è la produzione di contenuti stereotipati di sfondo, che non integrano un *hate speech* manifesto ma collocano stabilmente alcune persone in ruoli subordinati, ridicoli o passivizzati (es. la donna come oggetto, la persona migrante come problema di ordine pubblico, la persona con disabilità come mero destinatario di assistenza, ecc.). Il danno, in questi casi, è meno percepibile ma la diffusione di tali rappresentazioni della realtà possono contribuire ad un fenomeno sociale di ampia portata¹³. A ciò si aggiunge il ruolo dei sistemi di raccomandazione delle piattaforme, che tendono a privilegiare contenuti polarizzanti ed emotivamente forti e la combinazione tra architettura delle piattaforme, logiche di *engagement* e capacità generativa di produrre quantità industriali di messaggi e/o immagini può dare luogo a una vera e propria infrastruttura dell’odio, nella quale la libertà di espressione viene piegata a dinamiche di spettacolarizzazione del conflitto. La lesione è spesso plurale e incrocia più fattori di vulnerabilità (genere e origine etnica, status migratorio e orientamento sessuale, disabilità e condizione socio-economica). Progetti come la mappa dell’intolleranza¹⁴ monitorano la diffusione di tale linguaggio e presidiano l’utilizzo dell’IA generativa per l’analisi e il contrasto dei fenomeni d’odio online.

¹³ Come riferimento per il monitoraggio degli stereotipi e del linguaggio d’odio online cfr. il progetto “Mappa dell’intolleranza” proposta dall’associazione Vox Diritti, giunta alla sua ottava edizione il cui report è disponibile sul sito <https://www.voxdiritti.it/ottava-edizione-mappa-intolleranza/>.

¹⁴ <https://www.voxdiritti.it/ottava-edizione-mappa-intolleranza/>, cit.

4. *L'IA generativa come alleato e non come minaccia.*

Appare sin qui evidente che l'alternativa non è tra accettare o rifiutare l'IA, ma tra lasciarne lo sviluppo alle sole logiche di mercato o orientarlo in modo coerente con i diritti fondamentali. Non è casuale che anche il legislatore nazionale, con la legge 23 settembre 2025, n. 132, abbia fatto propri i principi di uso antropocentrico, trasparente e non discriminatorio dell'IA, vincolando le future scelte regolatorie a una verifica sistematica dell'impatto sui diritti. In questa chiave, i sistemi generativi possono essere letti non solo come fonti di rischio, ma come potenziali strumenti di illuminazione delle dinamiche discriminatorie e di rafforzamento delle capacità di controllo delle istituzioni. La stessa tecnologia che produce *output* sessisti o razzializzati può essere impiegata, se opportunamente ripensata, per individuare stereotipi, sbilanciamenti lessicali e *pattern* ricorrenti di rappresentazione degradante, rendendo più visibili fenomeni che, altrimenti, resterebbero dispersi nel flusso comunicativo. Il quadro regolatorio europeo offre, da questo punto di vista, un'occasione nel senso che gli obblighi di documentazione, valutazione del rischio, governance dei dati e audit previsti dall'AI Act possono essere interpretati come dispositivi per "costituzionalizzare" il ciclo di vita dei sistemi, a condizione che siano effettivamente attuati e coordinati con le garanzie in materia di libertà di espressione e di contrasto all'odio previste dal DSA¹⁵ e dagli strumenti del Consiglio d'Europa. Affinché la generativa possa fungere da alleato occorrono, tuttavia, alcune condizioni minime. Gli obiettivi di progetto devono includere esplicitamente la tutela dell'eguaglianza e non soltanto l'accuratezza o l'engagement. Intorno ai modelli deve esistere un'"umanità competente", capace di definire tassonomie di discriminazione e soglie di intervento; le catene di responsabilità devono essere chiare, soprattutto quando i sistemi sono impiegati per individuare contenuti potenzialmente discriminatori o per suggerire alternative linguistiche.

5. *Il progetto dell'identificazione degli stereotipi e del linguaggio d'odio attraverso l'Agent for Anti Discriminatory Language (AAL).*

Su questo sfondo si colloca il progetto AAL (*Agent for Anti-Discriminatory Language*), che mira a utilizzare un agente di IA generativa per coadiuvare la

¹⁵ Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio, del 19 ottobre 2022, relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (Digital Services Act), in *GUL* 277 del 27 ottobre 2022, 1-102.

ricerca e, potenzialmente, le istituzioni nel monitoraggio degli stereotipi e del linguaggio d'odio nel discorso pubblico italiano. Le esperienze maturate con le "Mappe dell'intolleranza" hanno mostrato la possibilità di geolocalizzare e quantificare l'odio online, individuando le categorie maggiormente colpite e le evoluzioni nel tempo, ma hanno evidenziato anche il carattere fortemente manuale di molti passaggi, difficilmente scalabile rispetto al volume dei contenuti oggi circolanti. AAL muove dall'idea che, per un'analisi costituzionalmente significativa, non basti contare gli insulti espliciti: occorre cogliere la trama più sottile di stereotipi e narrazioni che naturalizzano la disegualianza.

Il progetto combina tre livelli. Anzitutto, la definizione di una tassonomia delle forme di odio e di discriminazione conforme ai parametri costituzionali e sovranazionali, che distingua tra categorie protette, tipi di espressione (insulto diretto, incitamento, disumanizzazione, paternalismo, negazione di diritti), intensità del discorso e contesto comunicativo. Il secondo livello è lo studio e l'utilizzo di strumenti sviluppati in sede internazionale (dalle raccomandazioni del Consiglio d'Europa ai toolkit sul hate speech delle organizzazioni delle Nazioni Unite) per offrire indicazioni utili per calibrare soglie di gravità e differenziare le risposte (contro-narrazione, misure educative, rimedi civili, sanzioni penali). In terzo luogo, si è proceduto alla costruzione di un *corpus* di testi rappresentativi del discorso pubblico online, come post e commenti sui social media, articoli di testate digitali, interventi su piattaforme di dibattito *et similia*, opportunamente anonimizzati. Su un sottoinsieme di questo corpus un gruppo interdisciplinare di esperti ha proceduto ad un'annotazione manuale, classificando le espressioni secondo la tassonomia definita e individuando, in chiave intersezionale, le combinazioni di fattori di discriminazione presenti in ciascun caso. A questo punto, ha avuto luogo l'addestramento di un modello linguistico in grado di apprendere da queste annotazioni.

L'agente che ne è risultato non è concepito come classificatore binario "odio/non odio", ma come strumento di supporto con tre funzioni principali: segnalare, con un forte grado di probabilità, la presenza di espressioni potenzialmente discriminatorie, indicando le categorie coinvolte; fornire una breve spiegazione delle ragioni per cui il contenuto può risultare lesivo della dignità o alimentare esclusione, richiamando in forma sintetica i parametri rilevanti; suggerire possibili riformulazioni che perseguano la stessa finalità comunicativa senza ricorrere a schemi stigmatizzanti. Questo assetto corrisponde a un'impostazione dichiaratamente "umanocentrica": l'agente non sostituisce il giudizio umano, ma lo assiste, offrendo una prima mappatura dei contenuti problematici e un supporto argomentativo. In contesti istituzionali (amministrazioni che vogliano verificare la qualità non discriminatoria della propria comunicazione,

autorità indipendenti chiamate a monitorare l'odio online etc.) AAL può funzionare come filtro preliminare, capace di ridurre il carico di lavoro e di indirizzare l'attenzione verso i casi più critici. In ambito di ricerca, consente analisi longitudinali su grandi quantità di testo, utili a osservare l'evoluzione di stereotipi e narrativa d'odio su specifici gruppi. Un tratto qualificante del progetto è l'integrazione dell'approccio intersezionale nella struttura del modello. Le etichette non si limitano a indicare una categoria bersaglio, ma registrano l'intreccio tra genere, origine etnica o razziale, religione, orientamento sessuale, disabilità, status migratorio. In tal modo l'agente è in grado di evidenziare non solo che un'espressione è discriminatoria, ma come colpisca in modo specifico soggetti situati all'intersezione di più fattori di vulnerabilità, restituendo con maggiore fedeltà l'ampiezza del danno costituzionale prodotto dall'odio e dagli stereotipi.

Resta, tuttavia, un doppio limite allo studio, da un lato la sovra-inclusione, quando vengono qualificati come d'odio discorsi che rientrano nel legittimo dissenso e, dall'altro, la sotto-inclusione, quando non vengono intercettate forme sofisticate di discriminazione implicita, ironica o codificata. Per questo il progetto prevede fasi iterative di valutazione, nelle quali i risultati dell'agente sono costantemente confrontati con il giudizio degli esperti e con le categorie giuridiche e linguistiche di riferimento. In questa prospettiva, AAL non propone un'automazione della tutela dei diritti, ma un uso dell'IA generativa come ausilio al lavoro umano di riconoscimento e decostruzione degli stereotipi. Il suo valore costituzionale non sta nella promessa di sostituire la valutazione del giudice o del garante, bensì nella capacità di rendere più visibili le zone d'ombra del discorso pubblico, fornendo elementi di conoscenza che possano essere impiegati per politiche di prevenzione, strategie di contro-narrazione e, ove necessario, interventi regolatori mirati.

6. Conclusioni.

L'emersione dei sistemi generativi ha reso evidente che la discriminazione algoritmica non si esaurisce nelle decisioni automatizzate che negano l'accesso a un beneficio o aggravano un controllo, ma si iscrive anche nelle forme del parlare e del rappresentare. In questo spazio ibrido, nel quale si incontrano linguaggi d'odio, *bias* nei dati e architetture di piattaforma, il costituzionalismo è chiamato a misurarsi con una dimensione simbolica della disegualianza che la tecnologia può al tempo stesso occultare e rendere più leggibile. Il quadro europeo, con l'AI Act e il Digital Services Act, e quello interno, con la legge 23

settembre 2025, n. 132, offrono primi strumenti per riportare la progettazione e l'uso dell'IA entro l'orbita dei diritti fondamentali; ma la loro effettività dipenderà dalla capacità di declinarne le clausole generali in pratiche concrete di design, audit e rimedio. In questa direzione si collocano progetti sperimentali come AAL, che provano a utilizzare la stessa tecnologia generativa per mappare e problematizzare gli stereotipi, nella consapevolezza che nessun modello può sostituire il giudizio umano, ma che senza strumenti adeguati la scala e la velocità dei contenuti rischiano di rendere inattuabile il progetto costituzionale di pari dignità. Se l'art. 3, comma 2, Cost. impone alla Repubblica di rimuovere gli ostacoli che, limitando di fatto la libertà e l'eguaglianza, impediscono il pieno sviluppo della persona, è difficile negare che oggi una parte di tali ostacoli si annidi anche negli algoritmi che selezionano, producono e raccomandano parole, immagini, narrazioni. La sfida non è "umanizzare" la macchina, ma continuare a umanizzare il diritto, rendendo l'IA oggetto di responsabilità e di scelta politica, piuttosto che lasciarla agire come sfondo inevitabile. L'uso costituzionalmente orientato della IA generativa, anche come strumento di analisi critica del linguaggio, può costituire un tassello, sicuramente parziale, ma non marginale, di questo processo.

PARTE II
GOVERNANCE INCLUSIVA
DELL'INTELLIGENZA ARTIFICIALE:
PRATICHE ISTITUZIONALI, FORENSI E AZIENDALI

CAPITOLO XVI

INTELLIGENZA ARTIFICIALE, DISCRIMINAZIONI
DI GENERE E PUBBLICHE AMMINISTRAZIONI

Elena Buscemi

Il progetto PRIN *AiGeDi*, “*L’intelligenza artificiale da causa a strumento di contrasto delle discriminazioni di genere*”, è segno di un percorso lungimirante tracciato da chi, già da alcuni anni, avvertendo le potenzialità e i rischi dell’intelligenza artificiale, ha deciso di affrontare temi cruciali dello sviluppo tecnologico promuovendo un dialogo serrato tra la realtà accademica e i tanti soggetti, professionali e istituzionali, del tessuto locale, ossia del Comune di Milano.

Le giornate conclusive di questo progetto sono testimonianza di questa cooperazione tesa ad arricchire di un sapere collettivo tutti coloro che possano beneficiare di queste conoscenze in tema di intelligenza artificiale e di contrasto alle discriminazioni, con particolare riguardo a quelle di genere, sia nello svolgimento della propria attività lavorativa sia nella vita quotidiana.

I temi che qui sono chiamati in causa, infatti, non sono solamente tecnologici, ma riguardano questioni politiche – nel senso alto del termine – e, pertanto, giustificano la necessità di un dibattito franco e plurale, arricchito da punti di vista fra loro differenti ma ugualmente indispensabili per un’analisi compiuta e approfondita.

È chiaro che per una città come Milano l’intelligenza artificiale significhi molte cose: per chi si occupa di pubbliche amministrazioni, infatti, la poliedricità di questa tecnologia impone di guardare alle sue multiformi applicazioni prestando attenzione anche ai molti rischi, spesso poco evidenti, che possono generarsi, specie per le minoranze e per chi si trova in una posizione di maggiore vulnerabilità.

Proprio questa estrema varietà di applicazioni pratiche, che sembrano presentare l’IA come panacea di tutte le difficoltà che attanagliano la P.A., impone di indagare diverse direttrici.

– La prima comprende i servizi che riguardano la c.d. *smart city*, ossia la gestione dei flussi, del traffico o degli spostamenti, e che promettono un miglioramento significativo per la vita quotidiana delle persone che usufruiscono dei servizi della città. È chiaro che qui l'impiego dell'IA è teso a garantire una maggiore efficienza e una maggiore qualità dei servizi offerti dalla P.A., ma il rischio è che ci si muova in ordine sparso e non coordinato nella corsa all'adozione di strumenti di intelligenza artificiale. Anche per questo motivo è utile un'opera di ricognizione attraverso cui il Comune possa comprendere come si stiano muovendo le sue società partecipate.

– La seconda direttrice consiste nell'impiego dell'IA nei servizi di *welfare*. Qui il tema noto e alquanto delicato è il trattamento dei dati personali, particolarmente sensibili nel caso dell'assistenza alla persona. L'interoperabilità tra sistemi diversi rischia qui di amplificare gli effetti odiosi delle discriminazioni che i singoli sistemi, replicando la realtà sociale, possono produrre, tramutando i *bias* sociali in *bias* algoritmici. È infatti noto da tempo, e questo progetto lo dimostra, che l'IA non è neutrale, ma anzi può divenire “megafono” per i pregiudizi, che, trovando un pubblico inconsapevole, possono tramutarsi in discriminazioni.

– La terza direttrice è quella che consiste dell'integrazione dell'IA nei processi della P.A., al fine di migliorarne l'efficienza, riducendo i costi e gli sprechi e massimizzando la produttività. La prospettiva, per quanto sicuramente promettente e allettante, non è scevra da rischi, soprattutto in termini di responsabilità e di formazione del personale, che, specie per quanto riguarda le discriminazioni algoritmiche, deve essere adeguatamente preparato.

– Un'altra direttrice ancora, forse trasversale a quelle finora individuate, è rappresentata dalla consapevolezza che debbano essere le istituzioni tutte, dalle università, custodi del sapere e della ricerca, fino alle istituzioni politiche che rappresentano i cittadini, a farsi carico della formazione della cittadinanza sui rischi dell'impiego dell'IA. Sono, infatti, fondamentali momenti di “alfabetizzazione” specificamente cuciti sulle diverse esigenze delle differenti platee per diffondere una cultura consapevole delle potenzialità e dei rischi dell'IA che sia in grado poi di tradursi in applicazioni concrete responsabili e coscienti. È attraverso la c.d. *AI literacy*, infatti, che una conoscenza approfondita dell'intelligenza artificiale può offrire “argini” in grado di resistere a possibili tentativi di “impetuosa” deregolamentazione, di cui si sta facendo esperienza anche oggi, fuori dall'Unione Europea. Come già accaduto con i *social network*, infatti, gli interventi di formazione attuati a valle anziché a monte non si sono mostrati in grado di arginare fenomeni tanto noti quanto diffusi come l'*hate speech* online, di cui poi sono vittime principali le donne.

In conclusione, ritengo che la consapevolezza – che accompagna anche il progetto PRIN AiGeDi, “L’intelligenza artificiale da causa a strumento di contrasto delle discriminazioni di genere” – della necessità di “far rete”, fra istituzioni accademiche, tessuto professionale e produttivo, e pubbliche amministrazioni, sia da custodire e diffondere, per favorire un impiego dell’IA che sia realmente consapevole delle potenzialità e dei rischi di questa tecnologia così pervasiva. L’impatto dell’intelligenza artificiale sui diritti, sulle persone più vulnerabili e sulle democrazie è infatti da tempo al centro della ricerca accademica: ora occorre che le istituzioni e le pubbliche amministrazioni tutte si impegnino per far sì che l’approdo a queste consapevolezze possa diventare patrimonio comune, per utilizzo dell’intelligenza artificiale che non solo non discrimini, ma anzi sia volto a contrastare le discriminazioni con strumenti inediti.

CAPITOLO XVII

INTELLIGENZA ARTIFICIALE E DISCRIMINAZIONI
DI GENERE: IL RUOLO DELLE ISTITUZIONI FORENSI
NEL GOVERNO DI UN ECOSISTEMA INCLUSIVO

Antonino La Lumia

SOMMARIO: 1. Premessa: perché parlare di IA, genere e avvocatura. – 2. IA e discriminazioni di genere: forme, casi, meccanismi. – 2.1. Dal pregiudizio sociale al *bias* algoritmico. – 2.2. Occupazione, selezione del personale e valutazione delle performance. – 3. L'*AI Act* europeo e la centralità del rischio discriminatorio. – 4. IA, genere e responsabilità delle istituzioni forensi. – 4.1. Dall'etica "dichiarata" alla responsabilità "praticata". – 4.2. IA e dovere di riservatezza: una questione anche di genere. – 5. Parità di genere e accessibilità economica ai sistemi di IA. – 6. IA come strumento di contrasto delle discriminazioni: dal rischio all'opportunità. – 7. Conclusioni: un'agenda condivisa per un'IA inclusiva.

1. Premessa: perché parlare di IA, genere e avvocatura.

Il progetto PRIN *AiGeDi* – “*L'intelligenza artificiale da causa a strumento di contrasto delle discriminazioni di genere*” – nasce dall'intuizione, condivisa dall'Università degli Studi di Milano e da un'ampia rete di studiose e studiosi, che l'intelligenza artificiale sia, allo stesso tempo, un fattore di rischio e un possibile alleato nella lotta contro le discriminazioni di genere.

I lavori del convegno “*IA e discriminazioni di genere. Prospettive future per una regolamentazione inclusiva*”, ai quali ho avuto l'onore di partecipare in qualità di Presidente dell'Ordine degli Avvocati di Milano, si collocano precisamente in questo crocevia: tra innovazione tecnologica, tutela dei diritti fondamentali e responsabilità delle istituzioni.

Per l'avvocatura, l'IA non è una semplice “nuova tecnologia”, ma un ambiente in cui si ridisegnano – spesso in modo silenzioso – le condizioni effettive di eguaglianza, di accesso alle opportunità, di esercizio delle libertà. Parlare di discriminazioni algoritmiche significa, in quest'ottica, interrogarsi sul modo in

cui l'ordinamento realizza, o manca di realizzare, la promessa di uguaglianza sostanziale.

La domanda di fondo è semplice, ma radicale: come garantire che sistemi decisionali automatizzati, addestrati e gestiti in contesti ancora profondamente attraversati da stereotipi e asimmetrie di potere, non diventino moltiplicatori di quelle stesse disuguaglianze che il diritto mira a correggere?

2. IA e discriminazioni di genere: forme, casi, meccanismi.

2.1. Dal pregiudizio sociale al bias algoritmico.

La discriminazione algoritmica non è un fenomeno “nuovo” in senso assoluto: è piuttosto la traduzione, in linguaggio matematico e statistico, di pregiudizi e squilibri radicati negli ordinamenti e nelle società. L'algoritmo non “inventa” il *bias*; spesso lo apprende, lo consolida e talvolta lo amplifica.

Nei sistemi di IA addestrati su grandi quantità di dati storici, il genere può entrare in gioco in modi diretti (ad es., quando il sesso è esplicitamente presente come variabile) o indiretti (ad es., quando il genere è inferibile da combinazioni di indicatori: professione, interruzioni di carriera, orari di lavoro, percorsi formativi tipicamente femminilizzati o maschilizzati, ecc.). La discriminazione indiretta risulta talvolta più insidiosa, perché meno visibile, ma non meno efficace nel produrre esiti asimmetrici.

2.2. Occupazione, selezione del personale e valutazione delle performance.

Uno degli ambiti più studiati è quello della selezione del personale e della gestione delle risorse umane, dove sistemi automatizzati vengono utilizzati per scremare *curricula*, valutare competenze, condurre video-colloqui, profilare i candidati.

Studi recenti hanno evidenziato come tali strumenti possano penalizzare, a parità di qualifiche, candidate donne, soggetti con carriere non lineari, persone con accenti non nativi o con disabilità che incidono sul parlato, nonché lavoratrici con periodi di inattività legati a maternità o cura familiare.

La combinazione di dati storici squilibrati (meno donne in ruoli apicali; carriere più frammentate; segregazione settoriale) e metriche algoritmiche centrate sull'“ininterrotta disponibilità” può determinare un effetto di esclusione sistemica, difficilmente riconoscibile dal singolo candidato, ma statisticamente rilevante su larga scala.

A ciò si aggiunge l'uso di sistemi di IA generativa per la preselezione o la valutazione automatizzata dei profili, che possono interiorizzare correlazioni spurie tra genere e "affidabilità", "leadership", "flessibilità" o "compatibilità culturale", riproducendo stereotipi maschili e femminili presenti nei dati testuali o nelle rappresentazioni mediatiche.

3. *L'AI Act europeo e la centralità del rischio discriminatorio.*

Il regolamento (UE) 2024/1689 sull'intelligenza artificiale (*AI Act*) introduce il primo quadro giuridico organico a livello europeo, basato su un approccio "*risk-based*".

Molte delle applicazioni che più direttamente incidono sulla parità di genere rientrano tra i sistemi ad "alto rischio", soggetti a obblighi particolarmente stringenti:

- sistemi di IA utilizzati per selezione del personale, analisi e filtraggio di candidature, valutazione delle performance e promozioni;
- sistemi impiegati per l'accesso a servizi essenziali (credito, assistenza sanitaria, prestazioni sociali), contesti nei quali *bias* di genere possono tradursi in esclusioni o trattamenti deteriori;
- sistemi utilizzati in ambito educativo e di formazione professionale, idonei a condizionare percorsi di studio e accesso a competenze strategiche.

La logica del regolamento è chiara: maggiore è il potenziale impatto sui diritti fondamentali, maggiore deve essere il livello di controllo, di trasparenza, di tracciabilità e di vigilanza *ex ante* ed *ex post*.

La dottrina che ha esaminato l'*AI Act* "attraverso la lente di genere" ha tuttavia messo in luce come il regolamento, pur facendo riferimento alla non discriminazione, non contenga una disciplina pienamente strutturata e autonoma in tema di parità di genere, demandando in larga misura la tutela a clausole generali sui diritti fondamentali e ai meccanismi di valutazione di impatto.

Ciò rende ancora più rilevante il ruolo:

- delle autorità nazionali (inclusi gli organismi per l'uguaglianza) nella raccolta di segnalazioni e nella capacità di far emergere pratiche discriminatorie;
- degli operatori del diritto, chiamati a far vivere, anche nei contesti algoritmici, i principi di eguaglianza e non discriminazione sanciti dalla Carta e dalle Costituzioni nazionali.

4. *IA, genere e responsabilità delle istituzioni forensi.*

4.1. *Dall'etica "dichiarata" alla responsabilità "praticata".*

Il tema delle discriminazioni in ambito IA è "il più delicato", perché investe il punto di incontro tra la pervasività tecnologica e la fragilità dei contesti umani in cui la tecnologia opera.

Per gli ordini forensi, ciò comporta un duplice impegno:

1. Formare avvocati e avvocate a comprendere il funzionamento, i limiti e i rischi dei sistemi di IA, non solo per utilizzarli, ma per saperli contestare in giudizio quando producano esiti discriminatori o lesivi dei diritti.
2. Dare contenuto concreto a principi etici e deontologici che, senza linee guida operative, rischiano di restare enunciazioni astratte.

In questa prospettiva, si colloca l'esperienza dell'Ordine degli Avvocati di Milano, che ha approvato una Carta di principi (Carta HOROS) per un uso consapevole dell'IA in ambito forense, ponendo al centro:

- la centralità della persona e dei diritti;
- la necessità che la decisione resti umana, con IA come strumento di supporto e non sostitutivo;
- l'attenzione alla qualità e alla rappresentatività dei dati;
- il dovere di proteggere la riservatezza delle informazioni affidate al professionista;
- l'impegno a vigilare su possibili effetti discriminatori, anche di genere, derivanti dall'uso di strumenti automatizzati.

Questi principi non intendono "aggiungere" un'etica parallela, ma declinare, nel contesto digitale, i doveri tradizionali dell'avvocato come garante dei diritti e della dignità delle persone assistite.

4.2. *IA e dovere di riservatezza: una questione anche di genere.*

Uno dei punti che ritengo più critici riguarda l'utilizzo di sistemi di IA generalisti per la redazione di atti, pareri o analisi, mediante l'inserimento – spesso inconsapevole – di dati, documenti, strategie difensive, informazioni sensibili dei clienti.

Questi sistemi, se non correttamente configurati, possono:

- incorporare tali dati nei propri processi di addestramento o miglioramento;
- trattarli in infrastrutture collocate in contesti giuridici differenti;
- renderli, in prospettiva, riutilizzabili in modi non pienamente controllabili.

Per le vittime di discriminazioni e violenze di genere, che si rivolgono all'avvocato in condizioni di particolare vulnerabilità, il rischio di una circolazione non governata di informazioni personali – spesso altamente sensibili – è inaccettabile: la fiducia nel sistema di giustizia passa anche dalla certezza che i propri dati non verranno “diluiti” in sistemi opachi.

Da qui la necessità, per gli ordini, di:

- promuovere l'uso di soluzioni tecniche “chiuse”, configurate nel rispetto dei principi di *privacy by design* e *by default*;
- chiarire nei codici deontologici i limiti all'uso di IA generaliste per contenuti affidati in regime di segreto professionale;
- sensibilizzare la categoria sulla differenza, giuridicamente rilevante, tra strumenti *consumer* e strumenti professionali.

5. Parità di genere e accessibilità economica ai sistemi di IA.

Quando discutiamo di pari opportunità in ambito IA, tendiamo istintivamente a pensare al “cosa” l'algoritmo decide. Ma c'è un piano ulteriore, che riguarda il “chi” può permettersi di usare gli strumenti più avanzati.

Negli ultimi anni, si è assistito a una rapida differenziazione dell'offerta: accanto a versioni base, ad accesso gratuito o a basso costo, sono stati introdotti servizi premium – spesso destinati a uso professionale – con capacità molto più elevate e con un costo sensibilmente maggiore.

Questo fenomeno incrocia, in modo non neutrale, le disuguaglianze economiche e di genere:

- se l'accesso a strumenti di IA ad alta prestazione diventa condizione per competere su mercati professionali sempre più digitalizzati, chi non può sostenerne il costo viene collocato in una posizione strutturalmente più debole;
- tale dinamica può riflettersi in settori, come quello legale o imprenditoriale, in cui si incontrano già ostacoli economici e organizzativi maggiori (ad es., per carichi di cura, minore accesso a capitali o reti, ecc.).

In questa prospettiva, la discriminazione non si manifesta solo negli output degli algoritmi, ma anche nella distribuzione delle capacità di usufruire delle tecnologie più avanzate.

Le istituzioni forensi, nella misura in cui promuovono formazione e strumenti per la categoria, possono contribuire a ridurre questo divario:

- negoziando condizioni di accesso eque a soluzioni professionali;

- favorendo piattaforme condivise, sviluppate in logica cooperativa;
- sostenendo iniziative pubbliche che mettano a disposizione delle professioni strumenti di IA conformi ai principi di parità e non discriminazione.

6. *IA come strumento di contrasto delle discriminazioni: dal rischio all'opportunità.*

Il titolo del PRIN *AiGeDi* richiama esplicitamente la possibilità di trasformare l'IA “da causa a strumento di contrasto” delle discriminazioni di genere.

Ciò implica un cambio di prospettiva: non limitarsi alla funzione difensiva – importantissima – di identificare e correggere i *bias*, ma utilizzare le potenzialità dei sistemi di IA per:

- mappare in modo più accurato e tempestivo le disparità esistenti (ad esempio, nelle retribuzioni, nelle progressioni di carriera, nell'accesso a posizioni apicali);
- supportare valutazioni di impatto delle politiche pubbliche e delle pratiche organizzative, in chiave di genere, sfruttando capacità analitiche avanzate;
- sviluppare strumenti di supporto alle vittime, dalla raccolta sicura delle segnalazioni di violenza o discriminazione fino alla costruzione di percorsi di tutela personalizzati.

In questo quadro, l'avvocatura può:

- contribuire a definire standard di “audit algoritmico” che tengano conto delle esigenze probatorie nei giudizi antidiscriminatori;
- partecipare a tavoli di lavoro con autorità indipendenti, università e imprese per definire metriche di equità e trasparenza utilizzabili anche in sede contenziosa;
- utilizzare l'IA stessa per migliorare l'accesso alla giustizia (ad esempio, strumenti di informazione giuridica di base, assistenti per la compilazione di istanze, supporti linguistici per vittime straniere), a condizione che siano progettati in modo inclusivo.

7. *Conclusioni: un'agenda condivisa per un'IA inclusiva.*

Le discriminazioni di genere nell'ambito dell'IA non sono “incidenti di percorso” destinati a scomparire da sole con l'avanzare della tecnologia. Al contrario, l'esperienza dimostra che, in assenza di scelte politiche, giuridiche e culturali

consapevoli, i sistemi automatizzati tendono a consolidare le asimmetrie esistenti, rendendole meno visibili ma più pervasive.

Per evitare questo esito, è necessaria un'agenda condivisa che coinvolga:

- il legislatore, chiamato a dare piena attuazione ai principi di uguaglianza e non discriminazione nei nuovi scenari algoritmici, anche integrando l'*AI Act* con strumenti specifici in materia di genere;
- le autorità di garanzia e gli organismi per l'uguaglianza, che devono essere messi in condizione di intercettare e sanzionare le discriminazioni algoritmiche, inclusi i casi in cui il pregiudizio si annida nei modelli di business o nelle logiche di *targeting*;
- l'università, luogo in cui si costruiscono le competenze critiche necessarie per comprendere l'IA, sia sul piano tecnico che su quello giuridico e filosofico;
- le istituzioni forensi, che devono assumersi la responsabilità di guidare la professione: con formazione mirata, carte dei principi, linee guida deontologiche e iniziative concrete a favore di un uso inclusivo della tecnologia;
- la società civile e le associazioni, in particolare quelle impegnate nei diritti delle donne, che rappresentano un osservatorio privilegiato sui modi in cui le nuove tecnologie incidono sulla vita reale delle persone.

L'intelligenza artificiale non è un destino, ma uno strumento. Se vogliamo che questo strumento sia coerente con i valori inscritti nelle nostre Costituzioni e nelle Carte dei diritti, dobbiamo presidiare i luoghi in cui si decidono i dati, le metriche, le regole del gioco.

L'avvocatura, per la sua funzione di tutela dei diritti e di difesa delle persone più esposte, ha il dovere – e l'opportunità – di essere protagonista in questo percorso, al fianco dell'università e delle istituzioni democratiche.

Solo in questo modo l'IA potrà davvero passare “da causa a strumento di contrasto” delle discriminazioni di genere, diventando parte di un progetto più ampio di giustizia, inclusione e dignità per tutte e per tutti.

CAPITOLO XVIII

INTELLIGENZA ARTIFICIALE RESPONSABILE,
UN VANTAGGIO COMPETITIVO: UN FRAMEWORK
AZIENDALE PER INNOVAZIONE, ETICA E SOSTENIBILITÀ

Gabriele Secol, Beatrice Zulianello

SOMMARIO: SEZIONE I: L'AI Act in azienda. – 1. Gli impatti delle regolamentazioni europee nelle società di consulenza. – 2. L'iniziativa aziendale. – SEZIONE II: Innovazione responsabile. – 3. La sostenibilità dell'innovazione. – 4. Algoritmi discriminatori. – 5. Processo di gestione dei *bias* nei sistemi di intelligenza artificiale. – 6. Sensibilità degli strumenti. – SEZIONE III: Impegno di Deloitte. – 7. Promozione dell'innovazione per il sociale.

SEZIONE I

L'AI ACT IN AZIENDA

1. *Gli impatti delle regolamentazioni europee nelle società di consulenza.*

L'AI Act impone a tutte le imprese di valutare e classificare i propri sistemi di intelligenza artificiale, inclusi quelli sviluppati internamente. Per adempiere a questo obbligo, è fondamentale che chiunque interagisca con tali sistemi sia adeguatamente formato, al fine di utilizzarli in modo ottimale e minimizzare i rischi per le persone e la sicurezza delle comunità.

In ambito aziendale, questi rischi sono spesso legati alla diffusione di materiale sensibile. In generale, le aziende di consulenza hanno accesso a diversi dataset che possono includere dati di identificazione personale (PII: *Personally Identifiable Information*) o dati di business. Infatti, in Deloitte il personale lavora su progetti cliente, di natura riservata, in cui vengono processati dati che l'azienda cliente fornisce in via del tutto confidenziale.

Nel contesto attuale, i dati rappresentano una risorsa fondamentale per le imprese e spesso sono considerati un vantaggio competitivo che deve essere

tutelato con la massima attenzione. La protezione di questi dati è essenziale non solo per mantenere la fiducia dei clienti, ma anche per garantire la crescita dell'impresa stessa e il buon successo dei progetti che affidano alla società di consulenza.

La responsabilità, come società di consulenza, è di proteggere tali dati mentre vengono processati, mantenendo la correttezza nel loro utilizzo e comunicandolo in modo trasparente. In questo processo, si fa uso di strumenti di intelligenza artificiale (IA) che supportano il personale nelle attività lavorative quotidiane, migliorando l'efficienza e l'accuratezza delle operazioni, ad una velocità maggiore. Questi sistemi, però, se non gestiti correttamente, possono esporre i dati sensibili a terze parti, compromettendo la riservatezza e l'integrità delle informazioni. In quest'ottica, dunque, l'AI Act si intreccia necessariamente con un altro pilastro normativo come il GDPR, dedicato alla protezione dei dati, ed occorre mantenere conformità contemporaneamente rispetto ai due regolamenti. Il GDPR rende chiara la collezione, la conservazione e l'utilizzo di tali dati, e l'IA deve necessariamente tenere conto della normativa privacy e della sicurezza, adottando strategie di minimizzazione dei rischi, ad esempio incorporando sistemi di crittografia, anonimizzazione dei dati e una conservazione sicura di essi, per evitare di incorrere in possibili rischi.

Inoltre, occorre stabilire delle procedure interne alle aziende, che partano dalle norme esistenti a livello italiano ed europeo, per garantire un impiego etico dell'IA, assicurandone in primo luogo l'utilizzo trasparente dei modelli e dei loro risultati. Solo così facendo può essere possibile sfruttare al meglio le potenzialità degli strumenti innovativi di IA e favorire allo stesso tempo la fiducia dei clienti, promuovendo una crescita sostenibile ¹.

2. *L'iniziativa aziendale.*

Come primo passo per rispondere al bisogno emerso, in Deloitte è stato deciso di affrontare tale necessità in modo proattivo. Sono stati creati degli strumenti interni e un centro dedicato all'IA generativa, Solaria Space, che risponde pienamente alla necessità di un approccio etico e sicuro per i nostri collaboratori e per i nostri clienti. Questo ci permette di operare al meglio, garantendo la protezione delle informazioni sensibili.

Solaria rappresenta un esempio di innovazione che grazie a figure formate e

¹ J. KURAT, *Ethical AI in Business Consulting: Navigating Challenges*, in *Predictive Analytics*, 2024.

competenti, e a team di lavoro dedicati allo studio, analisi e sviluppo della tecnologia e della regolamentazione, emerge come un modello di eccellenza nel campo dell'intelligenza artificiale applicata in azienda.

Solaria propone al personale interno di Deloitte diversi strumenti che possono facilitare ed efficientare il proprio operato, mantenendo livelli di sicurezza e di trasparenza nell'utilizzo dei dati sensibili elevati. Questo, però, da solo non è sufficiente per evitare il rischio di utilizzo non regolato di altri sistemi open source, ma l'introduzione in azienda di uno strumento quale Solaria ha come sicuro impatto positivo una maggiore consapevolezza acquisita dal personale nei confronti di una tecnologia innovativa che, se non governata, può portare a rischi che possono danneggiare le persone, Deloitte stessa e le aziende clienti.

Nell'ottica di supportare una formazione interna, e anche di supportare le imprese esterne, è nato un team interno a Deloitte, che si dedica a studiare e sviluppare soluzioni che valutano e monitorano i sistemi IA, identificando aree di non conformità e proponendo interventi correttivi. Da queste progettazioni si stanno sviluppando (di concerto con attori esterni anche accademici che possono portare un'ampiezza e una profondità di analisi totalmente sinergica con il valore generato dalle competenze delle società di consulenza) servizi di supporto delle imprese clienti, specialmente quelle con risorse limitate, nell'adeguarsi alle nuove linee guida dettate dall'AI Act, attraverso analisi e definizione di metodologie che misurano gli impatti dei sistemi IA sulle attività aziendali.

Questo approccio non solo aiuta le aziende a rispettare le normative e ad affrontare i cambiamenti normativi, ma anche a utilizzare l'intelligenza artificiale in modo responsabile e sicuro.

SEZIONE II

INNOVAZIONE RESPONSABILE

3. La sostenibilità dell'innovazione.

L'innovazione non può più essere considerata come un'attività che viene svolta "a porte chiuse". Come già visto, l'innovazione che oggi viene richiesta deve avere un carattere più responsabile. La regolamentazione, quindi, non deve essere vista come un ostacolo, ma come uno strumento determinante per una vera innovazione.

L'AI Act può essere vincolante sotto il punto di vista del puro innovatore

che lavora isolato, privo di input esterni e che vorrebbe non avere limiti nella creazione e ideazione di prodotti, per poter sfruttare al massimo la tecnologia.

D'altro canto, la rivoluzione che l'intelligenza artificiale sta portando nella società e la sua integrazione nei processi aziendali, sia a livello operativo sia a livello strategico-decisionale, pongono sfide etiche significative che non possono e non devono essere ignorate. Esse implicano la costruzione di un equilibrio che integri tre dimensioni essenziali nella gestione dell'IA: la tecnica, l'etica e la compliance. Questo approccio garantisce che la tecnologia non solo avanzi, ma che lo faccia in modo sicuro e sostenibile, rispettando le normative e i valori etici.

L'esperienza che emerge in Deloitte, in cui trattiamo l'innovazione in molte sue sfumature, e che una realtà come la nostra ha il dovere di portare sul mercato, è che innovare in modo responsabile e socialmente sostenibile non è più un'opzione, ma una necessità per chi desidera non solo competere, ma emergere nel lungo termine.

Per approfondire il significato di "innovazione responsabile", bisogna avere a mente che l'innovazione che dobbiamo portare avanti deve avere anche uno scopo socialmente sostenibile, e cioè, promuovere il benessere duraturo delle persone interessate; occorre, quindi, soffermarsi su un tema di attuale e fondamentale importanza come l'individuazione di impatti negativi che i sistemi di intelligenza artificiale possono provocare sui diritti fondamentali dell'uomo e, ad esempio, una conseguente discriminazione di genere.

4. Algoritmi discriminatori.

Alla luce di quanto emerso, risulta evidente che la gestione dell'IA non possa prescindere da una riflessione approfondita sulle sue implicazioni etiche e sociali. Infatti, se da un lato l'adozione di soluzioni di IA rappresenta un potente motore di progresso, dall'altro comporta la responsabilità di prevenire e mitigare i potenziali impatti negativi che tali tecnologie possano avere sui diritti fondamentali dell'uomo.

In particolare, uno degli aspetti più critici riguarda la possibilità che i sistemi di IA, se non adeguatamente progettati e monitorati, contribuiscano a perpetuare o addirittura amplificare forme di discriminazione già presenti nella società. Questi rischi non sono solo teorici, ma si manifestano concretamente in diversi ambiti, incidendo sulla parità di trattamento, sull'accesso alle opportunità e sulla tutela della dignità individuale.

L'intelligenza artificiale, infatti, non nasce neutra. Si dice spesso che i suoi

algoritmi sono caratterizzati da *bias* provenienti sia dalle basi dati sia dalle scelte di chi li definisce e che quindi questi algoritmi possono essere discriminatori.

L'IA, infatti, viene educata e allenata sui dati storici collezionati in specifici database che descrivono la società attuale e che, conseguentemente e inevitabilmente, contengono alcuni *bias* che, banalmente, riflettono una cultura non ancora libera da stereotipi sorti nel passato e che al tempo non venivano letti come tali. La cultura e la sensibilità sociale evolvono e, pertanto, dati passati possono rappresentare visioni attualmente superate se non addirittura distorte secondo il pensiero attualmente vigente. Oltre ad un tema temporale esiste un tema spaziale: i database non sono rappresentativi di tutta la popolazione mondiale e gli algoritmi potrebbero essere allenati su alcuni modi di pensare che non sono condivisi da tutte le culture esistenti e generare output non considerati etici in determinati ambiti, culture o religioni.

Per questo motivo, nello sviluppo degli algoritmi si innescano delle “contaminazioni” che portano all’amplificazione dei *bias*, andando a minacciare la reputazione delle imprese anche con il rischio di violare standard etici.

Reminiscenze come la discriminazione dei confronti delle persone afroamericane, o aspettative sedimentate circa i ruoli di uomini e donne nella società e nella famiglia, si possono ritrovare ancora e perpetuare negli algoritmi. Anche se il mercato del lavoro ha vissuto una forte evoluzione negli ultimi 10-15 anni, i dataset con cui oggi vengono allenati molti modelli IA, restano spesso radicati in una storia parziale e in graduale corso di superamento: l'uomo considerato come capofamiglia, la donna che si occupa della casa e dei figli. Di fatto, se non correttamente gestita e supervisionata vi è il rischio concreto che l'IA, invece di contribuire al superamento degli stereotipi, li amplifichi e li consolidi.

Per fare un altro esempio, in ambito aziendale si può incorrere in un rischio discriminatorio nei confronti di candidati nelle attività di recruiting in base al genere, razza o età, che partono dalla stesura delle job description, che involontariamente includono particolari terminologie, spesso appartenenti ad un linguaggio comune del passato, che possono risultare discriminatorie ².

In sintesi, il problema non sembra essere appartenere alla tecnologia in sé, ma al bagaglio di pregiudizi che gli individui rischiano di trasferirle se il training degli algoritmi non viene gestito attivamente e coscientemente. Emerge quindi, l'importanza di includere in azienda figure adeguatamente formate per monitorare questi *bias* e poter introdurre sistemi di mitigazione opportuni. A tal riguardo possono essere indicati tre pillar che devono essere considerati nella

² P. KASHEFI, Y. KASHEFI, A.G. MIRSARAEI, *Shaping the future of AI: balancing innovation and ethics in global regulation*, in *Uniform Law Review*, 3, 2024.

gestione dei sistemi di IA: correttezza negli strumenti di addestramento dei sistemi IA, trasparenza dei dati e “human oversight”, identificazione e mitigazione dei *bias*.

5. *Processo di gestione dei bias nei sistemi di intelligenza artificiale.*

Tali pillar emergono in un ipotetico processo di gestione dei sistemi IA, a partire dalla sua implementazione e addestramento.

Inizialmente, risulta opportuno ampliare il quanto più possibile il range di variabili che possono essere inseriti nei data set di allenamento dei sistemi IA. L’obiettivo consiste nel poter offrire agli algoritmi una base di apprendimento inclusiva e rappresentativa al meglio della popolazione. Vengono in questo caso utilizzati dei sandbox: ambienti virtuali e isolati dedicati al test di software in ambienti che simulano quello reale senza incorrere in possibili rischi.

Il training dei sistemi, inoltre, deve avvenire ad opera di un team multidisciplinare in cui la supervisione non sia solo ad opera di data scientist ma anche di profili esperti di etica, di diritto, di filosofia, di sociologia nonché di linguisti.

I punti precedenti pongono le basi per assicurare algoritmi e dunque sistemi di intelligenza artificiale che operino in modo corretto (rispettando il pillar della *fairness*) e assicurino degli output non discriminatori per un’equa considerazione degli individui e dei gruppi culturali.

Nel fare ciò si introduce un ulteriore pillar, ed è la necessità che i sistemi IA siano controllati e valutati nei loro processi decisionali, secondo il concetto di *human oversight*. Eticamente questo risulta un punto di dibattito a livello mondiale, in cui si sono inseriti negli anni numerose vedute, anche di natura distopica, in cui le attività di ricerca e sviluppo che l’uomo sta portando avanti nell’evoluzione e potenziamento di questa tecnologia sono ritenute pericolose.

Per questo motivo, è essenziale che il giudizio umano continui a svolgere un ruolo di supervisione attiva sui sistemi di IA, prevenendo la diffusione di pregiudizi e garantendo che le decisioni automatizzate siano sempre allineate ai principi di equità e rispetto dei diritti fondamentali. Solo attraverso un’efficace *human oversight* è possibile assicurare che l’IA operi come strumento di supporto e non di sostituzione del discernimento etico umano, ponendo così le basi per una governance responsabile e trasparente, in linea con le più recenti evoluzioni normative e con le aspettative della società.

E se il diritto deve arrivare a normare gli aspetti che a mano a mano diventano più rilevanti per la società aiutandola nella sua evoluzione responsabile ecco che, in parallelo ai dibattiti etici e filosofici, l’Unione Europea è intervenuta con la

pubblicazione dell'AI Act in cui è stata ribadita la centralità dell'uomo nella governance dell'IA, poiché la tecnologia non può sostituire il livello di giudizio etico che l'essere umano possiede. Occorre, quindi, assicurare trasparenza nel modo con cui i diversi sistemi di IA prendono le decisioni e come vengono processati i dati. Il ruolo che in questo contesto, le società di servizi professionali come Deloitte sono chiamate ad avere è quello di acquisire un solido expertise in questo ambito per essere in grado di fornire ai clienti soluzioni che permettano una chiara comprensione dei processi dei tool IA.

L'ultimo fondamentale punto consiste nel monitoraggio dei sistemi, per l'individuazione dei *bias* esistenti e per la loro correzione, prima che essi si radichino nei modelli. Possono essere costituiti degli strumenti di mitigazione basati sulle policy aziendali, sulle regolamentazioni nazionali e sulla collaborazione tra esperti delle diverse discipline. Ogni azienda, in collaborazione con enti esperti, possono realizzare il proprio strumento di monitoraggio e correzione, in base alle proprie caratteristiche e necessità, ma le sfide che sia dal punto di vista logico sia da quello applicativo si aprono sono notevoli, infatti, queste metodologie e strumenti possono solo mitigare quella che inevitabilmente sarà la nascita di sistemi di IA con una propria competenza oltre che sensibilità.

6. *Sensibilità degli strumenti.*

A questo punto si pone un ulteriore tema, ancora più sfidante per le imprese e tutti gli enti interessati nella gestione e utilizzo corretto dei sistemi IA: quanto si è in grado di misurare il livello di sensibilità o di discriminazione di un sistema IA nel confronto con altri algoritmi e modelli IA?

Siamo di fronte a un tema, come si diceva, sia applicativo sia logico o, secondo una differente prospettiva, tecnico e filosofico insieme: da un lato, servono metriche oggettive per valutare il grado di *bias* o di sensibilità di un algoritmo; dall'altro, dobbiamo porci domande nuove: cosa significa che un modello è "più sensibile"? Su quali variabili definiamo questa sensibilità? E con quali benchmark lo confrontiamo?

Esistono alcuni benchmark utilizzati per indicare le performance, le abilità e i punti deboli dei sistemi di IA, per confrontare i modelli, per monitorare i loro progressi. Tali benchmark vengono spesso usati in contesti giuridici, nella definizione di regolamentazioni proprio come nell'AI Act in cui vengono indicati livelli di rischio, alcuni requisiti come quelli di accuratezza, robustezza e sicurezza informatica (cybersecurity).

Una delle prime cause di differenza tra diversi benchmark è che ogni Stato

possiede le proprie linee guida con livelli di accuratezza ed esaustività difficilmente comparabili. Ne consegue che la prima problematica si evidenzia nella mancanza di linee guida globalmente riconosciute per la gestione delle discriminazioni generate dagli algoritmi di IA.

Ciononostante, l'obiettivo delle singole indicazioni o regolamentazioni emesse dagli Stati consiste nel fornire un benchmark e degli strumenti per rendere i modelli di IA più trasparenti e sicuri.

Queste, però, sono però privi di critiche, poiché le regolamentazioni, e i benchmark di riferimento contenuti al loro interno, possono risultare condizionate da diverse strategie politiche ed economiche, che si riverberano nello sviluppo e applicazione degli strumenti di IA.

Risulta chiaro come non solo gli stessi modelli di IA operino in modo differente tra di loro, ma gli stessi benchmark per la loro valutazione non siano uguali, e quindi probabilmente nemmeno del tutto neutri. Devono perciò essere applicati con cautela, e dovrebbero essi stessi essere sottoposti ai principi di trasparenza e correttezza, riportati in precedenza, che vengono usati per la verifica dei sistemi IA. Questo perché anche i dataset generati per queste attività, potenzialmente differenti rispetto a quelli su cui si allenano i sistemi IA, si aprono alle stesse difficoltà di composizione analizzate in precedenza in termini temporali e spaziali: creare un data set che sia comprensivo di tutte le culture e di tutto ciò che viene considerato etico in un dato momento dalla popolazione globale è molto elevata se non insuperabile (sicuramente da un punto di vista pratico forse anche da quello logico).

Può darsi che per far fronte a questa ancora instabile sistema di gestione e valutazione dei sistemi IA possa condurre ad una riflessione etico-filosofica, applicabile in tutti gli ambiti e settori. Quando l'uomo si ritrova ad avere un contatto con un sistema IA e si avvicina al suo utilizzo, il primo confronto può sempre essere fatto con la propria intelligenza umana, la propria cultura, che inevitabilmente non può essere comunemente condivisa con l'intera popolazione globale. Ma l'individuo ha la possibilità di riconoscere quando un output riportato da un sistema di IA va ad intaccare un diritto fondamentale e una certa categoria di persone.

Tale atto di giudizio non può diventare scontato o passare in secondo piano poiché si potrebbe incorrere nel rischio di delegare anche la totale determinazione della verità o del giudizio umano, ragionevole, ai sistemi IA. Non può, però, nemmeno lasciato ad un singolo, ma deve essere aperto al giudizio di un gruppo statisticamente rilevante (almeno secondo un subset di variabili considerate rilevanti). Oppure, come per creare e approvare leggi ci si affida a gruppi scelti di cittadini, per definire un concetto, una base di riferimento anche sui

temi IA ci si potrebbe affidare ad una coscienza umana collettiva. In definitiva, indipendentemente dalla traduzione operativa che verrà applicata per la mitigazione del rischio di valutazione discriminatoria dello strumento, anche in questo caso, questi sistemi di valutazione devono essere basati sulla coscienza collettiva e non su quella di un singolo individuo.

Inoltre, dovremmo costruire framework di audit e di monitoraggio continuo, che non si limitino a una validazione *ex ante* del sistema ma accompagnino l'intero ciclo di vita dell'IA, proprio per non perdere la dimensione temporale della tematica in esame.

Andando avanti col ragionamento si pone un ulteriore tema: questi sistemi essendo educati sulla base della nostra intelligenza, evolveranno fino ad iniziare a “riflettere su sé stessi” e a sviluppare una sensibilità sociale, sviluppando in qualche misura una forma di auto-consapevolezza del proprio ruolo e delle proprie scelte?

E ancora, a valle o in parallelo a questa maggior consapevolezza dello strumento, l'uomo arriverà a trattare i sistemi IA, già facilmente rappresentabili sotto forma di sembianze umane, come avatar, o umanoidi, allo stesso modo di altri individui umani.

Non è solo un tema tecnico, ma anche un nodo etico che, come mondo aziendale, dobbiamo prepararci ad affrontare con approcci proattivi e condivisi con il mondo accademico, istituzionale e legale.

SEZIONE III

IMPEGNO DI DELOITTE

7. Promozione dell'innovazione per il sociale.

Non possiamo, quindi, trattare il tema solo da un punto di vista tecnico ed è proprio questo l'approccio di una società come Deloitte: come visto, l'etica e la responsabilità sociale sono integrate sin dall'inizio nei processi di innovazione e non considerate come aspetti secondari. L'innovazione, infatti, non è solo progresso tecnologico, ma un motore di trasformazione per il business, la società e la cultura.

Infatti, per Deloitte l'innovazione è una leva per ridurre disuguaglianze e problematiche sociali. Oggi innovare non significa solo creare un nuovo prodotto ma introdurre un cambiamento nella società, nel sistema economico di

un territorio. Viene considerato come un cambiamento culturale in grado di generare un valore sostenibile.

Per perseguire questa innovazione integrata, l'approccio di Deloitte prevede il coinvolgimento attivo di tutti gli stakeholder (clienti, dipendenti, comunità, istituzioni, università e imprese partner) nei processi decisionali e di innovazione

Per citare alcuni esempi concreti di iniziative e progetti in cui Deloitte è implicata nella promozione di impatto sociale positivo attraverso l'innovazione, si può sottolineare in particolare la sua affiliazione con l'ecosistema d'innovazione MUSA, attraverso cui vengono seguiti progetti dedicati allo sviluppo di società sostenibili ed inclusive.

In particolare, con il suo team dedicato, Deloitte supporta l'Osservatorio MHEO nel monitoraggio e analisi degli istituti d'istruzione superiore della Città Metropolitana di Milano e le loro interazioni con i soggetti interessati, il centro di ricerca Human Hall cui è stato offerto un supporto strategico ed operativo al progetto di ricerca "Human Hall for AI" dedicato alla creazione di modelli per la valutazione della compliance dei sistemi di IA in relazione al rispetto dei diritti umani oltre ad aver svolto delle ricerche sulla valorizzazione degli Uffici di Trasferimento Tecnologico e sull'impiego o autoimpiego delle risorse a più elevata scolarizzazione.

CAPITOLO XIX

INTELLIGENZA ARTIFICIALE TRA ROTTURA
E TRANSIZIONE: DAL “*RED FLAG ACT*”
ALLO *HUMAN OVERSIGHT*

Giuseppe Vaciago

SOMMARIO: 1. Introduzione: la paura “sbagliata” e il rischio di una nuova grande rottura. – SEZIONE I: Governare senza frenare. – 2. L’*AI Act*: ambizione, cornice di tutela e rischio di *over-regolazione*. – 3. Lo *Human Oversight Officer*: profilo e ruolo operativo. – SEZIONE II: Poteri privati e diritto pubblico. – 4. Le piattaforme come nuovi regolatori privati. – 5. Trasparenza, *auditing* e tutela effettiva degli utenti. – SEZIONE III: Un lessico per il futuro. – 6. Dall’informatica giuridica al diritto dell’informatica. – 7. Proposte operative per i corpi intermedi e gli Ordini professionali. – 8. Conclusioni.

1. *Introduzione: la paura “sbagliata” e il rischio di una nuova grande rottura.*

Nel dibattito sull’intelligenza artificiale si alternano entusiasmi e paure, spesso con la stessa intensità con cui si accompagnano le grandi trasformazioni sociali. La paura, in sé, non è un errore: è un segnale utile, che indica la presenza di rischi e la necessità di comprenderli prima di affrontarli. Diventa però fuorviante quando si dirige verso l’oggetto sbagliato, ossia la tecnologia in quanto tale, oppure quando spinge verso risposte istituzionali sproporzionate, che rallentano l’innovazione senza realmente ridurre i pericoli. La trasformazione tecnologica che stiamo vivendo richiede, più che barriere, strumenti di accompagnamento: alfabetizzazione tecnica per comprendere i meccanismi dell’intelligenza artificiale, procedure di valutazione che garantiscano un controllo consapevole, responsabilità chiaramente definite, trasparenza nelle scelte e una costante capacità di revisione. In questo spazio di equilibrio tra innovazione e garanzia, il diritto può operare in modo realmente efficace, traducendo i principi

costituzionali in requisiti organizzativi e tecnici concreti, verificabili e capaci di adattarsi all'evoluzione continua delle tecnologie digitali.

Il confronto con la stagione rivoluzionaria del 1789 non è un esercizio retorico. Allora come oggi, la promessa di emancipazione convive con il rischio del Terrore: ogni ampliamento del potere, anche di quello computazionale, genera nuove forme di responsabilità e impone un ripensamento dei limiti. Il costituzionalismo contemporaneo non può ignorare che l'intelligenza artificiale potenzia l'azione amministrativa e privata, moltiplica le decisioni automatizzate e ridefinisce il perimetro dei diritti individuali. La lezione che possiamo trarre da quella stagione storica è che i principi non devono restare semplici enunciazioni di valore, ma devono tradursi in dispositivi operativi, proporzionati al rischio e capaci di adattarsi al cambiamento. Solo così il diritto potrà continuare a svolgere la sua funzione originaria: garantire libertà e giustizia anche nei nuovi spazi digitali.

Ed è proprio nel rapporto tra diritto e innovazione che emergono gli errori del passato come avvertimenti per il presente. Ogni volta che il legislatore ha reagito con eccessiva prudenza o ha cercato di contenere la novità con strumenti inadeguati, il risultato è stato quello di frenare il progresso senza accrescere la sicurezza. Emblematico, in questo senso, è il caso del "Red Flag Act" del 1865¹.

Il legislatore britannico impose allora che un uomo precedesse i primi veicoli a motore sventolando una bandiera rossa, con l'intento di garantire la sicurezza pubblica. Pur animata da una finalità legittima, la norma scelse un mezzo incompatibile con la direzione del progresso tecnico. Il risultato fu duplice: un rallentamento dell'innovazione e un'assenza di reale mitigazione del rischio. Questo paradigma di eccesso regolatorio rimane attuale ogni volta che la tutela si ancora a strumenti obsoleti o statici, trasformando la compliance in un rito privo di efficacia concreta. La vera lezione della "bandiera rossa" è che il diritto non deve cercare di normare la tecnologia in sé, ma i suoi usi e i suoi effetti. Le regole più efficaci sono quelle tecnologicamente neutrali e orientate al risultato, capaci di agire sui processi – progettazione, test, esercizio, audit – più che su adempimenti formali o su ruoli privi di reali poteri. In questa prospettiva, il diritto diventa non un freno all'innovazione, ma una guida che la orienta verso esiti sostenibili e coerenti con i valori fondamentali della società democratica².

¹J. AGNEW, *Steam engines on UK roads, 1862-1865: Banning orders, agricultural locomotives and the "red flag" Act*, in *The International Journal for the History of Engineering & Technology*, 90, 1, 53-74.

²L. LESSIG, *Code and Other Laws of Cyberspace*, New York, 2006.

SEZIONE I

GOVERNARE SENZA FRENARE

2. L'AI Act: *ambizione, cornice di tutela e rischio di over-regolazione.*

L'AI Act nasce con l'ambizione di coniugare innovazione e tutela, adottando una logica fondata sul rischio. La sua struttura a livelli di pericolosità lega gli obblighi normativi al tipo di utilizzo, riducendo così l'illusione di un controllo *ex ante* totale e onnisciente. Tuttavia, rimane aperto il rischio che, nella pratica, le prescrizioni si trasformino in adempimenti burocratici che offuscano le responsabilità reali. La differenza, come sempre, la fanno i processi: una documentazione viva e aggiornata, un testing periodico, una revisione critica costante, canali di escalation ben definiti e la capacità di sospendere o riportare a *fallback* i sistemi quando necessario.

Il cuore costituzionale dell'AI Act è rappresentato dallo *human oversight*, ossia dalla supervisione umana effettiva, informata e tempestiva, soprattutto nei sistemi ad alto rischio³. L'art. 14 del regolamento non istituisce un ufficio, ma delinea funzioni e competenze: comprendere i limiti del modello, riconoscere i segnali di allerta e intervenire con leve concrete, come soglie, interruttori di emergenza o procedure di *fallback*. Da questa impostazione è nata l'idea dello *Human Oversight Officer*, una figura analoga per visibilità al *Data Protection Officer*, ma con un mandato tecnico-giuridico più ampio. Tale ruolo ha senso solo se dotato di poteri reali e risorse adeguate: diversamente rischia di ridursi a una mera etichetta, una casella aggiuntiva nell'organigramma priva di efficacia.

3. Lo Human Oversight Officer: *profilo e ruolo operativo.*

Il profilo ideale del supervisore deve essere ibrido, con una prevalenza di competenze tecniche: conoscenza delle pipeline di dati, dei cicli *MLOps* e dei rischi specifici come *adversarial attack* o *prompt injection*; capacità di leggere e interpretare *model cards* e *datasheets*; sensibilità verso l'interazione uomo-macchina e attenzione ai fenomeni di *automation bias*. Dal punto di vista operativo, il supervisore dovrebbe coordinare attività di *red teaming*, *stress test* e analisi delle metriche di robustezza ed equità, in un processo continuo di miglioramento. Le responsabilità, inoltre, non possono essere individuali, ma

³ M. FINK, *Human Oversight under Article 14 of the EU AI Act, AI Act Commentary: A Thematic Analysis*, Londra, 2026.

devono distribuirsi tra diverse figure: *product owner*, *risk owner*, sicurezza informatica, ufficio legale e compliance, riuniti in una cabina di regia comune.

L'implementazione effettiva dell'*oversight* varia a seconda del contesto applicativo. Nel settore pubblico, ciò implica tracciabilità delle decisioni, registri degli incidenti, pubblicazione delle politiche di utilizzo dei modelli e canali trasparenti di contestazione. Nelle piccole e medie imprese la supervisione deve essere leggera ma sostanziale, basata su checklist operative, conservazione sicura dei log e audit mirati sui casi più critici. Le grandi organizzazioni, invece, devono integrare l'*oversight* con i sistemi di gestione esistenti – ISO, sicurezza, qualità – e con i comitati rischio, evitando sovrapposizioni. In ambiti delicati come sanità e lavoro, la supervisione deve mantenere margini di discrezionalità clinica e garanzie contro la discriminazione, ponendo particolare attenzione alla qualità dei dati, agli esiti differenziali e alla validazione in contesti reali.

Il principio di proporzionalità rappresenta il contrappeso necessario per evitare un nuovo “effetto bandiera rossa”. Gli oneri devono essere commisurati all'impatto, al contesto e alla gravità degli esiti potenziali. La calibrazione passa attraverso tre domande fondamentali: quali diritti sono in gioco? Quanto è reversibile l'errore? Quali misure di difesa sono realisticamente attuabili? Da queste valutazioni derivano obblighi differenziati in materia di documentazione, test, monitoraggio in produzione e canali di rimedio.

L'esperienza di altre normative offre importanti insegnamenti. Il *Data Protection Officer* ha mostrato l'importanza dell'indipendenza e dell'accesso diretto alle informazioni, mentre il regolamento DORA ha posto l'accento sulla resilienza operativa, sul testing e sulla gestione delle terze parti. Applicando queste lezioni all'intelligenza artificiale, emerge chiaramente che i presidi sono efficaci solo quando integrati in processi reali di esercizio, monitoraggio e risposta agli incidenti, con indicatori e soglie ben definite e con una catena di responsabilità trasparente.

La robustezza dei sistemi di IA dipende, in larga parte, da una corretta gestione tecnica: *logging* accurato, tracciabilità dei dataset, controllo del ciclo di vita dei modelli e piani di continuità operativa. La risposta agli incidenti deve essere strutturata, con ruoli e tempi di contenimento precisi, comunicazioni tempestive agli *stakeholder* e un processo di apprendimento che consenta di migliorare i modelli a seguito di ogni evento critico. Con i fornitori di modelli e servizi cloud è fondamentale negoziare clausole che garantiscano visibilità sui log, obblighi di cooperazione, audit condivisi e meccanismi chiari di aggiornamento e uscita, per evitare fenomeni di lock-in e opacità contrattuale.

L'esperienza del GDPR costituisce un ulteriore punto di riferimento. Pur avendo introdotto principi fondamentali come accountability, privacy by design e sicurezza, nella pratica molte organizzazioni hanno ridotto l'applicazione del

regolamento a un esercizio documentale. Nel dominio dell'intelligenza artificiale, questo approccio non è sufficiente: servono infrastrutture tecniche solide, controllo rigoroso della qualità dei dati, monitoraggio continuo dei modelli in produzione, tracciabilità delle fonti e meccanismi di revisione umana costante. La cultura dell'incidente – che comprende la segnalazione precoce, l'analisi delle cause e le azioni correttive – deve diventare parte integrante della tutela dei diritti.

I modelli generativi, infine, riportano al centro il tema della data governance. Occorre garantire la provenienza lecita dei dati, la corretta gestione delle licenze, il giusto equilibrio tra trasparenza e tutela dei segreti industriali, oltre a meccanismi di protezione contro l'esfiltrazione o la memorizzazione indesiderata delle informazioni. La qualità dei dati non è un concetto unico ma multidimensionale: include completezza, accuratezza, rappresentatività e assenza di discriminazioni. Da essa dipendono le performance e la *fairness* del modello. Strumenti come il *reinforcement learning from human feedback*, i filtri di sicurezza e i controlli sugli output possono mitigare i rischi, ma non sostituiscono una solida politica dei dati. Quest'ultima resta l'unico antidoto strutturale contro *bias*, derive algoritmiche e perdita di controllo umano sul processo decisionale⁴.

SEZIONE II

POTERI PRIVATI E DIRITTO PUBBLICO

4. *Le piattaforme come nuovi regolatori privati.*

Le grandi piattaforme digitali esercitano oggi un potere che si avvicina, per pervasività e capacità regolatoria, a quello delle istituzioni pubbliche. Non si limitano a fornire servizi o infrastrutture tecnologiche: stabiliscono le condizioni di accesso, definiscono regole di comportamento, impongono standard tecnici e disciplinano – in modo unilaterale – i meccanismi di visibilità e distribuzione dei contenuti. Attraverso termini di servizio, politiche di utilizzo, procedure di moderazione e scelte di design, esse modellano concretamente il modo in cui si formano l'informazione, l'opinione pubblica e persino la concorrenza economica. Le modifiche improvvisate alle policy, alle condizioni di accesso alle API o ai criteri di ranking dimostrano quanto il potere privato delle piattaforme possa, di fatto, anticipare o addirittura sostituire l'intervento normativo.

⁴ R. PARASURAMAN, *Humans and Automation: Use, Misuse, Disuse, Abuse*, in *Human Factors*, 39, 2, 1997, 230.

Di fronte a questa trasformazione, la risposta non può consistere in una contrapposizione ideologica o in un ritorno al controllo statale rigido, ma nella costruzione di un modello di co-regolazione maturo e dinamico. In tale prospettiva, le regole pubbliche e le regole private devono dialogare in modo strutturato, dando vita a un sistema di governance condiviso che coniughi autonomia d'impresa e tutela dell'interesse collettivo.

Gli strumenti di questa nuova architettura non sono solo le norme imperative, ma anche standard armonizzati, audit indipendenti, obblighi di trasparenza metodologica e meccanismi di responsabilità verificabile. L'obiettivo è creare un equilibrio tra innovazione e garanzia, dove la conformità non sia un atto formale, ma un processo continuo di responsabilizzazione tecnica e giuridica.

5. *Trasparenza, auditing e tutela effettiva degli utenti.*

La trasparenza, in questo contesto, non può ridursi alla semplice apertura del codice o alla pubblicazione di documenti tecnici incomprensibili. Deve piuttosto tradursi nella capacità di rendere intellegibili i processi decisionali, le basi di dati utilizzate e le metriche che orientano la distribuzione di contenuti o l'erogazione di servizi automatizzati. Solo una trasparenza "utile", capace di comunicare il significato delle scelte algoritmiche, consente agli utenti, ai ricercatori e alle autorità di comprendere e valutare l'impatto reale delle tecnologie digitali.

In parallelo, gli audit dovrebbero assumere un ruolo centrale come strumento di verifica effettiva, andando oltre la conformità formale per indagare la robustezza dei sistemi, la sicurezza delle infrastrutture, la qualità dei dati e gli impatti differenziali sulle diverse categorie di utenti. L'audit indipendente diventa così un presidio di accountability, capace di introdurre un controllo *ex post* fondato su evidenze e non su mere dichiarazioni di principio ⁵.

Agli utenti devono essere garantiti strumenti concreti di tutela: il diritto di contestare decisioni automatizzate, di ottenere una revisione umana in casi di rilievo, di accedere a informazioni chiare sulle logiche che determinano raccomandazioni, esclusioni o priorità. Questi diritti non rappresentano un ostacolo all'efficienza tecnologica, ma ne costituiscono la condizione di legittimità. Solo una partecipazione informata e un controllo effettivo possono ricostruire la fiducia nelle piattaforme digitali, riducendo le esternalità negative che oggi incidono sul discorso pubblico, sulla concorrenza e sull'integrità dei mercati informativi.

⁵ J. MÖKANDER, L. FLORIDI, *Auditing of AI: Legal, Ethical and Technical Approaches*, in *Digital Society*, vol. 2, Article 49, 2023.

SEZIONE III

UN LESSICO PER IL FUTURO

6. *Dall'informatica giuridica al diritto dell'informatica.*

La tradizione giuridica italiana ha sempre distinto due prospettive: da un lato l'“informatica giuridica”, intesa come l'uso delle tecnologie digitali al servizio del diritto; dall'altro il “diritto dell'informatica”, cioè l'insieme delle norme e dei principi che regolano lo sviluppo, la diffusione e l'impatto delle tecnologie dell'informazione. Oggi questa distinzione, pur utile dal punto di vista storico, appare sempre meno sufficiente a descrivere la complessità del presente. La digitalizzazione non è più uno strumento esterno al diritto, ma un ambiente in cui il diritto stesso si forma, si interpreta e si applica. In questo contesto, il diritto dell'informatica non è più una nicchia specialistica, ma il terreno su cui si costruisce la capacità delle istituzioni e dei professionisti di comprendere e governare la società algoritmica.

Ciò comporta un vero cambio di paradigma. I giuristi non possono più limitarsi a “interpretare” la tecnologia secondo categorie tradizionali, così come gli informatici non possono più sviluppare sistemi ignorando le implicazioni normative e valoriali delle proprie scelte. È necessario un dialogo metodico e costante tra saperi, in cui il linguaggio giuridico e quello tecnico si contaminano reciprocamente, generando una nuova forma di bilinguismo operativo. Questo bilinguismo significa saper leggere, con pari familiarità, un'architettura applicativa e una clausola di responsabilità contrattuale, un log di sistema e una norma sul trattamento dei dati personali, una curva ROC e un principio di proporzionalità. Solo attraverso questa integrazione linguistica e concettuale è possibile tradurre i principi in procedure effettive, rendendo il diritto non un ostacolo, ma un alleato della progettazione tecnica ⁶.

Questo nuovo approccio implica anche una trasformazione culturale: il diritto deve diventare capace di operare con strumenti di modellazione, versionamento, testing e validazione, così come la tecnologia deve interiorizzare la logica dei principi giuridici – legalità, trasparenza, proporzionalità, non discriminazione – come vincoli progettuali e non come correzioni *ex post*. In altre parole, la sfida non è “regolare la tecnologia”, ma costruire un ecosistema in cui la tecnologia stessa incorpori la normatività del diritto. Solo così il lessico giuridico potrà continuare a essere un linguaggio di garanzia anche in un mondo in cui le decisioni sono sempre più mediate da algoritmi.

⁶J. NAY, *Law Informs Code: A Legal Informatics Approach to Aligning Artificial Intelligence with Humans*, in *Northwestern Journal of Technology and Intellectual Property*, 20, 3, 2022.

7. *Proposte operative per i corpi intermedi e gli Ordini professionali.*

Tradurre questa visione in pratica richiede un impegno organizzativo preciso da parte dei corpi intermedi, delle associazioni professionali e degli Ordini. Non si tratta di immaginare grandi riforme astratte, ma di definire un programma minimo, concreto, verificabile, che permetta di costruire una cultura della responsabilità tecnologica. Il primo passo è la redazione di linee guida interne realmente “vive”, cioè soggette a revisione periodica, con esempi applicativi e criteri di valutazione aggiornati in base all’evoluzione tecnica e normativa. Tali linee guida dovrebbero essere affiancate da un comitato interdisciplinare per l’intelligenza artificiale, composto da giuristi, tecnologi, esperti di sicurezza e rappresentanti della deontologia professionale, con un mandato chiaro di valutazione e supervisione.

Il secondo passo riguarda la gestione del rischio. Occorre distinguere i processi di valutazione in base alla natura delle soluzioni adottate: modelli preconfezionati (*off-the-shelf*), modelli personalizzati tramite fine-tuning e modelli sviluppati internamente. Ogni categoria richiede metriche, test e controlli diversi. Nei contesti più sensibili, è necessario mantenere un sistema di human-in-the-loop, con soglie di intervento, ruoli definiti e procedure di *fallback* che consentano di sospendere automaticamente un modello in caso di risultati anomali. Tutte le attività devono essere accompagnate da audit tecnici indipendenti e da un registro degli incidenti di intelligenza artificiale, in cui ogni evento sia documentato, analizzato e seguito da un rapporto di *post-mortem* obbligatorio.

Un altro elemento centrale riguarda i rapporti con le terze parti. I contratti con fornitori di modelli o servizi cloud devono prevedere obblighi espliciti di conservazione dei log, di cooperazione nelle indagini e di apertura agli audit. La trasparenza non è un principio solo giuridico, ma anche operativo: senza accesso ai dati di performance e agli aggiornamenti dei modelli, nessuna organizzazione può garantire il rispetto dei principi di equità e accountability. In parallelo, è necessario adottare metriche di equità che traducano i principi giuridici in indicatori tecnici misurabili, collegando la *fairness* a criteri di proporzionalità, non discriminazione e impatto differenziale.

Per evitare che queste iniziative rimangano dichiarazioni di principio, è utile costruire una roadmap di adozione scandita per fasi temporali⁷. Nel primo trimestre, le organizzazioni dovrebbero concentrarsi sulla mappatura dei casi d’uso, sull’individuazione dei criteri di rischio e sulla definizione di clausole

⁷ A. VILLANI, C. MEDAGLIA, L. TANGI, *AI governance: a systematic literature review*, in *AI & Ethics*, 4, 2024, 563-582.

contrattuali minime per i fornitori. Nel secondo trimestre, avviare progetti pilota basati su documentazione by design, log centralizzati e metriche iniziali di qualità e *fairness*, accompagnati da percorsi formativi differenziati per giuristi e tecnici, così da ridurre le asimmetrie culturali. Nel terzo trimestre, istituire formalmente il comitato IA, sottoporre i progetti pilota a una verifica indipendente, attivare il registro degli incidenti e definire procedure di escalation e comunicazione verso gli stakeholder. Infine, nel quarto trimestre, procedere all'estensione controllata ai casi d'uso a medio rischio, integrando i controlli con i sistemi di gestione già esistenti – ISO, sicurezza, qualità – e prevedendo una revisione complessiva delle policy.

Questa roadmap, flessibile e adattabile alle diverse realtà organizzative, rappresenta un antidoto contro due pericoli opposti: da un lato, la paralisi operativa dovuta all'eccesso di cautela; dall'altro, l'adozione incontrollata di sistemi di intelligenza artificiale "senza rete". L'obiettivo è costruire un equilibrio sostenibile tra innovazione e garanzia, tra autonomia tecnica e responsabilità etico-giuridica, trasformando l'*oversight* da obbligo formale a pratica quotidiana e consapevole.

8. Conclusioni.

L'intelligenza artificiale evolve a una velocità che il diritto, per sua natura, fatica a eguagliare. Le regole si fondano su procedure lente, deliberative, costruite per garantire ponderazione e prevedibilità; l'IA, al contrario, cresce per iterazioni rapide, aggiornamenti continui, sperimentazioni incessanti. È in questo divario di tempi che si gioca la sfida più profonda del nostro tempo: non fermare l'innovazione, ma imparare ad accompagnarla. La scelta non è, e non deve mai diventare, una contrapposizione tra lassismo e paralisi. La vera alternativa è un diritto capace di stare accanto ai processi tecnologici, di comprenderne le logiche, di anticiparne i rischi e di tradurre i propri principi in strumenti operativi.

Accompagnare significa riconoscere che la normatività non risiede solo nella legge scritta, ma anche nei processi, nelle prassi e nelle responsabilità condivise. Significa adottare un modello di governance che intrecci *hard law* e *soft law*, combinando la forza vincolante delle norme con la flessibilità degli standard tecnici, delle linee guida, dei codici di condotta e delle pratiche di auditing indipendente. Questo equilibrio – proprio dell'approccio europeo – consente di coniugare certezza giuridica e adattabilità, evitando sia l'irrigidimento burocratico sia l'anarchia tecnologica. Il diritto, così inteso, non è una barriera all'innovazione, ma la sua infrastruttura di fiducia: la cornice che permette di sperimentare senza compromettere la dignità e la sicurezza delle persone.

L'accompagnamento implica anche una trasformazione organizzativa profonda. Non bastano regole astratte o principi generali: occorrono persone formate, ruoli riconosciuti e leve decisionali concrete. Lo *human oversight* ne è l'esempio più emblematico. Non si tratta di un'etichetta o di un obbligo formale, ma di una pratica viva, capace di coniugare competenze tecniche e sensibilità giuridica. Richiede supervisor che comprendano i limiti dei modelli, che sappiano leggere i segnali di rischio, intervenire in tempo reale, valutare l'impatto e, se necessario, fermare un sistema. È la traduzione operativa di un principio antico: il diritto come presidio di umanità nelle scelte automatizzate.

Accompagnare, tuttavia, significa anche costruire una cultura del miglioramento continuo⁸. Ogni incidente, ogni deviazione, ogni *bias* individuato deve diventare occasione di apprendimento collettivo. È questa la differenza tra un sistema che reagisce agli errori e uno che li previene. Il diritto deve promuovere processi di revisione sistematica, audit trasparenti, meccanismi di correzione che non siano percepiti come sanzione ma come occasione di crescita. In tal modo la conformità non si riduce a un adempimento rituale, ma diventa parte della strategia di qualità e di responsabilità sociale dell'impresa.

L'immagine della "bandiera rossa" resta un monito potente. Come nel 1865, il rischio non è tanto quello di proibire la tecnologia, quanto di regolarla in modo tale da impedirne il progresso. Oggi, più che mai, il diritto deve evitare di camminare davanti all'innovazione sventolando segnali di allarme, ma deve restare al suo fianco, guidandola con consapevolezza, prudenza e fiducia. Il compito delle istituzioni non è frenare il cambiamento, ma renderlo governabile, intelligibile, conforme ai valori costituzionali che fondano la nostra convivenza.

In questa prospettiva, il diritto torna a essere non la voce che dice "no", ma quella che indica "come". È lo strumento che consente alla società di procedere lungo la strada dell'innovazione senza smarrire la propria direzione etica. Accompagnare l'IA significa dunque riconoscere che il progresso tecnologico non è un destino da subire, ma un percorso da orientare. Solo un diritto capace di apprendere, dialogare e adattarsi potrà evitare la nuova "bandiera rossa", restando fedele alla sua missione originaria: tutelare la libertà dell'uomo anche quando a decidere non è più soltanto l'uomo.

⁸ C. NOVELLI, P. HACKER, J. MORLEY, J. TRONDAL, L. FLORIDI, *A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities*, in *European Journal of Risk Regulation*, 16, 2, 2025.

CAPITOLO XX

L'INTELLIGENZA ARTIFICIALE IN A2A

Patrick Oungre

SOMMARIO: 1. Introduzione: *overview* sull'utilizzo e sviluppo di sistemi di intelligenza artificiale in A2A. – 2. I pillar di A2A e i progetti. – 2.1. L'obiettivo. – 2.2. Agenti IA a supporto dei team. – 2.3. Tecnologie agentiche a supporto dei processi. – 2.4. L'IA fisica. – 2.4.1. Progetto Prometeo. – 2.4.2. Progetto Pandora: manutenzione predittiva della rete elettrica. – 2.4.3. Progetto Ottimizzazione calore erogato impianti di teleriscaldamento.

1. *Introduzione: overview sull'utilizzo e sviluppo di sistemi di intelligenza artificiale in A2A.*

Già dal 2021 sono stati sviluppati ed implementati i primi sistemi di intelligenza artificiale a supporto delle attività lavorative dell'azienda, per migliorare l'efficienza operativa, ridurre l'impatto ambientale e ottimizzare la gestione delle risorse

Ad oggi, più di 200 sistemi di IA risultano essere in produzione e vengono impiegati quotidianamente, spaziando da utilizzi nelle attività del *customer care* a quelle della manutenzione predittiva degli asset di rete, alla previsione di produzione degli impianti, all'ottimizzazione della raccolta dei rifiuti e pulizia delle strade, alla redazione e analisi di documentazione e all'interfacciamento semplificato con grossi moli di dati numerici e testuali.

2. *I pillar di A2A e i progetti.*

2.1. *L'obiettivo.*

L'intelligenza artificiale è uno dei principali motori di trasformazione per "A2A Life Ventures", perché permette al Gruppo di accelerare il processo di

decarbonizzazione e innovazione grazie a soluzioni tecnologiche avanzate, sia per le operazioni interne sia per le infrastrutture fisiche. La visione strategica di A2A Life Ventures punta a integrare l'IA in ogni area operativa, con l'obiettivo di ottimizzare e automatizzare i processi, rendere accessibili gli strumenti di IA a tutti, migliorare l'efficienza operativa e promuovere la sostenibilità.

2.2. Agenti IA a supporto dei team.

L'obiettivo che ci poniamo è quello di automatizzare e agentificare processi *time consuming* per A2A al fine di poter incrementare l'efficienza riducendo tutte le attività manuali, creando sempre più gruppi composti sia da persone che da agenti AI.

Due dei progetti dai quali siamo partiti, in forma autoreferenziale, sono:

1. "One plus One": piattaforma unificata, che integra diversi **agenti AI** per supportare le **attività operative e decisionali** dei team aziendali, alcune delle sue funzionalità sono la ricerca facilitata di contenuti relativi a startup e documenti rilevanti per la funzione di innovazione, predisponendo un monitoraggio continuo di trend di mercato;
2. "AI Superfactory" mira a diventare un modello di eccellenza per lo sviluppo di soluzioni AI, integrando agenti intelligenti in tutte le fasi del ciclo di produzione del software. L'obiettivo è realizzare una piattaforma più **scalabile, efficiente e innovativa**, in cui agenti AI operino come veri e propri **co-piloti**, automatizzando le attività ripetitive e a minor valore aggiunto, migliorando la **qualità del software**, incrementando **produttività e velocità di rilascio** e riducendo **errori, rischi e inefficienze**.

2.3. Tecnologie agentiche a supporto dei processi.

In un contesto tecnologico sempre più complesso, nasce la necessità di garantire continuità operativa, sicurezza e qualità dei servizi nell'ambito delle *operation*, anticipando i disservizi, riducendo i tempi di intervento, migliorando l'esperienza di utenti e operatori. Questo è un tema molto sentito all'interno di grosse aziende come la nostra.

Qualche spunto su attività progettuali:

- Gestione del triage e riduzione dei "multiple hops".
- Risoluzione assistita e automatizzata.
- Correlazione degli alert (AIOps).

2.4. *L'IA fisica.*

L'IA fisica è un ramo dell'intelligenza artificiale che si concentra sull'integrazione tra sistemi intelligenti e il mondo fisico.

Questi sistemi combinano sensori, attuatori, algoritmi di apprendimento automatico e capacità cognitive per svolgere compiti complessi come navigare in spazi reali, manipolare oggetti, collaborare con esseri umani o altri robot.

Essendo A2A un'azienda di infrastrutture, negli anni ci siamo concentrati molto nello sviluppo di soluzioni per i nostri asset; di seguito indico una serie di progettualità.

2.4.1. *Progetto Prometeo.*

Il progetto Prometeo, si concentra sulla previsione accurata del carico sulla rete elettrica di Milano tramite l'utilizzo di dati provenienti da diverse fonti, come tensioni e correnti aggiornate e previsioni meteo giornaliere al fine di anticipare situazioni critiche e supportare la creazione di un mercato di flessibilità locale. Dal momento che le tematiche trattate sono fortemente sperimentali ci siamo aggiudicati un finanziamento da ARERA.

Gli obiettivi principali del progetto sono:

- prevedere il carico elettrico sulla rete in media tensione di Milano a lungo, breve e brevissimo termine in modo da efficientare la gestione del carico e così evitare sovraccarichi, instabilità e inefficienze operative;
- ottimizzare l'attivazione dei servizi di flessibilità che prevedono la richiesta di riduzione dei consumi a certe tipologie di utente e di aumento della produzione in specifici momenti della giornata in cambio di una remunerazione economica.

L'approccio si basa sull'utilizzo di dati provenienti da diverse fonti, come tensioni e correnti aggiornate ogni 15 minuti e previsioni meteo giornaliere, che permettono a Prometeo di fornire previsioni aggiornate per ciascun asset monitorato in modo che gli operatori possano anticipare eventuali congestioni della rete e prendere decisioni per attivare i servizi di flessibilità. Questo approccio non solo migliora l'efficienza della rete, ma ne aumenta anche l'affidabilità e la stabilità.

La soluzione proposta dal progetto si basa sull'utilizzo dell'Intelligenza Artificiale per prevedere in modo accurato il carico elettrico degli asset monitorati. A questo scopo vengono impiegati diversi modelli di machine learning e modelli statistici di forecasting, che analizzano dati storici di consumo insieme a variabili meteorologiche, come temperatura e radiazione solare, oltre a informazioni di

calendario. Grazie a questa combinazione di dati, il sistema è in grado di anticipare il carico richiesto da ciascun asset.

L'attivazione ottimizzata degli asset è poi gestita tramite algoritmi di ricerca operativa, che permettono di individuare la sequenza di attivazioni più efficiente.

I benefici attesi includono una gestione più efficiente della rete elettrica, con conseguente riduzione dei costi operativi e dei fenomeni di sovraccarico. Inoltre, l'uso di queste tecniche avanzate contribuisce a migliorare la stabilità e l'affidabilità complessiva della rete.

2.4.2. Progetto Pandora: manutenzione predittiva della rete elettrica.

Il progetto Pandora mira a sviluppare un sistema di manutenzione predittiva per la rete elettrica di Milano. Questo progetto è nato dall'esigenza di migliorare l'affidabilità e l'efficienza della nostra rete, affrontando le sfide legate alla manutenzione predittiva. Gli obiettivi del progetto sono:

- Supportare la pianificazione degli interventi di diagnostica.
- Supportare la stesura del piano di resilienza nazionale comunicato all'autorità di regolazione per energia reti e ambiente (ARERA).
- Supportare la sala controllo nell'individuazione della tratta in guasto.

Pandora introduce un nuovo modello che integra diverse sorgenti di dati, tra cui dati anagrafici e topologici della rete, dati sui guasti e sulle interruzioni degli ultimi anni, variabili meteorologiche e carichi su ciascuna tratta.

Utilizzando tecniche avanzate di intelligenza artificiale, il progetto prevede il calcolo di un tasso di guasto dinamico, supportando le azioni sui servizi di flessibilità. La creazione di un mercato della flessibilità locale si pone come obiettivo quello di ottimizzare l'utilizzo della rete e bilanciarne domanda e offerta, questo approccio non solo migliora l'efficienza della rete, ma ne aumenta anche l'affidabilità e la stabilità.

Il progetto Pandora si distingue per l'originalità dei risultati attesi, valutati rispetto allo stato dell'arte tecnologico, di processo e organizzativo sia a livello nazionale che internazionale. L'innovazione proposta è significativamente diversa rispetto alle alternative esistenti, grazie all'uso di tecnologie avanzate di IA e all'integrazione di dati eterogenei.

I risultati preliminari evidenziano come l'applicazione del modello di manutenzione predittiva consenta di incrementare in modo significativo l'efficacia degli interventi di diagnostica sulla rete elettrica di Milano, con un miglioramento dal 23% al 43%, pari a un incremento dell'87%. Questo si traduce in

una maggiore capacità di individuare tempestivamente le criticità della rete, riducendo la probabilità che si trasformino in disservizi per la città.

2.4.3. *Progetto Ottimizzazione calore erogato impianti di teleriscaldamento.*

Il progetto di ottimizzazione del teleriscaldamento mira a migliorare l'efficienza energetica degli impianti di teleriscaldamento attraverso l'uso di tecnologie avanzate di intelligenza artificiale. Questo progetto è nato dall'esigenza di ridurre gli sprechi energetici e migliorare la pianificazione e l'esercizio degli impianti.

Gli obiettivi principali del progetto sono:

- Ottimizzazione dell'energia termica erogata: migliorare la pianificazione e l'esercizio degli impianti attraverso un forecast orario del carico termico prodotto dalle centrali.
- Riduzione degli sprechi energetici: implementare sistemi di intelligenza artificiale per prevedere con maggiore accuratezza il carico termico, riducendo così le emissioni di CO₂ e migliorando l'efficienza complessiva.

Attualmente il sistema di previsione è stato sviluppato su 10 impianti del Gruppo, con performance molto soddisfacente anche rispetto a *tool* di mercato.

Per ogni impianto si prevede un risparmio di CO₂ di 10.000 ton CO₂/y, dei 10 impianti 5 hanno anche il sistema di *optimal dispatching* che porta alla riduzione della CO₂.

CAPITOLO XXI

INTELLIGENZA ARTIFICIALE E IMPRESA:
TRA INNOVAZIONE TECNOLOGICA
E RESPONSABILITÀ CULTURALE

Alessandro Renna

SOMMARIO: 1. Premessa. – 2. Il vero problema non è la tecnologia. – 3. L'IA come leva competitiva delle organizzazioni: la nostra esperienza. – 4. L'IA come abilitatore di una nuova offerta tecnologica in ambito *legal*. – 5. Questioni interpretative aperte: il concetto di prevalenza. – 6. Conclusioni: un approccio europeo.

1. *Premessa.*

Prima di portare la mia opinione su alcuni temi che riguardano l'intelligenza artificiale e che ho affrontato nel mio lavoro di imprenditore, ritengo necessario un ragionamento di contesto che colleghi le sollecitazioni emerse durante i lavori del convegno, e in particolare le riflessioni sulla discriminazione algoritmica e la riproduzione culturale nelle tecnologie.

2. *Il vero problema non è la tecnologia.*

Il concetto di riproduzione culturale e radicalizzazione, analizzato con riferimento ai social network ma certamente estensibile all'intelligenza artificiale, ci pone di fronte a una verità scomoda: queste tecnologie tendono a restituirci ciò che siamo e che ci aspettiamo, amplificando i nostri *bias* culturali e personali.

Se questo è vero, la questione relativa a un utilizzo “etico” delle tecnologie non può esaurirsi nel controllo tecnico dei dataset di addestramento o nella vigilanza sulle discriminazioni algoritmiche, profili che – peraltro – sfuggono alla nostra concreta sfera di influenza.

Il centro della questione è a mio avviso a monte, e cioè nell'etica, nei valori e nella cultura dell'utente che utilizza questi strumenti. È anzitutto su noi stessi che dobbiamo e possiamo lavorare, insieme, se vogliamo che l'intelligenza artificiale dia risultati etici. È su questo che le istituzioni e le organizzazioni leader della business community devono assumere un impegno chiaro, coerente e onesto, un impegno che spesso non si riscontra tuttavia nell'attualità.

Preoccupa, in questa prospettiva, la carenza di leadership di gruppi globali che abdicano a qualsiasi coerenza etica in favore di comportamenti che potremmo certamente definire "riproduttivi" proprio nel senso di restituire all'esterno il messaggio che il contesto richiede e si aspetta, a prescindere da qualsiasi coerenza valoriale.

Si pensi al recente "caso Meta": un'azienda leader globale che aveva fatto della sostenibilità, dell'inclusione e dei principi ESG una cifra identitaria e che ha rapidamente abbandonato questi valori in seguito a mutamenti del contesto politico che li hanno resi invisibili al potere.

Il "riproduttivismo" non fa che amplificare e blandire lo status quo, ed è un modo di essere pericoloso in ambito tecnologico e non. Cultura, etica e consapevolezza sono il vero terreno su cui affrontare *bias*, pregiudizi e discriminazioni algoritmiche.

L'Università è una sede privilegiata in cui affermare il nostro impegno, come operatori tecnologici, verso un approccio critico e valoriale all'intelligenza artificiale.

3. L'IA come leva competitiva delle organizzazioni: la nostra esperienza.

Venendo agli aspetti più operativi, nel Gruppo 4C guardiamo all'intelligenza artificiale da diverse prospettive, tutte orientate alla creazione di un valore tangibile. La prima prospettiva è quella dell'IA come strumento di efficienza e competitività interna.

Per portare un'esperienza concreta, posso dire che in azienda utilizziamo sistemi di intelligenza artificiale in modo pervasivo in cinque aree: attività editoriali, comunicazione, servizi legali, *finance* e sviluppo software.

Chi ha una formazione giuridica, come nel mio caso dopo anni di pratica forense, può apprezzare quanto sia rivoluzionario non dover più redigere un contratto partendo da modelli preesistenti o addirittura da zero, ma utilizzando prompt ben strutturati e ottenendo un ottimo prodotto in qualche minuto. Nell'ambito legale, che comprende anche l'analisi di documenti complessi e lo

svolgimento di ricerche giuridiche, il software che abbiamo sviluppato e che proponiamo al mercato (4cAi X) è un'eccellente risorsa prima di tutto per noi.

L'intelligenza artificiale si rivela preziosa anche per affrontare ambiti nei quali abbiamo competenze meno puntuali. Quando dobbiamo preparare un pitch per investitori, elaborare una presentazione per una banca o costruire un prospetto economico-finanziario, l'intelligenza artificiale rende queste attività accessibili anche a chi come me non ha un background economico ma ha acquisito sul campo competenze che richiedono un "partner tecnico" per essere opportunamente supportate.

Oltre all'ambito *finance*, anche quello della comunicazione vede nell'IA un alleato strategico: i post su LinkedIn, le newsletter, i programmi dei convegni sono tutte attività che gestiamo con l'IA risparmiando tempo e generando output pienamente coerenti con la nostra cifra stilistica.

E da ultimo, anche lo sviluppo software merita grande considerazione. Le proiezioni indicano che in pochi anni la stragrande maggioranza dello sviluppo di codice sarà realizzata con il supporto dell'intelligenza artificiale. Noi abbiamo iniziato questo percorso e i risultati sono evidenti in termini di velocità e qualità dei prodotti.

Immaginare di lavorare senza IA mi sembra oggi impossibile: significherebbe sprecare tempo, spendere risorse dove non serve e dedicarsi ad attività noiose togliendo attenzione a quelle più strategiche. Significherebbe, in pratica, perdere competitività e scivolare inesorabilmente ai margini del mercato.

4. L'IA come abilitatore di una nuova offerta tecnologica in ambito legal.

Il secondo aspetto fondamentale in cui l'IA risulta strategica riguarda il nostro business core e la nostra offerta di servizi verticali dedicati al mercato legale. Il Gruppo 4C, con le sue divisioni 4cLegal e 4cAi, offre ad avvocati, studi legali e direzioni legali soluzioni tecnologiche ormai pienamente integrate con l'IA.

Si dice che l'IA sta "eliminando" il software tradizionale e questo è vero: le vecchie logiche di sviluppo sono superate dalla facilità con cui l'IA consente all'utente di ottenere ciò che vuole chiedendolo in linguaggio naturale e dal nuovo conseguente benchmark in termini di "esperienza utente".

In questa prospettiva si è reso necessario (e avvincente) ripensare i nostri software e modificare significativamente la nostra offerta commerciale, per incontrare una domanda di mercato che esprime un solido interesse verso soluzioni che abilitano risparmio di tempo, risparmio di costi e vera governance dei dati.

Questa evoluzione ci ha costretto ad affrontare concretamente alcune questioni di compliance che spesso vengono discusse in modo astratto.

Abbiamo classificato il nostro sistema di IA secondo lo schema dell'AI Act, verificando se rientrasse tra le pratiche vietate dell'art. 5, tra i sistemi ad alto rischio o in altre categorie. Nel nostro caso, siamo nella categoria del rischio limitato. Questo ci ha permesso di dimensionare gli obblighi di compliance e di muoverci abbastanza velocemente verso la commercializzazione.

Mentre ci occupavamo di questi adempimenti ci siamo chiesti quante aziende operino effettivamente su casistiche che riguardano l'alto rischio o le pratiche vietate, come tali maggiormente interessate dalle restrizioni normative. Si tratta certamente della stragrande minoranza.

Varrebbe allora la pena concentrare la nostra attenzione – più che su giudizi critici in merito all'AI ACT – su come facilitare l'adozione ampia e profonda dell'intelligenza artificiale da parte della stragrande maggioranza delle imprese, che operano in ambiti meno critici in materia di compliance e che potrebbero trarre da questa tecnologia non soltanto maggiore produttività ma anche nuove idee per migliorare la loro offerta di mercato.

Chi attacca eccessivamente l'AI Act probabilmente non fa i conti con questa realtà, oltre che dimenticare che questa normativa protegge – a mio avviso in modo ragionevole – diritti fondamentali di cui tutti noi siamo titolari.

5. Questioni interpretative aperte: il concetto di prevalenza.

Lascio ai professori presenti l'approfondimento di un tema che ritengo cruciale: l'interpretazione del concetto di "prevalenza del lavoro intellettuale", previsto nel disegno di legge [*N.d.R.: oggi legge*] come requisito limite dell'utilizzo dell'intelligenza artificiale da parte dei professionisti (art. 13).

Mi sembra particolarmente interessante guardare a questo concetto tenendo in considerazione la sentenza Schufa della Corte di Giustizia UE, del dicembre 2023, che offre a mio avviso spunti interpretativi interessanti in chiave analogica.

Il caso in questione riguarda un software di merito creditizio tedesco che aveva generato un punteggio negativo rispetto alla domanda di finanziamento presentata da un consumatore, portando al rifiuto della concessione del finanziamento. La banca argomentò che la decisione di rifiuto non era una «decisione basata unicamente sul trattamento automatizzato», vietato dall'art. 22 del GDPR, e che il software era solo uno strumento di supporto interno, senza effetto diretto sul cliente. La Corte stabilì invece che la banca decideva

prevalentemente sulla base dell'output del software, e quindi l'effetto era diretto, assimilabile a una decisione basata unicamente su un trattamento automatizzato.

Mutatis mutandis, trasponendo questo ragionamento all'attività professionale e all'obbligo che l'intelligenza artificiale sia utilizzata in tale ambito con «prevalenza del lavoro intellettuale oggetto della prestazione d'opera»: se un avvocato redige un parere utilizzando un prompt e l'intelligenza artificiale genera un testo che viene poi sostanzialmente recepito senza modifiche nei contenuti, cosa è prevalente? Come si interpreta il concetto di prevalenza in questo contesto? È una prevalenza in senso temporale o nel senso del “peso” causale rispetto all'output generato? In un contesto di costante perfezionamento dell'intelligenza artificiale, quanto l'avvocato procederà realmente a valutare in modo critico il parere generato dall'IA, così da far ritenere prevalente il lavoro intellettuale? Sono temi che a mio avviso potranno porsi in modi non molto diversi da quelli che hanno portato a ritenere – nel caso Schufa – che l'output finale fosse nella sostanza un output automatizzato (nonostante il possibile scrutinio dell'operatore bancario).

Questi sono temi su cui i corpi intermedi – ordini professionali, associazioni di categoria – potrebbero svolgere un ruolo importante fornendo linee guida agli operatori. La certezza interpretativa è essenziale per un'adozione responsabile e diffusa dell'intelligenza artificiale nelle professioni intellettuali.

6. Conclusioni: un approccio europeo.

L'intelligenza artificiale non è solo una sfida tecnologica o regolamentare: è un'opportunità per ripensare i processi, creare nuove linee di business e aumentare la competitività. Ma tutto questo deve avvenire senza perdere di vista i valori fondamentali e la cultura organizzativa che dovrebbe guidare l'innovazione. Non si tratta di rallentare lo sviluppo – come talvolta si teme – ma di dare corpo a una visione dell'IA valoriale e distintiva, in grado di “riprodurre” il meglio della profondità e della ricchezza che la nostra cultura europea è in grado di esprimere.

Finito di stampare nel mese di dicembre 2025
nella Stampatre s.r.l. di Torino – Via Bologna, 220

UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI GIURISPRUDENZA

PUBBLICAZIONI DEL DIPARTIMENTO DI DIRITTO PUBBLICO ITALIANO E SOVRANAZIONALE

Studi di diritto pubblico

Per i tipi di Giuffrè

1. VITTORIO ITALIA, *Libertà e segretezza della corrispondenza e delle comunicazioni*, 1963, pp. XXII-348.
2. ROBERTO GIANOLIO, *Le occupazioni d'urgenza*, 1963, pp. VII-226.
3. VITTORIO ITALIA, *La denominazione nel diritto pubblico*, 1966, pp. XVII-209.
4. VALERIO ONIDA, *Le leggi di spesa nella Costituzione*, 1969, pp. IV-892.
5. VITTORIO ITALIA, *Gli statuti nel diritto pubblico*. vol. I, *Potestà e norma statutaria*, 1974, pp. XXVI-430.
6. CARLO EMILIO TRAVERSO, *Il partito politico nella Costituzione italiana*, 1969.
7. PIETRO GIUSEPPE GRASSO, *Il principio nullum crimen sine lege nella Costituzione italiana*, 1972, pp. XII-372.
8. RICCARDO VILLATA, *L'esecuzione delle decisioni del Consiglio di Stato*, 1971, pp. VIII-636.
9. VITTORIO ITALIA, *Le disposizioni di principio stabilite dal legislatore*, 1970, pp. XVI-366.
10. GIANFRANCO MOR, *Le sanzioni disciplinari ed il principio nullum crimen sine lege*, 1970, ristampa 1974, pp. VIII-224.
11. VITTORIO ITALIA, *La deroga nel diritto pubblico*, 1977, pp. XII-257.
12. RICCARDO VILLATA, *Autorizzazioni amministrative e iniziativa economica privata*, 1974, pp. VIII-212.
13. GIANFRANCO MOR, *Profili dell'amministrazione regionale*, 1974, pp. VIII-244.
14. ALDO BARDUSCO, *La struttura dei contratti delle pubbliche amministrazioni*, 1974, pp. VIII-404.
15. GUIDO GRECO, *Provvedimenti amministrativi costitutivi di rapporti giuridici tra privati*, 1977, pp. IV-406.
16. CARLO EMILIO TRAVERSO, *La tutela costituzionale della persona umana prima della nascita*, 1977, pp. IV-252.
17. ALDO BARDUSCO, *Lo stato regionale italiano*, 1980, pp. IV-252.
18. RICCARDO VILLATA, «Disapplicazione» dei provvedimenti amministrativi e processo penale, 1980, pp. IV-176.
19. GUIDO GRECO, *L'accertamento autonomo del rapporto nel giudizio amministrativo*, 1980, pp. IV-256.
20. MARIA LUISA MAZZONI HONORATI, *Il referendum nella procedura di revisione costituzionale*, 1982, pp. VIII-168.
21. CARLO EMILIO TRAVERSO, *Partito politico e ordinamento costituzionale*, 1983, pp. IV-280.
22. ERMINIO FERRARI, *I servizi sociali*, vol. I, 1986, pp. XVI-276.
23. ENZO BALBONI-FABRIZIO D'ADDABBO-ANTONIO D'ANDREA-GIOVANNI GUIGLIA, *La difficile alternanza. Il sistema parlamentare italiano alla prova (1985-1987)*, 1988, pp. XVI-236.
24. CARLO ENRICO PALIERO-ALDO TRAVI, *La sanzione amministrativa*, 1988, pp. XVI-356.
25. MARCO SICA, *Effettività della tutela giurisdizionale e provvedimenti d'urgenza*, 1991, pp. XII-352.
26. GIOVANNI BOGNETTI, *La cultura giuridica e le facoltà di giurisprudenza a Milano nel secolo ventesimo*, 1991, pp. X-198.
27. SERENA MANZIN MAESTRELLI, *Il partito politico nella giurisprudenza del tribunale costituzionale federale tedesco*, 1991, pp. VIII-156.

28. MARTA CARTABIA, *La tutela dei diritti nel procedimento amministrativo*, 1991, pp. VI-138.
29. GIOVANNI BOGNETTI, *Europa in crisi*, 1991, pp. VIII-184.
30. MARILISA D'AMICO, *Giudizio sulle leggi ed efficacia temporale delle decisioni di incostituzionalità*, 1993, pp. XIV-182.
31. GIOVANNI BOGNETTI, *La costituzione economica italiana. Interpretazione e proposta di riforma*, 1993, pp. X-206.
32. MARILISA D'AMICO, *Donna e aborto nella Germania riunificata*, 1994, pp. VIII-286.
33. GABRIELLA MANGIONE, *La revisione del Grundgesetz in materia di asilo*, 1994, pp. X-106.
34. GIOVANNI BOGNETTI, *Costituzione, televisione e legge antitrust*, 1996, pp. VI-136.
35. LUCA ANTONINI, *Dovere tributario, interesse fiscale e diritti costituzionali*, 1996, pp. XII-446.
36. EUGENIO BRUTI LIBERATI, *Consenso e funzione nei contratti di diritto pubblico tra amministrazioni e privati*, 1996, pp. X-352.
37. MAURIZIO CAFAGNO, *La tutela risarcitoria degli interessi legittimi. Fini pubblici e reazioni di mercato*, 1996, pp. VIII-360.
38. MARCO BIGNAMI, *Costituzione flessibile, costituzione rigida e controllo di costituzionalità in Italia (1848-1956)*, 1997, pp. VIII-242.
39. GIOVANNI BOGNETTI, *Lo stato e i gruppi di interesse negli ordinamenti borghesi*, 1998, pp. XII-182.
40. MARGHERITA RAMAJOLI, *Attività amministrativa e disciplina antitrust*, 1998, pp. XII-524.
41. *Norme di correttezza costituzionale, convenzioni ed indirizzo politico*. Atti del Convegno organizzato in ricordo del Prof. Paolo Biscaretti di Ruffia, a cura di Gianfranco Mor, Stefania Ninatti, Quirino Camerlengo e Giulio Enea Vigevani, 1999, pp. VIII-194.
42. GABRIELLA MANGIONE, *Il diritto di asilo nell'ordinamento costituzionale tedesco*, 1999, pp. X-262.
43. ALESSANDRA CONCARO, *Il sindacato di costituzionalità sul decreto-legge*, 2000, pp. X-198.
44. MARIA ELENA GENNUSA, *La posizione costituzionale dell'opposizione*, 2000, pp. X-316.
45. LUCA ANTONINI, *Il regionalismo differenziato*, 2000, pp. XII-418.
46. *Percorsi e vicende attuali della rappresentanza e della responsabilità politica*. Atti del Convegno - Milano, 16-17 marzo 2000, a cura di Nicolò Zanon e Francesca Biondi, introduzione di Gustavo Zagrebelsky, 2001, pp. XVI-302.
47. MIRYAM IACOMETTI, *I Presidenti di Assemblea parlamentare*, 2001, pp. X-518.
48. *Studi in onore di Umberto Pototschnig*, voll. I e II, 2002, pp. X-1602.
49. *Le trasformazioni dello stato regionale italiano. In ricordo di Gianfranco Mor*, a cura di Vittorio Angiolini, Lorenza Violini, Nicolò Zanon, 2002, pp. X-488.
50. QUIRINO CAMERLENGO, *I fatti normativi e la certezza del diritto costituzionale*, 2002, pp. XIV-444.
51. GIUSEPPE MONACO, *Pubblico ministero ed obbligatorietà dell'azione penale*, 2003, pp. XIV-412.
52. WLADIMIRO TROISE MANGONI, *L'opposizione ordinaria del terzo nel processo amministrativo*, 2004, pp. X-350.
53. FRANCESCO GOISIS, *Contributo allo studio delle società in mano pubblica come persone giuridiche*, 2004, pp. X-396.
54. STEFANIA NINATTI, *Giudicare la democrazia? Processo politico e ideale democratico nella giurisprudenza della Corte di Giustizia Europea*, 2004, pp. XIV-324.
55. *L'incerto federalismo. Le competenze statali e regionali nella giurisprudenza costituzionale*, a cura di Nicolò Zanon e Alessandra Concaro, 2005, pp. VI-424.
56. *Itinerari di sviluppo del regionalismo italiano*. Primo Incontro di Studio "Gianfranco Mor" sul diritto regionale, a cura di Lorenza Violini, con la collaborazione di Quirino Camerlengo, 2005, pp. X-590.
57. *La giustizia costituzionale ed i suoi utenti*. Atti del Convegno internazionale in onore del prof. Valerio Onida - Milano, 15 aprile 2005, a cura di Pasquale Pasquino e Barbara Randazzo, 2006, pp. X-192.

58. QUIRINO CAMERLENGO, *Contributo ad una teoria del diritto costituzionale cosmopolitico*, 2007, pp. X-358.
59. MARCO CUNIBERTI, *Autorità indipendenti e libertà costituzionali*, 2007, pp. XVI-590.
60. MONICA DELSIGNORE, *La compromettibilità in arbitrato nel diritto amministrativo*, 2007, pp. XIV-306.
61. PAOLO PIZZA, *Le società per azioni di diritto singolare tra partecipazioni pubbliche e nuovi modelli organizzativi*, 2007, pp. XVI-698.
62. SARA VALAGUZZA, *La frammentazione della fattispecie nel diritto amministrativo a conformazione europea*, 2008, pp. XXXII-422.
63. LUCA BERTONAZZI, *Il ricorso straordinario al Presidente della Repubblica: persistente attualità e problemi irrisolti del principale istituto di amministrazione giustiziale*, 2008, pp. X-324.
64. BARBARA RANDAZZO, *Diversi ed uguali. Le confessioni religiose davanti alla legge*, 2008, pp. XX-456.
65. *Come decidono le Corti Costituzionali (e altre Corti) - How Constitutional Courts make decisions*. Atti del Convegno internazionale svoltosi a Milano, il 25-26 maggio 2007, a cura di Pasquale Pasquino e Barbara Randazzo, 2009, pp. VIII-232.
66. GIUSEPPE PERICU, *Scritti scelti*, 2009, pp. VI-956.
67. STEFANO CATALANO, *La "presunzione di consonanza". Esecutivo e Consiglio nelle Regioni a statuto ordinario*, 2010, pp. VIII-392.
68. IRENE PELLIZZONE, *Contributo allo studio sul rinvio presidenziale delle leggi*, 2011, pp. XVIII-318.
69. *Verso il decentramento delle politiche di welfare*. Incontro di studio "Gianfranco Mor" sul diritto regionale, a cura di Lorenza Violini, 2011, pp. VIII-504.
70. MONICA DELSIGNORE, *Il contingentamento dell'iniziativa economica privata. Il caso non unico delle farmacie aperte al pubblico*, 2011, pp. VIII-208.
71. SARA VALAGUZZA, *Società miste a partecipazione comunale. Ammissibilità e ambiti*, 2012, pp. X-214.
72. WLADIMIRO TROISE MANGONI, *Il potere sanzionatorio della CONSOB. Profili procedurali e strumentalità rispetto alla funzione regolatoria*, 2012, pp. VIII-248.
73. FRANCESCA BIONDI, *Il finanziamento pubblico dei partiti politici. Profili costituzionali*, 2012, pp. XIV-232.
74. BARBARA RANDAZZO, *Giustizia costituzionale sovranazionale. La Corte europea dei diritti dell'uomo*, 2012, pp. X-270.
75. GIUSEPPE ARCONZO, *Contributo allo studio sulla funzione legislativa provvedimentale*, 2013, pp. XIV-376.
76. LUCA PIETRO VANONI, *Laicità e libertà di educazione. Il crocifisso nelle aule scolastiche in Italia e in Europa*, 2013, pp. VIII-318.
77. BENEDETTA VIMERCATI, *Consenso informato e incapacità. Gli strumenti di attuazione del diritto costituzionale all'autodeterminazione terapeutica*, 2014, pp. X-346.
78. ELISA FAGNANI, *Tutela dei diritti fondamentali e crisi economica: il caso dell'istruzione. Stato di attuazione, funzioni amministrative e finanziamento del sistema*, 2014, pp. XII-412.
79. *Scritti scelti di Giovanni Bognetti*, a cura di Miryam Iacometti, 2015, pp. XXXVI-530.
80. PAOLO PROVENZANO, *I vizi nella forma e nel procedimento amministrativo. Fra diritto interno e diritto dell'Unione europea*, con prefazione di Diana-Urania Galetta, 2015, pp. XX-332.
81. *Il controllo preventivo dei trattati dell'Unione europea*. Atti del Convegno tenutosi a Milano il 28 maggio 2014, a cura di Nicolò Zanon, 2015, pp. XII-202.
82. STEFANIA LEONE, *Contributo allo studio dello scioglimento anticipato nel sistema costituzionale*, 2016, pp. X-394.
83. ALESSANDRA OSTI, *Teoria e prassi dell'access to Justice. Un raffronto tra ordinamento nazionale e ordinamenti esteri*, 2016, pp. X-238.
84. ANNALISA NEGRELLI, *Accesso al mercato e autorizzazioni amministrative nazionali*, 2016, pp. XLII-450.
85. ANTONIA BARAGGIA, *L'autonomia universitaria nel quadro costituzionale italiano ed europeo. Già e non ancora ...*, 2016, pp. XII-268.

86. SARA VALAGUZZA, *Il giudicato amministrativo nella teoria del processo*, 2016, pp. XIII-348.
87. BENEDETTA LIBERALI, *Problematiche costituzionali nelle scelte procreative. Riflessioni intorno alla fecondazione medicalmente assistita e all'interruzione volontaria di gravidanza*, 2017, pp. XVI-772.
88. FILIPPO ROSSI, *La costruzione giuridica del licenziamento. Legislazione, dottrina e prassi fra XIX e XX secolo*, 2017, pp. X-322.
89. *Il diritto all'acqua*, a cura di Lorenza Violini e Barbara Randazzo, 2017, pp. VI-282.
90. FEDERICO GAFFURI, *Il principio di non contestazione nel processo amministrativo*, 2018, pp. XVI-304.

Per i tipi di Giappichelli

91. GIADA RAGONE, *Eine empirische Wende? La Corte costituzionale e le sfide della complessità tecnico-scientifica*, 2020, pp. X-246.
92. LORENZA VIOLINI, *Una forma di Stato a regionalismo differenziato? Percorsi e argomenti per l'attuazione dell'art. 116, III comma, Cost.*, 2021, pp. XVIII-286.
93. *La Costituzione non odia. Conoscere, prevenire e contrastare l'hate speech on line*, a cura di Marilisa D'Amico e Cecilia Siccardi, 2021, pp. XXII-234.
94. GIULIA FORMICI, *La disciplina della data retention tra esigenze securitarie e tutela dei diritti fondamentali. Un'analisi comparata*, 2021, pp. XIV-434.
95. AA.VV., *Cinquant'anni dopo. L'esecuzione delle decisioni del Consiglio di Stato. Scritti di diritto processuale amministrativo dedicati a Riccardo Villata*, 2021, pp. XIV-274.
96. ELISABETTA CRIVELLI, *Il contributo dei Protocolli nn. 15 e 16 Cedu al processo di riforma della Corte di Strasburgo*, 2021, pp. X-142.
97. MARCO ANTONIOLI, *Rapporto senza potere e tutela dell'affidamento. Le nuove frontiere della responsabilità civile della P.A.*, 2022, pp. XVI-208.
98. BENEDETTA LIBERALI, *Un processo bifronte. Stato e Regioni davanti alla Corte costituzionale*, 2022, pp. XXXII-400.
99. *Diritto e valutazioni scientifiche*, a cura di Benedetta Liberali e Lavinia Del Corona, 2022, pp. XXVI-518.
100. LAVINIA DEL CORONA, *Libertà della scienza e politica. Riflessioni sulle valutazioni scientifiche nella prospettiva del diritto costituzionale*, 2022, pp. XVI-312.
101. ANTONIA BARAGGIA, *Stati Uniti e Irlanda. La regolamentazione dell'aborto in due esperienze paradigmatiche*, Seconda edizione, 2022, pp. X-214.
102. FILIPPO ROSSI, *Ragionevoli dubbi. Percorsi storici del recesso unilaterale*, 2022, pp. XIV-314.
103. RAFFAELLA BIANCHI RIVA, *Lo scandalo tra alto medioevo e prima età moderna. Itinerari tra dimensione giuridica, politica e sociale*, 2022, pp. XII-316.
104. ANTONIA BARAGGIA, *La condizionalità come strumento di governo negli Stati compositi. Una comparazione tra Stati Uniti d'America, Canada e Unione europea*, 2023, pp. X-270.
105. *One health. Dal paradigma alle implicazioni giuridiche*, a cura di Lorenza Violini, 2023, pp. VIII-168.
106. *Parole pericolose. Conflitto e bilanciamento tra libertà e limiti. Una prospettiva trasversale*, a cura di Sara V. Parini, 2023, pp. XXVI-214.
107. DANIELE CAMONI, *Diritto all'istruzione e opinioni dissenzianti in prospettiva comparata*, 2023, pp. XII-324.
108. ELISABETTA ROMANI, *L'esecuzione delle sentenze della Corte EDU nella giustizia amministrativa*, 2023, pp. XVI-512.
109. LUCA PIETRO VANONI, *«Never let a good crisis go to waste». Il principio della separazione dei poteri prima e dopo la pandemia*, 2023, pp. XX-348.
110. DANIELA BUCCOMINO, *Ingeniorum delectus. I dottori in diritto dello studio di Pavia (1525-1796)*, 2023, pp. XII-988.
111. *L'intervento nel processo amministrativo*, a

- cura di Margherita Ramajoli e Riccardo Villata, 2023, pp. XII-292.
112. *Scritti in onore di Miryam Iacometti*, a cura di Lorenza Violini, Antonia Baraggia, Alessandra Osti, Luca Pietro Vanoni, 2023, pp. XIV-258.
 113. PIETRO VILLASCHI, *Rappresentante e rappresentato al tempo della rivoluzione digitale*, 2024, pp. XX-444.
 114. *Il diritto amministrativo tra ordine sistematico e realtà economico-sociale. Scritti in ricordo di Giuseppe Pericu*, a cura di Eugenio Bruti Liberati, Maurizio Cafagno, Luigi Piscitelli, Margherita Ramajoli, Marco Sica, 2024, pp. XXXIV-142.
 115. STEFANIA LEONE, *La salute degli altri. Giustificazione e limiti degli obblighi vaccinali*, 2024, pp. XVIII-302.
 116. NANNEREL FIANO, *Le radici del male. Antisemitismo e Costituzione*, 2024, pp. XVI-272.
 117. *I diritti delle donne. Problematiche attuali e prospettive future. Women's rights. Current issue and future prospective*, a cura di Marilisa D'Amico, Benedetta Liberali, 2024, pp. XXVI-582.
 118. ANNA GIULIA CARAGLI-MICHELA LEGGIO, *Governance e regolazione del settore agroali-mentare verso il paradigma One Health*, 2024, pp. XVI-320.
 119. *Liber Amicorum per Guido Greco*, a cura di Franco Gaetano Scoca, Mario Pilade Chiti, Diana-Urania Galetta, 2024, pp. XXVIII-916.
 120. *Libertà di espressione e forme moderne di censura*, a cura di Alessandra Osti, Mirko Della Malva, 2024, pp. VIII-216.
 121. *Questioni vecchie e nuove di diritto costituzionale. Scritti in onore di Vittorio Angiolini*, a cura di Marilisa D'Amico, Lavinia Del Corona, 2025, pp. XXII-442.
 122. COSTANZA NARDOCCI, *Algoritmi, eguaglianza, discriminazione. Le sfide dell'intelligenza artificiale*, 2025, pp. XIV-546.
 123. ERMINIO FERRARI, *Scritti scelti*, 2025, pp. XII-1044.
 124. *Il diritto d'asilo delle donne. Principi, norme e procedure. Volume I. Dai risultati del progetto Defend women's rights: a legal and forensic analysis of asylum for victims of discrimination and violence*, a cura di Cecilia Siccardi, 2025, pp. XXIV-152.
 125. *IA e discriminazione di genere. Prospettive future per una regolamentazione inclusiva*, a cura di Marilisa D'Amico e Niccolò Panigada, 2025, pp. XXII-298.

