UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI GIURISPRUDENZA
Pubblicazioni del Dipartimento di Scienze Giuridiche “Cesare Beccaria”
39

INTELLIGENZA ARTIFICIALE

Diritto, giustizia, economia ed etica

a cura di

FABIO BASILE, MARCO BIASI, LUCIO CAMALDO
GAIA CANESCHI, BEATRICE FRAGASSO, DANIELA MILANI

G. Giappichelli Editore

VIA PO, 21 - TEL. 011-81.53.111
http://www.giappichelli.it

ISBN/EAN 979-12-211-1381-5
ISBN/EAN 979-12-211-6290-5 (ebook)

Volume pubblicato con fondo Unico Dipartimentale - Assegnazione 2025 – Dipartimento di Scienze Giuridiche "Cesare Beccaria", codice F_DOTAZIONE_2025_DIP_004.

Stampa: Stampatre s.r.l. - Torino

INDICE

pag.

INTELLIGENZA ARTIFICIALE E GIUSTIZIA PENALE

INTELLIGENZA ARTIFICIALE E DIRITTO

UNA LETTURA DELL'ARTIFICIAL INTELLIGENCE ACT: NORME, ETICA, ADEMPIMENTI, ATTUAZIONE

di GIOVANNI ZICCARDI

1. *L'avvento dell'Artificial Intelligence Act*

L'Artificial Intelligence Act (d'ora in avanti: "AI Act") ha rappresentato un passo epocale nella regolamentazione dell'intelligenza artificiale all'interno dell'Unione Europea; è, senza dubbio, il più importante atto legislativo dell'intera Legislatura che è terminata lo scorso anno, con una forte valenza anche *politica*, e non solo normativa.

La sua genesi, e il suo sviluppo, sono il risultato di un lungo e articolato processo di analisi e riflessione, volto a creare un quadro normativo capace di bilanciare le esigenze dell'innovazione tecnologica con la necessità di tutelare i *diritti fondamentali* dei cittadini europei [1].

[1] Per una panoramica introduttiva dal taglio europeo si veda, in lingua inglese, N.T. NIKOLINAKOS, *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies – The AI Act*, Cham, Springer, 2023. Per un primo approccio, sempre in lingua inglese, al tema e alla norma, si veda P. VOIGT, N. HULLEN, *The EU AI Act: answers to Frequently Asked Questions*, Springer, Berlin-Heidelberg, 2024. Per un'introduzione informatico-giuridica all'intelligenza artificiale ormai "classica" ma ancora un punto di riferimento, si veda G. SARTOR, *Intelligenza artificiale e diritto: un'introduzione*, Giuffrè, Milano, 1996. In lingua italiana, invece, si vedano: M. BIASI (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffrè Francis Lefebvre, Milano, 2024; G. VACIAGO (a cura di), *Intelligenza artificiale generativa e professione forense*, Giuffrè Francis Lefebvre, Milano, 2024; M.G. PELUSO, *Intelligenza artificiale e tutela dei dati*,

La crescente diffusione di sistemi di intelligenza artificiale (soprattutto negli ultimi due anni, con l'avvento di ChatGPT e la consapevolezza diffusa di una *dipendenza tecnologica esclusiva* dell'Europa da società nordamericane o cinesi) ha sollevato questioni etiche, legali e sociali, chiamando un intervento normativo che potesse garantire un utilizzo responsabile di tali tecnologie.

I primi segnali di un "interesse istituzionale" nei confronti della regolamentazione dell'intelligenza artificiale sono emersi nel 2018, quando la Commissione Europea ha iniziato ad adottare una *linea politica* (e una conseguente *strategia*) molto chiara dedicata a questo ambito, riconoscendo l'importanza di promuovere lo sviluppo e l'adozione dell'IA in Europa, ma con un'attenzione particolare alla tutela dei valori fondanti dell'Unione.

Tale strategia si proponeva di consolidare la competitività europea nel settore creando, al contempo, un ecosistema "di fiducia" che potesse favorire l'accettazione delle tecnologie da parte dei cittadini e delle imprese. Molta enfasi venne data, sin dall'inizio, all'idea di *trust* e di *trustworthy*, ossia di tecnologia percepita, da molti, come pericolosa, ma che potesse al contrario ispirare sempre più *fiducia* in tutti i cittadini.

In quello stesso anno venne istituito il "Gruppo di Esperti di Alto Livello sull'Intelligenza Artificiale" (di cui si parlerà nel Paragrafo seguente), un organismo composto da accademici e rappresentanti dell'industria e della società civile, con il compito di elaborare linee guida etiche per il corretto sviluppo (e utilizzo) dell'intelligenza artificiale.

Questo gruppo di lavoro ha pubblicato, nel 2019, le "Linee guida etiche per un'IA affidabile", un documento di fondamentale importanza che delinea i principi essenziali per uno sviluppo tecnologico rispettoso della dignità umana e che sarà incorporato nella parte iniziale (considerando 27) dell'AI Act.

Tra i principi individuati, tipici della tradizione dell'Unione Europea, emergono il rispetto della dignità e non discriminazione della persona, dello stato di diritto, della privacy e della protezione dei dati, dell'autonomia individuale, la prevenzione del danno e l'equità e la trasparenza, elementi ritenuti imprescindibili per garantire una crescita sostenibile e socialmente accettabile delle tecnologie basate sull'intelligenza artificiale.

Parallelamente, la Commissione Europea avviò una consultazione pubblica e approfondì l'analisi del contesto tecnologico, sociale ed economico dell'IA: nel 2020 venne, a tal fine, pubblicato un "Libro bianco sull'intelli-

---

Giuffrè Francis Lefebvre, Milano, 2023; O. Pollicino, F. Donati, G. Finocchiaro, F. Paolucci, *La disciplina dell'intelligenza artificiale*, Giuffrè Francis Lefebvre, Milano, 2025.

genza artificiale – Un approccio europeo all'eccellenza e alla fiducia", che segnava un ulteriore avanzamento verso l'elaborazione di un quadro normativo strutturato.

Questo documento proponeva un approccio regolatorio basato sulla *valutazione del rischio*, in base al quale i sistemi di IA dovevano essere sottoposti a differenti livelli di controllo e verifica a seconda del loro potenziale impatto sulla società. L'obiettivo principale era quello di individuare e mitigare i rischi più gravi, senza ostacolare l'innovazione e lo sviluppo tecnologico.

Come è certamente noto al lettore, l'approccio basato sul rischio adottato dall'AI Act è stato oggetto di *numerose critiche*, soprattutto per la sua rigidità, la difficoltà di attuazione e il rischio di frenare l'innovazione. Si tratta, d'altro canto, di un approccio che era già stato adottato in altri provvedimenti precedenti (si pensi al GDPR, il regolamento sulla protezione dei dati) e che aveva dato discreti frutti.

Esistevano, ovviamente, numerosi approcci alternativi che avrebbero potuto essere considerati nella creazione di una normativa per l'intelligenza artificiale. Ognuno di questi modelli presenta, comunque, vantaggi e svantaggi che determinano sempre un diverso equilibrio tra innovazione, tutela dei diritti fondamentali e sicurezza.

Si pensi, *in primis*, a un approccio basato sui *principi* (di solito definito "principle-based regulation"), ossia un modello che definisca una serie di principi generali a cui gli sviluppatori e gli utilizzatori di sistemi di IA si devono attenere, lasciando ampio margine di interpretazione e adattabilità (un esempio di questo approccio è, come noto, il quadro normativo degli Stati Uniti d'America, che si basa su *linee-guida flessibili* piuttosto che su regole rigide).

I vantaggi sarebbero stati, probabilmente, una maggiore adattabilità ai rapidi sviluppi tecnologici e un minore impatto sull'innovazione, oltre a evitare il rischio di una regolamentazione obsoleta a causa della velocità con cui evolve l'IA. Al contempo, però, un approccio basato sui principi può risultare troppo vago e di difficile applicazione concreta, lasciando spazio a interpretazioni divergenti e a una scarsa capacità di *enforcement*.

Ancora, si pensi a un possibile approccio *settoriale* (quella che viene definita comunemente "sector-specific regulation"): invece di classificare i sistemi di IA in base al rischio, si adotta un modello che regolamenta l'intelligenza artificiale in base al *settore di applicazione*, come avviene in alcuni territori come gli Stati Uniti d'America o la Cina. Ciò, forse, avrebbe portato a una maggiore precisione nel regolare l'IA nei settori critici (sanità, finanza e sicurezza pubblica), senza imporre vincoli eccessivi su settori

meno sensibili, ma avrebbe portato, allo stesso tempo, un forte rischio di *frammentazione normativa* (rischio che l'Unione Europea, attraverso l'uso di un *Regolamento*, voleva assolutamente evitare…), con la conseguente necessità di sviluppare molteplici regolamenti settoriali, rendendo assai difficile la gestione di tecnologie trasversali come l'IA.

Al contempo, un approccio basato sulla responsabilità (definito "accountability-based regulation") sarebbe stato un modello capace di porre un'enfasi maggiore sulla *responsabilità legale* di sviluppatori e utenti dell'IA, obbligandoli a dimostrare il rispetto di determinati criteri di sicurezza ed etica, con una maggiore flessibilità rispetto a un approccio rigido e con la possibilità per le aziende di adattare le proprie soluzioni senza sottostare a normative prestabilite. Allo stesso tempo, però, l'efficacia sarebbe dipesa dall'esistenza di *meccanismi di controllo* e di sanzioni adeguati, oltre che dalla capacità di verificare concretamente la conformità.

Un approccio (oggi molto di moda) basato sulla *certificazione volontaria* si sarebbe, a sua volta, presentato come un'alternativa meno vincolante (incentivando la creazione di standard e, appunto, certificazioni volontarie per garantire la sicurezza e l'affidabilità dell'IA, simile a quanto avviene con le certificazioni ISO), con il vantaggio di evitare il rischio di ostacolare l'innovazione e permettendo alle aziende di scegliere *se e come aderire* agli standard, incentivando la compliance senza imporla. Al contempo, però, vi era all'orizzonte, chiaro, il rischio di *scarsa adesione*, soprattutto da parte di aziende "meno etiche", e la difficoltà nel garantire un'applicazione uniforme a livello europeo.

Infine, un ipotetico approccio ispirato alla *regolamentazione dei prodotti* (definita "product safety regulation") avrebbe "trattato" i sistemi di IA come *prodotti fisici* soggetti a regolamentazioni sulla sicurezza e la conformità, proprio come avviene per i dispositivi medici o i macchinari industriali, legandosi così a un modello consolidato e già applicato a molte tecnologie (e magari permettendo di integrare l'IA in framework normativi esistenti). Si noti, però, che l'intelligenza artificiale non è un semplice "prodotto fisico", ma una tecnologia in continua evoluzione, rendendo difficile una regolamentazione *statica* basata su standard fissi.

In questa fase preliminare di "nascita" dell'AI Act, quindi, possiamo affermare che l'*approccio basato sul rischio* è stato scelto con la speranza di garantire un equilibrio tra innovazione e protezione dei cittadini, ma può presentare limiti significativi, soprattutto in termini di applicabilità pratica e di impatto sulla competitività europea.

Altri approcci, come quello basato sulla responsabilità o sulla certificazione volontaria, avrebbero forse potuto garantire maggiore flessibilità,

mentre un modello settoriale avrebbe permesso di affrontare con più precisione le specificità dei diversi usi dell'IA.

Tuttavia, nessun approccio è esente da criticità, e la sfida principale rimane quella di conciliare la necessità di regolamentare l'intelligenza artificiale senza *soffocare* il progresso tecnologico.

Dopo queste discussioni preliminari sull'approccio da adottare, e la scelta del *rischio*, il 21 aprile 2021 segnò una tappa cruciale con la presentazione della proposta legislativa dell'Artificial Intelligence Act da parte della Commissione Europea.

L'approccio basato sul rischio prevedeva, come conseguenza immediata, una suddivisione delle applicazioni dell'IA in diverse categorie.

I sistemi considerati *ad alto rischio*, come quelli impiegati in ambiti critici quali la sanità, l'occupazione, l'istruzione o il sistema-giustizia, erano sottoposti a obblighi stringenti di conformità e trasparenza.

I sistemi di IA con un rischio limitato, come i chatbot o i filtri di contenuti online, devono rispettare alcuni requisiti di trasparenza, mentre i sistemi con rischio minimo possono essere utilizzati senza particolari restrizioni.

Infine, alcune applicazioni dell'IA, considerate di rischio inaccettabile in quanto potenzialmente lesive dei diritti fondamentali, venivano *vietate*. Tra queste rientrano, ad esempio, i sistemi di sorveglianza biometrica indiscriminata da remoto in tempo reale in luoghi pubblici, o quelli in grado di manipolare il comportamento umano in modo dannoso.

L'AI Act prevede, a tal fine, una serie di obblighi specifici per i sistemi di IA ad alto rischio, tra cui l'adozione di misure di gestione del rischio, una documentazione tecnica dettagliata, la trasparenza verso gli utenti e la previsione di un monitoraggio umano che permetta di intervenire in caso di anomalie o malfunzionamenti.

Inoltre, viene introdotto un meccanismo di *governance* che prevede l'istituzione di un Comitato europeo per l'intelligenza artificiale, incaricato di garantire l'applicazione coerente del regolamento in tutti gli Stati membri e di facilitare la cooperazione tra le diverse autorità nazionali.

Dopo la presentazione della proposta, il processo legislativo dell'AI Act ha seguito le consuete fasi di discussione e negoziazione all'interno delle istituzioni europee, in alcuni momenti anche abbastanza accese e con una forte attività di lobbying da parte delle (poche) grandi multinazionali tecnologiche che producono IA.

Il Parlamento Europeo e il Consiglio dell'Unione hanno, in queste fasi, apportato modifiche e integrazioni al testo iniziale, con l'obiettivo di affinare il bilanciamento tra innovazione e tutela dei diritti fondamentali.

L'entrata in vigore del regolamento, il primo di agosto del 2024, segna un momento storico, in quanto si tratta del primo tentativo globale di regolamentare l'intelligenza artificiale in modo organico e sistematico, ponendo l'Unione Europea nel dibattito internazionale sulla governance dell'IA.

L'AI Act non solo stabilisce, a nostro avviso, un precedente normativo, ma rappresenta anche un *modello* potenzialmente replicabile a livello globale: il suo approccio basato sul rischio, unito a una forte attenzione alla protezione dei diritti fondamentali, potrebbe costituire un punto di riferimento per altri Paesi e organizzazioni internazionali che stanno valutando l'introduzione di normative analoghe.

Tuttavia, l'efficacia della regolamentazione dipenderà in larga misura dalla sua attuazione pratica e dalla capacità delle istituzioni europee di monitorare e far rispettare le disposizioni contenute nell'atto.

Il cammino verso un'intelligenza artificiale affidabile, sicura ed eticamente responsabile non si esaurisce, ovviamente, con l'adozione dell'AI Act, ma richiederà un costante aggiornamento delle norme e una collaborazione continua tra governi, imprese, ricercatori e società civile, per affrontare le sfide in continua evoluzione che l'intelligenza artificiale pone alla nostra società.

2. *L'incorporazione dei principi di computer ethics nell'AI Act*

L'evoluzione dell'intelligenza artificiale nell'Unione Europea, accanto alle considerazioni di politica legislativa che abbiamo illustrato nel paragrafo precedente, ha sempre seguito un percorso guidato anche da *principi etici*, e da un costante equilibrio tra innovazione tecnologica e protezione dei diritti fondamentali[2].

La volontà di garantire uno sviluppo responsabile dell'IA ha portato, nel 2018, alla costituzione del "Gruppo di Esperti di Alto Livello sull'Intelligenza Artificiale" ("AI HLEG"), incaricato di delineare un insieme di principi che potessero fungere da base per la regolamentazione futura.

Il risultato principale di questo lavoro è stato la pubblicazione, nel 2019, delle "Linee guida etiche per un'IA affidabile", un documento che identifica i principi fondamentali a cui qualsiasi sistema di intelligenza arti-

[2] Sul tema dell'etica dell'intelligenza artificiale si veda P. Benanti, S. Maffettone, *Noi e la macchina. Un'etica per l'era digitale*, LUISS University Press, Roma, 2024. Si veda anche P. Benanti, *Human in the loop. Decisioni umane e intelligenza artificiale*, Mondadori, Milano, 2022.

ficiale dovrebbe attenersi per essere ritenuto affidabile, sicuro ed eticamente sostenibile.

Secondo l'AI HLEG, affinché un sistema di IA possa essere considerato affidabile, esso deve rispettare tre componenti essenziali: i) deve essere *legale*, rispettando tutte le normative vigenti; ii) deve essere *etico*, garantendo il rispetto dei principi morali e dei diritti fondamentali; iii) deve essere *solido* dal punto di vista tecnico, evitando malfunzionamenti o vulnerabilità che potrebbero generare danni.

All'interno di questa cornice generale, sono stati individuati quattro principi fondamentali che devono guidare lo sviluppo e l'implementazione dei sistemi di IA.

Il primo principio è il *rispetto dell'autonomia umana*, che implica la necessità di garantire che i sistemi di IA supportino le capacità decisionali degli esseri umani, senza sostituirle o manipolarle. Questo principio pone l'accento sulla necessità di preservare la libertà di scelta degli individui, evitando che le tecnologie possano essere utilizzate per influenzare le decisioni umane in modi che riducano il libero arbitrio.

L'AI Act ha incorporato questo principio prevedendo obblighi di trasparenza e garanzie contro la manipolazione ingannevole, in particolare per i sistemi di IA che interagiscono *direttamente* con gli utenti o che impiegano tecniche di persuasione avanzate, come la personalizzazione algoritmica e il riconoscimento delle emozioni.

Il secondo principio è la *prevenzione del danno*, che richiede che i sistemi di IA siano progettati per garantire la sicurezza, evitando qualsiasi tipo di danno fisico o psicologico agli esseri umani.

Questo principio si riflette nell'AI Act attraverso una *categorizzazione* dei sistemi di IA in base al rischio che comportano, vietando quelli considerati di rischio inaccettabile e imponendo rigorosi requisiti di sicurezza per quelli classificati come ad alto rischio. Il regolamento europeo stabilisce, ad esempio, che i sistemi impiegati in ambiti critici come la sanità, i trasporti o la giustizia debbano essere sottoposti a valutazioni di conformità dettagliate, dimostrando la loro affidabilità prima di poter essere utilizzati.

Il terzo principio riguarda l'*equità e la non discriminazione*, elementi essenziali per evitare che i sistemi di IA rafforzino o amplifichino pregiudizi esistenti.

Poiché gli algoritmi di apprendimento automatico si basano su dati storici che possono contenere *bias*, è fondamentale che i modelli siano progettati per ridurre al minimo le distorsioni e per garantire trattamenti equi per tutti gli utenti. L'AI Act affronta questa problematica imponendo obblighi

specifici in termini di trasparenza e auditing per i sistemi di IA utilizzati in settori sensibili come l'occupazione, il credito e l'istruzione, assicurando che le decisioni automatizzate siano *spiegabili* e prive di discriminazioni sistematiche.

Il quarto principio è la trasparenza, che prevede che i sistemi di IA siano comprensibili e accessibili agli utenti e agli stakeholder coinvolti. La *trasparenza* implica che sia chiaro quando una persona interagisce con un sistema di intelligenza artificiale e che siano disponibili informazioni sul funzionamento del modello e sui criteri che guidano le decisioni automatizzate.

Nell'AI Act, questo principio si traduce in obblighi di *spiegabilità e tracciabilità* per i sistemi di IA ad alto rischio, che devono fornire documentazione tecnica dettagliata, consentendo la verifica e il controllo delle loro decisioni. Inoltre, vengono introdotte misure per garantire che gli utenti siano sempre informati quando interagiscono con sistemi di IA, specialmente in ambiti come la sorveglianza biometrica o la creazione di contenuti generati artificialmente.

La *spiegabilità* di un sistema di intelligenza artificiale ("AI explainability") è un aspetto fondamentale anche per i giuristi, che merita un rapido approfondimento. Consiste, in particolare, nella capacità di rendere *comprensibile* agli esseri umani il funzionamento di un modello di IA, permettendo di comprendere come e perché un sistema prenda determinate decisioni.

Questo concetto è essenziale per garantire la trasparenza, la fiducia e la responsabilità nell'uso dell'intelligenza artificiale, soprattutto nei settori ad alto impatto sociale come la sanità, la finanza e la giustizia.

La spiegabilità, semplificando molto, può essere suddivisa in quattro diverse componenti chiave: i) la *interpretabilità* (riguarda la possibilità di comprendere le relazioni tra input e output di un modello di IA, e alcuni modelli, come le reti neurali profonde, sono difficili da interpretare, mentre altri, come gli alberi decisionali, sono più trasparenti); ii) la *trasparenza* (si riferisce alla capacità di esaminare il *funzionamento interno* di un sistema di IA, come i pesi e i parametri di un modello di apprendimento automatico, e anche in questo caso alcuni algoritmi sono considerati "scatole nere" perché il loro processo decisionale è opaco e complesso); iii) la *giustificabilità* (implica la possibilità di fornire motivazioni coerenti e logiche che spieghino una decisione presa dal sistema, in modo comprensibile per gli esseri umani), e iv) la *riproducibilità* (significa che un sistema di IA dovrebbe produrre risultati simili quando vengono forniti dati e condizioni simili, permettendo di verificare e validare il suo funzionamento).

Oltre a questi quattro principi fondamentali, il Gruppo di Esperti di Alto Livello ha individuato *sette requisiti chiave* per l'implementazione di un'IA affidabile, tra cui la governance dei dati, la robustezza e la sicurezza, la supervisione umana e la responsabilità.

Si legga, a tal proposito, il testo del Considerando 27:

> "Sebbene l'approccio basato sul rischio costituisca la base per un insieme proporzionato ed efficace di regole vincolanti, è importante ricordare gli orientamenti etici per un'IA affidabile del 2019 elaborati dall'AI HLEG indipendente nominato dalla Commissione. In tali orientamenti l'AI HLEG ha elaborato sette principi etici non vincolanti per l'IA che sono intesi a contribuire a garantire che l'IA sia affidabile ed eticamente valida. I sette principi comprendono: 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. Fatti salvi i requisiti giuridicamente vincolanti del presente regolamento e di qualsiasi altra disposizione di diritto dell'Unione applicabile, tali orientamenti contribuiscono all'elaborazione di un'IA coerente, affidabile e antropocentrica, in linea con la Carta e con i valori su cui si fonda l'Unione. Secondo gli orientamenti dell'AI HLEG con «intervento e sorveglianza umani» si intende che i sistemi di IA sono sviluppati e utilizzati come strumenti al servizio delle persone, nel rispetto della dignità umana e dell'autonomia personale, e funzionano in modo da poter essere adeguatamente controllati e sorvegliati dagli esseri umani. Con «robustezza tecnica e sicurezza» si intende che i sistemi di IA sono sviluppati e utilizzati in modo da consentire la robustezza nel caso di problemi e resilienza contro i tentativi di alterare l'uso o le prestazioni del sistema di IA in modo da consentire l'uso illegale da parte di terzi e ridurre al minimo i danni involontari. Con «vita privata e governance dei dati» si intende che i sistemi di IA sono sviluppati e utilizzati nel rispetto delle norme in materia di vita privata e protezione dei dati, elaborando al contempo dati che soddisfino livelli elevati in termini di qualità e integrità. Con «trasparenza» si intende che i sistemi di IA sono sviluppati e utilizzati in modo da consentire un'adeguata tracciabilità e spiegabilità, rendendo gli esseri umani consapevoli del fatto di comunicare o interagire con un sistema di IA e informando debitamente i deployer delle capacità e dei limiti di tale sistema di IA e le persone interessate dei loro diritti. Con «diversità, non discriminazione ed equità» si intende che i sistemi di IA sono sviluppati e utilizzati in modo da includere soggetti diversi e promuovere la parità di accesso, l'uguaglianza di genere e la diversità culturale, evitando nel contempo effetti discriminatori e pregiudizi ingiusti vietati dal diritto dell'Unione o nazionale. Con «benessere sociale e ambientale» si intende che i sistemi di IA sono sviluppati e utilizzati in modo sostenibile e rispettoso dell'ambiente e in modo da apportare benefici a tutti gli esseri umani, monitorando e valutando gli impatti a lungo termine sull'individuo, sulla società e sulla democrazia.

> L'applicazione di tali principi dovrebbe essere tradotta, ove possibile, nella progettazione e nell'utilizzo di modelli di IA. Essi dovrebbero in ogni caso fungere da base per l'elaborazione di codici di condotta a norma del presente regolamento. Tutti i portatori di interessi, compresi l'industria, il mondo accademico, la società civile e le organizzazioni di normazione, sono incoraggiati a tenere conto, se del caso, dei principi etici per lo sviluppo delle migliori pratiche e norme volontarie".

Questi elementi elencati come "non vincolanti" nel Considerando citato sono, poi, stati incorporati "indirettamente" nell'AI Act attraverso una serie di disposizioni che mirano a garantire che lo sviluppo dell'intelligenza artificiale avvenga in modo etico e responsabile.

Ad esempio, il regolamento prevede che i sistemi di IA ad alto rischio siano soggetti a *verifiche periodiche* e a *meccanismi di controllo continuo*, mentre per le aziende che sviluppano o utilizzano tali tecnologie vengono introdotti obblighi di *conformità* e *tracciabilità*.

L'integrazione dei principi etici dell'AI HLEG nell'AI Act rappresenta una delle caratteristiche distintive dell'approccio europeo alla regolamentazione dell'intelligenza artificiale.

A differenza di altri modelli normativi, che si concentrano prevalentemente sugli aspetti *economici* e di *sicurezza*, il quadro europeo pone al centro della regolamentazione la protezione dei *diritti fondamentali*, cercando di prevenire gli impatti negativi dell'IA sulla società. Questo approccio ha influenzato anche il dibattito internazionale sulla governance dell'intelligenza artificiale, spingendo altre giurisdizioni a considerare l'adozione di *misure simili* per garantire uno sviluppo tecnologico etico e sostenibile.

L'AI Act si presenta, quindi, non solo come una norma giuridica, ma vuole rappresentare un vero e proprio tentativo di definire un modello di riferimento per *l'uso responsabile dell'intelligenza artificiale* a livello globale.

L'incorporazione di principi etici nel regolamento europeo non solo contribuisce a creare un ecosistema di fiducia, ma aiuta anche a promuovere un'innovazione che sia al servizio dell'umanità, piuttosto che una mera espressione di progresso tecnologico privo di considerazioni morali.

La sfida principale, tuttavia, rimarrà quella *dell'attuazione pratica* di queste disposizioni e della capacità delle istituzioni europee di monitorare e garantire il rispetto dei principi etici nel lungo termine, in un contesto tecnologico in continua evoluzione.

3. *Un esempio pratico di carta per la governance dell'AI: il Decalogo della Statale*

Un esempio interessante di fusione di principi etici, giuridici, tecnologici e di buon senso in una *carta programmatica* per una realtà molto complessa quale una grande università pubblica è un documento di governance dell'AI presentato ai primi di febbraio del 2025 dal gruppo di lavoro sull'intelligenza artificiale dell'Università di Milano. Si tratta di un vero e proprio "decalogo" teorico-pratico che si propone di orientare tutta la popolazione dell'Ateneo nell'uso di questi strumenti così innovativi.

Come si legge nel preambolo, l'Università degli Studi di Milano muove dall'idea di *sostenere* l'uso di strumenti di intelligenza artificiale a sostegno di *tutte* le sue attività, mantenendo fermo, prima di tutto, il principio di porre *la persona* al centro di ogni iniziativa. Lo scopo, dunque, non è quello di *vincolare* ma, piuttosto, di promuovere un utilizzo etico, legittimo e consapevole di strumenti di intelligenza artificiale, recependo i più recenti documenti normativi in materia e le *buone pratiche* già promosse da molte università e centri di ricerca a livello nazionale e internazionale.

I principi generali del decalogo saranno, poi, corredati da specifiche linee guida in tema di ricerca e terza missione, didattica, e attività amministrative, che comprenderanno anche indicazioni sugli strumenti il cui utilizzo è avvallato dell'ateneo (tali linee guida, in particolare, saranno definite attraverso *modalità partecipative* che coinvolgeranno, per ciascun ambito, rappresentanze di tutte le componenti dell'ateneo).

Il primo punto trattato è quello della "AI Literacy" (l'alfabetizzazione in tema di IA resa *obbligatoria* dall'art. 4 del regolamento, adempimento già in vigore dal febbraio 2025). In particolare, vi è la previsione di *percorsi di alfabetizzazione* in tema di intelligenza artificiale e di eventi culturali, tecnici, giuridici, scientifici e divulgativi al fine di garantire un innalzamento del livello di alfabetizzazione e di competenze digitali in tema di intelligenza artificiale nell'Università degli Studi di Milano. Ciò è stato programmato al fine di accrescere le conoscenze di tutti, di rispettare gli obblighi di legge e le migliori pratiche e di sviluppare lo spirito critico della comunità accademica, con l'obiettivo ambizioso di portare l'intera comunità universitaria, entro il 2030, a un livello di conoscenza che consenta un uso affidabile, responsabile e condiviso degli strumenti di intelligenza artificiale.

Viene affrontato, subito dopo, il tema della *data protection*, prevedendo un obbligo costante di protezione dei dati personali e di data governance. In questo caso si richiede un utilizzo di strumenti di intelligenza artificiale

che avvenga sempre nel rispetto dei principi, delle buone pratiche e delle norme in materia di tutela dei dati personali e di riservatezza delle persone e delle informazioni a loro riferite.

In particolare, viene evidenziato il principio di *minimizzazione dei trattamenti*, che impone grande cautela nel trasmettere agli strumenti di intelligenza artificiale *unicamente* le informazioni necessarie per ottenere i risultati attesi. Allo stesso tempo, ogni strumento di intelligenza artificiale che non sia stato preventivamente *avallato* o "certificato" dall'Ateneo e dagli uffici preposti alla ricognizione tecnologica, agli acquisti e alla amministrazione dei sistemi, richiederà apposite valutazioni (da parte del Comitato Etico, di commissioni *ad hoc* o in base a norme contenute in specifici regolamenti) prima di poter essere utilizzata.

Nel prosieguo del documento viene illustrato, poi, il già visto (e fondamentale) principio di *trasparenza e tracciabilità* delle operazioni. Più precisamente, si stabilisce come l'utilizzo di strumenti di intelligenza artificiale per lo svolgimento delle rispettive attività (amministrative, di ricerca, di studio e redazione elaborati o tesi) debba essere dichiarato e trasparente. In particolare, è necessaria l'indicazione (tra gli altri) dei seguenti elementi: *i*) la base di dati utilizzata, *ii*) le chiavi di ricerca immesse o i percorsi di interrogazione fatti, *iii*) gli strumenti concretamente utilizzati. Deve anche essere garantita, come richiesto dalla regolamentazione europea, la *tracciabilità* delle operazioni effettuate dal sistema di intelligenza artificiale.

Particolare attenzione, in linea con i contenuti del piano strategico di Ateneo, è poi data alla *sostenibilità ambientale*: gli strumenti di intelligenza artificiale sono in grado di svolgere molteplici funzioni utili, ma sono anche molto *energivori*. Ciò comporta che, prima di adottare o adoperare uno strumento di intelligenza artificiale, sia opportuno *valutare l'impatto* ambientale dello stesso, limitandone l'utilizzo ai casi in cui vi sia un'effettiva esigenza di supporto ad attività lavorative, di studio o di ricerca.

Con riferimento alla *qualità dei dati* utilizzati e alla inclusività, i risultati dell'attività di elaborazione svolta dagli strumenti di intelligenza artificiale devono sempre essere *attentamente analizzati* al fine di evitare o mitigare possibili risultati discriminatori correlati ad elementi caratteristici di un soggetto o di un gruppo di persone, soprattutto se rientranti nella categoria dei *soggetti vulnerabili*. A tal fine, fondamentale è la garanzia della qualità dei dati (in ingresso, di training, di validazione) utilizzati dal sistema.

Centrale, poi, è la questione di *cybersecurity*: come qualsiasi ritrovato informatico, anche gli strumenti di intelligenza artificiale possono essere vittime di *attacchi informatici*. Per questo, l'attenzione alla protezione dei dispositivi, alla legittimità delle basi di dati, alla possibilità di avvelenamento

degli stessi e alla affidabilità del produttore o del distributore del sistema di IA deve sempre essere elevata.

Nella parte finale del Decalogo ci si concentra, invece, su *accountability* e *supervisione umana*. L'Ateneo si impegna a rispettare tutti gli obblighi di legge, con particolare riferimento a quelli relativi alla protezione dei dati, alla cybersecurity, al rispetto del diritto d'autore, alla sicurezza delle infrastrutture, alla regolamentazione europea dell'intelligenza artificiale e al plagio di contenuti altrui. Si chiede, a tal fine, la *collaborazione* di tutti affinché non vi siano utilizzi dell'intelligenza artificiale che non siano rispettosi di tali principi e delle regole professionali e deontologiche applicabili.

Per tutti i sistemi di intelligenza artificiale usati in Ateneo dovrà poi essere garantita, come richiesto dalla normativa vigente, una *supervisione/sorveglianza umana* in grado di correggere i risultati, re-interpretarli o ribaltarli, disattivare o bloccare il sistema di intelligenza artificiale.

Con riferimento, invece, alla *governance*, l'Ateneo s'impegna a promuovere i processi *data-driven* nel rispetto dei valori etici e giuridici. In particolare, è *proibita* l'immissione in sistemi di intelligenza artificiale di dati personali "sensibili" e di dati che possono porre dei rischi per l'Ateneo (verbali riservati, informazioni soggette a riservatezza, informazioni con elevato valore economico, informazioni legate alla tutela della proprietà intellettuale e industriale dell'Ateneo), a meno che non sia strettamente necessario per lo svolgimento dei propri compiti. In questo caso, l'Ateneo svolge delle apposite valutazioni preliminari prima di consentire tali procedure.

Infine, la valutazione circa l'*opportunità* di adottare strumenti di intelligenza artificiale, e l'adozione degli stessi, viene lasciata, in un'ottica di responsabilità, a ciascun utente, che deve impegnarsi a un uso consapevole e attento dello strumento e dei suoi risultati.

4. *Gli adempimenti previsti e le concrete difficoltà*

Muovendo, ora, a un approccio meno teorico e più pratico e informatico-giuridico, si anticipava come l'Artificial Intelligence Act abbia previsto un quadro normativo *rigoroso* per la regolamentazione dei sistemi di intelligenza artificiale nell'Unione Europea, con un approccio basato sulla valutazione del rischio.

La conformità alle disposizioni dell'AI Act varia, quindi, in base alla classificazione del rischio dei sistemi di IA, con obblighi più *stringenti* per

le applicazioni considerate ad alto rischio e requisiti più *leggeri* per quelle a rischio limitato o minimo [3].

I sistemi di intelligenza artificiale classificati come ad alto rischio sono soggetti, in particolare, a un regime di conformità rigoroso, in quanto il loro utilizzo potrebbe avere un impatto significativo sui diritti e sulla sicurezza delle persone. Questi sistemi includono, ad esempio, quelli impiegati nei settori della sanità, dell'occupazione, dell'istruzione, delle infrastrutture critiche e del sistema-giustizia.

Per tali applicazioni, l'AI Act prevede una serie di adempimenti specifici, tra cui, *in primis*, l'obbligo di implementare un sistema di gestione del rischio che consenta di individuare, valutare e mitigare potenziali minacce derivanti dall'utilizzo dell'IA.

Questo processo deve essere *continuo* e *aggiornato*, in modo da garantire che i rischi emergenti vengano prontamente affrontati. Il livello di difficoltà di tale adempimento, a nostro avviso, si può presentare *elevato*, in quanto richiede la definizione di strategie di gestione del rischio, la formazione del personale e l'integrazione di strumenti di monitoraggio avanzati.

Un altro requisito essenziale per i sistemi ad alto rischio riguarda la *documentazione tecnica*, che deve essere dettagliata e dimostrare la conformità del sistema alle disposizioni dell'AI Act.

Tale documentazione deve contenere informazioni sul funzionamento dell'algoritmo, sui dati utilizzati per l'addestramento e sui meccanismi di mitigazione del rischio. Questo adempimento, sebbene tecnicamente complesso, può essere, a nostro avviso, gestito attraverso un'integrazione efficiente di procedure di *audit* e *certificazione*, garantendo che ogni fase dello sviluppo del sistema sia tracciabile e verificabile.

La trasparenza è, si diceva poco sopra, un ulteriore pilastro della conformità per i sistemi ad alto rischio.

Gli utenti devono essere informati in modo chiaro e comprensibile sul funzionamento dell'IA e sulle modalità con cui le decisioni vengono prese, e questo obbligo si traduce, in pratica, nella necessità di fornire spiegazioni dettagliate, utilizzando un linguaggio accessibile e strumenti grafici, per consentire agli utenti di comprendere come, e perché, una determinata decisione sia stata adottata dal sistema. Questo adempimento, pur non essendo tecnicamente complesso, può presentare difficoltà nel garantire una *comunicazione* efficace tra i fornitori di IA e gli utenti finali.

---

[3] Per una guida interessante agli adempimenti si veda, in lingua inglese, T. Myklebust, T. Stålhane, D.M.K. Vatn, *The AI Act and The Agile Safety Plan*, Springer, Cham, 2025.

L'AI Act richiede, inoltre, che i sistemi di IA ad alto rischio siano sottoposti a *supervisione umana*, in modo da garantire che le decisioni automatizzate possano essere *controllate* e *corrette* in caso di errori o risultati imprevisti. Questo requisito impone alle aziende di implementare *procedure* che permettano l'intervento umano in tempo reale, il che può risultare complesso in applicazioni ad alta automazione. L'adozione di interfacce di monitoraggio intuitive e di protocolli di revisione periodica rappresenta la soluzione più efficace per adempiere a questo obbligo, che vediamo come il *più complesso* del Regolamento.

Infine, la *robustezza* e la *sicurezza* dei sistemi ad alto rischio devono essere garantite attraverso test approfonditi e verifiche periodiche, che dimostrino l'affidabilità delle soluzioni adottate e la loro resistenza a potenziali attacchi o malfunzionamenti.

Questo adempimento, che potremmo definire genericamente di "cybersecurity", è altamente tecnico e può richiedere *ingenti risorse*, sia in termini di tempo che di investimenti tecnologici, rendendo fondamentale la collaborazione tra sviluppatori, esperti di sicurezza e organismi di certificazione.

Per quanto riguarda i sistemi di IA a rischio *limitato* o *minimo*, l'AI Act prevede requisiti *meno stringenti*, volti principalmente a garantire la trasparenza nei confronti degli utenti.

Questi sistemi comprendono, ad esempio, chatbot, assistenti virtuali e algoritmi di raccomandazione, che non presentano un impatto significativo sui diritti fondamentali o sulla sicurezza delle persone.

Per tali applicazioni, l'adempimento principale riguarda l'*obbligo di informare gli utenti* del fatto che stanno interagendo con un sistema di intelligenza artificiale. Questa misura di trasparenza è, a nostro avviso, relativamente semplice da implementare (richiama molto l'idea di *informativa tanto* cara alla protezione dei dati) e può essere soddisfatta attraverso messaggi chiari e visibili all'interno delle interfacce utente.

Un ulteriore requisito per i sistemi a rischio limitato riguarda la prevenzione della *manipolazione ingannevole* (anche con tecniche subliminali), che impone ai fornitori di IA di garantire che i loro sistemi non influenzino in modo subdolo il comportamento degli utenti.

Questo obbligo può essere rispettato attraverso la progettazione di interfacce-utente etiche, che evitino pratiche di persuasione occulta o oscura, anche note come "dark patterns". Anche se questo adempimento non presenta particolari difficoltà tecniche, richiede un'attenzione specifica nella fase di progettazione dei sistemi di interazione.

Le modalità migliori per garantire la conformità all'AI Act dipendono,

in definitiva, dalla tipologia del sistema di IA e dal livello di rischio ad esso associato.

Per le applicazioni ad alto rischio, è consigliabile adottare un approccio *integrato* che combini audit interni ed esterni, strumenti di monitoraggio automatico e verifiche periodiche condotte da enti indipendenti. L'adozione di *standard internazionali* per la certificazione della sicurezza e dell'affidabilità dell'IA può rappresentare un ulteriore vantaggio competitivo, facilitando la dimostrazione della conformità alle autorità regolatorie.

Per i sistemi a rischio limitato, invece, la conformità può essere garantita attraverso l'implementazione di meccanismi di *trasparenza* e *informazione* all'utente, che non richiedono necessariamente un controllo esterno ma devono essere integrati efficacemente all'interno dell'esperienza d'uso del sistema. Un'adeguata formazione dei gruppi di sviluppo e l'adozione di *linee guida* chiare per la progettazione etica possono contribuire significativamente alla riduzione del rischio di non conformità.

Sebbene il livello di complessità degli adempimenti vari in base al rischio associato al sistema, la chiave per garantire la conformità risiede in un *approccio proattivo*, che integri la gestione del rischio, la trasparenza e la supervisione umana fin dalle prime fasi di sviluppo dell'IA. In tal modo, sarà possibile non solo soddisfare i requisiti normativi, ma anche rafforzare la fiducia degli utenti e promuovere un'adozione responsabile delle tecnologie basate sull'intelligenza artificiale.

5. *Alcune conclusioni: i limiti dei tempi di attuazione e l'attuale situazione geo-politica*

L'Artificial Intelligence Act, si è visto, rappresenta indubbiamente un ambizioso tentativo di regolamentare l'intelligenza artificiale a livello europeo, ma non è privo di *limiti* e *criticità*, che si manifesteranno con sempre maggior evidenza nei prossimi cinque anni (termine previsto per un'attuazione *completa* di tutte le parti del Regolamento).

Nonostante il suo impianto normativo dettagliato, e la sua enfasi sulla protezione dei diritti fondamentali, l'efficacia della sua attuazione e il suo impatto globale sono oggetto, in questi mesi, di un dibattito assai vivace. In un contesto geopolitico in cui l'intelligenza artificiale è divenuta un terreno di *competizione* tra grandi potenze, il modello regolatorio europeo si distingue nettamente dagli approcci adottati da Stati Uniti e Cina, sollevando però, al contempo, interrogativi sulla sua capacità di *incidere* efficacemente nel panorama internazionale.

Uno dei principali limiti dell'AI Act riguarderà, probabilmente, la sua *implementazione pratica*. La complessità degli adempimenti richiesti, soprattutto per i sistemi di IA ad alto rischio, potrebbe rappresentare un *ostacolo* significativo per le aziende, in particolare per le piccole e medie imprese. Il regolamento impone una serie di *oneri burocratici*, tra cui la documentazione dettagliata, i meccanismi di trasparenza e le verifiche periodiche, che potrebbero risultare gravosi per le realtà con minori risorse a disposizione.

A ciò si aggiunge il rischio, sollevato da più parti, che l'AI Act possa *rallentare l'innovazione* in Europa, creando un ambiente normativo troppo rigido rispetto ad altri contesti internazionali più permissivi.

Questo aspetto è stato oggetto di critiche non solo da diverse parti politiche in Europa ma, anche, da parte dell'*industria tecnologica*, che teme un'eccessiva regolamentazione capace di ostacolare la competitività delle imprese europee rispetto ai giganti tecnologici statunitensi e cinesi.

A livello geopolitico, l'AI Act si colloca, oggi, in un panorama frammentato, caratterizzato da approcci differenti alla regolamentazione dell'intelligenza artificiale e da un uso bellico dell'IA legato alle *guerre* in corso (soprattutto i conflitti Russia-Ucraina e Israele-Hamas).

Gli Stati Uniti d'America, ad esempio, hanno adottato un modello basato principalmente sull'autoregolamentazione e sull'intervento *mirato* solo in determinati settori, come la sicurezza nazionale e la protezione dei dati personali. Un simile approccio, meno vincolante rispetto a quello europeo, consente un'accelerazione dell'innovazione e una maggiore flessibilità per le aziende ma, al contempo, presenta il rischio di una minore protezione per gli utenti.

La Cina, dal canto suo, ha sviluppato una strategia che combina una forte regolamentazione statale con un uso intensivo dell'IA per scopi governativi e di sorveglianza. Il governo cinese ha introdotto norme stringenti su alcune applicazioni dell'IA, come il *riconoscimento facciale* e gli *algoritmi di raccomandazione*, ma con un obiettivo principalmente orientato al controllo sociale piuttosto che alla tutela dei diritti individuali.

Questa divergenza negli approcci normativi solleva una questione di fondo sulla capacità dell'AI Act di influenzare il panorama globale, in un periodo storico di *insofferenza diffusa*, nella classe politica, dell'idea di regolamentazione e per le attività delle autorità di controllo indipendenti.

Se, da un lato, il regolamento europeo si propone giustamente come modello di riferimento per una governance etica dell'IA, con un intento *nobile* che è chiarissimo, dall'altro viene ventilato il rischio che l'Europa rimanga *isolata* nel definire standard che non vengano adottati da altre potenze tecnologiche.

La mancanza di un quadro normativo armonizzato a livello internazionale potrebbe, così, portare a una *frammentazione* del mercato dell'IA, con imprese europee costrette a rispettare norme più severe rispetto ai concorrenti stranieri, con conseguenti svantaggi competitivi.

Un'altra critica mossa all'AI Act nei primi mesi di attuazione riguarda la sua capacità di affrontare le *sfide emergenti* dell'intelligenza artificiale in un contesto in *continua evoluzione*.

La rapidità con cui si sviluppano nuove tecnologie basate sull'IA rende difficile una regolamentazione che possa anticipare tutti i possibili rischi e scenari futuri: si pensi all'avvento *improvviso* prima di ChatGPT (nella primavera del 2023) e, poi, della IA cinese DeepSeek agli inizi del 2025.

Il regolamento, pur adottando un approccio basato sul rischio, potrebbe risultare inadeguato di fronte a innovazioni impreviste e così rapide, creando la necessità di continui aggiornamenti normativi. Un'eccessiva rigidità nelle regole potrebbe portare a problemi di applicazione, rendendo il sistema normativo rapidamente obsoleto rispetto ai progressi tecnologici.

Un ulteriore punto critico riguarda, a nostro avviso, la (futura) capacità delle istituzioni europee di *far rispettare* le disposizioni dell'AI Act.

L'attuazione completa del regolamento richiede la creazione di autorità di vigilanza nazionali e meccanismi di controllo che potrebbero rivelarsi difficili da gestire, soprattutto considerando le differenze tra gli Stati membri in termini di risorse e competenze tecniche. Inoltre, il rispetto degli obblighi di conformità dipenderà in gran parte dalla capacità delle aziende di adeguarsi alle nuove regole, il che potrebbe generare ritardi e incertezze nella fase di implementazione.

Nonostante queste possibili criticità, che saranno valutate con attenzione da tutti nei prossimi cinque anni, l'AI Act rappresenta un passo fondamentale verso una regolamentazione responsabile dell'intelligenza artificiale.

La sfida principale per l'Unione Europea sarà quella di trovare un *compromesso* tra la necessità di proteggere i cittadini e il desiderio di promuovere l'innovazione, evitando che il quadro normativo diventi un ostacolo alla crescita del settore.

Il futuro dell'AI Act dipenderà, in definitiva, dalla sua capacità di adattarsi ai rapidi cambiamenti tecnologici e di influenzare il dibattito globale sulla governance dell'intelligenza artificiale, cercando di promuovere un modello che possa essere adottato anche a livello internazionale.

L'Unione Europea si trova, in conclusione, di fronte a un *bivio*. Da un lato, ha concretamente la possibilità di diventare un punto di riferimento globale per una regolamentazione etica e responsabile dell'IA. Dall'altro, appare all'orizzonte il rischio di *restare indietro* rispetto a potenze tecnolo-

giche che privilegiano un approccio più flessibile e orientato alla competitività.

La sfida, a nostro avviso, sarà quella di garantire che l'AI Act non si trasformi in un freno all'innovazione, ma che possa invece fungere da stimolo per la creazione di un ecosistema di intelligenza artificiale sicuro, affidabile e conforme ai valori europei.