ISSN 1826-3534

federalismi.it
RIVISTA DI DIRITTO PUBBLICO ITALIANO, COMPARATO, EUROPEO

19 APRILE 2023

La prospettiva del controllo nell'era dell'Intelligenza Artificiale: alcune osservazioni sul modello *Human In The Loop*

di Iole Pia Di Ciommo
Dottoranda di ricerca in Diritto pubblico, comparato e internazionale
Sapienza – Università di Roma

La prospettiva del controllo nell'era dell'Intelligenza Artificiale: alcune osservazioni sul modello *Human In The Loop*[*]

di Iole Pia Di Ciommo
Dottoranda di ricerca in Diritto pubblico, comparato e internazionale
Sapienza – Università di Roma

Abstract [It]: Lo scritto analizza alcuni aspetti etici e giuridici dell'Intelligenza Artificiale, evidenziando l'esigenza prioritaria di garantire un controllo umano effettivo sugli algoritmi (Human In The Loop). A questo riguardo, la recente Proposta di Regolamento europeo sull'IA declina la prospettiva del controllo in un approccio normativo basato sul rischio. Tale strategia regolatoria solleva tuttavia alcune perplessità e sembra privilegiare una visione incentrata sul mercato piuttosto che sull'utente.

Title: Control perspective in the age of Artificial Intelligence: few reflections upon the model *Human In The Loop*
Abstract [En]: The paper aims to analyze some ethical and legal aspects of the Artificial Intelligence systems, stressing the importance to ensure an effective human control over the algorithms (*Human In The Loop*). In the recent Proposal for a European AI Regulation, this kind of human control is premised on a risk-based approach. Despite its advantages, this regulatory strategy raises some concerns and seems to privilege a market-centered view rather than a user-centered one.

Parole chiave: Intelligenza Artificiale; governance; controllo umano; diritti; rischio; regolazione
Keywords: Artificial Intelligence; governance; human control; rights; risk; regulation

1. Intelligenza artificiale: *governance without government*?

Uno dei temi che oggi maggiormente interessa gli studiosi delle nuove tecnologie attiene all'elaborazione di una *governance* dell'Intelligenza Artificiale. Il termine «*governance*» sta ad indicare un metodo di analisi e comprensione delle dinamiche decisorie interne e sovranazionali ed è ormai invalso nel campo pubblicistico, sebbene difetti di una definizione univoca e condivisa[1]. L'attenzione che la tematica suscita è in realtà connaturata all'esigenza – avvertita sin dai primi studi sull'informatica – di governare quelle

[*] Articolo sottoposto a referaggio.
[1] Sulla *governance* e sui profili etici dell'Intelligenza Artificiale, cfr. L. Floridi, *Soft Ethics and the Governance of the Digital*, 2018. L'Autore rileva che "the real challenge is no longer digital innovation, but the *governance* of the digital" (p.3) e definisce la «governance digitale» come "the practice of establishing and implementing policies, procedures, and standards for the proper development, use and management of the infosphere". Egli precisa, inoltre, che spesso il termine «governance» viene adoperato come sineddoche, per far riferimento anche all'etica e alla regolazione digitale.

nuove forme di potere che la postmodernità ci consegna ad un ritmo spaventoso, impedendo che queste prendano il sopravvento e segnino il definitivo trionfo di una deriva antiumanista.

La necessità di regolamentare l'emersione dei nuovi domini digitali giustifica le riflessioni contenute nel presente scritto, incentrate sulla valorizzazione delle possibilità del controllo umano nell'era dell'Intelligenza Artificiale.

Di fronte alla realtà aumentata che scaturisce dall'IA, la visione antropocentrica dominante sin dai tempi dell'Illuminismo sembra cedere il passo ad un dominio dispotico ed insindacabile, che mette a repentaglio l'autonomia dell'individuo ed incrina la tradizionale tassonomia delle libertà individuali. L'idea stessa della democrazia, intesa come "governo del potere visibile"[2] vede sfumare i propri tratti caratterizzanti per assumere piuttosto le sembianze di un potere opaco, che agisce nella penombra. In questa dialettica tra visibile e invisibile, tra trasparenza e opacità[3], anche il principio personalista radicato nei moderni assetti costituzionali sembra progressivamente incrinarsi, sotto la spinta di sistemi assai pervasivi che marginalizzano la centralità dell'individuo. Nella Dichiarazione europea sui diritti e i principi digitali per il decennio digitale COM (2022)28, approvata nel gennaio 2022, viene chiaramente manifestato l'intento di elevare le aspirazioni delle persone e la loro libertà di scelta a criterio fondante il modello europeo per la transizione digitale[4].

L'obiettivo di rileggere la dicotomia umano/digitale non più in chiave distopica bensì collaborativa (*human-centric* and *trustworthy AI*) ha da sempre ispirato gli studi sull'informatica. Nel 1948 Norbert Wiener parla per la prima volta di «cibernetica», disciplina intesa a studiare i processi di controllo sulle macchine[5]. Un'esigenza analoga, quella del controllo, ispira tutt'oggi il dibattito pubblico sui sistemi di IA; non è un

[2] La visione della democrazia come "governo del potere visibile", contrapposta all'autocrazia e alle varie forme di potere eterogeneo, è il frutto di una celebre riflessione di Norberto Bobbio, comparsa per la prima volta in un articolo pubblicato su "La Stampa" il 23 novembre 1980.

[3] Di una opacità multiforme parla F. Faini, *Intelligenza artificiale e regolazione giuridica: il ruolo del diritto nel rapporto tra uomo e macchina*, in *Federalismi.it*, 2/2023, p. 1 e ss. Sottolinea brillantemente l'Autrice che l'approccio statistico e probabilistico su cui si basa l'IA è connotato da "una sorta di congenita opacità che caratterizza gli algoritmi, che si declina in un'opacità strutturale, derivante dal funzionamento degli stessi e dal fatto che resta non comprensibile persino ai programmatori l'iter logico seguito dalla macchina per giungere al risultato partendo dai dati a disposizione, cui si sommano un'opacità linguistica, dal momento che il linguaggio è informatico e non è quello naturale delle norme giuridiche, e una possibile opacità giuridica, nel caso in cui gli algoritmi e le soluzioni di intelligenza artificiale siano oggetto di diritti di proprietà intellettuale o industriale riconosciuti agli ideatori".

[4] Alla salvaguardia della libertà di autodeterminazione del soggetto è dedicato l'intero Capitolo III della Dichiarazione europea sui diritti e i principi digitali per il decennio digitale COM (2002)208, ove si evidenzia che "l'intelligenza artificiale dovrebbe fungere da strumento per le persone, con l'obiettivo ultimo di aumentare il benessere umano".

[5] N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine,* Cambridge, 1948. Sulla relazione tra cibernetica e diritto cfr. le illuminanti riflessioni di V. Frosini, *Cibernetica, diritto e società*, Milano, 1973. Cfr. anche A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in A. D'Aloia (a cura di*), Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Franco Angeli, Milano, 2020, p. 66, il quale ripropone l'affinità tra le due discipline, poiché entrambe "mirano a studiare e a rendere prevedibili i modelli di comunicazione e controllo dei comportamenti collettivi".

caso che il termine *governance*, di derivazione anglosassone, condivida la matrice etimologica di cibernetica, da *kybernao*, "dirigere una nave".

Nel dibattito politico e accademico degli ultimi venti anni, il lemma *governance* viene spesso adoperato come sinonimo di *government* inteso quale istituzione, organizzazione, apparato[6]. Questa assimilazione, tuttavia, non può valere nel campo dell'Intelligenza Artificiale e, più in generale, delle nuove tecnologie, settori nei quali l'idea stessa dello Stato-apparato subisce un radicale ridimensionamento ed anche il principio di separazione dei poteri conosce una significativa rimodulazione. È allora più opportuno che la necessità di regolamentare l'evoluzione tecnologica in atto sia descritta come una forma di «*governance without government*»[7], una strategia antropocentrica di governo della complessità tecnica.

L'universalità e la singolarità dell'Intelligenza Artificiale involgono ogni sfera del pensiero umano, senza risparmiare neanche il diritto, da sempre chiamato a conformarsi alla realtà concreta, pur con l'innegabile difficoltà di dover regolare entità sempre più mutevoli, che evolvono ad una velocità incontrollabile[8]. Come ha rilevato il Garante nazionale per la protezione dei dati personali nel Parere reso sulla Proposta di Regolamento UE sull'Intelligenza Artificiale COM 2021 (206), sia l'IA che la riservatezza sono materie trasversali, "accomunate dal rappresentare la sfida, attuale e futura, lanciata dalla tecnica al diritto e alla sua possibilità di regolamentare anche ciò che appare, nella sua evoluzione incessante, più refrattario alla norma"[9].

Una tale ritrosia potrebbe condurre a pensare che si stia affermando una nuova concezione del diritto, non più in grado di regolare i fenomeni sociali dall'esterno, ma connaturato ai medesimi e connotato da un approccio essenzialmente proattivo[10]. Il giurista si trova inevitabilmente a dover affrontare domande nuove: quali aspetti regolare dell'IA e quali tralasciare, consentendo al "disordine ordinato"[11] di fare il suo

---

[6] Su questo tema cfr. le considerazioni di P. Benanti, *Human in the loop. Decisioni umane e intelligenze artificiali*, Mondadori, Milano, 2022, p. 135. L'Autore sottolinea che il termine *governance* è privo di un sostantivo corrispondente nella lingua italiana ed è generalmente utilizzato da economisti, politologi ed esperti di relazioni internazionali per marcare una distinzione ed una contrapposizione con il *government*. La nozione di *governance* si è diffusa anche a seguito dell'utilizzo del termine in alcuni documenti europei ufficiali, tra cui meritano di essere segnalati: *La governance, Un libro bianco*, COM (2001), 25 luglio 2001 e *OECD Economic Glossary. English-France*, OECD, Paris, 2006. Su questi ed altri aspetti, cfr. F. Scamardella, *La governance: genesi, diffusione e disavventure di un lemma fortunato*, in *Riv. Filosofia del Diritto*, 1/ 2021, p.11 e ss.

[7] L'espressione è mutuata dal titolo del volume di J. Rosenau e E. Czempiel, *Governance without Government: Order and Change in World Politics* (Cambridge Studies in International Relations), Cambridge University Press, 1992, DOI https://doi.org/10.1017/CBO9780511521775.

[8] Sul complesso rapporto tra IA e diritto si rinvia alla vasta letteratura di riferimento tra cui, *ex plurimis*, A. D'Aloia (a cura di), *Intelligenza artificiale e diritto*, *cit.*; G. Alpa, *Diritto e intelligenza artificiale. Profili generali, soggetti, contratti, responsabilità civile, diritto bancario e finanziario, processo civile*, Pacini Giuridica, Pisa, 2020; G. Cerrina Feroni, C. Fontana, E.C. Raffiotta (a cura di), *AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Il Mulino, Bologna, 2022, G. Sartor, *L'intelligenza artificiale e il diritto*, Giappichelli, Torino, 2022.

[9] Il testo del Parere è liberamente consultabile sul sito del Garante.

[10] Su queste riflessioni, cfr. A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Mondadori Università, Città di Castello, 2020, p. 26.

[11] L'eloquente espressione, riferita al mondo virtuale e all'*Internet of Things*, si deve a T.E. Frosini, *Il costituzionalismo nella società tecnologica*, *cit.*, reperibile online.

corso? Che tipo di regole adottare, soltanto giuridiche o anche etiche e tecniche e quale grado di vincolatività imporre?

Le riflessioni che seguono si focalizzano su alcuni aspetti etici e giuridici dell'IA, osservati attraverso la lente del controllo. Nella prima parte dello scritto, la prospettiva del controllo si dipana nelle trame dell'etica, traducendosi nella garanzia «*to keep human in the loop of AI*», che trova i suoi corollari nei principi di trasparenza, conoscibilità e non esclusività, valorizzati anche dai recenti orientamenti della giurisprudenza amministrativa sul cd. provvedimento algoritmico.

Nella seconda parte, il testo si concentra su alcuni profili regolatori dell'Intelligenza Artificiale, declinando la prospettiva del controllo nel contesto normativo europeo, fondato sul modello della gestione del rischio. La citata Proposta di Regolamento sull'IA mostra invero di condividere la prospettiva antropocentrica di governo della tecnica, rendendo però arduo circoscrivere i margini dell'intervento umano sui sistemi algoritmici. Seguono alcune riflessioni conclusive sulle reali modalità di valorizzazione del controllo in un'epoca caratterizzata da una nuova morfologia del potere di fronte all'emersione della sovranità digitale.

2. Le dimensioni del controllo tra etica e diritto: *human in the loop of AI*?

Nel 1958 Isaac Asimov, noto per l'elaborazione delle Leggi sulla robotica, scrive un racconto dal titolo "*Feeling of Power*", tradotto in italiano con "Nove volte sette". È la storia di un piccolo ometto dalle abilità intellettive sovraumane che risolve operazioni matematiche alla velocità di un calcolatore elettronico. Il racconto non ha un lieto fine perché il piccolo tecnico si determina al suicidio, schiacciato dal timore di perdere il controllo su tutto ciò che di potente è in grado di realizzare. La paura del robot che sovrasta l'umano è da sempre connaturata all'affermazione della tecnologia, come dimostra il fatto che le prime due leggi della robotica (comparse per la prima volta in un altro racconto scritto nel 1942, dal titolo "Circolo vizioso" contenuto nella medesima raccolta "Io, robot") siano finalizzate, direttamente o indirettamente, a garantire il primato dell'uomo.

Prima Legge (Non malvagità): Un robot non può recar danno a un essere umano né può permettere che, a causa del suo mancato intervento, un essere umano riceva danno;

Seconda Legge (Controllo umano): Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non vadano in contrasto alla Prima Legge;

Terza Legge (Giustizia): Un robot deve proteggere la propria esistenza, purché la salvaguardia di essa non contrasti con la Prima o con la Seconda Legge.

L'età postmoderna è caratterizzata da un pervasivo bisogno di controllabilità, inteso come il potere di rendere la realtà "visibile, accessibile, gestibile e utile"[12], in una parola governabile. Lo sviluppo dell'Intelligenza Artificiale incrina la prospettiva del controllo, sia sul versante pubblicistico che su quello privatistico.

Dal primo punto di vista, entra in crisi la dimensione tradizionale del costituzionalismo, quale dottrina del limite al potere, e si afferma un nuovo costituzionalismo (cd. costituzionalismo post-moderno o digitale[13]), la cui funzione risulta inevitabilmente conformata dall'esigenza di limitare il potere algoritmico, per definizione non più concentrato in un luogo, ma naturalmente de-spazializzato e quindi difficile da limitare. L'impossibilità di definire i confini del mondo nuovo rende ardua l'individuazione di una qualche forma di regolazione giuridica, atteso che il diritto è per definizione legato ad uno spazio, al «dove»[14]. In questo contesto, l'Unione Europea interviene con discipline stratificate, improntate alla co-regolazione, nella consapevolezza che solamente un approccio normativo condiviso e partecipativo possa consentire una regolazione efficace nell'era della complessità[15].

Dal punto di vista privatistico, l'avvento dell'IA, dei Big Data e delle nuove tecnologie viene seriamente ad impattare sull'autonomia dell'individuo, tanto da rendere imprescindibile l'esigenza "*to keep the human in the loop of AI*", ossia di garantire che gli individui siano attori informati delle scelte che li riguardano[16].

In verità, la scissione tra prospettiva pubblicistica e privatistica è probabilmente il frutto di una concezione separatista e tralatizia del diritto, non più attuale di fronte a fenomeni nuovi che richiedono un approccio unitario ed interdisciplinare, in grado di coniugare ambiti diversi della scienza, inclusa quella giuridica. Controllo del potere (*rectius* dei poteri) e autonomia dell'individuo sono infatti due facce della medesima medaglia, destinate a interferire nel campo dell'IA, ove è ineluttabile il coinvolgimento delle libertà individuali, connotato insopprimibile della persona umana e valore fondante dello Stato di diritto.

Gli effetti pervasivi degli algoritmi sulla società sono alla base della dimensione etica sottesa allo sviluppo dei sistemi di IA, una dimensione difficile da perimetrare, alla luce dei rischi inevitabili ed imprevedibili

[12] La citazione è tratta da H. Rosa, *Unverfügbarkeit* (Indisponibilità), Salisburgo-Vienna, Residenza, 2018, richiamato in G. Lo Storto – D. Manca, Prefazione a F. Pasquale, *Le nuove leggi della robotica. Difendere la competenza umana nell'era dell'Intelligenza Artificiale*, Luiss University Press, Roma, 2021, p.9.

[13] Cfr. G. Di Gregorio, *Digital Constitutionalism in Europe*, Cambridge University Press, Cambridge, 2022.

[14] Cfr. N. Irti e E. Severino, *Dialoghi su diritto e tecnica*, Roma-Bari, Edizioni Laterza, 2001.

[15] In particolare, la Proposta di Regolamento sull'IA, di cui si dirà *infra*, si inscrive in un quadro di regolamenti che delineano la cd. strada europea per l'IA: il *Data Governance Act* (Regolamento UE 2022/869), il *Digital Services Act* (Regolamento UE 2022/2065), il *Digital Markets Act* (Regolamento UE 2022/1925) e il *Data Act* (proposto il 23 febbraio 2022). Sugli approcci regolatori in questo settore si veda il recente scritto di A. Simoncini, *La co-regolazione delle piattaforme digitali*, in *Rivista Trimestrale di Diritto Pubblico*, 4/2022, pp. 1031-1050.

[16] Cfr. A. Pajno, M. Bassini, G. De Gregorio, M. Macchia, F.P. Patti, O. Pollicino, S. Quattrocolo, D. Simeoli, P. Sirena, *Intelligenza Artificiale: criticità emergenti e sfide per il giurista*, in *BioLaw Journal*, 3/2019, p. 205 ss.

per i diritti umani. Questi aspetti sono oggetto di studio da parte dell'«algoretica»[17], neologismo utilizzato per indicare quella branca dell'etica - comunemente nota come «etica applicata» - che analizza le implicazioni economiche, sociali e culturali di un certo fenomeno ed i modelli comportamentali di gestione[18]. Nel campo dell'IA, la letteratura suole spesso riferirsi anche all'«etica digitale», dedicata allo studio ed alla valutazione delle questioni morali relative a dati, informazioni, algoritmi e alle pratiche corrispondenti, al fine di formulare soluzioni moralmente valide[19].

Nel 2018, un Gruppo di Esperti di alto livello sull'IA, nominato dalla Commissione europea, ha redatto un catalogo dei principi etici per promuovere un'IA affidabile durante l'intero ciclo di vita del sistema, basato essenzialmente su tre componenti: legalità; eticità e robustezza[20]. Nello specifico, tra gli imperativi etici che devono guidare lo sviluppo ed il funzionamento di qualsiasi sistema di IA figurano quello di prevenzione dei danni (*Do Not Significant Harm*, DNSH), di equità, di esplicabilità e del rispetto dell'autonomia umana[21].

L'IA antropocentrica si fonda quindi su principi che hanno già un precipuo riconoscimento giuridico in altri atti e documenti europei (si pensi *in primis* alle norme in materia di protezione dei dati personali o a tutela dei consumatori) e tuttavia, come espressamente chiarito dal Gruppo di esperti, "l'adesione ai principi etici va oltre il rispetto formale del diritto vigente"[22].

Una delle prime accezioni in cui si declina lo *human in the loop* (HITL) è dunque connessa al rispetto dell'autonomia umana, che si traduce nell'esigenza di "garantire la sorveglianza ed il controllo dei processi operativi nei sistemi di IA da parte di esseri umani"[23]. Lo HITL opera come principio «a geometria

[17] Il lemma «algoretica» è stato coniato dal teologo Paolo Benanti per indicare una disciplina nata in risposta all'«algocrazia», cioè al dominio degli algoritmi, e volta ad indagare i risvolti etici derivanti dall'uso massiccio delle tecnologie informatiche. Su questi aspetti cfr. l'opera di P. Benanti, *Oracoli. Tra algoretica e algocrazia*, Luca Sossella editore, Roma, 2018.

[18] Su questi ed altri profili di carattere etico cfr. U. Pagallo, *Etica e diritto dell'Intelligenza Artificiale nella governance del digitale: il Middle-out Approach,* in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, Giuffrè Francis Lefebvre, Milano, 2020. Anche secondo le Linee Guida elaborate dall'High-Level Expert Group on AI (*Ethics guidelines for trustworthy AI*, disponibili online), l'etica applicata rappresenta una delle partizioni dell'etica accanto a metaetica, etica normativa ed etica descrittiva.

[19] L. Floridi, *op. cit*., p. 3. L'Autore ricostruisce il rapporto tra etica e *governance* secondo una relazione di complementarietà, per cui l'una conforma l'altra ma entrambe restano diverse e non totalmente sovrapponibili.

[20] High-Level Expert Group on AI, *Ethics guidelines for trustworthy AI*, *cit*., consultabili online.

[21] I quattro principi etici si ricollegano ad analoghi diritti fondamentali riconosciuti dalla Carta di Nizza. In particolare, la prevenzione dei danni è strettamente connessa alla protezione dell'integrità fisica e psichica (sancita dall'art.3). L'equità è strettamente connessa ai diritti alla non discriminazione, alla solidarietà e alla giustizia (artt.21 e ss.). L'esplicabilità e la responsabilità sono strettamente connesse ai diritti relativi alla giustizia (art.47). Infine, il rispetto dell'autonomia umana è strettamente connesso al diritto alla dignità umana e alla libertà (sanciti dagli artt. 1 e 6 della Carta).

[22] High-Level Expert Group on AI, *Ethics guidelines for trustworthy AI, cit.,* p.13.

[23] Secondo il Gruppo di Esperti, *ibidem*, "gli esseri umani che interagiscono con i sistemi di IA devono poter mantenere la propria piena ed effettiva autodeterminazione e devono poter essere partecipi del processo democratico. I sistemi di IA non devono subordinare, costringere, ingannare, manipolare, condizionare o aggregare in modo ingiustificato gli esseri umani. Al contrario, devono essere progettati per aumentare, integrare e potenziare le abilità cognitive, sociali e

variabile» e trova differenti declinazioni a seconda del settore di incidenza del sistema di IA. L'esigenza di assicurare la sorveglianza umana risulta invero maggiormente avvertita quanto maggiore è il grado di autonomia delle macchine[24].

Secondo la definizione contenuta nella Risoluzione del Parlamento Europeo del 2017, l'autonomia si sostanzia nella capacità della macchina "di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o un'influenza esterna"[25]. La prospettiva del controllo rappresenta dunque la chiave di volta per arginare l'autonomia decisionale ed operativa delle macchine, considerata dall'UE la vera rivoluzione indotta dall'Intelligenza Artificiale[26].

In particolare, l'apporto del supervisore umano sulla macchina può tradursi in prima battuta nell'«approccio con intervento umano» (*Human In The Loop in senso proprio*), che è il modello più garantista e prevede una costante interazione uomo/macchina. Questo approccio è tipico di settori ad alto rischio, quali la diffusione di auto a guida autonoma, in cui all'uomo deve essere riconosciuto un controllo costante, che gli consenta anche di arrestare il sistema qualora non lo ritenga sicuro.

Diverso è l'«approccio con supervisione umana» (*Human On The Loop*) che assicura un controllo minimo, solo *ab estrinseco*, in fase di progettazione o di monitoraggio del sistema. È diffuso, ad esempio, in ambito medico, per consentire all'operatore sanitario di intervenire a valle di una radiografia eseguita dalla macchina.

Da ultimo, l'«approccio con controllo umano» (*Human In Command*) consente un monitoraggio costante sul sistema e sui suoi effetti, lasciando ampia discrezionalità al supervisore che potrebbe finanche decidere di non servirsene in una certa situazione o di ignorare la decisione assunta mediante l'IA[27].

La letteratura in materia e la recente legislazione europea riferiscono la garanzia della sorveglianza umana unicamente ai processi decisionali che adoperano sistemi di *machine learning* e, in particolare, di *deep learning*[28], contesti cioè nei quali la decisione algoritmica assume la fisionomia di una "*black box*"[29]. Il

culturali umane. La distribuzione delle funzioni tra esseri umani e sistemi di IA dovrebbe seguire i principi di progettazione antropocentrica e lasciare ampie opportunità di scelta all'essere umano".

[24] Cfr. L. Rinaldi, *Intelligenza artificiale, diritti e doveri nella Costituzione italiana*, in *DPCE online*, 1/2022, p. 201 ss.

[25] Parlamento europeo, Risoluzione del 16 febbraio 2017, 2015/2103(INL). Il documento, recante "Raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica", riveste un'importanza prioritaria nel quadro normativo europeo sul digitale, atteso che per la prima volta si dà atto delle implicazioni di carattere etico e giuridico derivanti dalla diffusione dell'IA e dei rischi sottesi in punto di prevedibilità e trasparenza di questi sistemi.

[26] Cfr. A. Amidei, *La* governance *dell'Intelligenza Artificiale: profili e prospettive di diritto dell'Unione Europea*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica, cit.,* p. 571 ss.

[27] Per un'analoga classificazione cfr. Commissione europea, *Libro Bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia* (COM 2020/65 final), 19 febbraio 2020, pp. 21-23.

[28] Per un inquadramento sugli aspetti tecnici v. P. Traverso, *Breve introduzione tecnica all'Intelligenza Artificiale*, in *DPCE online*, 1/2022, pp. 155-168. Cfr. anche F. Lagioia – G. Sartor, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica,* in *Federalismi.it,* 11/2020, p. 88 e ss., i quali distinguono tre principali approcci di apprendimento automatico: apprendimento supervisionato, apprendimento per rinforzo e apprendimento non supervisionato.

[29] La fortunata espressione si deve all'opera di F. Pasquale, *The black box society: the secret algorithms that control money and society*, Harvard University Press, 2016.

pensiero va in primo luogo alle reti neurali artificiali, costruite per imitare, mediante l'autoapprendimento, un'operazione tipica della mente umana e compierla ad una velocità assai superiore. Tali sistemi si autodeterminano in base a regole interne non conoscibili né *ex ante* né *ex post*, rendendo così difficile governare la macchina ed eventualmente correggerne *output* errati o comunque indesiderati.

Uno scenario, quello attuale, non troppo lontano da quello che immaginava il padre dell'Intelligenza Artificiale, Alan Turing, il quale negli anni Cinquanta osservava: "Una macchina capace di apprendere realizza gli obiettivi che le sono affidati, senza che l'uomo abbia indicato alla macchina come procedere, e anzi senza che egli abbia consapevolezza di ciò che accade all'interno della macchina"[30].

3. La conoscibilità tra legge ed algoritmo: *Code is Law*?

Alla luce del quadro tracciato, appare evidente il rischio che dinamiche decisorie sorrette da tecnologie di IA possano facilmente infrangere il principio di legalità e quei corollari che rappresentano il cardine dello Stato di diritto: trasparenza, conoscibilità e non esclusività della decisione algoritmica.

In questo contesto, il modello di Stato tradizionale vede dissolversi i suoi elementi costitutivi: "la produzione, diffusione, circolazione, elaborazione dei dati per definizione non incontra limiti territoriali, non conosce un popolo di riferimento (…) e, soprattutto, non accetta, non riconosce, non subisce l'autorità dello Stato che non è più in grado di esercitare la sua sovranità su un territorio e su un popolo"[31]. Ne deriva una "sensazione diffusa di anomia, che diventa pervasiva e lascia le nostre società attonite"[32].

L'esigenza di regolare e di orientare finalisticamente i comportamenti umani sembra l'aspetto che accomuna il formante legislativo ("*ubi societas ibi ius*") ed il fenomeno algoritmico ("*ubi data society ibi ius*"[33]). Come afferma Paolo Benanti in una recente intervista, questa affinità è in verità più apparente che reale: sia l'algoritmo che la legge condividono forse le finalità, ma divergono significativamente nella fisionomia, atteso che la legittimità della legge proviene dalla sua conoscibilità, universalità e generalità e nessuno di questi aspetti trova riscontro nell'algoritmo[34].

---

[30] L'osservazione è riportata da G. Sartor e F. Lagioia, *Le decisioni algoritmiche tra etica e diritto*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica, cit.,* p. 69.

[31] B. Caravita, *Letture di diritto costituzionale*, Giappichelli Editore, Torino, 2021, p. 64.

[32] *Ibidem*.

[33] La locuzione, assai eloquente, è di F. Faini, *Data society. Governo dei dati e tutela dei diritti nell'era digitale*, Giuffrè Francis Lefebvre, Milano, 2019.

[34] Il testo dell'intervista è reperibile online. Secondo il teologo, "tra i principi formali dello Stato di diritto ci sono le leggi. Gli algoritmi hanno una funzione molto simile a quella delle leggi, che sono fatte per orientare i comportamenti dei cittadini. Ma una legge è legittima se conoscibile, se è universale e se è generale (…). Gli algoritmi sono universali e conoscibili? Il codice non è conoscibile, perché è protetto da *copyright*, ma anche se fosse *open source* nel momento in cui viene compilato da una macchina il compilatore può iniettare qualsiasi cosa nel codice e renderlo non più conoscibile. È universale? No, perché profila e sceglie lui a chi rivolgersi. È generale? No, perché obbedisce solo al soggetto che rimane in quel *server*. Allora ecco che questa nuova disposizione degli algoritmi, nati per motivi commerciali e ora capaci di orientare l'opinione pubblica, generano una tensione a un livello molto profondo tra loro e lo Stato di diritto. La questione è perciò questa: oggi ci troviamo di fronte a un bisogno di dare alla democrazia dei nuovi anticorpi di fronte

La questione della conoscibilità non rappresenta un problema soltanto etico ma ha la sua rilevanza anche sul piano giuridico, come conferma la circostanza che questi principi - oltre ad avere un esplicito riconoscimento normativo nella disciplina europea sulla protezione dei dati personali - hanno di recente rappresentato anche l'oggetto di alcune importanti pronunce del giudice amministrativo nazionale. In quella occasione, il Consiglio di Stato – dopo aver superato le obiezioni mosse in punto di conformità del provvedimento algoritmico al principio di legalità dell'azione amministrativa e al rispetto degli *hearing rights*[35] - mostra di ritenere fondamentale la garanzia dell'interazione uomo-macchina anche nell'ambito dei rapporti amministrativi, verificando in che misura sia ipotizzabile un controllo umano sul procedimento. Invero, richiamando la disciplina europea a tutela della privacy, il giudice amministrativo declina il principio *human in the loop* in una duplice prospettiva volta ad assicurare, da un lato, la piena conoscibilità dei criteri applicati e del modello adoperato (principio di trasparenza) e, dall'altro, l'imputabilità della decisione all'organo titolare del potere, il quale deve essere in grado di verificare la logicità e la legittimità della scelta e degli esiti, intervenendo per validare ovvero smentire la decisione automatizzata[36].

Sul piano normativo, come accennato, l'esigenza di conoscibilità ispira in primo luogo la disciplina europea sulla protezione dei dati personali (Regolamento UE 2016/679, *General Data Protection Regulation* - GDPR) ed implica la comprensibilità (artt. 13 - 14) e la non esclusività della decisione automatizzata (art. 22)[37], in un contesto di *legal protection by design and by default*.

queste nuove sfide". Per considerazioni analoghe cfr. A. Nuzzo, *Algoritmi e regole*, in *Analisi Giuridica dell'Economia*, 1/2019, p. 39, il quale sottolinea che la "legge algoritmica si trova priva di punti di incontro con la legge statuale" e "l'algoritmo – figlio dell'autonomia privata – s'impone oggi come nuovo potere in grado di esercitare una forza uguale e contraria, comunque autonoma rispetto a quella della norma statuale".

35 Questi profili hanno inizialmente determinato un atteggiamento di chiusura della giurisprudenza amministrativa (soprattutto di merito) rispetto al ricorso a procedure informatiche, che avrebbe comportato "la deleteria prospettiva orwelliana di dismissione delle redini della funzione istruttoria e di abdicazione a quella procedimentale" (Tar Lazio, Sez. III *bis*, sentenza n. 9224 del 10 settembre 2018). La svolta si è avuta con alcune sentenze del Consiglio di Stato rese nella nota vicenda della "buona scuola". Cfr., da ultimo, Cons. Stato, Sez. VI, sentenza n. 881 del 4 febbraio 2020, che richiama i precedenti in termini di Sez. VI, n.2270 dell'8 aprile 2019 e n.8472 del 13 dicembre 2019. Di recente, sulla distinzione tra impiego di un algoritmo e utilizzo di sistemi di Intelligenza Artificiale, v. Cons. Stato, Sez. III, sentenza n. 7891 del 25 novembre 2021.

36 Per un'analisi della giurisprudenza amministrativa e del principio di trasparenza nelle procedure automatizzate cfr. N. Muciaccia, *Algoritmi e procedimento decisionale: alcuni recenti arresti della giustizia amministrativa*, in *Federalismi.it,* 10/2020. Cfr. anche G: Lo Sapio, *La trasparenza sul banco di prova dei modelli algoritmici*, in *Federalismi.it*, 11/2021, p. 242 ss. e I.A. Nicotra e V. Varona, *L'algoritmo, intelligente ma non troppo*, in *Rivista AIC*, 4/2019. Cfr. anche M. Luciani, *La decisione giudiziaria robotica*, in A. Carleo (a cura di), *Decisione robotica*, Il Mulino, Bologna, 2020, p. 63 e ss.

37 Su questi aspetti, si vedano i contributi contenuti in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, Torino, 2018. Si veda anche G. Finocchiaro, *Intelligenza Artificiale e protezione dei dati*, in *Giurisprudenza italiana,* 7/2019, p. 1657 ss. e M. Palmirani, *Interpretabilità, conoscibilità, spiegabilità dei processi decisionali automatizzati*, in U. Ruffolo (a cura di), *XXVI lezioni di Diritto dell'Intelligenza artificiale*, Giappichelli, Torino, 2020, p. 66 ss.

Risulta tuttavia evidente che il «diritto alla spiegabilità»[38] dell'algoritmo spesso trascende il campo della protezione dei dati personali per intersecare quello più generale della tutela della persona umana nella sua complessità[39]. La comprensibilità (o *explainability*) si traduce nella pretesa del soggetto destinatario alla conoscenza dell'esistenza di decisioni algoritmiche che lo riguardano e, al contempo, nel dovere in capo al responsabile del trattamento dei dati di informare l'interessato: si traduce, in una parola, nella trasparenza del sistema di IA[40]. È tuttavia dibattuto se la trasparenza operi soltanto *ex ante*, ossia in un momento antecedente al trattamento automatizzato, ovvero sia tale da imporre al titolare del trattamento di fornire una specifica giustificazione sulla decisione finale assunta dal sistema, in funzione di garanzia *ex post*[41].

Nella realtà dei fatti, tuttavia, è ingenuo pensare che, inseriti determinati *input*, l'algoritmo generi sempre esiti comprensibili o che sia sempre conoscibile l'iter in base al quale determinati *output* sono stati generati: il rischio è dunque quello di produrre un risultato sostanzialmente opaco che incrina la fiducia degli individui verso la decisione algoritmica, impattando anche sulle libertà e sui diritti fondamentali delle persone[42]. Per questo motivo, in funzione compensativa, opera il principio di non esclusività della decisione. L'art. 22 del GDPR prevede infatti che, nel caso in cui una decisione automatizzata «produca effetti giuridici che riguardano o che incidano significativamente su una persona», questa ha diritto a che tale decisione non sia basata unicamente su tale processo automatizzato.

Deve cioè comunque esistere nel processo decisionale un contributo umano capace di controllare, validare ovvero smentire la decisione automatica. Questa affermazione, chiara declinazione dello HITL,

[38] V. L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, e E. Vayena, *AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and Machines,* 2018, 28 (4), pp. 689-707.

[39] Cfr. U. Pagallo, *Algoritmi e conoscibilità*, in *Rivista di filosofia del diritto*, 1/2020, p. 93 ss.

[40] Il tema della trasparenza può essere indagato da molteplici profili che attengono, in via esemplificativa, alla tracciabilità dei dati, all'accesso al codice sorgente posto a fondamento dell'algoritmo e alla connessa protezione della proprietà intellettuale. Su questi aspetti si veda M. Palmirani, *Big Data e conoscenza*, in *Riv. Filosofia del Diritto*, 1/2020, p. 74 e ss. e V. Zeno-Zencovich, *Dati, grandi dati, dati granulari e la nuova epistemologia del giurista*, in *Rivista di diritto dei media*, 2/2018, pp. 32-38.

[41] Sui termini del dibattito cfr. S. Wachter, B. Mittelstadt e L.Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law* 2017, 7 (2), pp. 76-99. V. anche E. Longo, *I processi decisionali automatizzati e il diritto alla spiegazione*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?,* Vol. I, Il Mulino, Bologna, 2022, p. 349 ss. L'Autore evidenzia come "un mero diritto alla spiegazione che si basi solo sul rendere trasparente la logica utilizzata dagli algoritmi di intelligenza artificiale potrebbe non bastare. Per realizzarsi nella sua pienezza, in molte circostanze l'informazione fornita dovrebbe includere anche la conoscibilità dei dati che hanno dato vita alla decisione automatizzata e soprattutto si dovrebbe consentire a chi gestisce la macchina di agire verso il sistema in ogni momento. Un intervento umano senza che siano noti il set di dati e i fattori che possono comportare le loro inesattezze sarebbe inutile".

[42] Su questi aspetti cfr. l'interessante scritto di M. Ebers, *Regulating AI and Robotics: Ethical and Legal Challenges*, in M. Ebers, S. Navas Navarro (eds.), *Algorithms and Law*, Cambridge University Press, Cambridge, 2019, disponibile in *open access*.

rischia tuttavia di tradursi in una mera enunciazione di principio, per due ordini di ragioni[43]: da un lato, perché lo stesso articolo contempla una serie di eccezioni di ampissima portata che, di fatto, finiscono per svuotare di contenuto la regola (si pensi per esempio al consenso esplicito dell'interessato, paradigma oggi decisamente inadatto a fornire un livello adeguato di protezione all'utente in rete). Dall'altro lato, il principio di *accountability* impone di non sottoporre i privati a decisioni basate *unicamente*[44] sul trattamento automatizzato, non operando perciò ogniqualvolta vi sia una cooperazione (anche minima) tra paradigma decisorio umano e algoritmico[45]. Tutti i principi evocati rispondono, in ultima istanza, alla necessità di garantire all'essere umano la possibilità di capire e tenere traccia delle decisioni prese dal sistema: una tale aspirazione rischia tuttavia di rimanere meramente ideale laddove sia destinata ad operare in un contesto che evolve ad una velocità strabiliante[46].

Alla luce del quadro tratteggiato, l'attuazione della garanzia di conoscibilità sembra lasciare irrisolti numerosi nodi problematici. Il diritto all'intervento umano potrebbe infatti non essere effettivo, in un ambito, come quello dell'IA, che si nutre continuamente di dati, per definizione non intellegibili all'uomo[47]. Anche qualora venisse consentito l'accesso al dataset di *input* del sistema, permane il dubbio che questo non sia sufficiente ad assicurare la dimensione costituzionale nell'ecosistema virtuale[48]. Ci sono infatti alcuni settori in cui il paradigma dello HITL appare difficilmente attuabile, stante il rischio paventato in dottrina del cd. *effetto moutunnier*[49], e cioè la tendenza del decisore umano ad appiattirsi sul

[43] Cfr. E. Falletti, *Decisioni automatizzate e diritto alla spiegazione: alcune riflessioni comparatistiche*, in *Il Diritto dell'informazione e dell'informatica*, 2/2019, in cui l'Autore parla di questa norma in termini di "barriera difensiva simbolica". V. anche B. Marchetti, *La garanzia* dello human in the loop *alla prova della decisione algoritmica amministrativa*, in *BioLaw Journal*, 2/2021, p. 367 ss.

[44] In merito all'utilizzo dell'avverbio «unicamente» è emerso un dibattito tutt'oggi non sopito, i cui termini sono riassunti da E. Pellecchia, *Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation*, in *Leggi civili commentate*, 2018, 5, p. 1224.

[45] Cfr. D. Martire, *Intelligenza artificiale e Stato costituzionale*, in *Diritto Pubblico*, 2/2022, p. 397 ss.

[46] Si pensi alla recente diffusione di sistemi di IA generativa sul modello di chatGPT, che mettono a dura prova la tenuta del sistema.

[47] Eloquente è la definizione di *Big Data* che troviamo sul sito dello *European Data Protection Supervisor* (EDPS): "Big Data means large amounts of different types of data produced at high speed from multiple sources, whose handling and analysis require new and more powerful processors and algorithms. Not all these data are personal, but many players in the digital economy increasingly rely on the large-scale collection of and trade in personal information. As well as benefits, these growing markets pose specific risks to individual's rights to privacy and to data protection". Sui *Big Data* e sul dibattito definitorio v. G. De Minico, *Big Data e la debole resistenza delle categorie giuridiche. Privacy e lex mercatoria*, in *Diritto Pubblico* 1/2019, p. 89 e ss.

[48] Sul tema della conoscibilità e sulla rilevanza anche delle cd. ontologie informatiche, cfr. M. Palmirani, *Big Data e conoscenza*, *cit.*

[49] L'espressione, invocata con riguardo al settore della giustizia, si deve a A. Garapon, J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Paris, PUF, 2018, p. 239.
Il rischio di un *anchoring effect* è messo in luce anche da D.U. Galetta, *Human-stupidity-in-the-loop? Riflessioni (di un giurista) sulle potenzialità e i rischi dell'Intelligenza Artificiale*, Editoriale del 22 febbraio 2023, in *Federalismi.it*, 5/2023. Con specifico riferimento all'assunzione di decisioni amministrative, l'Autrice sottolinea che una delle limitazioni più note della mente umana attiene ai cd. pregiudizi cognitivi, cioè "errori sistematici che influenzano il nostro giudizio e il nostro processo decisionale", e tra cui rientrano i cd. *bias* di conferma, in base ai quali "gli esseri umani tendono cioè a cercare e validare

ragionamento algoritmico e a non discostarsene reputandolo tendenzialmente infallibile. Per queste ragioni, non è isolata la tesi di dar vita ad un *nuovo* diritto alla spiegabilità nel contesto algoritmico[50], che prenda atto dell'assunto per cui oggi, nella dimensione digitale, anche il codice informatico detta legge ("*code is law*"[51]).

Lo scenario che si prospetta è compendiato in un *framework* normativo in cui coesistono fonti assai eterogenee, dal diritto positivo - nazionale e sovranazionale - alle norme sociali e consuetudinarie, con un ruolo di spicco riconosciuto alle forze di mercato e ai poteri privati. La complessità del quadro normativo e la carenza di conoscibilità dei dati che compongono il processo decisionale automatizzato necessitano di essere "bilanciate" da adeguati meccanismi di responsabilità, al fine di predisporre una regolazione sostanzialmente uniforme che non risenta delle diverse impostazioni giuridiche nazionali[52].

4. La Proposta di Regolamento dell'Unione europea: il controllo del rischio

La garanzia di un controllo umano sul dominio algoritmico è stata presa in considerazione anche nell'ambito della recente Proposta di Regolamento sull'IA (COM/2021/206 *final*, noto come *AI Act*), presentata dalla Commissione il 21 aprile 2021. In un'ottica antropocentrica, il legislatore europeo sembra compiere un importante passo in avanti e, nel solco delle indicazioni contenute nel Libro Bianco sull'intelligenza artificiale[53], decide di condividere la declinazione più rigorosa del principio *human in the loop*. Il principio del controllo, insieme a quello di prevenzione (*Do Not Significant Harm*) e cooperazione istituzionale, rappresenta uno dei cardini su cui si fondano le diverse tecniche di protezione adoperate dal legislatore europeo[54].

Consapevole della insopprimibilità dei rischi insiti nelle nuove tecnologie per la salute, la sicurezza e i diritti fondamentali, l'art.14 dell'*AI Act* prevede che l'essere umano debba essere posto nella condizione di comprendere le potenzialità ed i limiti del sistema, in modo da rilevare ed affrontare tempestivamente possibili anomalie, potendo giungere finanche ad interrompere il funzionamento della macchina mediante uno *stop button*. Si adotta dunque un modello di sorveglianza del tipo *human in command*, ritenuto il più adeguato al fine di impedire la temuta deriva antiumanista. Tale modello risulta però riferito unicamente

quelle informazioni che supportano le loro convinzioni; e ad ignorare invece le informazioni che le contraddicono". Sulla base di ciò, difficilmente il funzionario umano si discosterà dalle risultanze dell'algoritmo.

[50] V. M. Fasan, I *principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative*, in *DPCE online*, 1/2022, p.181 ss.

[51] Il tema è approfondito nel celebre libro di L. Lessig, *Code and Other Laws of Cyberspace*, New York, Basic Book, 1999.

[52] Sui vari modelli di responsabilità prospettabili cfr. F. Faini, *Intelligenza artificiale e regolazione giuridica, cit.,* pp. 13-14.

[53] Commissione europea, *Libro Bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia* (COM 2020/65 final), 19 febbraio 2020, p. 23.

[54] G. Alpa, *Quale modello normativo per l'intelligenza artificiale*? *cit.*, p. 1011.

ai «sistemi ad alto rischio», potenzialmente idonei ad avere un impatto notevole sugli individui e perciò soggetti a regole specifiche e più stringenti rispetto alle restanti applicazioni di IA.

Come noto, il legislatore europeo ha infatti prescelto un modello di regolazione duttile, prevedendo obblighi la cui rigidità viene parametrata al maggiore o minore livello di rischio del sistema di IA[55], desumibile dalla funzione svolta e dalle «finalità e modalità specifiche di utilizzo»[56]. Una classificazione di questo tipo risponde all'esigenza di tutelare i diritti fondamentali dell'individuo, evitando che il grado di rischio diventi intollerabile ed identificando, a tal fine, un punto di equilibrio tra sviluppo economico e protezione della sfera individuale.

In particolare, il Regolamento individua quattro livelli di rischio[57]. In primo luogo, si distinguono sistemi «a rischio inaccettabile» (quali quelli che utilizzano tecniche subliminali al fine di distorcere il comportamento umano, come i sistemi di *social scoring*) vietati perché in contrasto con i valori dell'Unione, salvo che sussistano alcune situazioni eccezionali che ne giustifichino l'utilizzo (art. 5). Queste deroghe, perlopiù rispondenti a ragioni di ordine pubblico, lasciano in realtà un ampio margine di discrezionalità in capo allo Stato[58]. Ciò, se da un lato favorisce la naturale flessibilità di una disciplina volta a regolamentare sistemi in continua evoluzione[59], dall'altro apre al rischio di un'eccessiva frammentazione

[55] Per «rischio» si intende, in senso ampio, la possibilità che si verifichino impatti negativi di qualsiasi genere su un individuo o sulla società. Una regolazione fondata sul *risk-based approach* comporta l'individuazione di differenti classi di rischio e, specularmente, l'individuazione di regimi normativi differenziati. Cfr. punto 26.1 delle *Policy and investiment reccomendations for thrusty AI,* pubblicate il 29 giugno 2019 e disponibili qui. Lo stesso approccio era stato condiviso dal Libro Bianco della Commissione sull'Intelligenza Artificiale del 19 febbraio 2020, che impone il modello *risk-based* unicamente per i sistemi ad alto rischio, destinatari di una regolazione di carattere prescrittivo, a differenza delle altre applicazioni, per cui si ritiene sufficiente un sistema di etichettatura su base volontaria (cfr. p. 27).
Nella Relazione della Commissione che illustra la Proposta di Regolamento, la delimitazione del concetto di rischio viene operata in connessione ai possibili pregiudizi per la sfera individuale e collettiva, specificando che «the same elements and tecniques that power the socio-economic benefits of AI can also bring about new risks or negative consequences for individuals or the society».
Analogamente cfr. Considerando 4, ove si ribadisce che «artificial intelligence may generate risks and cause harm to public interests and rights that are protected by Union law».
La dottrina censura il fatto che la nozione di rischio non sia direttamente desumibile dalla lettura della bozza di Regolamento, ove vi è soltanto il collegamento con "una scala di criticità identificata dalla Commissione sulla base delle tecnologie di AI e a seconda degli ambiti nei quali esse vengono utilizzate" (così O. Pollicino, G. De Gregorio, F. Paolucci, F. Bavetta, *Regolamento AI, la "terza via" lascia troppi nodi irrisolti: ecco quali*, 21 maggio 2021).

[56] Relazione illustrativa alla Proposta, p. 14.

[57] Per un'analisi approfondita e per applicazioni esemplificative dei vari sistemi di IA suddivisi in base al livello di rischio, cfr. D. Messina, *La proposta di regolamento europeo in materia di Intelligenza Artificiale: verso una "discutibile" tutela individuale di tipo consumer-centric nella società dominata dal "pensiero artificiale",* in *Rivista di diritto dei media*, 2/2022, pp. 196 e ss.

[58] Una deroga si lega, per esempio. alla possibilità di utilizzo di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, per individuare potenziali vittime specifiche di reato. In generale, sulla derogabilità del divieto di utilizzare sistemi «a rischio inaccettabile» e, in particolare, sui meccanismi di riconoscimento facciale, cfr. T. Madiega e H. Mildebrath, *Regulating facial recognition in the EU,* 2021, reperibile online.

[59] Cfr. S. Ranchordas, *Constitutional Sunsets and Experimental Legislation: A Comparative Perspective*, Edward Elgar, Cheltenham, 2014. L'Autrice sottolinea che "the rapid social and technological acceleration of society (…) does not interact well with slow-going legislators, resistance to legal change, obsolete laws, excessive administrative burdens and a perception of the rule of law as a law of permanent rules".

normativa a livello nazionale che vanifichi la finalità di armonizzazione perseguita dalla regolazione europea[60].

Il fatto che la Commissione europea abbia deciso di imporre un divieto assoluto per determinate tecnologie - una sorta di "*red line*" basata su concetti giuridici indeterminati[61] - rappresenta un'innovazione significativa nella regolazione del digitale, considerando che questo approccio diverge significativamente da quello del documento redatto dal Gruppo di Esperti, il quale si limita a segnalare le criticità insite in alcuni sistemi, senza addivenire ad un definitivo divieto[62].

Il Titolo III del Regolamento è invece dedicato ai richiamati sistemi di IA «ad alto rischio» (art. 6), che «hanno un impatto nocivo significativo sulla salute, la sicurezza e i diritti fondamentali delle persone nell'Unione»[63] e che rappresentano l'asse portante della disciplina[64]. Questi sistemi sono elencati nell'Allegato III e concernono otto settori tassativi, tra i quali la «identificazione e categorizzazione biometrica delle persone fisiche», la «gestione della migrazione, dell'asilo e del controllo delle frontiere», la «amministrazione della giustizia e processi democratici».

Una tale tecnica normativa – che rinvia ad un atto annesso per la definizione dell'ambito di applicazione oggettivo del divieto – è funzionale all'esigenza di consentire un rapido aggiornamento di uno schema regolatorio che rischia altrimenti di andare incontro ad una fisiologica obsolescenza[65]. Per questa ragione, alla Commissione è consentito aggiornare l'elenco mediante atti delegati non legislativi, senza però poter incidere su settori diversi da quelli espressamente indicati[66].

---

[60] Su questo aspetto v. A. Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, Springer, 2022, p.167, il quale sottolinea la difficoltà di elaborare una disciplina europea di armonizzazione in assenza di discipline nazionali: "The typical harmonisation goal of EU regulations – not applicable here in the absence of national laws on AI – is therefore replaced by a clear industrial strategy objective embodying a stronger and more centralised regulatory approach by the Commission which is reflected in the AIA Proposal".

[61] Lo sottolinea T. E. Frosini, *L'orizzonte giuridico dell'intelligenza artificiale*, in *Il diritto dell'informazione e dell'informatica*, in *BioLaw Journal*, 1/2022, p. 14, il quale rileva la presenza nel documento di "regole discrezionali, che sfruttano le potenzialità della IA quale potere coercitivo non solo e non tanto per reprimere reati e crimini, piuttosto per imporre la gestione della IA in capo allo Stato e ai suoi orfani di controllo. La discrezionalità del potere esercitato tramite IA potrebbe degenerare in arbitrio".

[62] Su questi aspetti cfr. il brillante commento di C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'Unione europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 3/2021, p.415 ss.

[63] Proposta di Regolamento, Considerando 27.

[64] I sistemi ad alto rischio vengono distinti in due categorie: i) quelli destinati ad essere utilizzati come componenti di sicurezza di prodotti soggetti a valutazione di conformità da parte di terzi; ii) sistemi specificamente elencati nell'Allegato III, che presentano implicazioni principalmente in relazione ai diritti fondamentali.

[65] Il rischio dell'obsolescenza che connota la dimensione regolatoria dell'infosfera è stato messo in luce anche dal documento della Commissione, *Tutela dei diritti fondamentali nell'era digitale – Relazione annuale 2021 sull'applicazione della Carta dei diritti fondamentali dell'Unione europea del 10 dicembre 2021*, COM(2021) 819 final, p. 19: "Uno specifico sottoinsieme di applicazioni di IA può subire un continuo adattamento, anche durante l'utilizzo, e cambiare ed evolvere in modo imprevisto senza poter essere facilmente monitorato. Ciò comporta un certo grado di imprevedibilità che può incidere sulla sicurezza o sui diritti fondamentali".

[66] Questo meccanismo di modifica, previsto dagli art. 7 e 73 del Regolamento, solleva qualche perplessità in relazione al potere rimesso alla Commissione. Invero, la rilevanza che i diritti fondamentali assumono nelle decisioni relative all'integrazione dell'elenco presuppone una delimitazione di questo potere di revisione, che dovrebbe essere limitato ad

I sistemi «ad alto rischio», come accennato, sono soggetti a regole uniformi volte ad evitare che arrechino pregiudizi agli interessi pubblici riconosciuti e tutelati dal diritto dell'Ue: si prevede così che l'immissione sul mercato dell'Unione e la messa in servizio avvenga solo previa verifica circa la sussistenza di determinati requisiti obbligatori e dopo aver svolto con esito positivo una valutazione di conformità[67]. Inoltre, la Proposta richiede che venga attuato un sistema di monitoraggio e dei meccanismi di vigilanza successivi all'immissione sul mercato, per garantire il rispetto degli oneri previsti dal regolamento (artt. 61 e 63). Queste previsioni risultano funzionali a limitare eventuali violazioni dei diritti umani, predeterminando "una sorta di «proceduralizzazione» del rischio che, mediante requisiti, certificazioni e controlli, dovrebbe essere ridotto entro un livello ritenuto accettabile"[68].

Da ultimo, i sistemi di IA possono essere presentare un «rischio limitato» (es. chatbot) o un «rischio minimo» (es. servizi di traduzione automatica, filtri antispam nelle mail): nel primo caso, il loro utilizzo è libero, salvi gli obblighi di trasparenza sanciti dall'art. 52, finalizzati a rendere l'utente consapevole dell'interazione con una macchina algoritmica. I sistemi di IA classificabili «a rischio minimo» rappresentano invece una categoria residuale, cui non si applica l'*AI Act* e le regole ivi contenute. Di conseguenza, possono essere utilizzati e sviluppati senza sottostare a particolari oneri, ferma restando l'opportunità dell'adozione volontaria di codici di condotta (cfr. art. 69).

Come anticipato, la garanzia dello *human oversight* è riferita unicamente ai sistemi ad alto rischio, per i quali deve essere assicurata un'interfaccia uomo-macchina che sia non solo adeguata ma anche costante, operante cioè per tutto il periodo in cui il sistema di IA è in uso (art. 14, par. 1). Un primo profilo problematico attiene alla concreta individuazione di questi strumenti di interfaccia, individuazione resa ancor più problematica dalla necessità di predisporre dei meccanismi idonei a monitorare la relazione utente-macchina durante l'intero ciclo di vita del sistema, dalla progettazione al funzionamento. In questo senso, il paragrafo 3 dell'art. 14 sembra offrire una risposta "a maglie larghe", chiarendo che la sorveglianza umana può concretizzarsi, alternativamente, in misure individuate dal fornitore prima della immissione sul mercato o messa in servizio, ove possibile, ovvero in misure individuate sempre *ex ante* dal fornitore ma adatte ad essere attuate dall'utente.

Si instaura così una sinergia - per ora più ideale che reale - tra fornitori e utenti, che rischia tuttavia di pregiudicare l'attuazione del principio *human in the loop*, considerando che non sempre l'utente è in

un accertamento di carattere tecnico. La valutazione rimessa alla Commissione presuppone, invece, una verifica di compatibilità dei sistemi ad alto rischio con i diritti fondamentali, e ciò probabilmente imporrebbe la partecipazione democratica del Parlamento. Su questo aspetto, cfr. A. Adinolfi, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno alla loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?, cit.,*Vol. I, p. 127 ss.

[67] Su questi profili v. L. Tosoni, *Intelligenza artificiale, i punti chiave del regolamento europeo*, 21 aprile 2021.

[68] A. Adinolfi, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno alla loro promozione: cit.,* p.157.

possesso di un bagaglio di conoscenze e competenze sufficiente ad interagire con la macchina. Da ultimo, il paragrafo 4 della disposizione in commento predispone una sorta di "manuale di istruzioni" per il soggetto che si relaziona con un sistema di IA ad alto rischio, elencando le azioni che l'utente deve poter compiere nell'ottica di un approccio antropocentrico[69].

In questa cornice regolatoria, la strategia normativa europea intende la garanzia del controllo umano come strumento precauzionale, volto a prevenire o ridurre al minimo i rischi per i diritti fondamentali; rischi che ben possono permanere anche dopo che il sistema di IA ha esaurito la procedura di verifica della conformità imposta dalla legge. In questo senso l'art.14, sebbene costituisca – come è stato notato[70] – il tentativo più avanzato, sul piano normativo, di contenere i limiti insiti nell'inarrestabile processo di autonomizzazione delle macchine, lascia permanere un alone di scetticismo circa le concrete modalità con cui l'interazione uomo-macchina possa essere effettivamente realizzata.

4.1. Uno sguardo d'insieme: quale spazio per lo *human in the loop* nell'*AI Act*?

Da una lettura complessiva della Proposta presentata dalla Commissione, sembra che il legislatore europeo abbia ben presente la necessità di assicurare la governabilità dei sistemi di IA. L'obiettivo di realizzare una *governance* efficace per un'intelligenza artificiale affidabile paga tuttavia un duplice scotto, legato a due ordini di considerazioni.

In primo luogo, la regolazione basata sul *risk approach* incide inevitabilmente sulla garanzia del controllo, poiché la riferisce unicamente ai sistemi classificati ad alto rischio.

Lo schema della gestione del rischio non rappresenta una novità nel panorama normativo europeo, anzi costituisce il modello prescelto nell'ambito della Strategia per il Mercato Unico Digitale in Europa. Sia il Regolamento sulla protezione dei dati personali[71] sia il *Digital Services Act*[72] condividono l'approccio basato

[69] Tra le azioni contemplate dall'art. 14, par. 4, compaiono, tra le altre, quelle finalizzate a comprendere appieno le capacità e i limiti del sistema di IA ad alto rischio ed essere in grado di monitorarne debitamente il funzionamento (lett. a), ovvero misure che consentano di non fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("distorsione dell'automazione" (lett. b) e, ancora, le azioni che mettano il soggetto nella condizione di decidere, in qualsiasi situazione particolare, di non usare il sistema di IA ad alto rischio o altrimenti di ignorare, annullare o ribaltare l'output del sistema di IA ad alto rischio (lett. d).

[70] D. Martire, *Intelligenza artificiale e Stato costituzionale, cit.,* p. 434.

[71] Sugli aspetti di comunanza e differenziazione tra GDPR e *Ai Act* cfr. G. Finocchiaro, L. Greco, *Il ruolo di titolare, responsabile e contitolare del trattamento nei trattamenti di dati personali mediante intelligenza artificiale*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?, cit.,*Vol. I, p. 313 ss. Anche O. Pollicino, G. De Gregorio, F. Paolucci, F. Bavetta, *Regolamento AI, la "terza via" europea lascia troppi nodi irrisolti: ecco quali, cit.*, rilevano che "a parte la scelta dalla Commissione di usare un regolamento come già fatto nel 2016 in materia di dati personali, la proposta adotta un sistema di *risk-based* statico che, a differenza del GDPR, non prevede una definizione flessibile di accountability rispetto al contesto in cui opera il titolare del trattamento" e che un tale atteggiamento "non sembra essere attuale e performante in una società dove il consenso costituisce sempre più una base poco affidabile".

[72] Il recente Regolamento sui Servizi Digitali, noto come *Digital Services Act* (DSA) ed approvato il 5 luglio 2022, fonda la sua disciplina sul rischio e prevede obblighi appositamente conformati in base ai destinatari della regolazione. Su questi aspetti v. G. De Gregorio, P. Dunn, O. Pollicino, *A partire dalla Strategia per il Mercato Unico Digitale, l'approccio basato sul*

sul rischio, quale strumento finalizzato ad incentivare una maggiore assunzione di responsabilità da parte dei soggetti pubblici e privati operanti sul mercato digitale. Ci si può ragionevolmente chiedere se la strada intrapresa dall'Unione fosse l'unica in astratto perseguibile ovvero se fossero immaginabili modelli di regolazione alternativi, non graduati in base al livello di rischio della tecnologia. Una parte della dottrina rileva che l'approccio basato sul rischio, seppure validamente perseguito in altri settori affini, sconta una sorta di "incompiutezza inevitabile" laddove sia riferito all'ambito dell'IA e propende piuttosto per un modello fondato sul più restrittivo principio di precauzione[73].

La scelta di graduare il meccanismo regolatorio sul livello di rischio atteso risponde presumibilmente all'obiettivo perseguito dall'Unione di agire da *first mover* nella regolazione dell'IA: questo fenomeno, noto come *Brussels effect*[74], produce effetti sia all'interno dei confini europei, sollecitando i Paesi membri ad allinearsi allo schema predisposto dall'Unione, sia all'esterno, influenzando le strategie regolatorie dei *competitors* internazionali[75].

Inoltre, sebbene altre iniziative normative adottate nell'ambito della Strategia Digitale europea condividano l'impostazione basata sul rischio, l'approccio fatto proprio dalla Proposta di Regolamento della Commissione presenta comunque alcune peculiarità. Infatti, al fine di sottrarre un significativo margine di discrezionalità in capo agli operatori, viene adottato un modello cd. *top-down*, in cui la qualificazione del livello di rischio del sistema è fatta dalla Commissione in sede legislativa, "sulla base di un automatismo imposto dall'alto"[76].

Diversa è invece la logica *bottom-up* che connota il GDPR e il DSA, in cui sono gli operatori del settore a definire il margine di rischio in concreto rilevante. Una predeterminazione del rischio rigidamente operata *ex ante* rischia di essere smentita dalla realtà dei fatti, ragion per cui potrebbe essere più opportuno

rischio è stato applicato in tutte le principali normative UE: analizziamo le analogie e le differenze tra GDPR, DSA e AI Act, 15 Settembre 2022.

[73] A. Oddenino, *Intelligenza artificiale e tutela dei diritti fondamentali: alcune notazioni critiche sulla recente proposta di regolamento della IA con particolare riferimento all'approccio basato sul rischio e al pericolo di discriminazione algoritmica*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?, cit.,*Vol. I, p.196. Sottolinea l'Autore che una tale incompiutezza, dovuta alla mobilità del sostrato tecnologico su cui la valutazione del rischio si innesta, suggerisce allora soluzioni differenti, fondate sul principio di precauzione, "tipicamente evocato in ambiti caratterizzati dalla impossibilità di una piena conoscibilità e conseguente gestione del rischio: tuttavia il richiamo è stato evitato, presumibilmente perché percepito come portatore di un impatto eccessivo sulle prospettive di mercato".

[74] L'espressione è stata coniata nel 2020 da una professoressa della Columbia Law School, Anu Bradford, *The Brussels Effect. How the European Union Rules The World*, Oxford University Press, 2020.

[75] Sulla strategia europea e sulle differenti iniziative adottate da Usa e Cina, cfr. A. Moreschini, *La proposta di Regolamento sull'intelligenza artificiale nel contesto globale*, in A. Lalli (a cura di), *L'amministrazione pubblica nell'era digitale*, Giappichelli, Torino, 2022, p. 145 ss. Sulla regolazione del rischio in prospettiva comparata cfr. M. Graziadei, *La regolazione del rischio e il principio di precauzione: Stati Uniti ed Europa a confronto*, in *Sistemi intelligenti*, 2017, p. 499 e ss.

[76] G. De Gregorio, P. Dunn, O. Pollicino, *A partire dalla Strategia per il Mercato Unico Digitale, l'approccio basato sul rischio è stato applicato in tutte le principali normative UE: analizziamo le analogie e le differenze tra GDPR, DSA e AI Act*, *cit.* Come già notato, la determinazione della Commissione è comunque "mobile", poiché idonea ad essere modificata tramite la revisione degli Allegati.

richiedere valutazioni in concreto, caso per caso. L'approccio prescelto dal legislatore europeo, meritevole di apprezzamento nell'ottica del principio di legalità, rischia tuttavia di innalzare le soglie di rischio e tradurle in una rigidità che non lascia margini di valutazione agli operatori su come adoperarsi nella pratica[77].

La seconda considerazione riguarda l'assetto complessivo che traspare dalla scelta regolatoria compiuta dall'Ue. Tralasciando in questa sede le due questioni preliminari relative all'opportunità di regolare un settore in continua evoluzione e alla scelta del livello di regolazione (se sovranazionale, nazionale o addirittura locale)[78], si può affermare con ragionevole convinzione che la strategia seguita dall'Ue si innesti su un impianto regolatorio tendenzialmente flessibile[79].

Quanto alla flessibilità, tale caratteristica sembrerebbe smentita dalla scelta dello strumento del Regolamento in luogo della Direttiva: la base giuridica dell'intervento normativo in commento si rinviene nell'art. 114 TFUE[80], che attribuisce alle istituzioni la competenza ad adottare misure di ravvicinamento delle discipline interne aventi incidenza sul funzionamento o sulla instaurazione del mercato comune[81].

---

[77] O. Pollicino, G. De Gregorio, F. Paolucci, F. Bavetta, *Regolamento AI, la "terza via" europea lascia troppi nodi irrisolti: ecco quali*, *cit.* In senso critico anche F. Donati, *Diritti fondamentali e algoritmi nella proposta di Regolamento sull'intelligenza artificiale*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?,* Vol. I, *cit.*, p. 119. L'Autore critica il fatto che il Regolamento si basi "su una identificazione del grado di rischio effettuata in astratto e in via preventiva, che potrebbe in certi casi rivelarsi inadeguata alla prova dei fatti". Il rischio sotteso all'adozione del modello *top-down* è tuttavia destinato ad essere ridimensionato grazie alla collaborazione tra attori pubblici e privati, che trova piena realizzazione nelle norme armonizzate (*harmonized standard*). Sul tema cfr. C. Marenghi, *La proposta di regolamento UE sull'intelligenza artificiale e la regolazione privata: spunti critici in tema di norme tecniche armonizzate*, in *Diritto comunitario e degli scambi internazionali*, 3-4/2021, pp. 563-583.

[78] Sulla opportunità di astenersi da tentativi regolatori per non intaccare il "disordine ordinato" della rete, v. T.E. Frosini, *Il costituzionalismo nella società tecnologica*, *cit.*, reperibile online. Un dibattito analogo, polarizzato sull'opportunità di adattare le normative vigenti al contesto dell'IA piuttosto che sulla creazione di un nuovo quadro regolatorio, ha preceduto anche l'intervento del legislatore europeo.

[79] Secondo G. De Gregorio, F. Paolucci, O. Pollicino, *L'intelligenza artificiale made in Ue è davvero "umano-centrica"? I conflitti della proposta*, 22 luglio 2021, il modello normativo europeo si colloca in una terza via tra il *laissez-faire* che ha contraddistinto la prima decade tecnologica degli anni duemila e quell'approccio di stampo autoritaristico, volto alla salvaguardia dei valori fondamentali alla base del costituzionalismo digitale. Anche A. Adinolfi, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali*, *cit.*, p. 130, mette in luce che "le peculiari caratteristiche dell'intelligenza artificiale richiedono, in ogni caso, che qualsiasi schema normativo generale presenti un sufficiente grado di flessibilità, giacché la continua evoluzione delle applicazioni tecnologiche – sia mediante soluzioni innovative sia a motivo di adattamenti nel corso del loro utilizzo – rischia di rendere rapidamente obsoleta una disciplina che detti regole puntuali".

[80] La norma attribuisce all'Unione il potere di adottare misure relative al ravvicinamento delle legislazioni degli Stati membri che hanno per oggetto l'instaurazione e il funzionamento del mercato interno. La Proposta di Regolamento si inserisce infatti nell'ambito della Strategia dell'Unione per il mercato unico digitale e mira a regolare lo sviluppo, l'immissione sul mercato e l'utilizzo dei sistemi di IA nell'Unione, in linea con quanto annunciato da Ursula von der Leyen, durante il discorso che ha seguito la sua elezione come primo presidente donna della Commissione europea (Discorso sullo stato dell'Unione 2020, "Un'Europa all'altezza del digitale" disponibile online). Tra le linee politiche della Presidente vi è quella di presentare nei primi cento giorni del suo mandato una proposta legislativa per un approccio europeo coordinato alle implicazioni etiche e umane dell'IA.

[81] Cfr. A. Adinolfi, *L'Unione europea dinanzi allo sviluppo dell'intelligenza artificiale: la costruzione di uno schema di regolamentazione europea tra mercato unico digitale e tutela dei diritti fondamentali*, in S. Dorigo (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pacini Giuridica, Pisa, 2020, p. 13 e ss. Critico rispetto all'utilizzo dell'art. 114 TFUE come fondamento giuridico della proposta è Van Cleynenbreugel, *EU By-Design Regulation in the Algorithmic Society*, in H.W. Micklitz, O. Pollicino, A. Reichman, A. Simoncini, G. Sartor e G. De Gregorio (a cura di), *Constitutional Challenges in the*

Infatti, come chiarisce l'*Explanatory memorandum* (par. 2.4), la scelta di un regolamento è strumentale all'esigenza di ridurre la frammentazione giuridica, evitando normative nazionali divergenti che ostacolino lo sviluppo di un mercato unico per sistemi di IA leciti, sicuri e affidabili[82].

Si potrebbe tuttavia dissentire sulla opportunità di una tale scelta normativa, volta a costruire una dimensione europea dell'IA conforme ai valori costituzionali e ai diritti fondamentali dell'Unione sposando, però, una logica di mercato. In altre parole, lo scopo di fondare la fiducia nell'IA sull'antropocentrismo come volano di protezione dei diritti umani sembra cozzare con l'obiettivo inconciliabile di accrescere la competitività del mercato europeo.

In verità, a dispetto della forma, la Proposta di Regolamento sull'IA non solo adopera volutamente concetti ampi, generici ed indeterminati – basti considerare la definizione indubbiamente ampia di Intelligenza Artificiale contenuta nell'art. 3 n.1[83] –, ma prevede altresì diversi meccanismi di riesame della disciplina: si pensi alla modificabilità degli Allegati – cui si è accennato *supra* - , al generale obbligo di revisione del regolamento con cadenza quinquennale (cfr. Relazione illustrativa, punto 5.1) e all'innovativo strumento della *regulatory sandbox*, una metodologia di verifica del prodotto in condizioni di sicurezza rafforzata[84]. Ne risulta un assetto tutt'altro che rigido, che traspare anche da ulteriori elementi[85]: in primo luogo, il fatto che ai sistemi «a rischio basso o minimo» non si applichi il Regolamento è una scelta coraggiosa, soprattutto considerando che questi rappresentano la maggior parte delle applicazioni

Algorithmic Society, secondo il quale l'art. 114 avrebbe un ambito limitato, poiché "essentially aims at harmonising Member States regulatory provisions rather than imposing specific design obligations on algorithmic designers".

[82] Questa strategia sta diventando sempre più nota ed è conosciuta come *acti-fication*. Sul tema, cfr. Papakonstantinou e P. De Hert, *EU lawmaking in the Artificial Intelligence Age: Actification*, *GDPR mimesis, and regulation*, in *European Law Blog*, 8 luglio 2021, disponibile online.

[83] A tenore della disposizione citata, per "sistema di intelligenza artificiale" (sistema di IA) si intende «un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono». Si tratta dunque di una nozione *mobile*, per circoscrivere la quale occorre far riferimento all'Allegato I, il quale richiama tre categorie di approcci: quello automatico (cd. *deep learning*), quello basato sulla logica *(logic based)* e sulla conoscenza (*knowledge based*) e, infine, gli approcci statistici, la stima baynesiana, i metodi di ricerca e l'ottimizzazione.

In senso critico rispetto all'ampiezza della definizione, cfr. M. U. Scherer, *Regulating Artificial Intelligence Systems: risks, challenges, competencies, and strategies*, in *Harward Journal of Law and Technology*, Vol. 29, Number 2, 2016, p. 354 ss. L'Autore sottolinea la profonda difficoltà di regolare ciò che non è chiaramente definito.

[84] Per questi aspetti cfr. C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'Unione europea in materia di intelligenza artificiale, cit.*

Sulle *regulatory sandboxes* v. anche L. Tosoni, *Intelligenza artificiale, i punti chiave del regolamento europeo, cit.,* e A. Merlino, *Regulatory Sandbox. Una nuova prospettiva ordinamentale*, Napoli, ESI, 2022. Stando allo studio elaborato dal Parlamento europeo e pubblicato online a settembre 2020, *Regulatory Sandboxes and Innovation Hubs for FinTech: impact on innovation, financial stability and supervisory convergence*, «regulatory sandboxes can be seen as a way of regulatory experimentation, which allows the supervisor to test a certain customized regulatory approach to an innovative service, product or business model, instead of regulating, potentially prematurely or inadequately».

[85] È la stessa Commissione che, al Considerando 63 della Proposta, pur ritenendo espressamente «necessaria una risposta normativa specifica», sottolinea l'intenzione di intervenire solo laddove strettamente necessario, nell'ottica della proporzionalità e rifuggendo da qualsiasi eccesso di regolamentazione.

di IA diffuse nella pratica (es. sistemi di raccomandazione). Una scelta di questo tipo sembra rispondere all'esigenza di conciliare una regolazione sufficiente con la necessità di evitare oneri eccessivamente gravosi in capo agli attori del processo tecnologico[86].

In secondo luogo, anche l'ambito territoriale di applicazione della normativa viene delimitato utilizzando come parametro il criterio della localizzazione del destinatario dell'offerta produttiva: il Regolamento si applica a condizione che il sistema di IA venga messo in servizio nell'Unione, a prescindere dal luogo in cui il fornitore è stabilito (quindi anche se ha la sua sede all'estero[87]).

Queste due considerazioni finali rappresentano un'arma a doppio taglio. Sia la scelta di lasciare fuori dall'ambito di applicazione del Regolamento i sistemi a rischio minimo, sia quella di delimitarne in senso ampio i confini territoriali potrebbero invero agevolare i produttori nel sottrarsi all'applicazione della normativa vigente, mediante semplici accortezze tecniche, così vanificando la prospettiva di un effettivo diritto allo "*human in the loop*".

In realtà, la garanzia di un approccio antropocentrico rimane centrale nelle intenzioni del legislatore europeo, come si desume analizzando il piano della *governance* a livello nazionale[88]: il Titolo VI della Proposta riconosce alle autorità nazionali di vigilanza del mercato il potere di monitorare il rispetto dei requisiti essenziali per tutti i sistemi di IA ad alto rischio immessi sul mercato (e solo per questi!).

A tal fine, l'art. 64 consente alle anzidette autorità «pieno accesso ai set di dati di addestramento, convalida e prova utilizzati dal fornitore». Diversamente, l'accesso al codice sorgente è garantito solo previa richiesta motivata e ove necessario per valutare la conformità del sistema di IA ad alto rischio ai requisiti prescritti dal Regolamento. Accanto ai poteri di vigilanza, le suddette autorità possono altresì intervenire con misure appropriate laddove i sistemi non fossero conformi alla disciplina o, seppure conformi, presentino

---

[86] Questo aspetto è evidenziato da O. Pollicino, G. De Gregorio, F. Paolucci, F. Bavetta, *Regolamento AI, la "terza via" europea lascia troppi nodi irrisolti: ecco quali*, *cit*., i quali rilevano che "tale approccio comporta la totale esclusione d'intere categorie di sistemi di AI – quelle considerate a basso rischio – da qualsiasi tipologia di controllo, comportando un vacuum normativo non indifferente. Infatti, laddove si riconosca l'esistenza di possibili rischi per i diritti fondamentali e data l'imprevedibilità del funzionamento di tali tecnologie, una simile esclusione potrebbe risultare quantomeno discutibile sul piano della protezione dei diritti dell'individuo, soprattutto in quanto non vi è alcuna analisi ex ante del singolo caso concreto".

[87] Sull'ambito di applicazione oggettivo cfr. in generale l'art. 2.1 della Proposta, a tenore del quale «il presente regolamento si applica: a) ai fornitori che immettono sul mercato o mettono in servizio sistemi di IA nell'Unione, indipendentemente dal fatto che siano stabiliti nell'Unione o in un paese terzo; b) agli utenti dei sistemi di IA situati nell'Unione; c) ai fornitori e agli utenti di sistemi di IA situati in un paese terzo, laddove l'output prodotto dal sistema sia utilizzato nell'Unione.

Parla espressamente di «*aterritoriality*» L. Floridi, *The European Legislation on AI: a brief Analysis of its Philosophical Approach, cit.,* p.220.

[88] Il Regolamento istituisce un sistema di governance multilivello (titolo VI): accanto al ruolo rimesso alle autorità nazionali dall'art. 59 - secondo il modello dell'amministrazione comunitaria indiretta -, viene infatti prevista (art. 56) la creazione di un apposito Comitato europeo (*European Artificial Intelligence Board*, EAIB), con il compito di sorvegliare la corretta applicazione del Regolamento nei vari Stati membri e di elaborare linee-guida in materia. Sulla *governance* dell'IA v. A. Amidei, *La governance dell'intelligenza artificiale: profili e prospettive di diritto dell'Unione europea*, in U. Ruffolo (a cura di*), Intelligenza artificiale. Il diritto, i diritti, l'etica, cit*., pp. 571-588.

un rischio per la salute o la sicurezza delle persone, per i diritti fondamentali o per altri aspetti della tutela dell'interesse pubblico.

Il ruolo delle Autorità specializzate quale garanti del rispetto del principio dell'"*under-user control*" rappresenta, nell'ottica del legislatore europeo, un chiaro indice della preminenza della visione antropocentrica su quella mercantilista. Tuttavia, l'ampia discrezionalità rimessa agli Stati circa l'individuazione del soggetto concretamente deputato a tutelare i diritti fondamentali a livello nazionale può seriamente mettere a repentaglio gli obiettivi della *governance* europea, facendo paventare un duplice rischio: da un lato, quello di alimentare "sacche di impunità" e, dall'altro, di dar luogo a pericolose duplicazioni procedimentali e sanzionatorie, in violazione del principio di matrice convenzionale del *ne bis in idem*[89].

5. Alcune riflessioni conclusive: aspettando Godot?

Lo scenario così delineato consente di tracciare alcune, per quanto sommarie, riflessioni conclusive. La dialettica tra la pervasiva diffusione dell'Intelligenza Artificiale ed i connaturati limiti etici e giuridici influenza in modo inevitabile la strategia normativa europea ed alimenta la tensione tra tecnocentrismo e antropocentrismo che da sempre caratterizza lo studio di queste tematiche.

La Proposta di Regolamento varata dall'Unione europea persegue due obiettivi innegabilmente ambiziosi, quello di tutelare i diritti fondamentali dell'individuo e, al contempo, quello di difendere i valori fondanti dell'Unione europea in una logica di mercato[90]. Queste due componenti rappresentano i *twin objectives* della strategia digitale globalmente intesa[91] e, almeno nell'intento del legislatore europeo, dovrebbero trovare un punto di incontro nell'approccio antropocentrico all'intelligenza artificiale.

Il modello regolatorio prescelto, graduato secondo il livello di rischio, mira espressamente a ridurre i pericoli per la sfera individuale e per la collettività, paralizzando i potenziali effetti negativi dell'Intelligenza artificiale su diritti umani, democrazia e Stato di diritto. Ciononostante, a dispetto delle intenzioni della Proposta di essere al contempo "efficace rispetto ai diritti fondamentali ed efficiente rispetto al mercato"[92], la sensazione che traspare è che difficilmente si possa riuscire a realizzare un equilibrio tra due anime tanto diverse.

[89] Su questo aspetto cfr. le riflessioni contenute nell'articolo di E. Raffiotta, *Quale autorità governerà l'Intelligenza Artificiale?*, pubblicato su *Il Sole 24Ore* il 27 marzo 2023 e reperibile qui.
[90] Un esplicito richiamo agli *European values* figura, ad esempio, nei Considerando 1 e 15 della Proposta.
[91] Nella Relazione Illustrativa all'*AI Act*, la Commissione collega esplicitamente il rischio tanto alla sfera individuale quanto a quella collettiva, rilevando che «the same elements and techniques that power the socio-economic benefits of AI can also bring about new risks or negative consequences for individuals or the society».
[92] A. Oddenino, *Intelligenza artificiale e tutela dei diritti fondamentali, cit.*, p. 196.

La strategia normativa europea è degna di plauso nella parte in cui, superando il modello statunitense e quello cinese, cerca di realizzare una sintesi equilibrata tra la tutela dei diritti fondamentali e l'efficienza del mercato comune. Tuttavia, la scelta di livellare diritti e obblighi in base al parametro del rischio risponde primariamente all'esigenza di uniformare e salvaguardare il mercato unico digitale e riesce decisamente meno a tutelare i singoli utenti che operano in quel mercato[93].

Il *Risk Management system* previsto dall'art. 9 del Regolamento delinea un meccanismo di sorveglianza volto a realizzare un controllo costante e sistematico sulle diverse applicazioni di IA, di cui apparentemente sono garanti tutti i soggetti coinvolti nella "catena", dal fornitore al distributore, dall'importatore all'utente. Proprio il ruolo dell'utente risulta, tuttavia, di difficile comprensione. Egli è il destinatario finale del sistema di IA e, sempre più spesso, risulta coinvolto in complesse dinamiche algoritmiche di cui difficilmente riesce a comprendere il funzionamento.

Diversamente dal consumatore che si trova a maneggiare un prodotto difettoso (artt. 114-127 del D.lgs. 206/2005, cd. codice del consumo), il fruitore di applicazioni di IA non solo non riesce a percepire immediatamente il "difetto" del sistema con cui entra in contatto (ed anzi non è neanche detto che se ne avveda), ma nemmeno dispone di forme di tutela paragonabili a quelle che il codice del consumo prevede per i prodotti difettosi [94].

In definitiva, la prefissata svolta antropocentrica non sembra trovare un effettivo recepimento nel testo della Proposta, da cui risulta obiettivamente non facile individuare chi è l'umano dietro la "*Humancentric*

[93] Evidenziano questa criticità G. De Gregorio, F. Paolucci, O. Pollicino, *L'intelligenza artificiale made in Ue è davvero "umano-centrica"? I conflitti della proposta*, 22 luglio 2021, e A. Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI, cit.*, p. 189. Quest'ultimo rileva come "the AIA Proposal marginalises the role of the AI users. They play no part in the risk management process and have no obligations in this regard, even though AI providers market solutions that are customisable by users. AI users may independently increase or alter the risks of harm to health and safety by their particular use of the systems, especially in terms of impact on individual and collective rights, given their variety and context dependence". Anche L. Floridi, *The European Legislation on AI: a brief Analysis of its Philosophical Approach*, *Philos. Technol.* 34/2021, p. 219, evidenzia che "this risk-based approach seems convincing (it is a common approach for internal market-based legislation) and aligned with the view that ethics benefits the market, not vice versa. But precisely for this reason, one may argue that the AIA could do much more to protect consumers' rights and be much more incisive about providing measures to redress the possible harms or losses that AI systems may cause. This is the part where one may expect and welcome more improvements in the proposal".

[94] La disciplina oggi contenuta nel Codice del consumo è frutto del recepimento della Direttiva 85/374/CEE in materia di "responsabilità sul prodotto". Gli artt. 114 e ss. del D.lgs. 206/2005 delineano un sistema di responsabilità presunta del produttore di prodotti difettosi, che prescinde dall'accertamento della colpevolezza e presuppone unicamente la prova del difetto e del nesso eziologico tra il difetto e il danno arrecato al consumatore.

AI"[95] e di quali strumenti egli dispone, attesa anche la mancanza di meccanismi rimediali in caso di danno direttamente arrecato all'utente[96].

L'individuazione di un adeguato regime di responsabilità in caso di danni cagionati dai sistemi artificiali potrebbe in tal senso rappresentare il primo tassello nell'elaborazione di una *governance* algoritmica globale dell'IA[97]. L'intervento regolatorio europeo sconta inevitabilmente il prezzo di una radicata disomogeneità e di una naturale frammentazione, dovuta ad una pluralità di fattori: da un lato, vi è la difficoltà di definire l'oggetto stesso della normazione - sembra essere proprio questa la ragione per cui l'iter normativo sta incontrando numerosi differimenti, alla luce della diffusione di sistemi sempre più complessi, come l'IA generativa di ChatGPT -; dall'altro, la pluralità di attori che operano nell'ecosistema digitale comporta una naturale stratificazione di discipline settoriali ed eterogenee, spesso tra loro incompatibili[98].

Tutti questi elementi non sono comunque idonei a smentire la validità di un approccio regolatorio trasversale e multilivello, basato sulla considerazione per cui "*technology that is not human-centered will not be a solution*"[99].

---

[95] Secondo L. Floridi, *op.ult.cit.,* p.219 "unfortunately, the AIA uses an anachronistic terminology to define this approach as «human-centric», that is, as an approach that places humanity at the centre of technological development. Yet this is both trivially true and dangerously ambiguous. On the one hand, it is obvious that any technology, AI included, must be at the service of humanity, its values, and needs. On the other hand, one must also consider the environment as crucially important, yet «humancentric» seems to be synonymous with «anthropocentric», and we know how much the planet has suffered from humanity's obsession with its importance and centrality, as if everything must always be at its service, including every aspect of the natural world, no matter at what costs and losses".

[96] Evidenziano la mancanza di meccanismi cd. di *redress* G. Di Gregorio, F. Paolucci, O. Pollicino, *L'intelligenza artificiale made in Ue è davvero "umano-centrica"? I conflitti della proposta*, 22 luglio 2021. Secondo gli Autori, tale vuoto normativo "riflette una generale mancanza di messa a fuoco sull'individuo, il quale, dalle dichiarazioni di indirizzo, avrebbe dovuto essere il principale perno di tutto l'assetto normativo. In mezzo a un certo affastellamento di articoli riferiti a obblighi, limiti e contro limiti per le imprese è davvero difficile scorgere l'approccio che, per lo meno nelle intenzioni, aspirava a porre la persona umana al centro della proposta. La scelta di fondo sembra essere quella di voler sposare il sistema del rischio per frenare l'espansione incontrollata del settore, senza però sforzarsi di rendere concreta la spinta costituzionalistica che viene promessa negli intenti della proposta". Analogamente anche D. Messina, *IA umanocentrica: cosa manca al nuovo regolamento europeo*, 14 gennaio 2022.

[97] Sui profili legati alla responsabilità, cfr. la Proposta di Direttiva sulla responsabilità per i danni da IA. Il testo della Proposta, adottata dalla Commissione europea a settembre 2022, e l'iter legislativo possono essere consultati sul sito web della Commissione. La tematica della responsabilità dei sistemi di IA è al centro dell'attenzione del legislatore europeo da tempo: si veda la Risoluzione del Parlamento europeo recante «Raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica» del 16 febbraio 2017 e la successiva Risoluzione «Raccomandazioni alla Commissione su un regime di responsabilità civile per l'Intelligenza Artificiale» del 20 ottobre 2020.

[98] Questo aspetto è ben evidenziato da W. Wallach e G. Marchant, *Toward the Agile and Comprehensive International Governance of AI and Robotics*, *in Proceedings of the IEEE*, Vo. 107, no. 3, 2019, nel passo seguente: "Rapidly emerging technologies, such as AI and robotics, present a serious challenge to traditional models of government regulation. These technologies are advancing so quickly that in many sectors, traditional regulation cannot keep up, given the cumbersome procedural and bureaucratic procedures and safeguards that modern legislative and rulemaking processes require. Consequently, regulatory systems will predictively fail to put in place appropriately tailored regulatory measures by the time new applications of fast-moving technologies begin to affect society. Perhaps even worse, if a regulatory system does somehow manage to rush into place new regulations for an emerging technology, they will likely be obsolete by the time the ink dries on the enactment. Given this socalled "pacing problem," traditional regulatory approaches will either produce no regulation or bad regulation".

[99] S. Dangel, M. Hagan, J.B. Williams, *Designing Today's Legal Education for Tomorrow's Lawyers: The Role of Legal Design, Technology and Innovation*, September 2018, disponibile qui.