INTELLIGENZA ARTIFICIALE E DIRITTO: UNA RIVOLUZIONE?

A CURA DI
ALESSANDRO PAJNO, FILIPPO DONATI E ANTONIO PERRUCCI

VOLUME I
DIRITTI FONDAMENTALI, DATI PERSONALI E REGOLAZIONE

SOCIETÀ EDITRICE IL MULINO

Alla pubblicazione di questa ricerca ha contribuito il Gruppo AlmavivA, che Astrid vivamente ringrazia

ISBN 978-88-15-29967-3

Redazione e produzione: Edimill srl - www.edimill.it

Capitolo quinto

INTELLIGENZA ARTIFICIALE E TUTELA DEI DIRITTI FONDAMENTALI: ALCUNE NOTAZIONI CRITICHE SULLA RECENTE PROPOSTA DI REGOLAMENTO DELLA UE, CON PARTICOLARE RIFERIMENTO ALL'APPROCCIO BASATO SUL RISCHIO E AL PERICOLO DI DISCRIMINAZIONE ALGORITMICA

1. *Cenni introduttivi sulla sfida regolativa dell'IA in relazione ai diritti fondamentali*

La sfida per una regolazione efficace dell'intelligenza artificiale in relazione alla tutela dei diritti fondamentali è uno dei banchi di prova più significativi su cui si misura la tenuta del paradigma giuridico contemporaneo. Interrogarsi sul ruolo del diritto nel governo dei sistemi di intelligenza artificiale (d'ora in avanti IA) pare dunque imprescindibile, e tale necessità consegue alla presa d'atto che la rivoluzione digitale trasforma non soltanto aspetti concreti della nostra quotidianità, ma altresì le nostre opinioni, i nostri valori, le nostre priorità e dunque, in ultima analisi, il nostro «essere umani». Il quesito su come regolare questo tipo di innovazione tecnologica significa, in altre parole, determinare l'orizzonte normativo dello sviluppo umano e orientare il tipo di innovazione che si ritiene sostenibile e socialmente preferibile[1].

La crescente diffusione e ubiquità dell'uso di algoritmi fa sì che l'intelligenza artificiale si proietti già oggi in una dimensione cibernetica immateriale che è capillarmente diffu-

Questo capitolo è di Alberto Oddenino.

[1] In tema sia consentito rinviare a A. Oddenino, *Sviluppo sostenibile, sviluppo umano e nuove tecnologie digitali oltre l'Agenda 2030 delle NU: alla ricerca di una visione integrata*, in Società italiana per l'organizzazione internazionale (SIOI), *Le Nazioni Unite di fronte alle nuove sfide economico-sociali 75 anni dopo la loro fondazione*, Quaderno n. 23 di «La Comunità internazionale», Napoli, 2021, pp. 103 ss. e bibliografia ivi citata.

sa[2]. Nel mondo dei cd. oggetti intelligenti (IOT), gli algoritmi sono divenuti una mediazione continua e sostanzialmente necessaria di ogni attività umana, non solo informativo-cognitiva, ma anche di interazione con l'ambiente circostante[3]. Per questo la diffusione dei processi decisionali algoritmici ha destato una crescente attenzione, che va di pari passo con quella per gli oggetti che incorporano forme, più o meno evolute, di intelligenza artificiale[4]. L'ampliamento progressivo del loro utilizzo e del loro campo di azione ne evidenzia sempre più l'attitudine a incidere direttamente sulle sfere individuali e tale circostanza pone in diretto collegamento il tema della *governance* algoritmica con quello della tutela dei diritti fondamentali[5].

Non si vuole con ciò disconoscere l'indubbio potenziale della IA nel favorire svariati aspetti di rilievo economico e sociale, da cui discende lo stesso più efficace esercizio dei diritti umani[6]. Ciò che interessa invece approfondire sono i legami fra la proliferazione di applicazioni di IA e l'aumento di un pericolo concreto di registrare impatti negativi

[2] In tema cfr. B. Bodo, N. Helberger, K. Irion, F. Zuiderveen Borgesius, J. Moller, B. van Es e C. de Vreese (a cura di), *Tackling the Algorithmic Control Crisis - The Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents*, in «Yale Journal of Law and Technology», 2018, in part. p. 171, ove si sottolinea la tendenza alla ubiquità degli agenti algoritmici, in un mondo immateriale e globalmente interconnesso.

[3] Cfr. M. Durante, *Potere computazionale. L'impatto delle ICT su diritto, società, sapere*, Milano, 2019, pp. 17 ss.

[4] *Ex multis* cfr. M. Ziewitz, *Governing algorithms: Myth, mess, and methods*, in «Science, Technology, & Human Values», 2016, pp. 3 ss. e T. Zarsky, *The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making*, in «Science, Technology, & Human Values», 2015, pp. 118 ss.

[5] Per questo il tema relativo a quella che viene talora definita come vera e propria «algocrazia» (così J. Danaher, *The threat of algocracy: Reality, resistance and accommodation*, in «Philosophy and Technology», 2016, pp. 245 ss.), deve necessariamente essere affrontato in prospettiva multidisciplinare, con un collegamento fra profili giuridici, tecnologici ed economici, ma anche etico-filosofici.

[6] Si tratta di una prospettiva, particolarmente evidente in relazione ai diritti economici e sociali, dalla quale necessariamente si prescinde nella presente indagine.

sui diritti umani. È questa una circostanza da cui discende una pressione regolativa crescente sulla progettazione ed il governo dell'IA per assicurarne l'equità, la trasparenza e la responsabilità[7].

Peraltro, le configurazioni giuridiche atte a fare fronte a questa esigenza sono ancora, particolarmente nella prospettiva del diritto internazionale, ad uno stadio embrionale, che vede solo una certa proliferazione di strumenti programmatici e di *soft law* che tendono a fondere la prospettiva giuridica con quella dell'etica delle macchine[8].

A fronte di ciò prospera altresì una *regulatory competition* fra le maggiori potenze tecnologiche, che si colora di tinte indiscutibilmente geopolitiche, stante l'assoluta strategicità, non solo economica, del tema. D'altronde è proprio il valore economico dello sviluppo dell'intelligenza artificiale a sconsigliare approcci manichei. L'IA è infatti lo strumento con cui si può estrarre valore economico dalla crescente mole di dati resa disponibile dall'uso sempre più diffuso e costante della Rete e dalla sua integrazione con gli oggetti intelligenti.

È pertanto evidente come una regolazione troppo rigorosa dell'intelligenza artificiale in relazione alla tutela dei diritti umani rischi di determinare un sostanziale stallo nello sviluppo di prodotti basati sulla IA stessa, causando uno svantaggio competitivo potenzialmente assai oneroso. Al contrario, una regolazione oculata può avere una influenza benefica nel rinsaldare il clima di fiducia nell'intelligenza artificiale, come presupposto per sviluppare il mercato della stessa e dei prodotti che la incorporano.

[7] Cfr. C. Cath, *Governing Artificial Intelligence: ethical, legal and technical opportunities and challenges*, in «Philosophical Transactions of the Royal Society A», 2018, pp. 376 ss.

[8] Si può osservare come sia proprio la difficoltà nel trovare principi internazionalmente condivisi per un approccio normativo ad aver determinato una certa proliferazione di strumenti etici, con conseguente sovrabbondanza di spunti e di principi in un panorama non sempre agevole da ricondurre ad unità. In tema cfr. L. Floridi, *A Unified Framework of Five Principles for AI in Society*, in «Harvard Data Science Law Review», 2019, pp. 1 ss.

In questo senso, il tema della regolazione dell'intelligenza artificiale appare cruciale non solo nella prospettiva di tutela dei valori costituenti della nostra tradizione giuridica, ma anche, al contempo, come volano di sviluppo delle nostre società, capace di creare i presupposti per il godimento di diritti fondamentali di contenuto economico che risultano intimamente legati al progresso del mercato[9].

In un simile quadro generale si situa l'attuale spinta regolativa della UE, che sembra poter trarre vantaggio dalla mancanza di modelli di *governance* regolamentare a respiro globale che si pongano in competizione con essa, dal momento che, in particolare da parte statunitense e cinese, l'attitudine appare ad oggi quella di astenersi dall'adozione di norme generali e stringenti[10]. Chiaramente, il massimo vantaggio derivante all'UE dall'essere sostanzialmente *first mover* in questo campo si concretizzerà solo a patto che l'intervento normativo sia al contempo efficace (per la tutela dei diritti) ed efficiente (per lo sviluppo del mercato). Ciò che appare chiaro è altresì che la regolazione UE avrà naturale propensione ad avere effetti che esorbitano i confini territoriali della stessa[11]: circostanza inevitabile sol

[9] In proposito si manifesta un profilo paradossale che indica la difficoltà di trovare un giusto bilanciamento regolatorio: è il rischio di comprimere il ricorso alla Intelligenza artificiale in nome dei diritti fondamentali per poi dover registrare un indebolimento dei medesimi proprio a causa dello sviluppo non adeguato della IA.

[10] Su un piano ancor più generale, sul quale non si può indugiare in questa sede, si situa il dibattito relativo alla opportunità o meno di una centralizzazione della *governance* dell'intelligenza artificiale, a fronte dell'attuale panorama contraddistinto da grande frammentazione. Si tratta di un tema delicatissimo, che segna una insanabile contrapposizione fra gli Stati, ansiosi di affermare la propria primazia sovrana su un tema tanto strategico, tanto da rendere del tutto improbabile che si trovi un consenso per costruire una *governance* condivisa. Sul tema, che a ben vedere riecheggia quello irrisolto di una *governance* autenticamente internazionale della rete, cfr. P. Cihon, M.M. Maas e L. Kemp, *Fragmentation and the Future: Investigating Architectures for International AI Governance*, in «Global Policy», 2020, pp. 545 ss.

[11] Non sfugge in merito la portata non solo economica ma anche geopolitica di un'azione legislativa comune dell'UE. Essa ha un enorme potenziale per fornire all'industria europea vantaggi competitivi e

che si consideri che l'intelligenza artificiale si innesta su un sostrato tecnologico, quello della Rete, che è per sua natura meta-nazionale e che essa si alimenta di una materia prima, quella dei dati nella loro dimensione tanto discreta quanto aggregata (*big data*), che è essa stessa potentemente proiettata in chiave transnazionale e globale[12].

Merita ricordare brevemente un punto ulteriore legato al quadro internazionale in cui tale spinta si inserisce. Di questo quadro, assai composito, non è possibile dare compiutamente conto in questa sede, ma giova sottolineare come esso risulti principalmente orientato a fornire linee guida etiche per orientare l'IA verso valori «buoni» per l'umano. Questo perché, come è stato ben argomentato, l'IA non si presenta come aprioristicamente buona ma piuttosto come neutra rispetto ai diritti fondamentali[13], e questo richiede che essa sia orientata al fine di servirli[14].

con essi imprimere uno straordinario impulso al mercato interno; al contempo l'adozione di standard dell'UE suscettibili di proiettarsi a livello globale garantirebbe che lo sviluppo e la diffusione dell'IA siano imperniati su valori, principi e diritti tutelati nell'UE e espressione di una «specificità» UE. In tema 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*, Pisa, 2020, pp. 13 ss.

[12] Da questo punto di vista la sfida regolativa dell'IA può essere vista come complementare a quella che riguarda la *Data Protection*, nella quale l'UE ha già caratterizzato la propria azione in chiave di regolatore globale attraverso l'adozione del GDPR, e in relazione alla quale si pongono delicati problemi di tutela dei diritti fondamentali (sul punto si rinvia all'ampio studio di G. Della Morte, *Big Data e protezione internazionale dei diritti umani*, Napoli, 2018).

[13] Cfr. AA.VV., *The role of artificial intelligence in achieving the Sustainable Development Goals*, in «Nature Communications», 2020. Ciò che emerge è l'opportunità di caratterizzare meglio le forme di IA in modo da renderla servente rispetto ai SDGs, dal momento che un ricorso indiscriminato alla medesima rischia al contrario di acuire discriminazioni e diseguaglianze.

[14] Su questa precisa linea argomentativa si segnalano varie iniziative per una IA al servizio dei *Sustainable Development Goals* dell'Agenda 2030 delle NU: cfr. in particolare, come efficace esempio, l'elaborazione in seno al progetto AI4SG del Digital Ethics Lab dell'Oxford Internet

Altro punto qualificante del dibattito internazionale meritevole di menzione è quello relativo al tipo di strumento da impiegare, e in particolare alla opportunità di affiancare soluzioni di natura schiettamente normativa alla adozione di strumenti non vincolanti, come *guidelines*, codici etici e raccomandazioni rivolte a progettisti e sviluppatori di IA. Senza entrare nell'ampio dibattito sul ruolo e le potenzialità della *soft law*, particolarmente in un ambito che è contrassegnato da una certa proliferazione di iniziative[15], è evidente che essa presenti una apprezzabile vantaggiosità in termini di costi di negoziazione, e questo ne ha determinato il tradizionale successo anche in dimensione europea[16].

Le opzioni normative basate su strumenti di *hard law* appaiono certo più incisive dal punto di vista delle garanzie di tutela, ma occorre ricordare che sono le stesse peculiari caratteristiche dell'intelligenza artificiale a richiedere, 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[17]. Sono queste tutte

Institute, su cui L. Floridi, J. Cowls, T.C. King e M. Taddeo, *How to design AI for Social Good: Seven Essential Factors*, in «Science and Engineering Ethics», 2020, pp. 1771 ss.

[15] In tema di ricerca di unità in un panorama davvero assai vasto cfr. L. Floridi, *A Unified Framework of Five Principles for AI in Society*, in «Harvard Data Science Law Review», 2019, p. 1, ove si analizza la declinazione dei cinque principi di *Beneficence*, *Nonmaleficence*, *Autonomy*, *Justice* e *Explicability* all'interno di molteplici strumenti di codificazione dei principi.

[16] Cfr. *ex multis* la raccomandazione del Consiglio dell'OECD sull'intelligenza artificiale OECD (*Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 2019) e le *Draft Ethics Guidelines for Trustworthy AI* del 18/12/2018, elaborate dal Gruppo di esperti della Commissione europea sull'intelligenza artificiale.

[17] Come ben espresso nel recente documento della Commissione *Tutela dei diritti fondamentali nell'era digitale - Relazione annuale 2021 sull'applicazione della Carta dei diritti fondamentali dell'Unione europea*, del 10/12/2021, COM(2021) 819 final, p. 19, uno «specifico sottoinsieme di applicazioni di IA può subire un continuo adattamento, anche durante

considerazioni assai pertinenti per l'esame della Proposta di Regolamento della UE in tema di IA che sarà svolta nei prossimi paragrafi.

Ciò che preliminarmente si può trarre è che, a livello internazionale, il quadro di *governance* sia solo embrionale e assai poco lineare, con interazione fra tre plessi di regole di natura diversa, e in particolare tecniche, etiche e giuridiche. Si tratta peraltro di un quadro la cui complessità è in certo senso inevitabile. La risultante di queste tre forze determina a sua volta la complessità della *governance* della intelligenza artificiale[18], la cui «capacità di tenuta» sembra poter discendere proprio dal buon bilanciamento reciproco fra i menzionati plessi di regole, e nello specifico a quanto peso specifico dovrebbero avere le regole giuridiche in un contesto che, pervaso dalla forza per così dire intrinseca degli standard tecnologici, appare tradizionalmente improntato a valorizzare prevalentemente i precetti di *digital ethics*[19].

2. *L'emersione di una specificità della regolazione UE nel senso di un approccio antropocentrico e basato sul rischio*

A fronte di questo quadro di *governance* internazionale imperfetto e fluido, si staglia con chiarezza la scelta del legislatore UE a favore di una regolazione giuridica ambiziosa, portatrice di un metodo specifico: metodo che, se è orientato a fare pienamente propria l'indicazione, già consolidatasi in alcuni strumenti internazionali, per una

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».

[18] Il punto è assai ben evidenziato in L. Floridi, *Soft Ethics and the Governance of the Digital*, in «Philosophy and Technology», 2018, pp. 4 ss., ove si indaga precisamente l'interazione fra *Digital Ethics*, *Digital Governance* e *Digital Regulation* rispetto all'obiettivo di dare forma attuale e futura all'intelligenza artificiale.

[19] In questa direzione cfr. G. Resta, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in «Politica del diritto», 2019, pp. 199 ss., in part. p. 220.

valorizzazione dell'umano, affianca a tale connotazione antropocentrica una chiara preferenza per un approccio basato sul rischio.

Prima di addentrarci nell'analisi di tale approccio, per evidenziarne caratteristiche e limiti, appare opportuno ricostruirne rapidamente l'enucleazione per tappe successive.

Un primo riferimento alla nozione di rischio in relazione all'intelligenza artificiale si ha nella Risoluzione del Parlamento europeo del 16/2/2017, recante *Raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica*[20]. In essa il Parlamento UE riscontra come dall'impiego della robotica, intesa come suscettibile di incorporare varie forme di intelligenza artificiale, derivino un insieme di tensioni o rischi da valutare seriamente anche da un punto di vista della sicurezza, della salute, della libertà, della vita privata, dell'integrità, della dignità, dell'autodeterminazione e della non discriminazione, oltreché della protezione dei dati personali.

In parallelo, il Comitato economico e sociale europeo (CESE) ha adottato il 31/5/2017 un parere sull'intelligenza artificiale[21], con il quale è stata raccomandata l'adozione di un approccio antropocentrico, teso ad assicurare all'uomo il controllo della macchina. A tale fine è stata auspicata tanto l'introduzione di un codice etico, a garanzia della compatibilità dello sviluppo e dell'uso dell'IA con i diritti umani fondamentali, e segnatamente la dignità umana, l'integrità, la libertà, la privacy e la diversità culturale e di genere, quanto la definizione di uno strumento normativo per il controllo dei sistemi di IA collegato allo sviluppo di un'infrastruttura di IA europea e di sistemi di IA definiti «responsabili» e provvisti di certificazione e etichettature europee.

[20] Cfr. la Risoluzione 2018/C 252/25, recante *Raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, finalizzate a sollecitare l'elaborazione di aggiornamento della disciplina europea in materia*.

[21] Cfr. Parere del Comitato economico e sociale europeo sull'intelligenza artificiale, *Le ricadute dell'intelligenza artificiale sul mercato unico (digitale), sulla produzione, sul consumo, sull'occupazione e sulla società*, 2017/C 288/01 del 31/5/2017.

Nell'aprile 2018 i rappresentanti di 25 Stati membri hanno sottoscritto una Dichiarazione di cooperazione sull'IA nella quale hanno indicato fra gli obiettivi da perseguire quello di garantire un idoneo quadro giuridico ed etico, fondato sui diritti fondamentali della persona[22]. La Commissione ha altresì proposto agli Stati membri di elaborare congiuntamente un Piano coordinato basato sulla citata dichiarazione di cooperazione, in cui viene sottolineata la necessità di consolidare la diffusione e l'eccellenza di tecnologie dell'IA affidabili e di elaborare orientamenti etici con una prospettiva globale, garantendo un quadro giuridico favorevole all'innovazione[23].

Sulla scorta di questa strategia, la Commissione, nel giugno 2018, ha nominato un gruppo di esperti indipendenti e ha, nell'aprile 2019, presentato i suoi orientamenti etici sull'IA[24]. In essi, ancora una volta, è centrale l'elemento

[22] Cfr. Comunicazione della Commissione al Parlamento europeo, al Consiglio europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *L'intelligenza artificiale per l'Europa*, COM(2018) 237 final, del 25/4/2018, ove, muovendo dalla raccolta delle varie istanze precedentemente ricevute, la Commissione ha proposto un primo documento organico sull'intelligenza artificiale, di natura programmatica, finalizzato a realizzare un approccio unitario fra gli Stati membri con l'espresso scopo di «mettere la forza dell'IA al servizio del progresso umano» e sottolineando la necessità di predisporre un opportuno quadro etico e giuridico basato sui valori dell'Unione e coerente con la Carta dei diritti fondamentali dell'UE.

[23] Il Piano coordinato è stato presentato nel dicembre 2018 e approvato dal Consiglio nel febbraio 2019. Centrale in esso è l'affermazione secondo la quale affinché la società accetti l'IA, è necessaria una maggiore fiducia, che solo una tecnologia prevedibile, responsabile, verificabile ed etica, rispettosa dei diritti fondamentali potrebbe assicurare.

[24] Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *Creare fiducia nell'intelligenza artificiale antropocentrica*, Doc. COM(2019) 168 final del 8/4/2019, ove sono indicati sette requisiti fondamentali dell'IA – delineati per una IA affidabile basata sui valori dell'UE – tra i quali la necessità dell'intervento e della sorveglianza umana su ogni sistema di IA, in maniera tale che siano sicuri (nel senso di verificabili) e trasparenti (nel senso di tracciabili, nel loro intero processo) secondo un meccanismo di *accountability*, che ne garantisca la verificabilità anche esterna.

della fiducia, e l'obiettivo di rinforzarla a fronte della circostanza che l'IA consente alle macchine di apprendere e di prendere decisioni ed eseguirle anche senza l'intervento umano, e che tali decisioni possono basarsi su dati incompleti e non affidabili, manomessi, condizionati o anche solo errati: ciò che emerge in sostanza è che una applicazione acritica e non adeguatamente regolata della tecnologia potrebbe condurre ad una riluttanza nell'accettazione della medesima da parte dei cittadini, e alla sua percezione come lontana, se non antitetica, proprio rispetto alle esigenze dello sviluppo umano.

È questo il fulcro del legame che la Commissione individua fra costruzione di un clima di fiducia e approccio antropocentrico all'IA. Ciò conduce ad affermare la necessità di integrazione dei valori fondativi della UE nelle modalità di sviluppo dell'IA e in questo senso deve intendersi il riferimento al vasto plesso di diritti compendiati nella Carta dei diritti fondamentali dell'UE, che, nella visione della Commissione, sono suscettibili di servire come quadro normativo destinato a diventare lo standard internazionale per l'IA antropocentrica.

Nelle fonti che si sono brevemente passate in rassegna emerge dunque una particolare enfasi sul legame fra antropocentrismo e costruzione di un clima di fiducia propedeutico allo sviluppo della IA. Manca invece in esse una piena focalizzazione sul tema del rischio, cui si è invece pervenuti con l'elaborazione del *Libro bianco sull'intelligenza artificiale*, documento fondamentale in cui è confluita la strategia della Commissione. Pubblicato nel febbraio 2020, esso coniuga gli obiettivi di costruzione di un ecosistema di eccellenza e di fiducia per l'IA[25].

In questo contesto emerge significativamente l'approccio incentrato sul rischio per i diritti fondamentali, che viene descritto in piena continuità rispetto a quello già formalizzato in tema di *data protection* e incorporato nel GDPR. Da questo punto di vista, si deve ritenere che i due approcci

[25] Cfr. *Libro bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia*, COM(2020) 65 final, 19/4/2020.

si presentino come complementari, ma che la frontiera di rischio legato alla IA risulti più avanzata, come conferma la circostanza che il *Libro bianco* riconosce ad essa una dimensione precipua, derivante da difetti nella progettazione complessiva dei sistemi di IA o dall'uso di dati senza che ne siano state corrette le eventuali distorsioni[26].

Si tratta di una importante presa d'atto di come l'intelligenza artificiale possa impattare su diritti già in prima battuta protetti nel prisma della *data protection* e ben oltre il tema della protezione dei (soli) dati personali. Nel documento emerge poi la consapevolezza che le distorsioni e discriminazioni che riflettono un rischio già intrinseco del processo decisionale umano, assumono, quando riferite alla IA, una scala quantitativa ad esso incommensurabile, dotata di effetti esponenzialmente moltiplicati rispetto ad una platea di soggetti molto più vasta, e richiedono per questo forme specifiche e adeguate di controllo[27].

Occorre infine menzionare che l'approccio basato sul rischio non è proprio della sola UE, dal momento che anche in seno al Consiglio d'Europa ha visto la luce il 17/12/2020, ad opera del Comitato ad hoc sull'intelligenza artificiale (CAHAI), uno studio di fattibilità che riprende tale nozione. Pur esulando l'analisi di tale documento dall'economia del presente contributo, basterà richiamare che esso è incentrato sulla ricerca degli elementi di un quadro giuridico per lo sviluppo, la progettazione e l'applicazione dell'intelligenza artificiale, che sia basato sugli standard del Consiglio d'Europa in materia di diritti umani. Centrale nel metodo di lavoro del CAHAI è proprio l'analisi dei rischi e delle opportunità derivanti dalla progettazione, dallo sviluppo e dall'applicazione dell'intelligenza artificiale, con

[26] È questo un tema delicatissimo, legato al problema della qualità dei dati usualmente designato con l'espressione «garbage in – garbage out», su cui si tornerà nel seguito della trattazione.

[27] Le esigenze esposte dalla Commissione sono state ribadite anche nel Report 2020 della European Agency For Fundamental Rights e nello studio dell'EPRS, *European Framework on ethical aspects of artificial intelligence, robotics and related technologies, European added value assessment*, PE 654.179, del settembre 2020.

una particolare attenzione al suo impatto sui diritti umani, sulla democrazia e sulla *rule of law*[28].

3. *La centralità dell'approccio basato sul rischio nella Proposta di Regolamento sulla IA e la sua articolazione*

Il contesto storico e sistematico che si è brevemente richiamato costituisce lo sfondo su cui è maturata la recente Proposta di Regolamento sull'intelligenza artificiale dell'Unione europea, pubblicata nell'aprile 2021, come ambizioso tentativo di regolare le molteplici applicazioni di IA alla luce di un approccio *risk based* che si pone in continuità con quanto elaborato nel *Libro bianco*[29].

La proposta predispone una tassonomia di tali applicazioni alla quale corrisponde un impianto normativo graduato, che vieta alcuni usi dell'IA, ne regolamenta incisivamente altri, considerati altamente rischiosi, e prevede un regime meno stringente e per così dire «di chiusura» per gli utilizzi restanti, essenzialmente qualificati come dotati di rischio limitato o minimo, e in cui l'adesione a un sistema di *accountability* è essenzialmente rimessa alla scelta volontaria degli operatori.

[28] Cfr. Ad Hoc Committee on Artificial Intelligence (CAHAI), *Feasibility study on a legal framework on AI design, development and application based on CoE standards*, adottato il 17/12/2020, reperibile su https://rm.coe.int/cahai-2020-23-final-engfeasibility-study-/1680a0c6da. Nello studio si perviene alla conclusione dell'opportunità di un quadro giuridico completo che combina strumenti giuridici vincolanti e non vincolanti, finalizzati a sviluppare uno strumento vincolante di carattere «orizzontale» volto a consolidare principi generali comuni per l'ambiente dell'intelligenza artificiale. Tale strumento dovrebbe basarsi, ancora una volta, su un approccio basato sul rischio, includendo al contempo disposizioni più granulari in linea con i diritti, i principi e gli obblighi individuati nello studio medesimo, e potrebbe combinarsi con ulteriori strumenti settoriali del Consiglio d'Europa a carattere «verticale», finalizzati ad affrontare sfide poste dai sistemi di intelligenza artificiale in particolari ambiti.

[29] Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione (sec [2021], 167 final - swd[2021] 84 final - swd[2021] 85 final), 21/4/2021.

Uno degli aspetti innovativi della proposta consiste nell'introduzione di una valutazione di conformità *ex ante* dei sistemi di IA ad alto rischio come requisito per la loro messa sul mercato. A tale requisito si accompagna un sistema di monitoraggio post-mercato per la rilevazione e limitazione di eventuali problematiche derivanti dalle applicazioni di sistemi di IA[30].

In questo senso la Proposta della Commissione riflette l'approccio «antropocentrico» all'IA di cui già si diceva, ponendo in essere un sistema di controlli e certificazioni finalizzate ad evitare la messa in commercio e, comunque, l'utilizzo di sistemi di IA che possano pregiudicare i diritti fondamentali. Non sfugge in proposito, anche alla luce di quanto riportato come sfondo di elaborazione, come la stessa predicazione dei diritti fondamentali sia in certo senso funzionale e servente alla connotazione della IA come antropocentrica, e che parimenti lo sia la strutturazione in fasce di rischio[31].

[30] In tema cfr. M. MacCarthy e K. Propp, *Machines Learn that Brussels Writes the Rules: The EU's New AI Regulation*, in «Lawfare», 28/4/2021, p. 1.

[31] Il rischio è quindi valutato in relazione all'entità delle conseguenze negative che può comportare alla salute e alla sicurezza o ai i diritti fondamentali delle persone. Più in dettaglio, i quattro livelli sono definiti come: rischio inaccettabile che comporta il divieto di una serie molto limitata di usi dell'IA particolarmente dannosi, che contravvengono ai valori dell'UE perché violano i diritti fondamentali (in particolare l'attribuzione di un punteggio sociale da parte dei governi, lo sfruttamento delle vulnerabilità dei minori, l'uso di tecniche subliminali e – soggetti ad eccezioni limitate – i sistemi di identificazione biometrica remota in tempo reale utilizzati in spazi accessibili al pubblico nelle attività di contrasto); rischio alto, che è considerato in relazione ad un numero limitato di sistemi di IA definiti nella proposta stessa, e che sono suscettibili di avere ripercussioni negative sulla sicurezza delle persone o sui loro diritti fondamentali, come tutelati dalla Carta dei diritti fondamentali dell'UE; rischio limitato, che comporta obblighi di trasparenza specifici; e il rischio minimo, che si intende categoria di chiusura e quantitativamente preminente e rispetto alla quale la regolazione non interviene. In tema cfr. C. Casonato e B. Marchetti, *Prime osservazioni sulla Proposta di Regolamento dell'Unione europea in materia di intelligenza artificiale*, in «BioLaw Journal - Rivista di BioDiritto», 2021, pp. 415 ss.

La sola parte della proposta che realizza una valutazione *ex ante* del rischio come inaccettabile è quella coperta dalla previsione dell'art. 5, che conduce ad un divieto dell'impiego della IA in specifici ambiti[32].

In relazione ai sistemi definiti ad alto rischio, la Proposta fa riferimento all'Allegato III, che è deputato ad annoverare

[32] Non è possibile approfondire in questa sede il tema, ma giova senz'altro fare riferimento alla formulazione dell'art. 5 comma 1, della Proposta, secondo cui: «sono vietate le pratiche di intelligenza artificiale seguenti: *a*) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole al fine di distorcerne materialmente il comportamento in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico; *b*) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di uno specifico gruppo di persone, dovute all'età o alla disabilità fisica o mentale, al fine di distorcere materialmente il comportamento di una persona che appartiene a tale gruppo in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico; *c*) l'immissione sul mercato, la messa in servizio o l'uso di sistemi di IA da parte delle autorità pubbliche o per loro conto ai fini della valutazione o della classificazione dell'affidabilità delle persone fisiche per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali o della personalità note o previste, in cui il punteggio sociale così ottenuto comporti il verificarsi di uno o di entrambi i seguenti scenari: *i*) un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di interi gruppi di persone fisiche in contesti sociali che non sono collegati ai contesti in cui i dati sono stati originariamente generati o raccolti; *ii*) un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di interi gruppi di persone fisiche che sia ingiustificato o sproporzionato rispetto al loro comportamento sociale o alla sua gravità; *d*) l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, a meno che e nella misura in cui tale uso sia strettamente necessario per uno dei seguenti obiettivi: *i*) la ricerca mirata di potenziali vittime specifiche di reato, compresi i minori scomparsi; *ii*) la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o di un attacco terroristico; *iii*) il rilevamento, la localizzazione, l'identificazione o l'azione penale nei confronti di un autore o un sospettato di un reato di cui all'articolo 2, paragrafo 2, della decisione quadro 2002/584/GAI del Consiglio, punibile nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà della durata massima di almeno tre anni, come stabilito dalla legge di tale Stato membro».

i medesimi secondo una logica di flessibilità e di periodica revisione volta ad allineare la previsione normativa con gli sviluppi e le esigenze future. Può forse sorprendere che l'obiettivo di garantire la fiducia e un livello uniforme ed elevato di protezione siano demandati ad una categorizzazione mobile, che ad oggi può apparire come una scatola vuota. Ai limiti di tale tecnica normativa, fa da contraltare l'individuazione di requisiti obbligatori per i sistemi ad alto rischio, che spaziano dalla qualità dei dati utilizzati, alla documentazione tecnica e alla conservazione dei dati medesimi, dalla trasparenza alla fornitura di informazioni agli utenti, dalla sorveglianza umana alla robustezza, dall'accuratezza alla cybersicurezza[33].

Il rispetto dei requisiti minimi deve essere dimostrato mediante una dichiarazione europea di conformità, emessa dalla competente autorità designata da ciascun Stato membro. La proposta prevede in proposito l'introduzione di una *governance*, sia a livello europeo, con la creazione di un comitato *ad hoc*, sia a livello nazionale, con l'istituzione di apposite autorità amministrative indipendenti e un regime di sanzioni.

La logica di tale costruzione è che, in caso di violazione, i requisiti consentano alle autorità nazionali di avere accesso alle informazioni necessarie per indagare se l'uso del sistema di IA sia stato o no conforme alle norme, e il quadro proposto si propone una coerenza al contempo con la Carta dei diritti fondamentali dell'Unione europea e con gli impegni commerciali internazionali dell'Unione. Si tratta di previsioni che a ben vedere costituiscono, in certo senso, il cuore della nuova disciplina in relazione alla gestione del rischio.

In relazione invece ai sistemi a rischio limitato, sono imposti specifici obblighi soprattutto nel senso della trasparenza, ad esempio laddove esista un evidente rischio di manipolazione. È questo in particolare il caso dell'uso di *chatbot* e in questi casi gli utenti dovrebbero essere consapevoli del fatto che stanno interagendo con una macchina.

[33] Si rinvia in part. agli artt. 8 ss. della Proposta.

Infine, per quanto attiene ai sistemi a rischio minimo, che nella logica di chiusura sono tutti quelli che non rientrano nelle altre categorie, essi possono essere sviluppati e utilizzati nel rispetto della legislazione vigente senza ulteriori obblighi giuridici. Deve essere sottolineato che si tratta della grande maggioranza dei sistemi di IA attualmente utilizzati nell'UE e che i fornitori di tali sistemi possono scegliere di applicare, su base volontaria, i requisiti per un'IA affidabile e aderire a codici di condotta volontari.

Elemento caratterizzante del sistema delineato nella Proposta è quindi il fatto che la valutazione di conformità dei sistemi di IA ad alto rischio debba avvenire in via preventiva, ed è questo un elemento su cui si tornerà in seguito, per qualche osservazione critica[34]. Questa dimensione di valutazione *ex ante* appare assai ambiziosa ed è completata da una dimensione *ex post*, costruita in chiave ancillare e basata su meccanismi di monitoraggio successivi all'immissione sul mercato. Ad essa è dedicato il titolo VII della proposta, dalla puntuale analisi del quale occorre prescindere nell'economia del presente lavoro. Basterà sottolineare, insieme alla portata di complemento rispetto alla logica *ex ante*, il fatto che esso sembra da un lato implicare una significativa presa d'atto che la mitigazione dei rischi *ex post* sia in certa misura ineludibile; ma anche che tale fase sia in larga misura demandata agli stessi produttori e senza meccanismi troppo stringenti o un adeguato rapporto con il tema della responsabilità civile dei medesimi.

[34] Cfr. Regolamento (EU) 2016/679 del Parlamento europeo e del Consiglio del 27/4/2016 (GDPR), che è stato un precursore di questo approccio avendo introdotto il concetto di *privacy by design* allo scopo di integrare le tutele nei trattamenti. Si tratta, come noto, di un processo che coinvolge diverse componenti tecnologiche e organizzative, volto all'implementazione dei principi della privacy e della protezione dei dati fin dalla progettazione di tecnologie e sistemi informatici. Questo fine è perseguito, innanzitutto, attraverso l'obbligo previsto sia dal GDPR che dalla nuova Proposta di Regolamento della Commissione di effettuare una valutazione di impatto nel momento della progettazione di ogni nuova tecnologia o servizio che possa rappresentare un rischio elevato per le persone a cui si rivolge la decisione.

La sensazione è insomma che, per quanto l'impianto della proposta sia fortemente sbilanciato in favore di una gestione *ex ante*, non sfugga nella complessità del testo la consapevolezza che il tema del rischio riguardi l'intero ciclo di vita di un sistema di IA, con il coinvolgimento di diversi operatori, che vanno dallo sviluppatore al produttore al distributore o ancora all'importatore: circostanza da cui discende la non lieve difficoltà nell'individuare il momento e l'operatore che si trova nella posizione più indicata per affrontare eventuali rischi attuali o potenziali.

Un ulteriore elemento da porre in relazione al rischio è poi quello relativo ai dati che alimentano l'IA, sulla centralità dei quali non vi è dubbio, senza che però la proposta individui nessi convincenti, dal momento che essa si concentra sul tema della intrinseca programmazione della intelligenza artificiale e non sul tipo e la qualità dei dati che la medesima utilizza. È questo tema delicatissimo, che segna la necessità di una piena saldatura con i principi fondamentali in tema di protezione dei dati, e particolarmente accuratezza, correttezza, minimizzazione e necessità e la correlata adeguata qualità dei dati e attendibilità delle fonti. Sembra difficile prescindere da una considerazione dei medesimi ben oltre l'impianto del GDPR e nel prisma dell'impianto basato sul rischio proprio della proposta, che, se sembra postulare la necessità di una qualche forma di incorporazione di detti principi nella IA, resta sul punto silente.

Queste ultime due considerazioni conducono ad interrogarsi se l'esaminato impianto basato sul rischio come base principale di tutela dei diritti fondamentali sia davvero adeguato: l'analisi sarà condotta a partire da alcune caratteristiche intrinseche della intelligenza artificiale e alla considerazione del pericolo di discriminazione algoritmica.

4. *L'impianto basato sul rischio a delicato confronto con la mobilità della nozione di intelligenza artificiale*

Quanto esposto consente di dedicarsi ora a qualche osservazione critica all'impianto del Regolamento, con

particolare riferimento proprio all'approccio *risk based*. Per fare ciò occorre richiamare brevemente alcune distinzioni e alcune conseguenti caratteristiche che sono proprie degli algoritmi, per evidenziare come vi sia una tensione intrinseca fra l'approccio *risk based*, posto al centro della proposta, e la stessa nozione di IA nella sua declinazione più attuale.

Il nodo, ancor prima che giuridico, è definitorio[35]. Se infatti si abbraccia una definizione di algoritmo come quella che è disponibile nella corrente percezione del fenomeno dell'intelligenza artificiale[36], l'impianto della proposta potrebbe sembrare privo di sostanziali criticità. Si prenda ad esempio la definizione corrente di algoritmo tratta dal vocabolario Treccani nei termini di

[35] Da sempre il nodo definitorio è un punto delicato in tema di intelligenza artificiale. Così sottolinea la varietà e vaghezza definitoria R.K. Hill, *What an algorithm is*, in «Philosophy and Technology», 2015, p. 36. Sulla pretesa neutralità e sulla correlata distinzione fra *managed, policy-directed algorithms* e *policy-neutral algorithms*, con i secondi che offrirebbero risultati non manipolati né orientati a valori predeterminati, cfr. O. Tene e J. Polonetsky, *Taming The Golem: Challenges of Ethical Algorithmic Decision-Making*, in «North Carolina Journal of Law & Technology», 2018, p. 132.

[36] D'altronde, a ben vedere e su un piano ancor più generale, è lo stesso riferimento al concetto di intelligenza ad apparire fuorviante. Se infatti è stato a lungo popolare creare un parallelo con l'intelligenza umana (come avviene ad esempio nella definizione di S. Chopra e L.F. White, *A Legal Theory of Autonomous Artificial Agents*, Ann Arbor, 2011, 5), in termini di «science of making machines do things that would require intelligence if done by persons», è stata ben evidenziata la scollatura fra un piano sintattico, che discende dalla capacità computazionale delle macchine, ed un piano semantico che invece continua a sfuggire ad una reale comprensione da parte delle stesse. Pertanto, come da ultimo ribadito argomentando sulla distinzione fra *agency* e intelligenza, non propriamente di intelligenza si tratterebbe, ma di un mezzo con cui svolgere compiti ripetitivi e «codificabili». In tema cfr. L. Floridi e F. Cabitza, *Intelligenza artificiale. L'uso delle nuove macchine*, Milano, 2021, in part. pp. 148 ss., in cui ben si argomenta la possibilità di una non problematica scissione fra l'*agere* e l'*intelligere*. Il punto è rilevante anche in relazione alla adeguatezza dell'approccio basato sul rischio, in relazione al quale rileva proprio che l'*agency* della l'IA artificiale sia in grado di modificarsi, pur in assenza di una comprensione semantica, e risulti suscettibile di impattare i diritti fondamentali ben al di là del prevedibile e del codificato.

qualunque schema o procedimento matematico di calcolo; più precisamente, un procedimento di calcolo esplicito e descrivibile con un numero finito di regole che conduce al risultato dopo un numero finito di operazioni, cioè di applicazioni delle regole[37].

Essa appare ben lungi dall'abbracciare la molteplicità di manifestazioni del fenomeno del quale tende ad evidenziare la definitezza, implicando in certa misura la conoscibilità dei passaggi computazionali che è terreno ideale per l'applicazione di strumenti di gestione del rischio. Ogni sistema di gestione del rischio postula infatti per essere efficace che si possa almeno in certa misura determinare preliminarmente i differenti livelli di rischio, per poi agire normativamente. Tale è d'altronde, come si è visto, la tipica caratterizzazione della Proposta di Regolamento.

Tuttavia, la citata definizione non abbraccia le forme più evolute ed attuali di intelligenza artificiale, e in particolare quelle basate sul *machine learning* e, ancor più, sul *deep learning.* Gli algoritmi di *machine learning* agiscono in modo peculiare rispetto agli algoritmi tradizionali, perché sono in grado di modificarsi e migliorarsi automaticamente attraverso l'esperienza e l'utilizzo di dati: questo avviene mediante la costruzione progressiva di un modello basandosi su dati di addestramento che consolidano la capacità della IA di effettuare predizioni anche senza che vi sia stata una esplicita programmazione a questo fine[38]. L'apprendimento certo può essere supervisionato, e in questo caso l'algoritmo necessita di dati di addestramento che siano già opportunamente etichettati, cioè contengano l'esplicita informazione della categoria a cui appartiene il dato utilizzato nell'apprendimento. Tuttavia esso può anche essere non supervisionato e in questo caso le correlazioni che determinano i modelli predittivi sono assai più libere,

[37] La definizione tratta dal vocabolario Treccani (www.treccani.it/vocabolario/algoritmo/) sembra sottolineare, e non appare un caso per quanto si analizzerà criticamente di seguito, la definitezza e quindi conoscibilità dei vari passaggi computazionali.

[38] Sulla centralità della problematica definitoria cfr. anche S. Russell e P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 2020.

e in certa misura imprevedibili, dal momento che i dati di ingresso non sono già categorizzati[39].

Alla base di molti algoritmi di *machine learning* di più recente generazione ci sono poi vere e proprie reti neurali artificiali, ispirate alle reti neurali che si ritrovano nel cervello biologico, che segnano uno stacco strutturale rispetto alla IA tradizionale. I nodi di una rete neurale artificiale sono infatti organizzati su numerosi strati, e i diversi strati della rete effettuano molteplici trasformazioni dei dati oggetto di immissione senza nessi evidenti e tracciabili. Ciò conduce ad una fase di apprendimento definibile come «profondo» (*deep learning*), che è impossibile ricostruire secondo logiche lineari: per questo si parla anche di intelligenza artificiale connessionista, e per questo *non-human readable*.

A fronte di questa evoluzione, pare evidente come lo schema di valutazione del rischio di violazione rispetto ai diritti fondamentali risulti per questi aspetti velleitario. Infatti, se gli algoritmi sono in grado di analizzare quantità elevatissime di dati, anche fra loro eterogenei sotto il profilo qualitativo, individuandone preziose correlazioni e procedendo ad un apprendimento secondo logiche profonde, ciò determina, nel quadro di una sempre crescente disponibilità di dati (il contesto dei cd. *big data* cui si faceva riferimento), inferenze che nessuna valutazione umana sarebbe in grado di comprendere e conseguentemente apprezzare in termini di livelli di rischio per i diritti fondamentali[40].

Detto in altre parole, si è di fronte ad una valutazione del rischio che avviene su un tappeto mobile, che non fa

[39] In tema cfr. M.U. Scherer, *Regulating artificial intelligence systems: risks, challenges, competencies and strategies*, in «Harvard Journal of Law & Technology», 2016, p. 365; G. D'Acquisto, *On conflicts between ethical and logical principles in artificial intelligence*, in «AI and Society», 2020, pp. 895 ss.

[40] Cfr. B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter e L. Floridi, *The ethics of algorithms: Mapping the debate*, in «Big Data & Society», 2016, p. 3, ove si osserva: «Analytics demonstrates how algorithms can challenge human decision-making and comprehension even for tasks previously performed by humans».

che rispecchiare quella che è stata ben definita come una ontologica differenza fra decisione algoritmica ed umana[41]. Si ripropone qui una aporia in tema di determinazione del rischio che sembra ricalcare quella relativa alla difficoltà di dotare i sistemi di intelligenza artificiale di una *accountability by design*[42]. Circostanza che, nella programmazione e messa in opera degli algoritmi predittivi, è riconosciuta ormai come assai difficile da perseguire rispetto a un processo decisionale, quello algoritmico, che ha come caratteristica l'attitudine a trascendere la comprensione umana[43]. Motivo per il quale la regolazione ispirata a principi di trasparenza o spiegabilità si è rivelata ingenua e difficilmente praticabile[44].

Si potrebbe pensare che la logica basata sul rischio nasca proprio come strumento per contrastare questa intrinseca incertezza, ma in realtà tale approccio ricade, come si vedrà ora, nello stesso circolo vizioso. Il dato strutturale da cui partire per tale dimostrazione è peraltro stato chiarito: il funzionamento degli algoritmi predittivi, basati su logiche di autoapprendimento, che prevedono la modifica della

[41] In tema cfr. F. Pasquale, *The Black Box Society: The Hidden Algorithms That Control Money and Information*, Cambridge, 2015, p. 187. Questo quadro dà ben conto di come la stessa nozione di discriminazione algoritmica sfugga ai classici strumenti del diritto anti-discriminatorio. Per questo, come è stato ben argomentato, tali strumenti sono spesso almeno in parte inefficaci e richiedono di essere integrati con tutti gli strumenti del plesso normativo dedicato alla *data protection*. Sul tema, su cui si tornerà nel prossimo paragrafo, cfr. P. Hacker, *Teaching Fairness to Artificial Intelligence*, in «Common Market Law Review», 2018, pp. 1146 ss.

[42] Cfr. J.A. Kroll, J. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson e H. Yu, *Accountable Algorithms*, in «University of Pennsylvania Law Review», 2017, p. 639.

[43] Cfr. A.D. Selbst e J. Powles, *Meaningful information and the right to explanation*, in «International Data Privacy Law», 2017, 233. In tema sia consentito rinviare anche a A. Oddenino, *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano*, in «DPCE online», 2020, p. 199 ss.

[44] In tema cfr. A. Chander, *The Racist Algorithm?*, in «Michigan Law Review», 2017, p. 1039, ove si osserva che «instead of transparency in the design of the algorithm, what we need is a transparency of inputs and outputs».

propria struttura dopo ogni decisione o analisi di dati[45]. Ciò ribadito occorre pensare che qualsiasi approccio normativo basato sul rischio richiede in certa misura di cristallizzare il rischio.

Per quanto quindi l'elemento del rischio rimandi alla nozione di incertezza, l'approccio basato sul rischio nasce proprio dal tentativo di standardizzare soglie di rischio da cui far discendere conseguenze regolatorie. Si tratta quindi di uno schema che muove dalla nozione di rischio per tentare di individuarlo *ex ante*, quasi nel tentativo di «catturarlo» attraverso la predisposizione di meccanismi normativi. Questo è d'altronde l'approccio della Proposta di Regolamento, che, come si è anticipato, si affida ad una ampia tassonomia basata su diversi livelli di rischio e appronta meccanismi basati su dichiarazioni di conformità (art. 43) e di certificazione (art. 44).

Una breve analisi di queste previsioni deve partire dal Considerando n. 66 della proposta, nel quale si evidenzia l'opportunità che un sistema di IA sia sottoposto a una nuova valutazione della conformità ogniqualvolta intervenga una modifica che possa incidere sulla conformità del sistema al Regolamento, oppure quando viene modificata la finalità prevista del sistema. Nel Considerando si precisa altresì che, per quanto riguarda i sistemi di IA che proseguono il loro apprendimento dopo essere stati immessi sul mercato, occorre prevedere regole atte a stabilire che le modifiche apportate all'algoritmo e alle sue prestazioni, predeterminate dal fornitore e valutate al momento della valutazione della conformità, non costituiscano una «modifica sostanziale».

Si tratta di una previsione ambiziosa ma a ben vedere essa non fa che spostare il problema, in modo surrettizio, su cosa debba intendersi per «modifica sostanziale». È precisamente su questo punto che l'analisi dell'art. 43 della proposta rivela

[45] Cfr Kroll, Huey, Barocas, Felten, Reidenberg, Robinson e Yu, *Accountable Algorithms*, cit., p. 659, ove si sottolinea che i sistemi più avanzati di *machine learning* «can update their model for predictions after each decision, incorporating each new observation as part of their training data».

qualche debolezza. Esso infatti nel disciplinare la procedura di valutazione della conformità, e nel prevedere che i sistemi di IA ad alto rischio siano sottoposti a una nuova procedura di valutazione della conformità dopo ogni modifica sostanziale, aggiunge che, per i sistemi di IA ad alto rischio che proseguono il loro apprendimento dopo essere stati immessi sul mercato o messi in servizio, le modifiche apportate al sistema di IA che sono state predeterminate dal fornitore al momento della valutazione iniziale della conformità, non costituiscono una modifica sostanziale[46].

La Commissione, dunque, muovendo dalla constatazione che vi sono sistemi di IA basati sull'autoapprendimento che sono in costante evoluzione, riconosce l'esistenza di una criticità rispetto al sistema valutativo e di certificazione, come peraltro già sottolineato sin dal *Libro bianco*, come si ricordava. Tuttavia, il sistema delineato nella Proposta di Regolamento pare demandare ad una valutazione dello stesso fornitore del sistema di IA sulla necessità o meno di procedere con una nuova valutazione e certificazione laddove intervenga una «modifica sostanziale». In particolare, poi, risulta delicata l'esclusione della sussistenza di una modifica sostanziale quando l'evoluzione sia stata «predeterminata» dal fornitore al momento della valutazione iniziale della conformità: circostanza che appare in aperta contraddizione con la natura strutturale degli algoritmi di *machine learning* che si è richiamata, e ancor più con la dinamica di *deep learning* che costituisce la frontiera più attuale di sviluppo dell'intelligenza artificiale.

[46] Giova sottolineare che tali informazioni fanno parte di quelle contenute nella documentazione tecnica di corredo, di cui all'Allegato IV, punto 2, lettera *f* della Proposta di Regolamento.

In tema di prospettive e criticità dei meccanismi di certificazione della IA cfr. A. Henriksen, S. Enni e A. Bachmann, *Situated Accountability, Ethical Principles, Certification Standards and Explanation Methods in Applied AI*, in «AIES 21, Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics and Society», 2021, p. 574. In tema, nella prospettiva delle regole etiche cfr. anche P. Cihon, M. Kleinaltenkamp, J. Schuett e S.D. Baum, *AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries*, in «IEEE Transactions on Technology and Society», 2021, pp. 200 ss.

Sulla scorta di tali considerazioni, appare discutibile affidare allo stesso fornitore il processo di revisione della certificazione della intelligenza artificiale ad alto rischio, a maggior ragione nella misura in cui esso si riveli, come si è visto, ancorato a presupposti tutt'altro che certi ed oggettivi.

Invero, è ragionevole prevedere un ampio dibattito sulla definizione, in concreto, del concetto di «modifica sostanziale» e sulla perimetrazione di ciò che è stato «predeterminato» dal produttore.

In definitiva, comunque, il nodo concettuale resta che, volendo approcciare la questione da un punto di vista anche solo logico ancor prima che giuridico, si rivela una aporia nel voler fare affidamento su una certificazione «stabile» rispetto ad un sistema che necessariamente evolve e muta, anche in maniera imprevista rispetto al produttore o comunque imprevedibile.

5. *La discutibile attrazione nell'alveo della logica della gestione del rischio del pericolo di discriminazione algoritmica*

Un altro aspetto che sorprende della Proposta di Regolamento è che essa non sembra considerare adeguatamente un pericolo concreto e trasversale rispetto ai vari ambiti che sono ricompresi nella sua ambiziosa mappatura del rischio in funzione di tutela dei diritti fondamentali: si tratta della discriminazione algoritmica. La ragione di questa mancanza è presumibilmente dovuta all'impostazione della proposta, che tende a preferire una mappatura per blocchi di materie, più agevole da inserire nella struttura tassonomica dei livelli di rischio. In effetti da questo punto di vista il solo riferimento indiretto può ritrovarsi nell'articolo 5, dedicato alle applicazioni ad altissimo rischio, che come si è visto sono proibite, risolvendo in tal modo il problema in radice.

Tuttavia, la discriminazione algoritmica si presenta come fattispecie assolutamente vasta e trasversale e, per la ragione che illustreremo, in certa misura come pericolo coessenziale al funzionamento dell'algoritmo.

Prima di addentrarci nella disamina di questo tema, cruciale, occorre richiamare brevemente la distinzione fra rischio e pericolo, che appare fondamentale per comprendere questo risvolto problematico della proposta.

È noto come l'attuale società sia stata caratterizzata dal grande sociologo tedesco Ulrich Beck come «società del rischio», il cui tratto caratterizzante è identificato precisamente nella crescente attitudine alla produzione e gestione del rischio, che diviene catalizzatore degli interessi non solo mediatici, politici e scientifici, ma anche economici e giuridici[47].

Questa circostanza rimanda a sua volta al tema della distribuzione delle risorse di gestione del rischio, che sono tanto risorse di investimento quanto risorse di conoscenza. Tale collegamento si riflette sulla circostanza che ad una diseguale distribuzione delle risorse di gestione del rischio corrisponde una forma di redistribuzione del potere e della ricchezza. Si tratta di una ricostruzione assai calzante e rilevante anche rispetto al tema del rischio di violazione dei diritti fondamentali da parte dell'IA che è posto al centro della proposta: essa, nel determinare l'impianto di gestione del rischio, sembra idonea a produrre rilevanti effetti redistributivi sulla catena di produzione e diffusione della IA.

Su questo sfondo si colloca una ulteriore e più ampia riflessione sulla tutela dei diritti fondamentali a fronte dello sviluppo della IA e della discriminazione algoritmica, che tenga conto di quest'ultima come vero e proprio pericolo piuttosto che come rischio. Deve sottolinearsi infatti una differenza sostanziale e fondamentale fra i due concetti, al di là di un uso spesso promiscuo nel linguaggio corrente. Con pericolo si designa infatti una proprietà o qualità intrinseca di un determinato fattore avente il potenziale di causare

[47] In tema risulta chiaro come il rischio presenti una natura strutturale, essendo legato ai meccanismi stessi di funzionamento delle società contemporanee, ed essa non può essere adeguatamente considerata solo sul piano individuale. In tema cfr. l'interessante indagine sulla regolazione del rischio in prospettiva comparata svolta in M. Graziadei, *La regolazione del rischio e il principio di precauzione: Stati Uniti ed Europa a confronto*, in «Sistemi intelligenti», 2017, pp. 499 ss.

danni, mentre con rischio si indica un fattore probabilistico di raggiungimento del livello potenziale di danno.

È chiaro, con ciò, che nel caso del pericolo, che è associato intrinsecamente ad un certo elemento, si dovrà sciogliere una stretta alternativa fra una presenza o una assenza di pericolo, che è trattato come una grandezza fondamentale e originale. Ben diversamente per il rischio, ove la valutazione è probabilistica e statistica: il rischio origina dalla presenza di un pericolo ma è legato alla probabilità che esso raggiunga la capacità di produrre un danno nonché all'entità del danno stesso. In questo senso il rischio è una grandezza complessa che deriva dalla combinazione di più elementi, fra i quali i comportamenti individuali, i fattori di contesto, nonché gli elementi tecnologici e organizzativi. È d'altronde proprio da questa circostanza che nascono le tendenze organizzative per la riduzione o minimizzazione del rischio ed è sempre in questa prospettiva che può essere visto e valutato l'impianto complessivo della proposta che si è esaminato.

Per contro la risposta normativa alla sussistenza del pericolo è, come noto, legata alla predisposizione di regimi di responsabilità specifici, stante il legame fra esposizione al pericolo e produzione di un danno, e la conseguente necessità di procedere al suo ristoro.

Su questa base è chiaro che la Proposta di Regolamento privilegia la categoria del rischio trascurando invece quella del pericolo. Essa avrebbe richiesto un diverso piano di considerazione che, per propria natura, non si presta ad una gestione *ex ante* se non attraverso la radicale proibizione o l'irrigidimento del regime di responsabilità, entrambi poco compatibili con lo spirito di favore complessivo che l'atto coltiva per lo sviluppo del mercato della IA.

Alla luce di quanto precede occorre interrogarsi sulla opportunità di qualificare la discriminazione algoritmica come rischio o invece come pericolo, nell'assenza tutto sommato sorprendente di un riferimento diretto ad essa nel testo della proposta. Se infatti il primo dichiarato obiettivo è quello di regolare l'IA in relazione ai diritti fondamentali, non può facilmente trascurarsi che essi trovano proprio nel

dovere di non discriminazione un volano rilevantissimo di applicazione concreta[48].

Sembra che la qualificazione della discriminazione algoritmica in termini di pericolo si possa giustificare alla luce di alcune circostanze che meritano ora una breve considerazione.

Due in particolare sono gli assi su cui si muove chi intenda confutare l'assunto che vorrebbe l'algoritmo impermeabile ai rischi discriminatori tipici del ragionamento umano. Da un lato vi è la questione dello strumento, ossia il codice algoritmico con cui i dati sono elaborati, dall'altro vi è la questione della materia prima su cui il codice si innesta, ossia i dati, nella loro consistenza aggregata.

Muovendo dalle considerazioni sullo strumento, occorre ulteriormente distinguere il caso in cui la discriminazione sia diretta, o per così dire intrinseca rispetto al codice dell'algoritmo per come esso è stato concepito, da quello in cui la discriminazione sia indiretta, ossia derivante da fattori non direttamente riconducibili al codice algoritmico.

Come è evidente la discriminazione può essere governata più facilmente con riferimento al primo caso, ossia quello della natura intrinsecamente discriminatoria dell'algoritmo: non vi sono dubbi, infatti, che il processo decisionale algoritmico non possa essere basato su parametri direttamente discriminatori, essendo la discriminazione diretta di per sé vietata, e ciò anche a prescindere dal fatto che essa sia posta in essere da un essere umano o da un algoritmo.

Si tratta peraltro di fattispecie di minore interesse teorico e che al tempo stesso appaiono residuali sotto il profilo pratico proprio perché è raro che gli sviluppatori di IA pongano in essere violazioni tanto plateali: la formalizzazione del processo di funzionamento degli algoritmi mediante la scrittura del codice condurrebbe facilmente a lasciare tracce indelebili di un intento direttamente discriminatorio, e questo profilo è proprio quello che potrebbe agevolmente essere

[48] In tema cfr. di recente R. Xenidis, *Tuning EU equality law to algorithmic discrimination: Three pathways to resilience*, in «Maastricht Journal of European and Comparative Law», 2020, pp. 736 ss.

governato con meccanismi di conformità e certificazione del tipo previsto nella Proposta di Regolamento.

Diverso è il discorso relativo alle discriminazioni indirette, che non sono come tali disciplinate dal punto di vista normativo e che sono fattispecie invero molto più diffuse nella prassi. La questione è se il processo decisionale algoritmico sia idoneo a ridurre tali discriminazioni, che non discendono dal codice algoritmico in sé considerato ma invece, e talora in forma non pienamente consapevole, da *a priori* logici e da presupposti di programmazione che incorporano pregiudizi inconsci[49].

In proposito, il problema tende ad essere minimizzato da chi considera il processo di scrittura del codice algoritmico come una operazione razionale e proceduralizzata, in cui può sembrare agevole contrastare la traslazione di eventuali pregiudizi inconsci e involontari dei programmatori. In realtà, è oggi crescente la consapevolezza della erroneità di questa impostazione: la non neutralità degli algoritmi, pur non formalmente concepiti come discriminatori, è individuata in ragione della sostanzialmente inevitabile traslazione delle discriminazioni indirette dal mondo analogico a quello digitale.

Se è vero, quindi, che in generale gli algoritmi predittivi risentono della realtà in cui operano e dei valori e dei condizionamenti di chi li ha concepiti, tale profilo è ancor più pronunciato se ci si muove verso forme più evolute di algoritmi. Solo nelle versioni più rudimentali di algoritmi, infatti, si ravvisa una relazione lineare e una corrispondenza diretta fra la programmazione e i risultati ottenuti. Si tratta di una circostanza che consente di isolare più facilmente gli elementi del condizionamento implicito o inconscio, meglio prevenendo le discriminazioni indirette. Ciò non avviene

[49] Cfr. *ex multis* I. Zliobaite, *Measuring discrimination in algorithmic decision making*, in «Data Mining and Knowledge Discovery», 2017, 1060, in part. 1067, ove emerge come l'esito discriminatorio sia veicolato da una soluzione corretta sotto il profilo dell'accuratezza statistica, suscettibile in quanto tale di confermare il buon funzionamento del modello. Sul punto cfr. anche A. Chander, *The Racist Algorithm?*, in «Michigan Law Review», 2017, pp. 1023 ss.

invece nella prospettiva degli algoritmi di nuova generazione, basati, come si è visto, sulla logica di *machine learning* nella quale l'intelligenza artificiale si volge all'autoapprendimento, amplifica l'effetto dei presupposti di programmazione e determina uno iato ben più marcato fra programmazione e risultati decisionali[50].

Muovendo ora alla seconda dimensione, ossia la considerazione del sostrato materiale dei dati, si evidenziano profili parimenti critici. I dati, giova ribadire, sono il sostrato materiale sulla base del quale la decisione algoritmica prende corpo, e sono sostrato essenziale alla concezione, alla verifica e infine alla operatività dell'algoritmo stesso[51].

Orbene, è giocoforza osservare che anche i dati, lungi dal garantire oggettività e neutralità, possono risultare fallaci per via di due ragioni fra loro ben differenti: in primo luogo, i dati possono essere stati raccolti secondo metodologie errate o imprecise che compromettono la corrispondenza fra la realtà descritta dai dati e il mondo reale (cd. *cognitive bias*); in secondo luogo, laddove i dati siano stati invece raccolti in modo corretto, e siano quindi idonei a rappresentare la realtà, essi non possono pertanto che riflettere anche i profili discriminatori esistenti nel mondo reale (cd. *statistical bias*)[52].

[50] Si osserva in effetti che «algorithms are inescapably value-laden as operational parameters are specified by developers and configured by users with desired outcomes in mind that privilege some values and interests over others» (cfr., anche per ulteriori riferimenti, Mittelstadt, Allo, Taddeo, Wachter e Floridi, *The ethics of algorithms: Mapping the debate*, cit., p. 1).

[51] Si sottolinea la duplice dimensione della discriminazione in Tene e Polonetsky, *Taming The Golem: Challenges of Ethical Algorithmic Decision-Making*, cit., p. 125, in part. 130, ove si legge «no algorithm is fully immune from the human values of its code and its creators. Algorithms are written by human designers, who could infuse them with their values, and are trained on human generated data, which carry their own biases».

[52] In tema, *ex multis* cfr. J.D. Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, in «Duke Law Journal», 2007, p. 345; A.G. Greenwald e L. Hamilton Krieger, *Implicit Bias: Scientific Foundations*, in «California Law Review», 2006, p. 945; A.J. Lee, *Unconscious Bias Theory in Employment Discrimination Litigation*, in «Harvard Civil Rights-Civil Liberties Law Review», 2005, p. 481; C.R.

Entrambi gli aspetti sono molto rilevanti, e contribuiscono a connotare il tema della discriminazione con tratti quasi paradossali. Se infatti è intuitivo dedurre che dati errati tendenzialmente conducono a decisioni errate, occorre valutare l'influenza giocata da dati raccolti correttamente sugli esiti del processo decisionale algoritmico. Tale nesso è ben rappresentato dal principio comunemente espresso con la locuzione «garbage in, garbage out», secondo cui le conclusioni del processo algoritmico possono al più raggiungere il livello di attendibilità e di neutralità che è proprio dei dati su cui esse si basano. Il valore di tale principio è stato recentemente riconosciuto anche dalla International Conference of Data Protection and Privacy Commissioners (ICDPPC), che ha provveduto al contempo ad individuare delle misure per ovviare all'effetto di condizionamento in chiave discriminatoria che è insito nel medesimo[53].

Questa ricostruzione rinforza la percezione che la presunta neutralità e oggettività valoriale dei dati sia in realtà nulla più che un mito. Se i dati riflettono il mondo reale essi non possono che replicare anche gli aspetti negativi che lo caratterizzano, e in particolare ineguaglianze e discriminazioni che si perpetuano e si consolidano nel tempo[54].

Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, in «Stanford Law Review», 1987, p. 317.

[53] Cfr. International Conference of Data Protection and Privacy Commissioners, *Declaration on Ethics and Data Protection in Artificial Intelligence*, 2018, principio 6, che pone gli obiettivi di: «*a*) ensuring the respect of international legal instruments on human rights and non-discrimination, *b*) investing in research into technical ways to identify, address and mitigate biases, *c*) taking reasonable steps to ensure the personal data and information used in automated decision making is accurate, up-to-date and as complete as possible, and *d*) elaborating specific guidance and principles in addressing biases and discrimination, and promoting individuals and stakeholders awareness».

[54] In questo senso è significativo che si pervenga ad evocare il concetto di «discriminazione istituzionale». Così S. Barocas e A.D. Selbst, *Big Data's Disparate Impact*, in «California Law Review», 2016, p. 671, in part. 673 ove si osserva che «while discrimination certainly endures in part due to decision makers' prejudices, a great deal of modern-day inequality can be attributed to what sociologists call "institutional" discrimination. Unconscious, implicit biases and inertia within society's

È evidente che a fronte di questa circostanza non si sia di fronte ad un rischio statistico gestibile *ex ante* in relazione alla non discriminatorietà del codice algoritmico, ma piuttosto ad un pericolo concreto: con il paradosso che per avere esiti non discriminatori e quindi rispettosi dei diritti fondamentali occorrerebbe intervenire modificando la portata (intrinsecamente discriminatoria) dei dati immessi, onde spezzare il circolo vizioso determinato dalla circostanza paradossale per la quale a dati veraci ed affidabili conseguono decisioni discriminatorie[55]. Questa lacuna in relazione al pericolo discriminatorio rivela insomma una sorta di vizio metodologico della proposta, pure a fronte di un impianto normativo assai ricco e ambizioso. Ciò conduce a svolgere qualche considerazione conclusiva, che potrebbe essere utile nel tentativo di meglio calibrare gli elementi della proposta in prospettiva di un iter legislativo che si preannuncia accidentato.

6. *Alcune considerazioni conclusive*

Il contributo ha evidenziato alcune criticità di una proposta che appare indiscutibilmente ambiziosa nell'intento di dare corpo normativo all'approccio antropocentrico alla IA, divenuto nel frattempo la bandiera e il tratto distintivo della attitudine della UE in un settore tanto strategico.

Certo, il problema della regolazione dell'IA non ha soluzioni semplici e ogni soluzione, come si ricordava in

institutions, rather than intentional choices, account for a large part of the disparate effects observed».

[55] Il paradosso, che ben si riflette nel fatto che la correttezza dei dati e la loro attitudine anti-discriminatoria appaiano, almeno in certa misura, escludersi reciprocamente, non fa che richiamare la tensione fra prescrizione e descrizione che bene è stata evidenziata in G. Della Morte, *Big Data e protezione internazionale dei diritti umani*, Napoli, 2018, in particolare nelle Conclusioni, che sottolineano la difficoltà di riconciliare una dimensione giuridica prescrittiva a tutela dei diritti e una dimensione predittiva dell'algoritmo, suscettibile di perpetuare le discriminazioni.

apertura, è inevitabilmente portatrice di scelte di società che si possono in ultima analisi definire politiche.

Volendo in questa prospettiva trarre alcune considerazione conclusive, pare opportuno concentrarsi sulla questione della efficienza e della efficacia della scelta posta alla base della proposta, ossia di affrontare la sfida di una regolazione che preservi il plesso dei diritti fondamentali che sono fondanti l'Unione europea attraverso un approccio che pone al centro la nozione di rischio e articola attorno ad essa il proprio apparato di regolazione.

La domanda è, in ultima analisi, se ciò sia efficace rispetto ai diritti fondamentali ed efficiente rispetto al mercato, ossia i due obiettivi che la Proposta di Regolamento si propone ambiziosamente di servire. Da questo punto di vista, la sensazione è che la proposta difficilmente possa essere un buon «servitore di due padroni», per citare la celebre commedia goldoniana.

È infatti convinzione di chi scrive che la risposta ad entrambi gli interrogativi rischi di dover essere negativa, per le ragioni che si sono evidenziate e per una ulteriore considerazione che da esse discende e che si sostanzia in una sorta di «incompiutezza inevitabile» dell'approccio basato sul rischio quando esso sia riferito all'ambito della IA.

Certo, non si può disconoscere l'importanza del rischio e della sua gestione nella moderna civiltà, già industriale e oggi post-industriale, e di massa. Non è dubbio che nell'età contemporanea la regolazione del rischio abbia acquisito crescente centralità, e che questa sia divenuta in modo sempre più diretto ed esplicito una funzione centrale del diritto. Funzione che peraltro si rinnova con continuità, grazie all'emersione di tecniche più sofisticate di calcolo del rischio e all'incremento delle conoscenze scientifiche relative ai fattori di rischio e alla loro incidenza[56].

[56] In tema cfr. Graziadei, *La regolazione del rischio*, cit., p. 500, ove si osserva come con l'espandersi delle funzioni dello Stato, all'idea di controllare il rischio attraverso le regole del diritto privato, si è via via affiancata l'idea di far spazio alla regolazione del rischio tramite l'azione dei poteri pubblici su larga scala.

In questo campo, e nei molteplici ambiti materiali che sono attraversati dal tema del rischio, l'UE spesso rivendica a sé la competenza di agire con efficacia di uniformazione delle regole, nella logica di un consolidamento del mercato interno. Pertanto, non può sorprendere se, a fronte dello sviluppo della IA, l'opzione del legislatore della UE sia stata quella di estendere a tale ambito l'impianto regolativo che era d'altronde già assurto ad ossatura normativa nel settore, assai collegato, della protezione dei dati personali.

La problematicità di questa estensione, come si è visto, si lega ad una sorta di falso sillogismo, che è quello di ritenere che l'approccio basato sul rischio, se è efficace per la tutela della *data protection*, debba esserlo anche in tema di IA e diritti fondamentali, perché quest'ambito è considerato in stretta continuità con quello della protezione dei dati. Si tratta certamente di un parallelo accattivante, così come lo è l'obiettivo di consolidare il ruolo regolativo che la UE si è ritagliata in chiave globale con il GDPR, estendendolo ad un ambito tanto strategico come l'IA[57].

Tuttavia, la fallacia di tale estensione, e la inevitabile incompiutezza dell'approccio che da essa discende, resta evidente per due ordini di ragioni[58].

Sotto un primo profilo occorre osservare che nel caso del GDPR tutto l'approccio *risk based* è stato costruito su un impianto di nuovi diritti codificati a beneficio dell'interessato. È il rischio di violazione di tali diritti a fungere da baricentro del sistema. Si tratta di un quadro definito e visibile, che si presta a schemi di *compliance* e minimizzazione del rischio come strumenti *ex ante*, anche perché i diritti sono stati elaborati con questo preciso obiettivo.

[57] In tema cfr. F. Ufert, *AI Regulation through the lens of Fundamental Rights: How Well Does the GDPR Address the Challenge Posed by AI*, in «European Papers», 2020, pp. 1087 ss.

[58] Per una utile individuazione del, comunque critico, rapporto fra rappresentazione del rischio e *data protection* cfr. M. Padden e A. Ojehag-Petterson, *Protected how? Problem representations of risk in the General Data Protection Regulation (GDPR)*, in «Critical Policy Studies», 2021, pp. 486 ss.

Nel caso dei diritti fondamentali applicati alla IA, questo aspetto non è stato altrettanto sviluppato, e ci si è limitati a saldare l'approccio basato sul rischio con la perseguita visione «antropocentrica» della IA, predicata come sufficiente a valorizzare i diritti compendiati nella Carta dei diritti fondamentali in relazione all'impiego della IA. Una scelta discutibile, se si considera l'efficacia profondamente trasformativa che la tecnologia ha sui diritti e la presenza, come si è visto, di un pericolo di discriminazione che si presenta come specificamente strutturale rispetto alle decisioni algoritmiche.

L'assenza di un riferimento esplicito a questo elemento nella ossatura della proposta è una lacuna grave, probabilmente frutto di una fideistica convinzione di poter gestire anche tale delicatissimo aspetto secondo una logica *ex ante*; o magari demandandolo ad una logica *ex post* esterna al Regolamento che rischierebbe però di essere basata sui soli e tradizionali strumenti del diritto antidiscriminatorio. Su questo punto, pertanto, la proposta resta pericolosamente sospesa e silente.

Un secondo profilo riguarda le soglie di rischio previste dalla proposta. È chiaro, su un piano generale, che la determinazione delle soglie di rischio da cui discendono i diversi regimi compendiati nella stessa è scelta che esprime una discrezionalità politica[59]. Da questo punto di vista, è lineare e condivisibile la decisione di individuare un nucleo di ambiti nei quali la soglia di rischio di violazione dei diritti fondamentali per uso di forme di intelligenza artificiale è ritenuta a priori troppo alta e dunque non accettabile, con conseguente divieto di impiego della medesima. Al di fuori di questo ristretto ambito, tuttavia, la determinazione delle altre soglie di rischio avviene come categorizzazione «in

[59] In tema, come ben osserva Graziadei, *La regolazione del rischio*, cit., p. 504, la regolazione del rischio come funzione dell'attività di governo è condizionata principalmente da vincoli attinenti al livello delle conoscenze scientifiche disponibili e al contempo all'appetito del pubblico per il rischio legato a determinate decisioni, la quale a sua volta dipende dalla propensione al rischio manifestata da una determinata popolazione, e dalla percezione del rischio diffusa in quella popolazione.

bianco» affidata alla estrema mobilità di un allegato e in cui la stessa scelta di discrezionalità politica sembra finire per dissolversi nella prospettiva velleitaria di una gestione prevalentemente *ex ante* del delicato ambito dell'alto rischio. Scelta che sembra trascurare la circostanza, che ben si è evidenziata sopra, che le forme di intelligenza artificiale sono strutturalmente mobili, in ragione della doppia circostanza della costante variabilità dei dati utilizzati e della capacità di apprendere ed evolvere autonomamente proprio in ragione dei medesimi.

A ben vedere, proprio a fronte di questa mobilità del sostrato tecnologico su cui la valutazione del rischio si innesta, si sarebbe potuto ricorrere ad una qualche formulazione del principio precauzionale, che è 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[60].

A ben vedere, l'unico profilo che appare allineato ad un approccio sostanzialmente, anche se non formalmente, precauzionale è proprio quello relativo alla proibizione delle applicazioni ad altissimo rischio, in relazione alle quali vi è stata a priori una considerazione di non accettabilità della soglia di rischio di violazione di diritti fondamentali. Anche sotto questo profilo emerge pertanto una sostanziale incompiutezza dell'approccio basato sul rischio.

In conclusione, dall'analisi svolta si trae la sensazione che la proposta, al di là della sua dichiarata ispirazione ad una logica antropocentrica, limitata peraltro a un troppo generico riferimento alla tutela dei diritti, presenti alcuni limiti di efficacia per la tutela dei diritti fondamentali proprio a

[60] Come è noto il principio di precauzione trova applicazione in condizioni in cui le informazioni scientifiche sono insufficienti, non conclusive o incerte e il decisore è chiamato adottare misure volte ad abbattere o a contenere il rischio di conseguenze avverse. In tema cfr. la sempre lucida e ampia analisi contenuta in L. Gradoni, *Il principio di precauzione nel diritto dell'Organizzazione mondiale del commercio*, in A. Bianchi e M. Gestri (a cura di), *Il principio precauzionale nel diritto internazionale e comunitario*, Milano, 2006, pp. 147 ss.

causa della dimostrata incompiutezza dell'approccio basato sul rischio e della sua non sufficiente considerazione diretta del pericolo di discriminazione algoritmica. Ne discende che l'essenza di tale protezione si esplicherà presumibilmente su piani differenti, e in particolare su quello della responsabilità civile, che resta una elaborazione normativa pendente, nella faticosa individuazione di un regime di tutela piena ed efficace che richiederebbe almeno un più diretto raccordo con la disciplina che si è qui analizzata[61].

Su un piano ulteriore, infine, ci si può chiedere se l'apparato regolatorio complesso e di non agevole gestione che viene delineato nella Proposta di Regolamento non sia foriero di inefficienza nella prospettiva dello sviluppo del mercato della IA, che pure è da sempre stato suo movente altrettanto rilevante rispetto alla tutela dei diritti. Deve in questa fase attentamente valutarsi se la normativa, una volta in atto senza sensibili correttivi, non finisca per costituire un aggravio insostenibile per i produttori di IA europea, e in particolare quelli di taglia media o piccola, per questo meno attrezzati ad assolvere al complesso meccanismo di conformità e certificazione previsto, e rispondere alla esigenza di una costante messa a giorno che ricorda in certa misura la fatica di Sisifo, stante la continua mutevolezza del sostrato tecnologico.

La sfida regolatoria della IA si scontra quindi con difficoltà rilevanti, e questo in certa misura si è riflesso, sino ad oggi, nella preferenza, cui si faceva cenno in apertura, per strumenti etici o di *governance* meno stringenti[62]. L'UE rivendica ora a sé, nel solco di quanto già fatto in tema di *data protection*, il ruolo, anche di valenza geopolitica, di *first mover* nel campo della regolazione vincolante, circostanza certo meritoria ma che indiscutibilmente la espone a rischi

[61] In tema cfr. E. Marchisio, *In support of «no-fault» civil liability rules for artificial intelligence*, in «Springer Nature Social Sciences», 2021, pp. 1-54.

[62] In tema M.C. De Vivo, *Digital Humanism between Ethics, Law and New Technologies*, in A. Caligiuri (a cura di), *Legal Technology Transformation. A practical assessment*, Napoli, 2020, pp. 65 ss.

evidenti di inefficacia e inefficienza. Ciò è particolarmente vero perché l'azione ha l'ambizione di servire al contempo l'eccellenza e competitività del mercato europeo relativo alla IA e la fiducia fondata sull'antropocentrismo come volano di protezione dei diritti umani. Il rischio, come spesso avviene per gli strumenti che si intendano funzionali a obiettivi non facilmente riconciliabili, è che l'esito applicativo si ritorca contro il suo stesso autore e finisca per servire interessi ancora ulteriori e differenti, accrescendo il divario con realtà produttive più competitive perché meno normativamente ipertrofiche, come quelle di estrazione statunitense o cinese.