Intelligenza artificiale: il rischio come opportunità

Artificial intelligence: risk as opportunity

Sara Tommasi*

Abstract

The word intelligence makes us think of human brain functions and recalling it for non-human machines is not adequate.

When we talk about artificial intelligence we refer to an automated decision-making process and that is to the fact that a user initially delegating a decision, partly or completely, to an entity by way of using software or a service; whereas that entity then in turn uses automatically executed decision-making models to perform an action on behalf of a user, or to inform the user's decisions in performing an action. This can be deduced from the European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014 (INL), which proposes to use the term "automated decision-making" rather than artificial intelligence, precisely to avoid the ambiguities of the latter expression. What has been said presents us with a difference: that of a decision delegated in part and that of a decision entirely delegated to entities that use software. As long as it can be said that the decision is not totally attributable to an artificial intelligence system - even if only because there are margins for human oversight - we can speak of risk. If, on the other hand, man cannot participate, albeit within the said limits, in the decision, then we must ask ourselves whether it is more appropriate to speak of danger. In this way, it is possible to verify that the risk is confirmed as an opportunity

1 Premessa

L'analisi delle recenti proposte di regolamentazione sull'Intelligenza artificiale a livello europeo delineano un approccio basato sul rischio.

La riflessione sui limiti di tale approccio e sui vantaggi che con lo stesso si intendono perseguire necessita preliminarmente di chiedersi se sia corretto parlare di intelligenza artificiale, che cosa si intenda per rischio e se il rischio sia aggettivabile e distinguibile in rischio inaccettabile, sistemico alto, limitato e minimo, come nelle proposte normative dell'Unione europea normative europee[1].

* Professore associato di Diritto privato presso l'Università del Salento. Ha conseguito l'abilitazione alle funzioni di I fascia in Diritto privato e in Diritto dell'economia, dei mercati finanziari e agroalimentari e della navigazione. Ha svolto attività di ricerca presso il Max Planck Institut für europäische Rechtsgeschichte di Francoforte e di docenza nella Facultad de Derecho dell'Universidad di Granada e nella Facultad de Ciencias Jurídicas dell'Universitat Rovira i Virgili di Tarragona. È autrice di opere monografiche (L'attività e le fonti delle obbligazioni, Pensa MultiMedia, Lecce, 2003; Pratiche commerciali scorrette e disciplina dell'attività negoziale, Cacucci, Bari, 2012; La tutela del consumatore nei contratti di credito immobiliare,ESI, Napoli 2018), di numerosi contributi in volumi collettanei e di saggi pubblicati in riviste italiane e straniere. È Delegata agli Accordi Istituzionali e partenariali, nonché alla concessione dei patrocini.

[1] Da ultima si veda la Proposta di Regolamento del 21 Aprile 2021, in https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELLAR%3Ae0649735-a372-11eb-9585-01aa75ed71a1. Ivi si rinvia per i riferimenti ai precedenti interventi europei sul tema, tra i quali si segnalano COM (2018) 237 final, La nuova agenda per le competenze per l'Europa; COM (2018) 795 final, *Piano coordinato sull'intelligenza artificiale; COM (2019) 168 final, Building Trust in Human-Centric Artificial Intelligence;* le Risoluzione del Parlamento europeo del 20 ottobre 2020: Legge sui servizi digitali: migliorare il funzionamento del mercato unico; Legge sui servizi digitali: adeguare le norme di diritto commerciale e civile per i soggetti commerciali che operano online; Atto sui servizi digitali e questioni sollevate in materia di diritti fondamentali; Quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate; Regime di responsabilità civile per l'intelligenza artificiale; Diritti di proprietà intellettuale per lo sviluppo di tecnologie di intelligenza artificiale, in https://www.europarl.europa.eu/doceo/document/TA-9-2020-10-20_IT.html. Cfr. S. Garreffa, La risoluzione del Parlamento europeo del 20 ottobre 2020 sul regime di responsabilità civile per l'intelligenza artificiale, in Persona e Mercato, 2020, Osservatorio OGID, p. 502; A. F. Uricchio, G. Riccio, U. Ruffolo (a cura di), Intelligenza Artificiale tra etica e diritti. Prime riflessioni a seguito del libro bianco dell'Unione europea, Bari, 2020; E. Calzolaio, Introduzione, in E. Calzolaio, La decisione nel prima dell'intelligenza artificiale, Milano, 2020, p. 1; A. Alpini, Sull'approccio umano-centrico all'intelligenza artificiale. Riflessioni a margine del "Progetto europeo di orientamenti etici per una IA affidabile", in www.comparazionedirittocivile.it; S. Tommasi, L'intelligenza artificiale antropocentrica: limiti e opportunità, in Juscivile, 2020, p. 853 ss.

2 Intelligenza artificiale o processo decisionale automatizzato?

La risposta alla prima domanda posta, e cioè se sia corretto parlare di intelligenza artificiale è negativa. La parola intelligenza ci fa pensare alle funzioni celebrali dell'uomo e richiamarla per le macchine non umane potrebbe significare evocare delle capacità umane[2].

Quando parliamo di intelligenza artificiale facciamo riferimento ad un processo decisionale automatizzato e cioè al fatto che un utente deleghi inizialmente una decisione, in parte o interamente, a un'entità utilizzando un *software* o un servizio; tale entità a sua volta utilizza modelli decisionali automatizzati per lo svolgimento di un'azione per conto di un utente. Lo si deduce dalla Risoluzione del Parlamento europeo del 20 ottobre 2020 su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL), che propone di utilizzare il termine "processo decisionale automatizzato" piuttosto che quello di Intelligenza artificiale, proprio per evitare le ambiguità di tale ultima espressione[3].

La prima riflessione ci pone subito davanti a una differenza: quella di una decisione delegata in parte e quella di una decisione delegata interamente a entità che utilizzano un *software*.

In entrambi i casi, ci troviamo di fronte non solo ad un rischio legato al possibile malfunzionamento del sistema di IA, cioè ad un rischio causale, ma ad rischio che può prescindere dal malfunzionamento del sistema ed essere legato proprio alle capacità decisionali del sistema. Il rischio causale può riguardare qualsiasi prodotto, e non è quello che qui ci interessa.

È il rischio legato alle capacità decisioni autonome dell'intelligenza artificiale sul quale occorre soffermarsi[4].

---

[2] Cfr. I. Giuffrida, F. Lederer E N. Vermerys, *A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law, in Case Western Reserve Law Review*, 2018, p. 750 ss., ove si afferma che ‹‹words like "language", "memory", "understand", "instruction", "read", "write", "command", and many others are in constant use. They are words which, in their primary meaning, have reference to cognitive beings. Computers are not cognitive››; J.R. Searle, *"Minds, brains and programs"*, in *Behavioral and Brain Sciences*, 1980, p. 417 ss, disponibile in: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/minds-brains-and programs/DC644B47A4299C637C89772FACC2706A; Id., *Menti, cervelli e programmi: Un dibattito sull' intelligenza artificiale*, trad. it. a cura di G. Tonfoni, Milano, 1984. Sulla circostanza che ‹‹sotto il profilo metodologico non è possibile costruire una soggettività, in termini strutturalistici, in forma unitaria, ma è d'obbligo intenderla in forma plurima, distinguendo i problemi della persona umana dai problemi di quei soggetti (sia pure diversificati per scopi e funzioni), che persone umane non sono: le cc.dd. persone giuridiche e ogni altro centro di imputazione soggettiva››, si rimanda a P. Perlingieri, *Il diritto civile nella legalità costituzionale secondo il sistema italo-comunitario delle fonti,* Napoli, 2006, p. 6 ss. Cfr. Id., *La personalità umana nell'ordinamento giuridico,* Napoli, 1972. Sui "controdiritti trans-soggettivi" come "strutture a fianco dei diritti soggettivi tradizionali" si veda P. Femia, *Il civile senso dell'autonomia,* cit., p. 9. Sull'idea di personalità/soggettività differenziate ‹‹come il frutto di un cauto processo di liberazione di nuove soggettività col loro accesso al mondo del giuridicamente rilevante quali centri d'imputazione di interessi protetti, nel che si risolve peraltro la titolarità di diritti soggettivi››, si rimanda a P.L. Portaluri, *L'articolazione delle figure soggettive: dalla personalità alla soggettività giuridica,* in www.ridiam.it, 2019, spec. p. 5. Cfr. R. Míguez Núñez, *Soggettivizzare la natura*? in *The Cardozo Electronic Law Bulletin*, 2019, p. 1 ss.; Id., *Le avventure del soggetto. Contributo teorico-comparativo sulle nuove forme di soggettività giuridica*, Milano-Udine, 2019, *passim.* Sul punto inevitabile il rimando a A. Falzea, *Il soggetto nel sistema dei fenomeni giuridici*, Milano, 1939, spec. p. 54; F. Alcaro, *Riflessioni critiche intorno alla soggettività giuridica. Significato di un'evoluzione*, Milano, 1976. Critica la tendenza ad «antropomorfizzare il fenomeno e a narrarlo come se le intelligenze artificiali coinvolte fossero emanazione di soggetti umani» G. Finocchiaro, *Il contratto nell'era dell'intelligenza artificiale*, in *Riv. trim. dir. proc. civ.*, 2018, p. 441; Cfr. G. Sartor, *Gli agenti software: nuovi soggetti del ciberdiritto?*, in *Contr. impr.*, 2002, p. 465. Indagano sulla fonte della diversità tra i paradigmi di operatività dei motori decisionali umani e quelli automatici, affermando che «bisogna riconoscere che i sistemi di AI flettono (nel vuoto) muscoli sintattici, ma non ragionano come la mente umana; non possono al momento (e forse non potranno mai) acquisire la capacità di saldare il piano semantico e quello sintattico che caratterizza la fluidità sintetica del ragionamento umano», R. Pardolesi, A. Davola, *Algorithmic legal decision making: la fine del mondo (del diritto) o il paese delle meraviglie?*, in «*Questione giustizia*», 2020, spec. 108. Sulla circostanza che i non umani siano in grado di entrare in comunicazione con gli umani cfr. B. Latour, *Politics of Nature: How to Bring the Sciences into Democracy*, Harvard, 2004, p. 71. In questa prospettiva tale comunicazione può avvenire in diversi modi. Il più semplice è quello in cui l'algoritmo è solo uno strumento del quale l'uomo si serve, ma ci può anche essere una sorta di interoperatività tra uomo e macchina, tale da potersi parlare di un ibrido. Sul punto cfr. G. Teubner, *Rights of Non-Humans? Electronic Agents and Animals as New Actors in Politics and Law*, in *Journal of Law and Society*, 2006, p. 500, ove si confronta il pensiero di N. Luhmann con quello di Latour; Id., *Ibridi e attanti. Attori collettivi ed enti non umani nella società e nel diritto,* trad. di L. Zampino, Milano-Udine, 2015, p. 21 ss. Sulla comunicazione come operazione che è fornita della capacità di autoosservarsi e, dunque, a parteciparvi possono essere soltanto i sistemi psichici, si rimanda a R. De Giorgi, N. Luhmann, *Teoria della società*, Milano, 2013 (1991), p. 27.

[3] Il testo della Risoluzione è rinvenibile in https://www.europarl.europa.eu/doceo/document/TA-9-2020-10-20_IT.html#sdocta9.

[4] Sul rischio decisionale come rischio di genere del tutto diverso rispetto a quello causale v. P. Femia (a cura di), *Gunther Teubner, Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, Napoli 2018, p. 117.

Fino a quando può dirsi che la decisione non è totalmente riconducibile al sistema- anche solo perché ci sono dei margini per l'*Human oversight* - può parlarsi di rischio[5].

Se invece questi margini non ci sono, se l'uomo non può partecipare, se pure nei limiti detti, alla decisione, allora dobbiamo chiederci se non sia più giusto parlare di pericolo.

Quanto osservato non può che confermare che il rischio non va visto in senso minaccioso e apocalittico, come nell'impostazione che lo contrappone alla sicurezza e lo vede come rottura di un ordine che, altrimenti, continuerebbe a sussistere.

Il rischio non è un dato, non è una sorta di realtà sotterranea che scorre occulta al di sotto della realtà che si produce con l'agire, né l'orizzonte lungo il quale valutare la rischiosità del rischio è la sicurezza, cioè «una condizione artificiale di stabilità e di certezza che si assume come razionale»[6].

Il rischio, in altri termini, è la possibilità di un evento dannoso che un'altra decisione avrebbe potuto evitare. La società moderna è società del rischio nel senso che «ha realizzato condizioni che le permettono di costruire futuri differenti, di mantenere alta la contingenza degli eventi, cioè di tenere aperte sempre più possibilità e, quando in conseguenza di una decisione si verifica un eventuale danno che si sarebbe voluto evitare, di sapere che un'altra decisione avrebbe potuto evitarlo»[7].

L'alternativa al rischio, dunque, non è la sicurezza ma il pericolo[8]. Tanto è vero che, quanto più si incrementano le misure di sicurezza più si incrementano i rischi. Basti pensare che i «sistemi di sicurezza costituiti da macchine controllate da macchine controllate da macchine moltiplicano al loro interno i rischi del controllo dei controllori»[9].

Il rischio è un'opportunità in quanto realizza condizioni che permettono di costruire futuri differenti e consente ai sistemi sociali di adattarsi alla complessità del loro ambiente.

Il rischio tiene aperte sempre più possibilità e mantiene alta la contingenza degli eventi.

Quello che rende rischiosa, nel senso positivo, la decisione di un processo decisionale automatizzato è la possibilità umana di partecipare alla decisione, quanto meno nella forma dell '*Human oversight*. Non si tratta di una chimera alla quale la tecnica deve imporci di rinunciare. A questo aspetto sono dedicate le osservazioni che seguono.

3 Intelligenza artificiale e *Human oversight*

La consapevolezza dell'importante ruolo dell' *Human oversight* emerge nella Proposta di Regolamento del 21 Aprile 2021, COM(2021) 206 final ove, adottando una formulazione diversa da quella della citata Risoluzione su un regime di responsabilità civile per l'intelligenza artificiale, si definisce sistema di intelligenza artificiale un *software* che, per un dato insieme di obiettivi definiti dall'uomo, genera risultati come contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagisce[10].

5 Sulla differenza tra rischio e pericolo si rimanda a N. Luhmann, Sociologia del rischio, Milano, 1996; R. De Giorgi, O risco na sociedade contemporânea, in Seqüência 21 anos, 1994, p. 45-54; Id., I*l rischio nella società del rischio, in Temi di filosofia del diritto*, vol. II, Lecce, 2015, p. 69 ss.

6 R. De Giorgi, *Il rischio nella società contemporanea,* in *Temi di filosofia del diritto*, cit., p. 61.

7 R. De Giorgi, *Il rischio nella società contemporanea,* in *Temi di filosofia del diritto*, cit., p. 62.

8 Sulla distinzione tra rischio e pericolo si rimanda a N. Luhmann, *Sociologia del rischio,* cit., p. 14 ss., ove si afferma che rispetto all'incertezza in riferimento ai danni futuri, ci sono due possibilità: o l'eventuale danno viene visto come conseguenza della decisione, cioè viene attribuito ad essa e parliamo allora di rischio, per la precisione di rischio della decisione; oppure si pensa che l'eventuale danno sia dovuto a fattori esterni e viene quindi attribuito all'ambiente: parliamo ora di pericolo.

9 R. De Giorgi, *Il rischio nella società contemporanea*, in *Temi di filosofia del diritto*, cit., p. 58.

10 Una definizione di Intelligenza artificiale si rinviene anche nella *Proposition de loi constitutionnelle relative à la Charte de l'intelligence artificielle et des algorithmes*, depositata il 15 gennaio 2020, in Francia, presso *l'Assemblée Nationale*, ove si legge che «la présente charte s'applique à tout système qui se compose d'une entité qu'elle soit physique (par exemple un robot) ou virtuelle (par exemple un algorithme) et qui utilise de l'intelligence artificielle. La notion d'intelligence artificielle est entendue ici comme un algorithme évolutif dans sa structure, apprenant, au regard de sa rédaction initiale. Un système tel que défini au précédent alinéa n'est pas doté de la personnalité juridique et par conséquent inapte à être titulaire de droits subjectifs. Cependant les obligations qui découlent de la personnalité juridique incombent à la personne morale ou physique qui héberge ou distribue ledit système devenant de fait son représentant juridique».

Il riferimento agli obiettivi definiti dall' uomo ne valorizza il ruolo a fronte di un percorso che era partito dall'idea di riconoscere la personalità giuridica a questi sistemi, proprio al fine di segnarne l'autonomia dall'uomo. Si pensi alla Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica[11]. Posizione che l'Unione europea ha progressivamente abbandonato. Nella citata Risoluzione su un regime di responsabilità civile per l'intelligenza artificiale, infatti, già è esplicitamente previsto che non è necessario conferire personalità giuridica ai sistemi di IA.

Non a caso, l'*Human oversight* è uno dei sette requisiti fondamentali per un'IA affidabile. Lo si evince già dalle comunicazione della Commissione Europea del 2019[12], ma il dato è reso più esplicito nella Risoluzione del Parlamento europeo del 20 ottobre 2020, sulla legge sui servizi digitali[13], che si pone lungo il percorso che porta al cd. *Digital services act (Dsa)* e al *Digital market act (Dma)*[14], e prevede espressamente che occorra rispettare il principio del controllo dell'uomo sulla macchina per prevenire l'aumento dei rischi per la salute, la sicurezza, la discriminazione, l'indebita sorveglianza, gli abusi e le potenziali minacce ai diritti e alle libertà fondamentali[15]. Conferme se ne traggano, altresì, dalla Risoluzione recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL)[16] che contiene una proposta di Regolamento, al cui art. 7, si propone un approccio all'intelligenza artificiale in base al quale le tecnologie dell'intelligenza artificiale ad alto rischio devono essere sviluppate, diffuse e utilizzate in modo da garantire il pieno controllo umano, in qualsiasi momento e in modo da consentire, ove necessario, la ripresa del pieno controllo umano, anche mediante la loro alterazione o disattivazione.

La Proposta di Regolamento del 21 Aprile 2021, all'art. 14, è esplicita nel riconoscere un ruolo centrale all' *Human oversight*. Si prevede, infatti, per alcuni high-risk AI systems, l'insufficienza anche del coinvolgimento di una sola persona umana, esigendosi che ‹‹ no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons››.

L'art. 14 è dettagliato nei riferimenti alla sorveglianza umana al fine di rendere possibile quanto si legge nel Considerando 48) della Proposta in oggetto e cioè che ‹‹High-risk AI systems should be designed and developed in such a way that natural persons can oversee their functioning. For this purpose, appropriate human oversight measures should be identified by the provider of the system before its placing on the market or putting into service. In particular, where appropriate, such measures should guarantee that the system is subject to in-built operational constraints that cannot be overridden by the system itself and is responsive to the human operator››.

Le misure legate alla sorveglianza umana sono tra quelle che devono essere comunicate ai fini di assolvere agli obblighi di informazione necessari per assicurare la trasparenza dei sistemi di intelligenza artificiale. È quanto emerge dall'art. 13, che è di particolare interesse per il richiamo ivi esplicitato alla necessità che gli *outputs* dei sistemi di IA siano interpretati e che agli utenti siano fornite tutte le informazioni necessarie per facilitare tale interpretazione.

---

[11] In https://eur-lex.europa.eu/legal-content/IT/ALL/?uri=CELEX%3A52017IP0051.

[12] Il riferimento è a COM(2019) 168 final, ove si afferma che la sorveglianza umana sui sistemi di IA può essere effettuata mediante diversi meccanismi che si possono sostanziare nel cd. "*human-in-the-loop*", ossia in un intervento umano in ogni ciclo decisionale del sistema; nel cd. "*human-on-the-loop*", cioè in un intervento umano durante il ciclo di progettazione del sistema e di monitoraggio del funzionamento del sistema, oppure in un controllo umano, cd. "*human-in-command*", che presuppone la capacità di sorvegliare l'attività complessiva del sistema di IA, definendo i livelli di discrezionalità umana e garantendo la possibilità di annullare una decisione adottata.

[13] *Risoluzione del Parlamento europeo del 20 ottobre 2020, recante raccomandazioni alla Commissione sulla legge sui servizi digitali: migliorare il funzionamento del mercato unico (2020/2018(INL)) in https://www.europarl.europa.eu/doceo/document/TA-9-2020-0272_IT.pdf.*

[14] Si tratta, quanto al *Digital services act (Dsa)* di COM(2020) 825 final, in *https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital;* Quanto al *Digital market act (Dma)* di COM(2020) 842 final, in *https://ec.europa.eu/info/sites/info/files/proposal-regulation-single-market-digital-services-digital-services-act_en.pdf.*

[15] Punto 41) della Risoluzione recante raccomandazioni alla Commissione sulla legge sui servizi digitali, cit.

[16] In https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_IT.html. In particolare al punto 10), si legge che «le decisioni prese o influenzate dall'intelligenza artificiale, dalla robotica e dalle tecnologie correlate dovrebbero rimanere soggette a un riesame, una valutazione, un intervento e un controllo significativi da parte dell'uomo. La complessità tecnica e operativa di tali tecnologie non dovrebbe mai impedire all'operatore o all'utente di poter, come minimo, disattivarle in modo sicuro, alterare o arrestare le loro operazioni ovvero tornare a uno stato precedente per ripristinare funzionalità sicure nei casi in cui sia a rischio il rispetto del diritto dell'Unione, dei principi etici e degli obblighi giuridici stabiliti nel presente regolamento».

4 Il rischio come opportunità: la possibilità di partecipazione umana alla decisione

Il ruolo certamente non secondario riconosciuto all'*human oversight* evidenzia i limiti di una impostazione che richiama l'opacità dei processi decisionali algoritmi come impeditiva, sempre, della partecipazione dell'uomo alla decisione[17].

Nello specifico, se l'opacità è assoluta, nel senso che impedisce ogni forma di comprensione del meccanismo della decisione e ogni tipo di intervento umano, allora non c'è possibilità di partecipazione umana alla decisione e, dunque, occorre chiedersi se ci si esponga ad un pericolo più che al rischio.

Se invece sono possibili meccanismi di trasparenza, l'opacità è un rischio dei processi decisionali algoritmici. Come dire che a rendere rischiosa l'Ia è la possibilità dell' uomo 1) di capire come il sistema prende le decisioni algoritmiche - anche se non è necessario comprendere ogni singola fase del processo decisionale; 2) di identificare i sistemi di IA come tali, e cioè che gli utenti sappiano che stanno interagendo con un sistema di IA[18].

Si potrebbe obiettare che i destinatari della decisione di IA, si trovano sempre in una situazione di pericolo rispetto ad una decisione che non prendono loro, sia pure se sottoposta ad una sorveglianza umana. Ciò è vero fino ad un certo punto in questo caso. È vero rispetto alla componente umana della decisione, ma non rispetto all'umanità di quella componente.

È vero rispetto alla componente umana perché è un altro uomo che sorveglia il sistema di Intelligenza artificiale e partecipa alla decisione. Non sono io che partecipo e, dunque, rispetto a questo, la situazione è di pericolo.

Non è vero rispetto all'umanità della componente umana, perché è quella umanità che consente di non affidare la decisione esclusivamente alla macchina, ma anche a quelle valutazioni che solo proprie solo dell'uomo.

In altri termini, non partecipo alla decisione come singolo, ma come umanità e consento all'umanità di avere a che fare con sistemi di IA rischiosi, anche molto, ma non pericolosi. Il rischio, dunque, si conferma un'opportunità.

5 Le aggettivazioni del rischio

A fronte dell'ampia gamma delle possibili applicazioni dei sistemi di IA, e nella consapevolezza che il vasto *corpus* di norme europee, sia pure in linea di principio pienamente applicabile all'IA, non copre tutti i nuovi rischi derivanti dalle tecnologie digitali emergenti, l'Unione Europea non rinuncia ad un quadro giuridico armonizzato, senza trascurare che tipi diversi di rischi necessitano di tipi diversi di controlli.

In questa prospettiva si spiega l'approccio basato sulla distinzione tra sistemi di IA ad alto rischio e sistemi che non rientrano in tale definizione.

La Commissione Europea ritiene che un'applicazione di IA dovrebbe essere considerata ad alto rischio se soddisfa due criteri cumulativi: 1) è utilizzata in un settore in cui, date le caratteristiche delle attività abitualmente svolte, si possono prevedere rischi significativi; 2) è utilizzata in modo tale da poter generare rischi significativi. Questo secondo criterio riconosce il fatto che non tutti gli usi dell'IA in dati settori comportano necessariamente

[17] Per ridurre le conseguenze legate all'opacità dei sistemi di intelligenza artificiale, la Commissione Europea ritiene necessario prevedere requisiti di trasparenza degli algoritmi, particolarmente importanti anche per il meccanismo *ex post* di controllo del rispetto della normativa e per rafforzare la fiducia nell'utilizzo di tali strumenti. L'importanza della trasparenza dell'algoritmo può dedursi anche dal cd. "*duty of explanation*" delle *Guidelines for Trustworthy Artificial Intelligence* dell'8 Aprile 2019, ove si fa rifermento alla finalità di rendere comprensibili non solo i criteri decisionali degli agenti algoritmici basati sull'intelligenza artificiale, ma anche i loro scopi e obiettivi. Su questo aspetto cfr.F . Z. Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence,* in *The International Journal of Human Rights,* 2020, p. 9, ove si legge che «even when assuming that people have a right to explanation regarding algorithmic decisions, it is often difficult, if not impossible, to explain the logic behind a decision, when an algorithmic system arrives at that decision after analysing large amounts of data. Moreover, in some circumstances, an explanation might not be of much help to people». Cfr. G. Alpa (a cura di), *Diritto* e *intelligenza artificiale*, Pisa, 2020; G. Alpa, *Fintech: un laboratorio per i giuristi,* in *Contr. impr.*, 2019, p. 377 ss; A. Longo, G. Scorza, *Intelligenza artificiale. L'impatto sulle nostre vite, diritti e libertà,* Milano, 2020; A. Santosuosso *Intelligenza artificiale e diritto: perché le tecnologie di IA sono una grande opportunità per il diritto*, Milano, 2020; U. Ruffolo (a cura di), *Intelligenza Artificiale. Il diritto, i diritti, l'etica,* Milano, 2020, *passim*; E. Gabrielli e U. Ruffolo (a cura di), *Intelligenza Artificiale e diritto*, in *Giur. it.*, 2019, p. 1657 ss.;

[18] G. Resta, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Politica del diritto*, 2019, p. 215 ss.

rischi significativi e che possono esistere anche casi in cui l'uso di applicazioni di IA per determinati scopi deve essere considerato ad alto rischio di per sé, ossia indipendentemente dal settore interessato[19].

La differenza tra sistemi di IA ad alto rischio e non ad alto rischio è alla base anche della citata Risoluzione su un regime di responsabilità civile per l'intelligenza artificiale, ove è contenuta una Proposta di Regolamento, il cui art. 3 prevede che, per alto rischio si intende ‹‹un potenziale significativo in un sistema di IA che opera in modo autonomo, di causare danni o pregiudizi a una o più persone in modo casuale e che va oltre quanto ci si possa ragionevolmente aspettare; l'importanza del potenziale dipende dall'interazione tra la gravità dei possibili danni o pregiudizi, dal grado di autonomia decisionale, dalla probabilità che il rischio si concretizzi e dalla modalità e dal contesto di utilizzo del sistema di IA››. Tale proposta invita a riflettere sull'opportunità di distinguere tra sistemi di Ia rischiosi e sistemi di IA pericolosi. Per questi ultimi, infatti, la Risoluzione prevede una responsabilità oggettiva tanto che l'operatore non può evitare di essere ritenuto responsabile sostenendo di aver agito con la dovuta diligenza o che il danno o il pregiudizio sia stato cagionato da un'attività, dispositivo o processo autonomo guidato dal sistema di IA[20]. Gli operatori non sono considerati responsabili soltanto se il danno o il pregiudizio è dovuto a cause di forza maggiore[21].

Invece, quando c'è la possibilità umana di incidere sulla decisione, quanto meno attivando degli strumenti di controllo, monitorando le attività e mantenendo l'affidabilità operativa mediante la periodica installazione di tutti gli aggiornamenti disponibili, si prevede una presunzione di colpa dell'operatore di sistema che non sarà responsabile se riesce a dimostrare che il danno o il pregiudizio arrecato non è imputabile a sua colpa[22]. Laddove il danno o il pregiudizio sia stato causato da un terzo che abbia interferito con il sistema di IA attraverso la modifica del suo funzionamento o dei suoi effetti, l'operatore è comunque tenuto a corrispondere un risarcimento se tale terzo è irrintracciabile o insolvibile, potendosi giovare solo dell'obbligo di cooperazione del produttore del sistema di IA. Quest'ultimo, infatti, è tenuto a cooperare con l'operatore o con la persona interessata e a fornire informazioni al fine di consentire l'individuazione delle responsabilità.

La Proposta di Regolamento del 21 Aprile 2021, COM(2021) 206 final introduce la distinzione tra AI that create (i) an unacceptable risk, (ii) a high risk, and (iii) low or minimal risk.

Il rischio inaccettabile è legato a sistemi di IA considerati una minaccia per la sicurezza, i mezzi di sussistenza e i diritti delle persone. L'elenco delle pratiche vietate nel titolo II comprende sistemi di IA il cui uso è considerato in contrasto con i valori dell'Unione. I divieti riguardano pratiche che possono, in modo significativo, manipolare le persone attraverso tecniche subliminali al di là della loro coscienza o sfruttare la vulnerabilità di specifici grupp come bambini o persone con disabilità, al fine di distorcere il loro comportamento in un modo che potrebbe causare loro danni psicologici o fisici.

[19] È quanto si legge in COM (2020) 65 final, cit., p. 19. Ivi, in particolare ci si muove nella direzione del divieto di rischi considerati dalla comunità socialmente inaccettabili, mentre le attività che non generano nessuno dei rischi rilevanti non dovrebbe ricadere nell' ambito intervento normativo. Tale approccio è stato criticato in quanto può riflette un visione indebitamente semplificata, secondo cui lo scopo della regolamentazione è esclusivamente quello di minimizzare i rischi "più significativi". Sul punto si rimanda a K. Yeung, *Response to European Commission White Paper on Artificial Intelligence*, in https://www.researchgate.net/profile/Karen_Yeung/publication/342199392, 2020, p. 1. L' A. propone ‹‹that the proposed regulatory framework adopt a much more fine-grained scale for the assessment of risk, rather than adopting a binary classification system for determining whether or not a particular ADM/AI application falls within the scope of the regulatory framework or not. In particular, I recommend a 5 point scale: this would provide a sufficient level of granularity for which an accompanying set of proportionately demanding regulatory requirements could be attached, along the following lines: • Level 1: no risk => negligible regulatory requirements • Level 2: low risk => modest, simple regulatory requirements (voluntary labelling?) • Level 3: medium risk => significant, but not unduly onerous regulatory requirements • Level 4: high risk => demanding regulatory safeguards including ex ante approval • Level 5: unacceptable risk => prohibited››.

[20] Sulla diversa allocazione dei rischi nei casi di responsabilità per colpa e responsabilità oggettiva si rimanda a P. Trimarchi, *Rischio e responsabilità oggettiva*, Milano, 1961, p. 13 ove, con riferimento allo sviluppo delle attività industriali, si faceva notare che l'esercizio di tali attività comporta necessariamente una serie di incidenti inevitabili, nonostante l'impego della massima diligenza e, data la complessità dell'organizzazione, la colpa si fraziona, in corrispondenza della divisione del lavoro, in mille piccole quote trascurabili, ciascuna insufficiente a giustificare una responsabilità del suo autore. Id., *La Responsabilità Civile: Atti Illeciti, Rischio, Danno*, Milano, 2019, p. 75 ss e spec. 85.

[21] È quanto si legge nell'art. 4 della Proposta di Regolamento contenuta nella citata Risoluzione del Parlamento europeo su un regime di responsabilità civile per l'intelligenza artificiale.

[22] L'art. 8 della Proposta di Regolamento contenuta nella citata Risoluzione del Parlamento europeo su un regime di responsabilità civile per l'intelligenza artificiale è chiaro sul punto.

I sistemi di intelligenza artificiale identificati come ad alto rischio sono quelli che utilizzano la tecnologia di intelligenza artificiale in: *critical infrastructures*, ossia infrastrutture che potrebbero mettere a rischio la vita e la salute dei cittadini; *educational or vocational training*, ossia tecnologie che possono determinare l'accesso all'istruzione e al percorso professionale della vita di qualcuno; *safety components of products*, cioè componenti di sicurezza dei prodotti, come per esempio applicazioni di AI nella chirurgia assistita da *robot*; *employment, workers management and access to self-employmen,* come *software* di smistamento CV per procedure di assunzione; *essential private and public services*, ossia servizi privati e pubblici essenziali, quali per esempio la valutazione del credito che nega ai cittadini l'opportunità di ottenere un prestito; sistemi di *law enforcemen* che possono interferire con i diritti fondamentali delle persone, come per esempio la valutazione dell'affidabilità delle prove; tecnologie applicate a *migration, asylum and border control management*, quali quelle di verifica dell'autenticità dei documenti di viaggio); sistemi di *administration of justice and democratic processes*.

Per i sistemi di IA che presentano questi tipi di rischi sono previsti rigorosi obblighi prima della loro immissione nel mercato. Per i sistemi cd. *low risk* sono invece previsti pecifici obblighi di trasparenza, in modo da consentire agli utenti di essere consapevoli che stanno interagendo con una macchina e prendere una decisione informata per decidere se continuare oppure no. Il progetto di regolamento non interviene sui sistemi con *minimal risk*, come videogiochi abilitati all'intelligenza artificiale o filtri antispam ritenuti non rischiosi per i diritti o la sicurezza dei cittadini.

Anche nel *Digital services act (Dsa)* si procede con un approccio basato sul rischio e con l'aggettivazione dei rischi. Si delinea anche in questo provvedimento una logica proporzionale e cumulativa nell'imposizione di obblighi, che aumentano e si sommano a mano a mano che i fornitori di servizi di intermediazione siano qualificabili come *hosting*, *piattaforme online* o *very large online platforms*. Il presupposto è che queste ultime, potendo reggiungere più persone, creano dei rischi cd. sistemici e sono tenute sia a rispettare obblighi aggiuntivi con riferimento alla trasparenza e alla pubblicità, ma anche sono tenute, almeno una volta all'anno, ad individuare, analizzare e valutare, *ex* art. 26, eventuali rischi sistemici significativi derivanti dal

funzionamento e dall'uso dei loro servizi nell'Unione, anche sulla base dei valori espressi dalla Carta dei diritti fondamentali dell'Unione europea

Ai sensi dell'art. 26, tali rischi sono riconducibili a tre categorie: a) la diffusione di contenuti illegali tramite i loro servizi; b) eventuali effetti negativi per l'esercizio dei diritti fondamentali al rispetto della vita privata e familiare e alla libertà di espressione e di informazione, del diritto alla non discriminazione e dei diritti del minore, sanciti rispettivamente dagli articoli 7, 11, 21 e 24 della Carta; c) la manipolazione intenzionale del servizio, anche mediante un uso non autentico o uno sfruttamento automatizzato del servizio, con ripercussioni negative, effettive o prevedibili, sulla tutela della salute pubblica, dei minori, del dibattito civico, o con effetti reali o prevedibili sui processi elettorali e sulla sicurezza pubblica.

L'attività di *risk assessment* è propedeutica all'adozione di misure di attenuazione ragionevoli, proporzionate ed efficaci, adattate agli specifici rischi sistemici individuati. I risultati della valutazione di tali rischi e delle relative misure di mitigazione formano oggetto di una relazione che le *very large online platforms* devono trasmettere al Coordinatore dei servizi digitali del luogo di stabilimento e alla Commissione, in adempimento alle obbligazioni di comunicazione trasparente di cui all'art. 33. Le piattaforme *online* di dimensioni molto grandi, inoltre, hanno l'obbligo di sottoporsi, a proprie spese e almeno una volta all'anno, ad *audit*, esterni e indipendenti, volti a verificare la conformità della loro condotta agli obblighi previsti. Ulteriori obblighi sono previsti dall'art. 29 con riferimento ai "sistemi di raccomandazione" (*recommender systems*, definiti nell'art. 2 lett. o). In particolare si prevede l'obbligo di specificare nelle condizioni generali, in modo chiaro, accessibile e facilmente comprensibile, i principali parametri utilizzati, nonché qualunque opzione messa a disposizione dei destinatari del servizio per consentire loro di modificare o influenzare i parametri principali.

Le complessità sono evidenti: non è da escludere che tra i rischi vietati dalla Proposta di Regolamento del 21 Aprile 2021, COM(2021) 206 final ci siano alcuni dei i rischi sistemici ai sensi del *Digital services act*, anche se quest'ultimo non vieta ciò che genera tali rischi, ma piuttosto li considera inevitabili, anche se da attenuare. E ancora

un rischio sistemico potrebbe non essere un alto rischio ai sensi della Proposta di Regolamento del 21 Aprile 2021, COM(2021) 206 final.

6 Conclusioni

Le aggettivazioni del rischio, per di più non accompagnate da una uniforme e chiara definizione di cosa si intenda per rischio, finiscono per lasciare dei nodi irrisolti e per affidarsi troppo alla costruzione di un complesso quadro burocratico costretto a lavorare su labili distinzioni.

Si lega l'alto rischio alla probabilità che si verifichino eventi considerati negativi. In questa prospettiva però si dà rilievo a problemi propri della statistica e non del rischio, eppure la statistica riguarda il passato, mentre il rischio è proiettato nel futuro.

La probabilità presuppone una comparazione e la comparazione ha senso, anche nella statistica, tra dati confrontabili. Il punto è che non sono confrontabili dati che riguardano le generalità con dati che riguardano il singolo, pretendendo, per esempio, che ciò che non è rischioso per gli altri debba anche io sentirlo come non rischioso, ma io non sono una probabilità e il rischio non è una misura.

Non è dunque l'aggettivazione del rischio e la distinzione tra tipi di rischi l'aspetto centrale di un tentativo di disciplina giuridica dell'intelligenza artificiale, ma piuttosto, la rinuncia ad applicare alle macchine le semantiche utilizzate per l'uomo, valutando, con quali differenze rispetto al diritto, i sistemi di intelligenza artificiale ci mettono di fronte al fatto che i vecchi requisiti di stabilizzazione delle aspettative non operano più. Anche i sistemi di intelligenza artificiale spesso sono imprevedibili a se stessi, anche se noi li percepiamo come oggettivamente misurabili, e rendono palese tutta la fragilità della costruzione della razionalità del futuro guardando al passato.

Nella prospettiva della rinuncia ad applicare alle macchine le semantiche utilizzate per l'uomo, si ritiene che l'intelligenza artificiale non debba essere utilizzata per evocare funzioni paragonabili a quelle del cervello umano.

Il riferimento ad un processo decisionale automatizzato piuttosto che all'intelligenza artificiale può creare minori ambiguità, così come la distinzione tra i diversi modi di funzionare di tali processi e di essere sottoposti all'*Human oversight*.