JUS CIVILE

SARA TOMMASI

Professore associato di Diritto privato – Università del Salento

L'INTELLIGENZA ARTIFICIALE ANTROPOCENTRICA: LIMITI E OPPORTUNITÀ

1. – L'intelligenza artificiale è già una realtà che fa parte integrante della nostra vita ed è destinata ad esserlo ancor di più. Le applicazioni sono innumerevoli e gli esempi si moltiplicheranno presto al di là dalla nostra attuale immaginazione, a tal punto che «il modo in cui ci relazioniamo all'IA determinerà il mondo in cui viviamo» [1].

[1] Si veda COM (2018) 237 final, in *www.ec.europa.eu*, p. 2. "*La nuova agenda per le competenze per l'Europa*", in *http://eur-lex.europa.eu. Ivi* si rimanda anche per una prima definizione europea di IA. Si legge, infatti, p. 1, che «Intelligenza artificiale (IA) indica sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi. I sistemi basati sull'IA possono consistere solo in *software* che agiscono nel mondo virtuale (ad esempio assistenti vocali, *software* per l'analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale), oppure incorporare l'IA in dispositivi *hardware* (per esempio in robot avanzati, auto a guida autonoma, droni o applicazioni dell'Internet delle cose)». Una definizione di Intelligenza artificiale si rinviene, da ultimo, 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». Per un primo tentativo di definire l'Intelligenza artificiale si rimanda a J. MARVIN, L. MINSKY, N. ROCHESTER e C.E. SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence:* August 31, 1955, pubblicato di recente in *AI Magazine*, 2006, p. 12, consultabile anche in *www.semanticscholar.org.*

Non pochi sono però gli interrogativi che si pongono di fronte a questa rivoluzionaria trasformazione tecnologica che, se non è possibile né utile arrestare, occorre non subire.

In questa sede, si proverà a dimostrare che gli ambiziosi obiettivi prefissati a livello comunitario possono essere concretamente, e più facilmente, raggiunti se non ci si fa accecare dall'enfasi per le opportunità che queste nuove tecnologie ci offrono, ma si riflette anche sui rischi delle stesse e sui limiti del diritto nell'affrontarli. In particolare, ci si interrogherà sul significato dell'espressione "Intelligenza Artificiale antropocentrica" che non dovrebbe evocare né l'esistenza di "surrogati della persona umana" dotati di medesima intelligenza, né una cd. *anthropocentric view* intesa come contrapposta ad una *ecocentric view*.

Occorre, dunque, prepararsi ai cambiamenti socioeconomici che ci aspettano, cercando di affrontare in modo proattivo e inclusivo i problemi che sorgeranno e che potrebbero accrescere le diseguaglianze [2]. Naturalmente, non si tratta soltanto di acquisire competenze digitali di base o di affidare i problemi posti dai sistemi dell'intelligenza artificiale a singole discipline specifiche, essendo l'IA un fenomeno interdisciplinare. A fronte di tanta complessità, scopo delle osservazioni che seguono è provare a delimitare il terreno su cui si giocano le nuove sfide da affrontare e mettere in guardia sulla necessità di evitare approcci che non consentirebbero l'approdo a conclusioni persuasive.

2. – Con l'espressione "intelligenza artificiale antropocentrica" si possono indicare due cose diverse: la tendenza a costruire tutto "ad immagine e somiglianza" dell'uomo, finendo anche per "umanizzare le macchine" dal punto di vista del linguaggio, dei concetti e delle problematiche connesse [3]; oppure un approccio all'intelligenza artificiale come strumento a servizio soltanto della persona umana [4]. Entrambi questi significati possono essere criticati.

Quanto al primo, può essere utile sgombrare il campo da fraintendimenti ed osservare 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» [5].

---

[2] Si vedano COM (2018) 237 final, cit., p. 12 e il *"Piano d'azione per l'istruzione digitale"*, in *https://eur-lex.europa.eu.*

[3] Cfr. E. CALZOLAIO, *Introduzione*, in E. Calzolaio, *La decisione nel prima dell'intelligenza artificiale,* Milano, 2020, p. 1.

[4] COM (2019) 168 final. "*Creare fiducia nell'intelligenza artificiale antropocentrica*", in *https://eur-lex.europa.eu,* p. 2.

[5] I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations*

L'idea che in questa sede vuole mettersi in discussione è quella dell'intelligenza artificiale come capacità di una macchina di imitare in modo intelligente il comportamento umano; si ritiene, piuttosto, che sarebbe meno forviante un approccio che descriva l'intelligenza artificiale partendo dagli elementi che la caratterizzano e, prima di tutto, dal cd. «*machine learning*», ossia dalla capacità di un *computer* di apprendere dai dati esistenti e di modificare la sua programmazione per tenere conto di nuovi dati, eseguendo modelli c.d. predittivi, che apprendono dai dati esistenti per prevedere comportamenti, risultati, e tendenze [6].

Se partiamo da questo assunto possiamo subito renderci conto che l'apprendimento automatico dipende dai dati. L'algoritmo può "imparare meglio" quanto più sono i dati ai quali può accedere. La qualità di tali dati, il modo in cui vengono immessi nel sistema e come il sistema è "addestrato" sono tutti fattori che possono avere effetti disastrosi sulla validità, accuratezza e utilità delle informazioni generate dall'algoritmo [7]. In altri termini, anche quando i dati sono precisi, l'IA potrebbe diffondere i propri pregiudizi nel sistema.

Pur ad ammettere che con riferimento agli algoritmi si possa parlare di «electronic agents actors» [8], dovremmo convenire che si tratta "di un'altra soggettività" rispetto a

---

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.

[6] Cfr. I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 751. Sulle potenzialità creative degli automi cfr. G. SPEDICATO, *Creatività artificiale, mercato e proprietà intellettuale,* in *Riv. dir. industrial,* 2019, p. 253 e ss.

[7] Significativi gli esempi riportati in I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 754. Ivi si legge che un algoritmo «not only fail to accomplish its set goals but may prove affirmatively harmful. For example, the algorithm used by Google to answer user questions erroneously declared that former president Barack Obama, a Christian, was a Muslim. In that case, the algorithm was not at fault. It simply gathered data from the Internet, "feeding" on websites that propagated false information. Its data pool was polluted, and the algorithm could not discern between "good" and "bad" data. Another example is that of the Microsoft chatbot, "Tay," which learned to interact with humans via Twitter. Within twenty-four hours, the chatbot "became racist," for lack of a better word, because "Internet trolls" had bombarded it with mostly offensive and erroneous data, i.e. inflammatory tweets, from which the Chatbot had "learned"».

[8] 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. *Ivi* ci si chiede: «does artificial intelligence create the new spiritual entities – the angels of our time – in the world of information processing? The question is whether these new actors – animals and electronic agents – fighting for their interests and even for full-fledged constitutional rights are nothing but social collectives who, rightly or wrongly, express their sympathies for non-human entities and ask to be formally accepted as legal actors (the anthropocentric view)? Or is social communication extending its capacities to include different autonomous processes in its environment and thus respect their eigenvalues (the ecocentric view)? Or are law and politics directly link-

quella umana [9], e che, dunque, ci troviamo di fronte a ‹‹rational actors» ma ‹‹in no way human» [10]. Almeno allo stato attuale non può parlarsi di surrogati artificiali della persona o delle relative volizioni [11].

Invero, la personificazione dei non umani è intesa come una strategia per affrontare e ridurre l'incertezza sull'identità dell'altro e come significato simbolico della capacità di partecipare alla comunicazione. Ciò presuppone che i non umani siano in grado di entrare in comunicazione con gli umani e che ciò possa 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 [12]. L'idea dell'ibrido non deve far pensare ad un meccanismo in cui

---

ing up with other 'living', 'pulsating', 'autonomous'processes which would steer their rule production into new directions (the juridicocentric or sociocentric view)?».

[9] Sull'idea di «personalità/soggettività differenziate: che operi non in danno, come un *minus* rispetto al parametro massimo della universale condizione giuridica umana e della sua capacità generale. Ma come un acquisto, 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. PORTALURI, *L'articolazione delle figure soggettive: dalla personalità alla soggettività giuridica,* in *www.ridiam.it*, 2019, spec. p. 5. Sul processo di moltiplicazione dei soggetti giuridici sviluppatosi negli ultimi decenni cfr. A. PISANO', *Diritti deumanizzati. Animali, ambiente, generazioni future, specie umana,* Milano, 2012; M.-A. HERMITTE, *L'intérêt d'une constitutionnalisation des normes relatives au vivant, in X. Bioy (dir.), Droits constitutionnels du vivant. Approches comparées de nouveaux objets constitutionnels: bioéthique et environnement,* Paris, 2019, p. 21 e ss.; R. MÍGUEZ NÚÑEZ, *Soggettivizzare la natura*? in *The Cardozo Electronic Law Bulletin*, 2019, p. 1 e ss.

[10] È quanto si legge nel documento "*Call for AI Ethics",* firmato a Roma il 28 Febbraio 2020, in *http://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/CALL%2028%20febbraio/AI%20Rome%20Call%20x%20firma_DEF_DEF_.pdf,* p. 3.

[11] J. SEARLE, *Intentionality. An Essay in the Philosophy of Mind*, Mass, 1983, p. 312, sostiene che l'intelligenza artificiale non dispone di atti autenticamente soggettivi, ma compie calcoli e riproduce atteggiamenti. In prospettiva diversa, ritiene che si possa parlare di possibili surrogati artificiali della persona o delle relative volizioni, qualificabili non più come oggetti ma come agenti, L. COPPINI, *Robotica e intelligenza artificiale: questioni di responsabilità civile* in *Politica del diritto*, 2018, p. 714 e ss. All'A. si rimanda per l'analisi delle diverse tipologie di robot secondo il Glossario tecnico della *Strategic Research Agenda (SRA) for robotics in Europe*. Segnatamente, oltre ai *Robot* cognitivi si classificano i: «*robot* teleoperati: sono composti da un set di parti mosse da motori controllati da persone fisiche tramite specifiche interfacce, come un *joy-stick* o anche uno *smartphone*. Le loro azioni sono completamente controllate dall'uomo, quindi, essi si possono configurare come semplici strumenti nelle mani dell'operatore. *Robot* autonomi: l'autonomia va intesa come l'abilità di svolgere un compito senza alcun intervento umano durante il processo. Il *Robot* è guidato, istruito tramite un programma che gli fornisce regole di comportamento. È opportuno parlare qui di autonomia debole».

[12] B. LATOUR*, Politics of Nature: How to Bring the Sciences into Democracy*, Harvard, 2004, p. 71, ritiene che «the psycho– systemic competence deficits of non-humans are adequately compensated by the distributed intelligence of social systems».

l'attore è sempre e solo umano, data la forte influenza anche della componente non umana. Piuttosto il processo comunicativo all'interno dell'ibrido è in grado di identificare eventi che possono essere compresi come espressioni del non umano, creando le condizioni di possibilità per l'ingresso di non umani in comunicazione [13].

In diversa prospettiva, la comunicazione è una operazione che è fornita della capacità di autoosservarsi e, dunque, a parteciparvi possono essere soltanto i sistemi psichici [14]. Gli algoritmi, dunque, sono strumenti attraverso ai quali si comunica, ma non partecipano alla comunicazione. Nel «processo elementare della comunicazione sociale, l'*altro* non è il destinatario della comunicazione – come spesso di dice– ma è il suo inizio: la funzione dell'*altro* consiste nel fatto che egli, attraverso la comprensione, rende possibile la comunicazione, la quale, a sua volta, solo così può essere attribuita a colui che ha attivato l'atto stesso del comunicare e che, perché ciò accada, deve essere egli stesso *osservato* dall'*altro* come colui che intende comunicare, quindi come *altro* dell'altro. E, infatti, se questi non coglie l'atto del comunicare come l'atto che intende comunicare che sta comunicando, se non coglie l'altro come altro che intende comunicare, non può esserci comunicazione. Nella comunicazione, allora, entrambi lottano per il riconoscimento, per usare un concetto hegeliano, ed entrambi ottengono il riconoscimento attraverso la comunicazione. Ma entrambi *si osservano come altro*» [15].

Le macchine sono in grado di fare cose che gli uomini non arrivano a fare, ma di contro, gli algoritmi non hanno una dimensione umana. In altri termini, a fronte di un pro-

---

[13] Sul punto cfr. G. TEUBNER*, Rights of Non-Humans?,* cit., p. 521.

[14] R. DE GIORGI, N. LUHMANN, *Teoria della società*, Milano, 2013 (1991), p. 27. G. TEUBNER*, Rights of Non-Humans?,* cit., p. 500 confronta il pensiero di N. Luhmann con quello di Latour ed afferma «however impressive this theoretical move appears, it does not go far enough to cover the far-reaching ambitions of the ecological movement and the cyber revolution. Collectives, i.e. social systems between human individuals, as actors – this is the point where Luhmann stops and where Latour begins. In Latour's account, a multitude of new actants and hybrids that cannot be identified with human individuals or with collective actors, are entering the scene and radically transforming today's political ecology». L'A. poi aggiunge che «regarding both Luhmann and Latour: Personification of non humans is best understood as a strategy of dealing with the uncertainty about the identity of the other, which moves the attribution scheme from causation to double contingency and opens the space for presupposing the others'selfreferentiality. Beyond Luhmann: There is no compelling reason to restrict the attribution of action exclusively to humans and to social systems. Personifying other nonhumans is a social reality today and a political necessity for the future. Beyond Latour: The admission of actors does not take place, as he suggests, into one and only one collective. Rather, the properties of new actors differ extremely according to the multiplicity of different sites of the political ecology».

[15] R. DE GIORGI, A. PRIZRENI, *Presentazione,* in R. DE GIORGI, A. PRIZRENI (a cura di), *Lo sguardo dell'altro,* Lecce, p. 10. Sul punto si rimanda a N. LUHMANN, *Introduzione alla teoria dei sistemi,* edizione italiana a cura di S. MAGNOLO, Lecce, 2018, p. 243 e ss.

cesso che conduce sempre più ad "umanizzare" le cose a tal punto da rendere assai difficile la loro considerazione in termini semplicemente di *res,* non può assistersi alla progressiva riduzione della persona ad un qualcosa di inanimato e in gran parte sostituibile dalle macchine [16].

Il rapporto tra umani e algoritmi è molto complicato e taluni rinvengono dei profili problematici paragonabili a quelli che riguardano il rapporto tra uomini e animali. Alcune differenze sono state puntualizzate in modo espresso. Si afferma, per esempio, che «paradoxically, they incorporate animals into human society in order to create defences against the destructive tendencies of human society against animals. The old formula of the social domination of nature is replaced by the new social contract with nature. For electronic agents, the exact opposite is true. Their legal personification, especially in an economic and technological context, creates aggressive new action centres as basic productive institutions. Here, their inclusion in society does not protect the new actors, just the opposite: it is society that needs to defend itself against the new actors. With the social inclusion of cyborgs and electronic agents, new problems of alienation appear on the horizon of the law» [17].

Un altro dato può essere utile alla riflessione. Gli algoritmi sono riducibili a mezzi dei quali l'uomo si serve, non così le entità non umane della natura. Lo suggerisce limpidamente Bruno Latour che riflette sulla circostanza che «it might be sufficient, strangely enough, to return to the definition that Kant gives of human morality, a definition that is so well known that people forgot to see that it is in fact wonderfully apposite for non-humans. Let us get back to this most canonical of all definitions: "Everything in creation which he wishes and over which he has power can be used merely as a means; only man, and, with him, every rational creature, is an end in himself. He is the subject of the moral law which is holy, because of the autonomy of his freedom. Because of the latter, every will, even the private will of each person directed to himself, is restricted to the condition of agreement with the autonomy of the rational being, namely, that it be subjected to no purpose which is not possible by a law which could have its origin in the will of the subject undergoing the action. This condition requires that the subject never be used simply as a means but at the same time as an end in itself» [18]. In questa prospettiva, con

[16] Cfr. P. STANZIONE, *Biodiritto, postumano e diritti fondamentali*, in *www.comparazionedirittocivile.it*, 2010.

[17] G. TEUBNER*, Rights of Non-Humans?,* cit., p. 521.

[18] B. LATOUR, *To modernize or to ecologize? That's the question,* in N. CASTREE, B. WILLEMS-BRAUN (a cura di), *Remaking Reality: Nature at the Millenium*, London– New York, 1998, p. 237.

riferimento alle componenti non umane della natura, ben si ritiene che «they do not at all say that we should not use, control, serve, dominate, order, distribute or study them, but that we should, as for humans, never consider them as simply means but always also as ends» [19]. Invece con riferimento all'intelligenza artificiale, si legge espressamente nella documentazione ufficiale dell'Unione Europea [20], che l'intelligenza artificiale non è fine a se stessa, ma è uno strumento a servizio delle persone che ha come fine ultimo quello di migliorare il benessere degli esseri umani [21]. Proprio qui viene in rilievo un altro aspetto critico della cd. intelligenza artificiale antropomorfica se intesa come strumento soltanto al servizio della persona.

Cosa deve intendersi per *"technology that works for people"*, come si legge in molti documenti comunitari? Se intendiamo che le macchine non possono e non devono essere meccanismi fini a se stessi o alla pari dell'uomo, il dato non può non essere condiviso. Riduttivo è però forse pensare all'intelligenza artificiale come uno strumento a servizio solo delle persone o degli essere umani, a meno che non si voglia pensare all'ambiente e a tutti gli essere viventi che lo popolano come a strumenti a servizio degli umani [22].

Le considerazioni che precedono inducono a ritenere che i problemi posti dall'intelligenza artificiale non debbano essere visti da un punto di vista esclusivamente antropocentrico, ma occorre riflettere su ciò che va anche oltre il solo essere umano, all'insegna di una IA sostenibile.

Il dato, d'altronde, emerge anche da COM (2019) 168 final, ove si prevede che per ottenere un'IA affidabile si dovrebbe tenere conto del suo impatto sull'ambiente e sugli altri esseri senzienti, che devono poter beneficiare della biodiversità e di un ambiente abitabile, per realizzare il quale l'IA stessa potrebbe avere un ruolo strategico [23].

---

[19] B. LATOUR, *To modernize or to ecologize?,* cit., p. 238.

[20] COM (2019) 168 final, in *www.ec.europa.eu*, p. 2.

[21] Cfr. Com (2018) 237 final, cit.; SWD (2018) 137 final. in *www.ec.europa.eu.*

[22] Sulla necessità improrogabile di impostare su basi nuove il rapporto tra uomo e natura in un'ottica di equilibrio ecologico si veda G. GRISI, *La lezione del Coronavirus*, in *Juscivile,* 2020, p. 190 e ss. L'A. nota, p. 207, che «bisogna che l'uomo comprenda che il pianeta è vivo, respira, soffre, si trasforma continuamente, si evolve; e noi, incoscientemente, lo abbiamo trattato come se fosse un oggetto inerte, da soggiogare al nostro dominio. La nostra scriteriata attività e l'ossequio prestato a logiche di tornaconto individuale da ricercare con ogni mezzo hanno compromesso gli ecosistemi e la biodiversità. La vita sul pianeta si basa su equilibri che l'uomo, dotato di funzioni cerebrali superiori ma non sempre ben sfruttate, ha sconvolto: con un consumo eccessivo delle risorse ha rapinato e violentato la natura e che questa dia, ciclicamente, segnali, più o meno forti, di insofferenza è – ci si perdoni il bisticcio – naturale, né c'è da sorprendersi che la reazione possa assumere anche le caratteristiche proprie delle pandemie». Sul punto cfr. C. LARRÈRE, R. LARRÈRE, *Penser et agir avec la nature. Une enquête philosophique*, Parigi, 2018, p. 338 e ss.

[23] È quanto si legge in COM (2019) 168 final, *cit.,* p. 7.

Alla "famiglia umana" fa riferimento anche il documento denominato "*Call for AI Ethics",* firmato a Roma il 28 Febbraio 2020 [24], ove ci è soltanto una timida apertura anche al rispetto di tutti gli ambienti naturali, nei termini che seguono: «new technology must be researched and produced in accordance with criteria that ensure it truly serves the entire "human family" (Preamble, Univ. Dec. Human Rights), respecting the inherent dignity of each of its members and all natural environments, and taking into account the needs of those who are most vulnerable» [25].

Lo stesso documento auspica una «vision in which human beings and nature are at the heart of how digital innovation is developed» e che «AI systems must be conceived, designed and implemented to serve and protect human beings and the environment in which they live» [26]. La natura e l'ambiente vengono, però, visti sempre come qualcosa da salvaguardare non in quanto tali, ma soltanto perché luoghi dove vivono gli essere umani.

Una visione un po' più matura traspare solo in quella parte del documento nel quale si sottolinea la necessità che il progresso tecnologico avvenga nel rispetto per il pianeta, anche se il pianeta è definito come "our common and shared home", facendo pensare, ancora una volta ad una prospettiva meramente antropocentrica. Qui, oltre ad una mera dichiarazione di principio, si individuano tre requisiti che devono essere soddisfatti per un progresso tecnologico sostenibile: 1*) it must include every human being, discriminating against no one; 2) it must have the good of humankind and the good of every human being at its heart; 3) finally, it must be mindful of the complex reality of our ecosystem and be characterised by the way in which it cares for and protects the planet (our "common and shared home") with a highly sustainable approach, which also includes the use of artificial intelligence in ensuring sustainable food systems in the future* [27].

---

[24] Sul punto cfr. L. CASALINI, *Rome Call for AI Ethics: Per un'intelligenza artificiale umanistica,* in *Persona e Mercato, Osservatorio OGID,* p. 8.

[25] Cfr. "*Call for AI Ethics",* cit., p. 3.

[26] Cfr. "*Call for AI Ethics",* cit., p. 4.

[27] Cfr. "*Call for AI Ethics",* cit., p. 4. Sul punto indicazioni interessanti si traggono altresì dalle *Ethics Guidelines for Trustworthy Artificial Intelligence* dell'8 aprile 2019, *published by the High-Level Expert Group (HLEG) on Artificial Intelligence (AI) appointed by the European Commission*, sulle quali si rimanda al sito *https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence. Ivi*, si legge, p. 42, che «non vi è dubbio che la lotta ai cambiamenti climatici debba essere una priorità assoluta per i responsabili politici di tutto il mondo, ma la trasformazione digitale e un'IA affidabile offrono grandi potenzialità in termini di riduzione dell'impatto umano sull'ambiente e di contributo a un uso efficiente ed efficace dell'energia e delle risorse naturali. Un' IA affidabile, ad esempio, può essere accoppiata ai *Big Data* per rilevare con maggiore precisione il fabbisogno energetico e favorire di conseguenza infrastrutture e consumi energetici più efficienti».

Il ruolo centrale dell'uomo di fronte alla trasformazione digitale alla quale assistiamo diventa, allora, quello di essere l'unico, tra gli esseri viventi, a poter "sorvegliare" i sistemi di IA, con un'importante responsabilità che trascende la propria "individualità", per il benessere dell'ambiente e di tutti gli essere viventi che lo popolano.

L'*Human oversight* è concetto che si ritrova nei documenti Europei dedicati all'IA e svela la possibilità di scenari molteplici e articolati, potendo riguardare un approccio con intervento umano in ogni ciclo decisionale del sistema, cd. *"human-in-the-loop*"; un approccio con supervisione umana, ossia caratterizzato dalla capacità di intervento umano durante il ciclo di progettazione del sistema e di monitoraggio del funzionamento del sistema cd. "*human-on-the-loop*", oppure un approccio con 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 durante l'uso del sistema e garantendo la possibilità di annullare una decisione adottata dal sistema [28].

In particolare in COM(2020) 65 final, *White paperon artificial intelligence – A European approach to excellence and trust,* nel *White paper on artificial intelligence,* sia pure a titolo soltanto esemplificativo, vengono individuate della modalità attraverso le quali la sorveglianza umana può esplicarsi: 1) *the output of the AI system does not become effective unless it has been previously reviewed and validated by a human*; 2) *the output of the AI system becomes immediately effective, but human intervention is ensured afterwards*; 3) *monitoring of the AI system while in operation and the ability to intervene in real time and deactivate*; 4) *in the design phase, by imposing operational constraints on the AI system* [29].

L'essere umano, dunque, ha un ruolo centrale e strategico, nell'affrontare le sfide po-

---

[28] COM(2019) 168 final, p. 5, nota 13. Sul punto si rimanda anche a COM(2020) 65 final, p. 25 ove si legge che «a European governance structure on AI in the form of a framework for cooperation of national competent authorities is necessary to avoid fragmentation of responsibilities, increase capacity in Member States, and make sure that Europe equips itself progressively with the capacity needed for testing and certification of AI-enabled products and services. In this context, it would be beneficial to support competent national authorities to enable them to fulfil their mandate where AI is used. A European governance structure could have a variety of tasks, as a forum for a regular exchange of information and best practice, identifying emerging trends, advising on standardisation activity as well as on certification. It should also play a key role in facilitating the implementation of the legal framework, such as through issuing guidance, opinions and expertise. To that effect, it should rely on a network of national authorities, as well as sectorial networks and regulatory authorities, at national and EU level. Moreover, a committee of experts could provide assistance to the Commission».

[29] COM (2020) 65 final, cit., p. 21. Sulla portata di COM (2020) 65 final, si veda S. ORLANDO, *Il Libro Bianco della Commissione Europea del 19 febbraio 2020 sull'Intelligenza Artificiale: "Eccellenza e Fiducia",* in *Persona e Mercato, Osservatorio OGID,* p. 7.

ste dall'Intelligenza artificiale, in quanto unico in grado di fare scelte qualitative e non soltanto quantitative [30] e unico ad avere la possibilità di controllare e correggere le decisioni del sistema di IA [31].

3. – Per non "subire" la trasformazione tecnologica alla quale assistiamo e renderla "accettabile" per i cittadini si fa riferimento all'affidabilità dei sistemi di intelligenza artificiale; affidabilità "misurabile" sulla base di dati ben precisi elaborati da un Gruppo di esperti nominato dalla Commissione europea [32]. Si tratta di dati basati su orientamenti non vincolanti e che non introducono nuovi obblighi giuridici.

Segnatamente, le *Ethics Guidelines for Trustworthy Artificial Intelligence* dell'8 Aprile 2019, presentate dal Gruppo di esperti prevedono che «per ottenere un'intelligenza artificiale affidabile sono necessari tre elementi: 1) l'IA dovrebbe rispettare la legge, 2) dovrebbe osservare i principi etici e 3) dovrebbe dimostrare robustezza» [33].

Ciascuna componente in sé è necessaria ma non sufficiente per realizzare un'IA affidabile.

Le *Ethics Guidelines for Trustworthy Artificial Intelligence* non affrontano esplicitamente la prima componente, ossia la legalità dell'IA, ma si propongono di offrire orientamenti per promuovere e garantire l'eticità e la robustezza dell'IA [34].

La premessa dalla quale parte il Gruppo di esperti è che «l'uso nella nostra società dei sistemi di IA, come di qualsiasi tecnologia potente, pone vari problemi etici, ad esempio in merito al loro effetto sulle persone, sulla società, sulle capacità decisionali e sull'incolumità» [35].

---

[30] Cfr. "*Call for AI Ethics*", cit., p. 3.

[31] Sono le indicazioni che si traggono dalla proposta, datata 6.2.2020, di Risoluzione del Parlamento europeo sui processi decisionali automatizzati: garantire la tutela dei consumatori e la libera circolazione di beni e servizi (2019/2915(RSP)), in *https://www.europarl.europa.eu/doceo/document/B-9-2020-0094_IT.pdf.*

[32] *Ethics Guidelines for Trustworthy Artificial Intelligence* dell'8 Aprile 2019, cit., p. 42.

[33] Cfr. COM(2019) 168 final, p. 5.

[34] Si veda L. FLORIDI, *Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical,* in *Philos. Technol.* 2019, p. 185. *Ivi* L'A. cerca di fornire una mappa per coloro che desiderano evitare o minimizzare alcuni dei rischi etici più evidenti e significativi, quando si passa dai principi alle pratiche dell'etica digitale e afferma che «understanding as early as possible that shortcuts, postponements, or quick fixes do not lead to better ethical solutions but to more serious problems, which become increasingly difficult to solve the later one deals with them, does not guarantee that the five malpractices analysed in this article will disappear, but it does mean that they will be reduced insofar as they are genuinely based on misunderstanding and misjudgements. Not knowing better is the source of a lot of evil. So, the solution is often more and better information for all».

[35] *Ethics Guidelines for Trustworthy Artificial Intelligence* dell'8 Aprile 2019, cit., p. 10.

Una particolare attenzione a questi aspetti emerge anche dal *White paperon artificial intelligence,* ove si è individuato come obiettivo strategico il cd. '*ecosystem of trust'*, ossia un ecosistema di fiducia «to do so, it must ensure compliance with EU rules, including the rules protecting fundamental rights and consumers'rights, in particular for AI systems operated in the EU that pose a high risk. Building an ecosystem of trust is a policy objective in itself, and should give citizens the confidence to take up AI applications and give companies and public organisations the legal certainty to innovate using AI» [36].

Gli esseri umani che interagiscono con i sistemi di IA devono poter mantenere la propria piena ed effettiva autodeterminazione e giovarsi di un aumento delle abilità cognitive. A tal fine, il Gruppo di lavoro pone, alla base di un'intelligenza artificiale affidabile, e come imperativi etici, il rispetto dell'autonomia umana, della prevenzione dei danni, dell'equità, e della esplicabilità [37].

Nell'ambito dei documenti che riguardano l'IA, l'etica, in verità, è intesa prevalentemente come *self-regulation* [38]. Non si è mancato di evidenziare alcune criticità [39].

È stato notato, anche da alcuni componenti del gruppo di esperti, che le *Ethics Guidelines for Trustworthy Artificial Intelligence* «move from a first, more theoretical *what* chapter, to a second, more practical *how* chapter, so to speak, is reasonable and commendable. However, in translating principles into practices, even the best efforts may be undermined by some unethical risks» [40].

Può verificarsi, per esempio, il cd. *Digital ethics shopping*, ossia la circostanza che si

---

[36] COM (2020) 65 final, p. 3.

[37] *Ethics Guidelines for Trustworthy Artificial Intelligence* dell'8 Aprile 2019, cit., p. 16.

[38] Cfr. B. WAGNER, *Ethics as an Escape from Regulation: From Ethics-washing to Ethics-shopping?*, in M. HILDEBRANDT (a cura di), *Being Profiled. Cogitas ergo sum*, Amsterdam, 2018, in *https://pdfs.semanticscholar.org*, p. 1.

[39] F.J. ZUIDERVEEN BORGESIUS, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in *The International Journal of Human Rights*, 2020, p. 12, afferma che «on the one hand, self-regulation is commendable. It can hardly be denied that ethical AI is preferable over unethical AI. Self-regulation could help mitigate discrimination, and could provide inspiration for legislators. On the other hand, there are serious problems with self-regulation. Most importantly: self-regulation is non-binding. Human rights protection cannot be left to voluntary measures. Enforcement is typically lacking of such ethics codes». Particolarmente fiduciosa ad un approccio etico ai problemi posti dai sistemi di IA è SK. KATYAL, *Private accountability in the age of artificial intelligence,* in *UCLA Law Review,* 2019, p. 56 e ss. In particolare, l'A esplora «the impending conflict between the protection of civil rights and artificial intelligence (AI). While both areas of law have amassed rich and well-developed areas of scholarly work and doctrinal support, a growing body of scholars are interrogating the intersection between them. This Article argues that the issues surrounding algorithmic accountability demonstrate a deeper, more structural tension within a new generation of disputes regarding law and technology. As I argue, the true promise of AI does not lie in the information we reveal to one another, but rather in the questions it raises about the interaction of technology, property, and civil rights».

[40] L. FLORIDI, *Translating Principles into Practices,* cit., p. 186.

finisca per elaborare un gran numero di principi, codici, linee guida o quadri etici, producendo un proliferare di documenti che generano incoerenza e confusione. Inoltre, spesso, gli attori pubblici e privati progettano propri codici per paura di apparire da meno degli altri, contribuendo così ulteriormente alla ridondanza delle informazioni. Si genera, in altri termini, il rischio che tutta questa iperattività finisca per essere un modo per giustificare i propri comportamenti, piuttosto che renderli coerenti con un quadro etico condiviso [41].

Il rischio del *Digital ethics shopping* può essere attenuato proprio dalle *Ethics Guidelines for Trustworthy Artificial Intelligence* la cui pubblicazione «is a significant improvement, given that it is the closest thing available in the European Union (EU) to a comprehensive, authoritative, and public standard of what may count as socially good AI» [42].

Un quadro di riferimento pubblico, autorevole e socialmente condiviso è utile anche per ridurre il cd. *Ethics bluewashing*, ossia la pratica scorretta «of making unsubstantiated or misleading claims about, or implementing superficial measures in favour of, the ethical values and benefits of digital processes, products, services, or other solutions in order to appear more digitally ethical than one is» [43]. Si tratta di una forma di disinformazione basata per lo più su una sponsorizzazione che mira a far apparire che si contribuisce a risolvere problemi che, in realtà, non possono essere affrontati dal singolo in modo adeguato [44].

Un'ulteriore criticità legata all'eccessiva fiducia nei principi etici per la regolamentazione dei problemi posti dall'IA è legata al rischio di *Digital ethics lobbying*, ossia della «malpractice of exploiting digital ethics to delay, revise, replace, or avoid good and ne-

---

[41] L. FLORIDI, *Translating Principles into Practices,* cit., p. 186.

[42] L. FLORIDI, *Translating Principles into Practices,* cit., p. 187; B. WAGNER, *Ethics as an Escape from Regulation,* cit., p. 1.

[43] L. FLORIDI, *Translating Principles into Practices,* cit., p. 187.

[44] L. FLORIDI, *Translating Principles into Practices,* cit., p. 188. L'Autore fa anche riferimento al *Digital ethics dumping*, inteso come «the malpractice of (a) exporting research activities about digital processes, products, services, or other solutions, in other contexts or places (e.g. by European organisations outside the EU) in ways that would be ethically unacceptable in the context or place of origin and (b) importing the outcomes of such unethical research activities» ed all'*Ethics shirking*, cioè alla «malpractice of doing increasingly less Bethical work (suchas fulfilling duties, respecting rights, and honouring commitments) in a given context the lower the return of such ethical work in that context is mistakenly perceived to be». Entrambe queste pratiche, continua l'A. «has historical roots and often follows geopolitical outlines. Actors are more likely to engage in ethics dumping and shirking in contexts where disadvantage populations, weaker institutions, legal uncertainties, corrupted regimes, unfair power distributions, and other economic, legal, political, or social ills prevail».

cessary legislation (or its enforcement) about the design, development, and deployment of digital processes, products, services, or other solutions» [45].

Critiche analoghe potrebbero essere mosse anche agli incentivi alla c.d. *voluntary labelling for no-high risk ai applications,* e cioè ad una etichettatura su base volontaria, che, pure, sono stati individuati come strumenti che potrebbero accrescere la fiducia dei cittadini nell'IA. Gli operatori economici interessati non soggetti alle prescrizioni obbligatorie potrebbero decidere, su base volontaria, di impegnarsi al rispetto di una serie specifica di prescrizioni, ottenendo in cambio un "*Quality Label"* per le loro applicazioni di IA.

Il marchio volontario consentirebbe agli operatori economici interessati di mettere in evidenza l'affidabilità dei loro prodotti e servizi basati sull'IA e consentirebbe agli utenti di riconoscere facilmente che i prodotti e i servizi in questione sono conformi a determinati parametri di riferimento standardizzati a livello dell'UE.

Il quadro per una intelligenza artificiale affidabile è completato dall'individuazione di alcuni requisiti dei sistemi di IA. Oltre alla *human oversight*, della quale si è detto, della non discriminazione e dell'*accountability* alle quali saranno dedicati specifici paragrafi, preme ora soffermarsi sulla robustezza tecnica, sulla sicurezza e sulla trasparenza, al fine di verificare il significato che questi termini assumono con riferimento ai sistemi di IA.

La robustezza è intesa come capacità dei sistemi di IA di essere resilienti sia agli attacchi palesi, sia a tentativi più subdoli di manipolazione dei dati o degli algoritmi, e di garantire l'esistenza di un piano di emergenza in caso di problemi, errori o incongruenza [46].

---

[45] L. FLORIDI, *Translating Principles into Practices,* cit., p. 188; B. WAGNER, *Ethics as an Escape from Regulation,* cit., p. 1, ove si afferma che «a strange confusion among technology policy makers can be witnessed at present. While almost all are able to agree on the common chorus of voices chanting 'something must be done,' it is very difficult to identify what exactly must be done and how. In this confused environment it is perhaps unsurprising that the idea of 'ethics' is presented as a concrete policy option. Striving for ethics and ethical decision-making it is argued, will make technologies better. While this may be true in many cases, much of the debate about ethics seems increasingly focussed on private companies avoiding regulation. Unable or unwilling to properly provide regulatory solutions, ethics is seen as the 'easy' or 'soft' option which can help structure and give meaning to existing self-regulatory initiatives. In this world, 'ethics' is the new 'industry self-regulation». F.J. ZUIDERVEEN BORGESIUS, *Strengthening legal protection against discrimination by algorithms,* cit., p. 12 afferma che «many self-regulatory AI principles are rather vague and fail to give detailed guidance. Wagner warns for 'ethics washing' in the context of AI. He cautions that firms may see ethics 'as the "easy" or "soft" option'. Indeed, self-regulation should not distract legislators from the possible necessity of new laws could be improved».

[46] COM (2019) 168 final, cit., p. 6. Si veda anche COM (2020) 65 final, cit., p. 20 ove la robustezza dei sistemi di IA è legata a: 1) *Requirements ensuring that the AI systems are robust and accurate, or at least correctly reflect their level of accuracy, during all life cycle phases*; 2) *Requirements ensuring that outcomes are reproducible*; 3) *Requirements ensuring that AI systems can adequately deal with errors or in-*

La sicurezza attiene, invece, alla verificabilità del modo di operare dei sistemi di IA, tenuto conto della necessità di garantire l'integrità fisica e mentale di tutte le persone coinvolte. Utili potrebbero essere prescrizioni per la tenuta di registri relativi alla programmazione dell'algoritmo e ai dati utilizzati per addestrarlo, in modo da poter risalire alle decisioni dei sistemi di IA e di verificarle [47].

La trasparenza è intesa, invece, come spiegabilità del processo decisionale degli algoritmi. I diversi portatori di interessi coinvolti devono essere in grado di rendersi conto delle capacità e dei limiti del sistema di IA che, inoltre, devono essere identificabili come tali, così che gli utenti sappiano che stanno interagendo con un sistema di IA e possano individuare le persone che ne sono responsabili [48]. I dispositivi, in altri termini, devono essere in grado di offrire agli individui informazioni sulla logica alla base degli algoritmi utilizzati per prendere decisioni. Si fa riferimento al cd. "*duty of explanation*" finalizzato a rendere comprensibili non solo i criteri decisionali degli agenti algoritmici basati sull'intelligenza artificiale, ma anche i loro scopi e obiettivi [49]. Su questo la dottrina è abbastanza scettica. È stato notato, infatti, 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» [50].

Tra l'altro, i sistemi di IA sono in costante evoluzione e agiscono in un ambiente dinamico, così che la realizzazione di un'IA affidabile è un processo continuo che pone numerose difficoltà [51], a fronte delle quali i requisiti indicati dovrebbero poter almeno essere "tradotti" in specifiche procedure e regole che il sistema deve sempre seguire e in

---

consistencies during all life cycle phases. 4) *Requirements ensuring that AI systems are resilient against both overt attacks and more subtle attempts to manipulate data or algorithms themselves, and that mitigating measures are taken in such cases.*

[47] COM (2020) 65 final, cit., p. 22.

[48] COM (2019) 168 final, cit., p. 7.

[49] Cfr. "*Call for AI Ethics",* cit., p. 4.

[50] F.J. ZUIDERVEEN BORGESIUS, *Strengthening legal protection against discrimination by algorithms,* cit., p. 9.

[51] In particolare, il Gruppo di esperti ha specificato che i sistemi di IA sono valutati secondo la prospettiva teorica del ciclo "percezione/pianificazione/azione". Perché questa architettura garantisca un'IA affidabile è necessario integrare i requisiti in tutte e tre le fasi del ciclo: i) nella fase di "percezione" si dovrebbe sviluppare il sistema in modo tale che riconosca tutti gli elementi ambientali indispensabili a garantire l'aderenza ai requisiti; ii) nella fase di "pianificazione" il sistema dovrebbe considerare solo i piani che aderiscono ai requisiti; iii) nella fase di "azione", le azioni del sistema dovrebbero essere limitate ai comportamenti che realizzano i requisiti".

determinati comportamenti o stati cui il sistema non deve mai trasgredire. L'adozione di tali requisiti, inoltre, deve essere proporzionata al contesto specifico in cui si applicano e all'impatto che possono avere. Basti pensare che un'applicazione di IA che suggerisce di leggere un libro non adatto comporta molti meno rischi rispetto a una che sbaglia una diagnosi di tumore e che, pertanto, non può non essere sottoposta a una vigilanza più rigorosa [52].

Un rischio molto alto, legato ai sistemi di IA, è che possano avere effetti discriminatori nei diversi ambiti di applicazione. A questi aspetti sono dedicate le osservazioni che seguono.

4. – Le distorsioni e le discriminazioni rappresentano un rischio intrinseco di qualunque attività sociale od economica. Anche il processo decisionale umano non è immune da errori e distorsioni, ma quando queste distorsioni nascono da sistemi di IA, gli effetti possono essere di dimensioni e di impatto incredibilmente maggiore [53].

Gli esiti discriminatori di un processo decisionale algoritmico possono derivare da più fattori e, in particolare, non solo da difetti nella progettazione originale, ma anche dal modo nel quale il sistema di IA "apprende" nel corso del suo funzionamento, generando risultati e rischi non sempre evitabili o individuabili in fase di progettazione [54].

---

[52] COM (2019) 168 final, cit., p. 5. Cfr. B. WAGNER, *Ethics as an Escape from Regulation,* cit., p.1 e ss.; si vedano, A. JABŁONOWSKA, M. KUZIEMSKI, A.M. NOWAK, H.W. MICKLITZ, P. PAŁKA, G. SARTOR *Consumer law and artificial intelligence: challenges to the EU consumer law and policy stemming from the business'use of artificial intelligence: final report of the ARTSY project*, in *EUI Law*, 2018, p. 17, ove si afferma che «there is not one AI problem, and there will be not one solution to challenges posed by AI. On the contrary, the response will be different across the sectors, both regarding the substance and form of the potential regulatory response».

[53] Cfr. F.J. ZUIDERVEEN BORGESIUS, *Strengthening legal protection against discrimination by algorithms,* cit., p. 2. *Ivi* l'A. afferma che sebbene il processo decisionale algoritmico possa sembrare razionale, neutrale e imparziale, può portare a ingiuste discriminazioni e, pertanto, occorre porsi due domanda fondamentali, e cioè: «(i) Which legal protection against algorithmic discrimination exists in Europe, and what are its limitations? (ii) How could that legal protection be improved?». Sul punto si rimanda anche a P. HACKER, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law*, in *Common Market Law Review*, 2018, p. 1143; S. WACHTER, *Affinity Profiling and Discrimination by Association in Online Behavioural Advertising*, in *Berkeley Technology Law Journal*, 2020, *https://doi.org/10.2139/ssrn.3388639.*

[54] Cfr. COM(2020) 65 final, cit., p. 13. *Ivi* si legge che alcuni algoritmi dell'IA, se usati per prevedere il rischio di recidiva di atti delittuosi, possono riflettere distorsioni legate alla razza e al genere, prevedendo probabilità di rischio di recidiva diverse per le donne rispetto agli uomini, oppure per i cittadini di un determinato paese rispetto agli stranieri. Sul punto si rimanda a S. TOLAN, M. MIRON, E. GOMEZ e C.

Per le prestazioni dei sistemi di IA è fondamentale la qualità dei *set* di dati utilizzati [55] che, infatti, deve essere sufficientemente ampio e rappresentativo di dimensioni di genere, etnia e altri possibili motivi di discriminazione vietata [56]. Dovrebbero, cioè, essere rappresentati tutti i bisogni interessati dall'intelligenza artificiale, al fine non solo di garantire che nessuno sia escluso, ma anche di espandere quelle aree di libertà che potrebbero essere minacciate dal condizionamento algoritmico [57]. Non è, però, questa una cosa da poco e facilmente realizzabile, tanto più in assenza di meccanismi di controllo sociale.

Quando si raccolgono dati, questi possono riflettere condizionamenti di tipo sociale o contenere inesattezze, errori e vizi materiali. Questo aspetto deve essere risolto prima di utilizzare una qualsiasi *set* di dati per addestrare un sistema di IA [58]. È già, dunque, nella fase di individuazione dei cd. "*training data"*, ossia dei dati di addestramento che occor-

---

CASTILLO, *Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia,* in *Best Paper Award, International Conference on AI and Law,* 2019. Alcuni programmi di IA per l'analisi facciale riflettono distorsioni legate al genere e alla razza, in quanto identificano con maggiore facilità il genere degli uomini di carnagione chiara mentre commettono più errori nel determinare il genere delle donne di pelle più scura. In argomento si vedano J. BUOLAMWINI, T. GEBRU, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Atti della 1ª conferenza sull'equità, sulla responsabilità e sulla trasparenza),* in *Proceedings of Machine Learning Research,* vol. 81, 2018, p. 77.

[55] Si vedano N. SCHMIDT, B.E. STEPHENS, *An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination,* in *ArXiv,* 2019, p. 144, ove si afferma che «recognizing the existence and nature of algorithmic bias, as well as the risks associated with reliance on newly available data, is a necessary step to designing and implementing approaches to ensure that these algorithmic decision-making processes are as fair as possible. Unlawful bias arising from the use of algorithms can be evaluated and mitigated, but doing so requires a deep understanding of these new methods, data, and the contexts and domains in which they are being used».

[56] Cfr. COM (2020) 65 final, cit., p. 21. Si vedano anche le *Ethics Guidelines for Trustworthy Artificial Intelligence* dell'8 Aprile 2019, cit., p. 24, ove si legge che «i *set* di dati utilizzati dai sistemi di IA (sia per l'addestramento che per il funzionamento) possono subire l'influenza di distorsioni storiche non intenzionali, dell'incompletezza e di modelli di cattiva *governance*. Se tali distorsioni permangono, determinati gruppi o persone potrebbero essere involontariamente oggetto di pregiudizi e discriminazioni dirette e indirette e ciò potrebbe aggravare il pregiudizio e l'emarginazione. Il danno può anche derivare dallo sfruttamento intenzionale di distorsioni (del consumatore) o dal praticare una concorrenza sleale, con mezzi quali l'omogeneizzazione dei prezzi tramite la collusione o un mercato non trasparente. Le distorsioni identificabili e discriminatorie dovrebbero essere eliminate, se possibile, nella fase di raccolta. Anche il modo in cui vengono sviluppati i sistemi di IA (ad esempio, la programmazione degli algoritmi) può subire l'influenza delle distorsioni inique. La soluzione a tal riguardo potrebbe risiedere nell'attuazione di processi di sorveglianza per analizzare e affrontare in modo chiaro e trasparente le finalità, i vincoli, i requisiti e le decisioni del sistema. L'assunzione di personale proveniente da contesti, culture e discipline diverse inoltre può garantire la diversità di opinioni e dovrebbe essere incoraggiata».

[57] Cfr. "*Call for AI Ethics*", cit., p. 3.

[58] COM (2019) 168 final, p. 6.

rono misure necessarie per garantire che siano rispettati i valori e le norme dell'UE [59].

Per esempio si sono riscontrati sistemi di IA discriminatori, non perché il sistema fosse di per sé "cattivo" ma perché ereditava comportamenti sbagliati che poi ripeteva. Si pensi all'uso di un'IA per la selezione del personale di un'azienda. Se il sistema si basa «prevalentemente su dati del passato ("negli ultimi 20 anni sono stati assunti solo dirigenti maschi sopra i 40 anni") può giungere a selezionare in maniera distorta i candidati (ossia, nell'esempio, l'IA selezionerà i *curricula* scartando tutte le donne e tutti coloro che hanno meno di 40 anni, a prescindere dalle loro competenze)» [60].

Si tratta dei rischi di una "eccessiva selettività" alla quale si affianca il fatto che l'IA, lavorando per obiettivi, non ha la sensibilità per gestire eventi e casi "extra" ordinari rispetto a quelli già contemplati dal sistema [61].

A rendere ancor più preoccupante la situazione è, oltre al maggior impatto che può avere una decisione elaborata da sistemi di intelligenza artificiale, l'opacità del processo decisionale algoritmico che può rendere più difficile scoprire la discriminazione stessa e la sua causa [62], specie quando si tratti di discriminazioni indirette.

Supponiamo che qualcuno faccia domanda per un prestito sul sito *Web* di una banca che usa un sistema algoritmico per decidere su tali richieste. Se la banca nega automaticamente il prestito sul suo sito *Web*, difficilmente il cliente riuscirà a capire come il sistema arrivi alle decisioni e, dunque, gli risulterà difficile scoprire se l'algoritmo è discriminatorio. I sistemi algoritmici sono spesso "scatole nere", per diversi motivi. Perfino gli esperti che hanno costruito un sistema algoritmico potrebbero non sapere come lo stesso si comporterà quando utilizza nella pratica i dati dei quali si alimenta [63].

I sistemi di intelligenza artificiale hanno, inoltre, due limiti che, in particolare sui di-

---

[59] COM (2020) 65 final, cit., p. 21.

[60] S. IANNITI, B. ORSINI, *Intelligenza artificiale e i principi di non discriminazione nei contratti assicurativi,* 2020, in *https://www.insuranceup.it*.

[61] Cfr. S. IANNITI, B. ORSINI, *Intelligenza artificiale e i principi di non discriminazione,* cit. *Ivi* si fa l'esempio del cliente che non ricade nel *target market* positivo: l'intermediario avrebbe comunque la possibilità di vendere il prodotto, assicurandosi che comunque (i) il cliente non ricada nel target market negativo e (ii) il prodotto soddisfi comunque le esigenze assicurative del cliente. Una distribuzione assicurativa totalmente automatizzata, che non sia in grado di compiere questa analisi di secondo livello, potrebbe difatti scartare *tout court* il cliente dal paniere in quanto *extra target market* positivo e si limiterebbe a non permettergli di acquistare il prodotto.

[62] Cfr. COM (2020) 65 final, cit., p. 15. Sul punto si veda J. BURRELL, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, in *Big Data & Society*, 2016, p. 1.

[63] F.J. ZUIDERVEEN BORGESIUS, *Strengthening legal protection against discrimination by algorithms,* cit., p. 6.

ritti fondamentali delle persone, possono avere un impatto particolarmente negativo, e cioè: «first, the logical rules in the systems did not always fit the messy reality of the world. In real life, things are not true or false, but have grades of truth: a person is not either old or not old, but oldness increases gradually with age. Second, experts had to provide the knowledge to put into the systems. That process costs a lot of time and money. With machine learning, the knowledge in the system does not have to be provided by experts. In contrast, machine learning systems are set a task and given a large amount of data to use as examples of how this task can be achieved or from which to detect patterns. The system then learns how best to achieve the desired output» [64].

In che modo, dunque, è possibile avere risultati concreti dal punto di vista della tutela dalla non discriminazione algoritmica?

Una prima importante considerazione poggia su un'osservazione di fatto: i sistemi discriminano perché riproducono discriminazioni da parte degli esseri umani. Questa constatazione può essere utilizzata per raggiungere un risultato positivo. Il processo decisionale algoritmico può anche essere usato per combattere la discriminazione, nella misura nella quale aiuta ad individuare delle discriminazioni che, altrimenti, potrebbero essere latenti. Un esempio può servire a chiarire. Supponiamo che un sistema algoritmico mostri che una raccolta di foto d'archivio contiene stereotipi di genere, «one interpretation is that the algorithmic system illustrates stereotyped behaviour that already exists. Hence, an algorithmic system could help to discover existing discrimination that would otherwise have remained hidden» [65].

Risultati concreti, sul versante della tutela dalla discriminazione algoritmica, possono aversi investendo sull'educazione e l'istruzione anche dei soggetti più vulnerabili, attraverso metodi accessibili a tutti e che possano offrire uguaglianza di opportunità, nella consapevolezza, però, che la non discriminazione non è riducibile alla parità di trattamento [66]. Non si tratta, cioè, solo di sviluppare competenze digitali, ma anche di sensibilizzare sulle opportunità e sulle possibili questioni critiche poste dall'intelligenza artificiale dal punto di vista dell'inclusione sociale e del rispetto della dignità individuale [67].

---

[64] È quanto si legge nel *Discrimination, artificial intelligence, and algorithmic decision-making. Study by Prof. Frederik Zuiderveen Borgesius*, 2019, p. 9.

[65] In questi termini F.J. ZUIDERVEEN BORGESIUS, *Strengthening legal protection against discrimination by algorithms,* cit., p. 5.

[66] Sul punto si rinvia all'insegnamento di A. GENTILI, *Il principio di non discriminazione nei rapporti civili*, in *Riv. crit. dir. priv.,* 2009, p. 207 e ss. Cfr. G. CARAPEZZA FIGLIA, *Divieto di discriminazione e autonomia contrattuale,* Napoli, 2013; S. TOMMASI, *La non discriminazione nel DCFR,* in *Riv. crit. dir. priv.,* 2011, p. 119 e ss.

[67] Numerosi gli studi promossi sul tema dell'incidenza dell'IA sui diritti umani. In particolare si ve-

L'educazione inclusiva significa usare l'IA per supportare e integrare ogni persona, offrendo aiuto e opportunità per la partecipazione sociale. Le nuove tecnologie, per esempio, possono rivelarsi estremamente utili nell'aiutare le persone con disabilità a imparare e diventare più indipendenti [68]. Particolare attenzione a tali profili emerge dalla direttiva 2019/882 sui requisiti di accessibilità dei prodotti e dei servizi, ove ci si pone l'obiettivo di creare le condizioni per una società più inclusiva che faciliti la vita indipendente delle persone con disabilità [69].

L'IA aumenta inoltre le possibilità di seguire e analizzare le abitudini quotidiane delle persone. Vi è ad esempio il rischio potenziale che l'IA venga utilizzata a fini di sorveglianza di massa, oppure dai datori di lavoro per osservare il comportamento dei loro dipendenti. Poiché permette di analizzare grandi quantità di dati e di individuare collegamenti tra di essi, l'IA può essere usata anche per ricostruire e deanonimizzare dati riguardanti le persone, il che implica nuovi rischi per la protezione dei dati personali, anche in relazione a *set* di dati che di per sé non contengono informazioni personali [70].

Da queste premesse emerge che, per la difesa delle persone contro la discriminazione algoritmica, gli strumenti giuridici più rilevanti sono, non solo le norme sulla non discriminazione, ma anche quelle sulla protezione dei dati [71].

Siamo esposti a rischi che spesso nemmeno percepiamo e legati, per esempio, a meccanismi discriminatori insiti nella elaborazione di dati statistici che riguardano i luoghi dove abitiamo o i negozi che frequentiamo o lo sfruttamento algoritmico del nostro

---

dano*Recommendation of the Council on Artificial Intelligenc*e del 22 maggio 2019, pubblicato *dall'Organisation for Economic Co-operation and Development (OECD*), in *https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449*; il *Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic s*ystems, Giugno 2019, in *https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-stateson-the-hu/168095eecf*; le *Ethics Guidelines for Trustworthy Artificial Intelligence* dell'8 Aprile 2019, cit.

[68] Cfr. "*Call for AI Ethics*", cit., p. 5.

[69] Si veda il Considerando 25) della direttiva (UE) 2019/882, sui requisiti di accessibilità dei prodotti e dei servizi.

[70] È quanto si legge in COM (2020) 65 fina, p. 13.

[71] Sulla profilazione delle persone e sulle sue conseguenze in termini di possibili discriminazioni che influenzano opportunità cruciali, come la capacità di ottenere credito, assicurazione, istruzione, lavoro o persino cure mediche, si veda A. DROZDZ, *Protection of Natural Persons with Regard to Automated Individual Decision-Making in the GDPR,* Netherlands*,* 2020; Per uno studio dal quale emerge che nessuna delle politiche sulla *privacy* analizzate soddisfa i requisiti del GDPR, si rimanda a G. CONTISSA, K. DOCTER, F. LAGIOIA, M. LIPPI, H.W. MICKLITZ, P. PRZEMYSLA, G. SARTOR, P. TORRONI, *CLAUDETTE meets GDPR: automating the evaluation of privacy policies using Artificial Intelligence*, in *European Consumer Organisation (BEUC) Study Report, 2018 Retrieved from Cadmus, European University Institute Research Repository,* in *http://hdl.handle.net/1814/60795.*

comportamento in veste di consumatori [72]. C'è il rischio di rimanere "bloccati" nei propri profili e che si differenzino i prezzi in base ad una profilazione dalla quale risulti la disponibilità del consumatore a pagare, ma anche la sua vulnerabilità [73].

Ad essere elaborati sono spesso dati e informazioni che recano le caratteristiche di un soggetto e permettono di identificarlo e di farne una rappresentazione *on line* [74]. Quella digitale diventa allora la più esposta e vulnerabile delle proiezioni della persona [75], anche in considerazione della scarsa consapevolezza delle conseguenze, da parte del soggetto, sia della elaborazione di dati che lo riguardano, sia della potenzialità lesiva della cd. memoria digitale [76], cioè delle informazioni magari ormai superate e non aggiornate, o anche inesatte, che rimangono comunque reperibili in Rete [77].

La libera costruzione della propria identità è continuamente esposto all'interferenza delle nuove tecnologie e l'io finisce per essere frammentato, a sua insaputa, in una mol-

---

[72] Sul punto si veda *Consumer Protection Implications of Algorithms, Artificial Intelligence, and Predictive Analytics*, Fed. Trade Comm'n (Nov. 13, 2018), in *https://www.ftc.gov/news-events/audio-video/audio/consumer-protection-implicationsalgorithms-artificial-intelligence*.

[73] Sul punto si rimanda a O. BAR-GILL, *Algorithmic Price Discrimination When Demand Is a Function of Both Preferences and (Mis)perceptions*, in *University of Chicago Law Review*, 2019, p. 217.

[74] Cfr. la proposta di Risoluzione del Parlamento europeo sui processi decisionali automatizzati, cit., ove si legge che sarebbe opportuno imporre «ai professionisti di informare i consumatori quando i prezzi dei beni o dei servizi sono stati personalizzati sulla base di processi decisionali automatizzati e di profilazione del comportamento dei consumatori che permettono ai professionisti di valutare il potere d'acquisto dei singoli consumatori».

[75] Cfr. M. TAMPIERI, *L'identità personale: il nostro documento esistenziale*, in *Europa e Diritto Privato*, 2019, p. 1195. G. ALPA, *Tecnologie e diritto privato*, in *Riv. it. scienze giuridiche*, 2018, p. 276, afferma che «la rivoluzione digitale e soprattutto la tecnologia informatica hanno cambiato la nozione di soggetto, non tanto nella sua qualità di titolare di diritti e doveri, quanto piuttosto nel modo di rappresentarlo, di tradurlo in termini giuridici. Oggi la persona presenta connotazioni nuove ed appare come un complesso di dati tradotti in algoritmi».

[76] Su questi aspetti si rimanda a G. RESTA *Identità personale e identità digitale,* in *Dir. informatica*, 2007, p. 511, ove si specifica che qualsiasi discorso sull'identità digitale deve toccare necessariamente due aspetti: quello della tutela dell'*identità* personale in rete (specie nei suoi profili reputazionali) e quello delle tecniche di *identificazione* del soggetto a mezzo di strumenti informatici.

[77] Si vedano *Court of Justice of the European Union*, 24 September 2019, C-507/17, *Google v Commission nationale de l'informatique et des libertés (CNIL)* e *Court of Justice of the European Union*, 3 October 2019, C-18/18, *Eva Glawischnig-Piesczek v. Facebook Ireland*, per un commento alle quali si rimanda a G. DE GREGORIO, *Google v. CNIL and Glawischnig– Piesczek v. Facebook: content and data in the algorithmic society,* in *Rivista diritto dei media,* 2020, p. 249. Sulla necessità di tutelare il consumatore che deve essere reso edotto dello scambio di prestazioni che è sotteso alla adesione ad un contratto per la fruizione di un servizio, quale è quello di utilizzo di un *social network*, si veda Tar Lazio, Sez. I, sentenza n. 261 del 18 dicembre 2019 – 10 gennaio 2020, n. 261. Per un primo commento a questa sentenza si veda M. GIANNINI, *Dati personali e valore economico – il Tar Lazio conferma l'importante provvedimento dell'Antitrust nel caso Facebook* in *Persona e mercato, Osservatorio OGID,* 2020, p. 7.

teplicità di banche dati [78]. Si ha «una raffigurazione parziale e potenzialmente pregiudizievole della persona, la quale verrebbe così ridotta alla mera sommatoria delle sue proiezioni elettroniche» [79].

Le persone saranno sempre più soggette ad azioni realizzate e a decisioni prese da sistemi di IA o con la loro assistenza [80] ed esposte al rischio di essere sovraccaricate di sollecitazioni artificiali. A questi profili dedica una particolare attenzione il cd. *Shaping europe's Digital Future*, approvato dalla Commissione europea il 19 febbraio 2020 [81], ove si prevedono gli obiettivi chiave sui quali la Commissione si concentrerà nei prossimi cinque anni per aiutare l'Europa a perseguire la propria strada verso una trasformazione digitale a beneficio delle persone e si auspica, in particolare, il raggiungimento di *balance between online and offline commerce*, facendo sì che le regole che si applicano *offline* valgano anche *online*.

Si occupa del processo decisione automatizzato, anche al fine di proteggere le persone dalla discriminazione, l'art. 22 del Regolamento generale sulla protezione dei dati (UE/2016/679), cd. GDPR [82]. In linea di principio, tale norma vieta determinate decisioni

---

[78] G. RESTA, *Identità personale*, cit., p. 518.

[79] G. RESTA, *Identità personale,* cit., p. 518 e ss. L'identità, afferma l'A., p. 524 «non viene più vista come dato preesistente (ossia come proiezione esterna di un patrimonio individuale già delineato nelle sue caratteristiche distintive), bensì come processo, costantemente in atto, aperto ad una pluralità di esiti e capillare e pervasiva, delle varie forme di potere sociale. Rispetto a tale processo l'ordinamento non si limita più ad una posizione di astensione e non interferenza − come predicato dalla lettura liberale classica dei diritti fondamentali − ma esercita un ruolo attivo di supervisione e controllo, con l'obiettivo di restituire il più possibile all'individuo la capacità di perseguire politiche dell'identità personale liberamente definite, sottraendolo al rischio della normalizzazione».

[80] Diventa ancora più attuale l'insegnamento in base al quale occorre una strategia istituzionale che metta tutti e ciascuno in condizione d'esercitare concretamente la propria autonomia. Il riferimento è a S. RODOTÀ, *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, II ed., Roma-Bari, 2004, spec. p. 22-23.

[81] Si tratta di COM (2020) 67 final, in *www.ec.europa.eu*. *Ivi*, p. 5 ci si propone di realizzare una nuova agenda per i consumatori "che consentirà ai consumatori di fare scelte informate e giocare un ruolo attivo nella trasformazione digitale", mettendoli in condizione di fidarsi dei prodotti e dei servizi digitali tanto quanto farebbero con qualsiasi altro prodotto e garantendo che il ruolo sistemico di alcune piattaforme *online* e il potere di mercato che acquisiscono non metta in pericolo l'equità del mercato. Particolare attenzione si pone anche per i «new digital business models – such as "free" services that users access while providing their data – and their implications for competitive constraints» e si aggiunge che «the ongoing fitness check of the Commission's 2014 Important Projects of Common European Interest (IPCEI) Communication is designed to assess whether an update is necessary to further clarify the conditions under which major Member State-led projects in key, strategic sectors for the digital and green future of Europe can proceed effectively».

[82] Cfr. I. MENDOZA, L.A. BYGRAVE, *The Right Not to Be Subject to Automated Decisions Based on Profiling*, in T. SYNODINOU, P. JOUGLEUX, C. MARKOU e T. PRASTITOU (a cura di), *EU Internet*

completamente automatizzate con effetti significativi. Si legge, infatti, al primo comma, che «l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona» [83].

La norma non è esente da critiche. A parte il dibattitto sulla possibilità e l'utilità della *legibility* del procedimento algoritmico [84], è stata evidenziata la minore tutela offerta dall'art. 22 GDPR rispetto all'art. 9 del *Data Protection Convention 108,* che non limita, ad una decisione basate unicamente sul trattamento automatizzato, il diritto «to obtain, on request, knowledge of the reasoning underlying data processing where the results of such processing are applied to him or her» [85].

Il dato è di particolare rilievo proprio con riferimento alla tutela dalla non discriminazione in quanto i rischi di discriminazione sono simili, sia per le decisioni totalmente automatizzate, sia per quelle che lo sono solo in parte, e questo a causa del cd. "pregiudizio dell'automazione" o "*automation bias*" e della tendenza umana a farsi influenzare dalla decisione "algoritmicamente preparata" [86].

Law, Cham, 2017, p. 77. *Ivi*, in particolare, viene svolta un'analisi critica dell'art. 22 del Regolamento generale sulla protezione dei dati dell'Unione Europea del 2016, nel confronto con l'art. 15 della Direttiva sulla protezione dei dati del 1995. L'articolo 22 impone limiti nel prendere decisioni completamente automatizzate basate sulla profilazione quando tali decisioni comportano effetti giuridici o conseguenze altrettanto significative per le persone a loro soggette e, secondo gli A., pur offrendo alle persone una protezione più forte rispetto all'art. 15 della direttiva del 1995, è dubbio che possa avere un impatto pratico significativo sulla profilazione automatizzata.

[83] Il divieto di decisioni basate unicamente sul trattamento automatizzato non è assoluto, ci sono delle eccezioni, specificate al comma 2 dell'art. 22, ossia qualora tale decisione a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato; c) si basi sul consenso esplicito dell'interessato.

[84] Su dibattitto innescato dall'art. 22 del GDPR riguardo la spiegabilità delle decisioni algoritmiche, si vedano G. MALGIERI e G. COMANDÉ, *Why a Right to Legibility of Automated Decision-making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, 2017, p. 243.

[85] M. VEALE e L. EDWARDS, *Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling*', in *Computer Law & Security Review,* 2018, p. 398. Cfr. F.J. ZUIDERVEEN BORGESIUS, *Strengthening legal protection against discrimination by algorithms,* cit., p. 11; E. PELLECCHIA, *Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation*, in *Leggi civ. comm.*, 2018, p. 1224; F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione,* Torino, 2018.

[86] Cfr. I.J. SKITKA, K.L. MOSIER, M. BURDICK, *Does automation bias decision-making?,* in *International Journal of Human-Computer Studies,* 1999, p. 991. Agli autori si rimanda per l'analisi di alcuni sistemi computerizzati di supporto e aiuto alle decisioni umane in alcuni contesti comuni, nel confronto con decisioni non supportate.

Il percorso che porta alla tutela dalla non discriminazione algoritmica sembra, dunque, ancora, molto lungo.

5. – Nell'uso dei sistemi di IA si riscontrano sia nuovi rischi, sia l'esponenziale proliferare di rischi tradizionali in settori delicatissimi della vita delle persone.

Diventa evidente, allora, la centralità dei problemi connessi ai risultati dell'attività di un algoritmo ed alle conseguenti responsabilità. Esistono essenzialmente tre modi per affrontare i dubbi legati alla responsabilità nei casi di utilizzo di dispositivi di AI. La soluzione più semplice sarebbe «AI-enabled devices can be treated as property and therefore be the responsibility of their users, owners, or manufacturers» [87]. Ciò, non è, comunque, senza difficoltà, in quanto occorre interrogarsi su «how to apportion liability among the manufacturer, programmer and trainer of the AI» [88]; cosa non facile nella pratica se solo si pensa alla impossibilità, spesso, di tracciare le fonti dei dati utilizzate da un algoritmo o alla circostanza che spesso i dati sono forniti da più soggetti o da terze parti indipendenti, molte delle quali derivate da altri dispositivi AI. La seconda opzione è che i sistemi di IA potrebbero essere trattati come entità semi-autonome, un pò come previsto per i soggetti che hanno una limitata capacità di agire; la terza è di considerare i dispositivi di intelligenza artificiale come entità autonome [89]. Quest'ultima opzione presuppone che si sia d'accordo nel ritenere che i dispositivi di intelligenza artificiale possano essere «"beings" deserving of independent legal status» [90], sol perché sono entità capaci di condurre analisi probabilistiche.

Il dato, come abbiamo visto, non è così scontato [91]. D'altronde per attribuire la respon-

---

[87] I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 763.

[88] I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 764.

I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 763.

[90] I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 765.

[91] Per un quadro delle diverse tesi sul punto, con riferimento specifico alle ricadute sul tema della responsabilità, cfr. A. SANTOSUOSSO, *Diritto, scienza, nuove tecnologie*, Padova, 2016, p. 321 e ss.; A. SANTOSUOSSO, C. BOSCARATO, F. CAROLEO, *Robot e diritto: una prima ricognizione*, in *Nuova giur. civ. comm*., 2012, p. 494 e ss.; E. PALMERINI, A. BERTOLINI, *Liability and risk management in robotics*, in R. SCHULZE e D. STAUDENMAYER (a cura di), *Digital Revolution: Challenges for Contract Law in Practice*, Baden, 2016, p. 239 e ss.

sabilità, anche agli essere umani, non basta certo "l'intelligenza" e, quindi, «"intelligence" is not enough for personhood, at least in most jurisdictions. Rather, the test for capacity is that of reason; a person has to be endowed with reason to be held civilly or criminally liable, to enter into a contract, or to exercise other forms of legal autonomy» [92].

La non riconducibilità del danno al sistema di intelligenza artificiale, che deve, comunque, essere in grado di eseguire gli ordini umani, sembra la direzione nella quale muove la *Proposition de loi constitutionnelle relative à la Charte de l'intelligence artificielle et des algorithmes*, depositata in Francia il 15 gennaio 2020. Il secondo articolo di questa proposta prevede, infatti, che «ne peut porter atteinte à un être ou un groupe d'êtres humains, ni, en restant passif, permettre qu'un être ou un groupe d'êtres humains soit exposé au danger; doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec le point précédent; – doit protéger son existence tant que cette protection n'entre pas en conflit avec les deux points précédents». Il dato è particolarmente interessante anche perché, coerentemente, muove in direzione opposta alla riconoscibilità della personalità giuridica ai sistemi di Intelligenza artificiale; possibilità in direzione della quale sembrava muoversi invece la Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica [93].

---

[92] Cfr. I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 766. *Ivi* si afferma che «that intelligence is no more than the capacity to conduct probabilistic analysis and that intelligence is perceived as the main criteria to stablish legal capacity». *Ivi* si aggiunge che «as Erich Fromm put it: Reason is man's faculty for grasping the world by thought, in contradiction to intelligence, which is man's ability to manipulate the world with the help of thought. Reason is man's strument for arriving at the truth, intelligence is man's instrument for manipulating the world more successfully; the former is essentially human, the latter belongs to the animal part of man. Whether or not one agrees with Fromm's postulate, it remains undeniable that reason and intelligence are intrinsically linked and that true "intelligence," for lack of a better word, is more than computing capacities, no matter how sophisticated. A case in point: individuals suffering from savant syndrome. Brought to public consciousness through Dustin Hoffman's character in the 1988 film "Rain Man" savant syndrome "is a rare, but extraordinary, condition in which persons with serious mental disabilities, including autistic disorder, have some 'island of genius' which stands in marked, incongruous contrast to overall handicap." Individuals afflicted with this condition will often display impressive calculating abilities, yet can still be considered legally incompetent».

[93] Sul punto si veda D. IMBRUGLIA, *Intelligenza Artificiale e Costituzione francese,* in *Persona e Mercato, Osservatorio OGID,* 2020, p. 8. Sulla Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica si rimanda a L. DI DONNA, *Diritto e tecnologia. Il contratto ai tempi dell'intelligenza artificiale*, in F. CAPRIGLIONE (a cura di), Liber amicorum *Guido Alpa*, cit., p. 235 e ss.; A. AMIDEI, *Robotica intelligente e responsabilità: profili e prospettive evolutive del quadro normativo europeo,* in U. RUFFOLO, *Intelligenza artificiale,* cit., p. 63 e ss.

Ciò posto, anche se esiste un vasto *corpus* di norme europee in materia di sicurezza dei prodotti e di responsabilità per danno da prodotti difettosi che, in linea di principio è pienamente applicabile anche all'IA [94], è importante valutare entro quali limiti tale normativa possa essere adeguatamente applicata per far fronte ai rischi specifici derivanti dai sistemi di IA [95], anche in considerazione del fatto che, tradizionalmente, i nostri sistemi legali tendono ad essere reattivi e non proattivi [96]. Tipi diversi di rischi dovrebbero essere oggetto di tipi diversi di controlli. Potrebbe non bastare, dunque, una disciplina generale, occorrendo anche delle specifiche disposizioni nei diversi settori [97], così da ga-

---

[94] Il quadro vigente in materia di sicurezza e di responsabilità in relazione a prodotti e ai servizi è ricostruito in COM (2020) 42 final.

[95] Cfr. COM (2020) 65 final, cit., p. 15. Si veda, altresì, Com (2018) 237 final, cit., p. 2, ove si legge che «the liability framework that is currently existing in the European Union – as will be described further in this document – is a stable framework that incites investment, innovation and risk-taking. Nevertheless, a reflection on future needs and developments is needed, not only from the perspective of the victim i.e. in order to ensure equitable remedies, compensation and allocation of responsibility, but also from the perspective of the innovators and companies operating in the EU as legal certainty is a key element for good business development». Su tali problematiche cfr. C. DE MEEUS, *The Product Liability Directive at the Age of the Digital Industrial Revolution: Fit for Innovation?,* in *Journal of European Consumer and Market Law,* 2019, p. 149 e ss.; G. HOWELLS, C. TWIGG-FLESNER, G. WILLETT, *Protecting the Values of Consumer Law in the Digital Economy: The case of 3D-printing,* in A DE FRANCESCHI, R. SCHULZE, M. GRAZIADEI, O. POLLICINO, F. RIENTE, S. SICA, P. SIRENA (a cura di), *Digital Revolution – Challenges for Law. Data Protection, Artificial Intelligence, Smart Products, Blockchain Technology and Virtual Currencies*, München, 2019, p. 214 e ss. *Ivi*, si propone, un framework «involves: (i) identifying the underpinning values of an area of law; (ii) identifying how these values manifest in specific rules; (iii) asking whether there is a reason to see these values changed by considering whether risks are lesser or greater than before, whether parties are more or less able to look after themselves and whether innovation is seriously threatened by sticking to the traditional values; (iv) if there are good reasons to stick to the traditional values, considering whether the existing specific rules can apply to new circumstances created by the innovations, either with no change or with adaptation or development that does not fundamentally change the legal regime; and (v) if radical regime change is needed to reflect traditional values, finding solutions that are workable and proportionate.

[96] Tre in particolare sono i punti cardine individuati per una legislazione che si adegui alle caratteristiche dell'IA: 1) an AI system must be subject to the full gamut of laws that apply to its human operator; 2) an AI system must clearly disclose that it is not human; 3) n AI system cannot retain or disclose confidential information without explicit approval from the source of that information. In questi termini O. ETZIONI, *How to Regulate Artificial Intelligence*, N.Y. Times, 2017, p. 1, in *www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html.*

[97] Nel settore dell'attività contrattuale, si ritiene che i sistemi di IA riducano i rischi legati all'esecuzione del contratto, in quanto si viene a creare una situazione simile a quella derivante dalla clausola *solve et repete*. Su queste problematiche si rinvia a D. DI SABATO, *Gli smart contracts: robot che gestiscono il rischio contrattuale*, in *Contratto e Impresa*, 2017, 378; T. PELLEGRINI, *Prestazioni auto-esecutive. Smart contract e dintorni*, in *Comparazione e diritto civile*, 2019, 843 e ss. Più in generale, sull'impatto delle nuove tecnologie nella formazione del contratto, si veda M. DUROVIC, A. JANSSEN, *The Formation of Blockchain-based Smart Contracts in the Light of Contract Law,* in *European Review of Private Law,* 2018,

rantire la proporzionalità dell'intervento normativo, distinguendo tra le diverse applicazioni di IA, quali siano o meno "ad alto rischio" [98].

Parte della dottrina ritiene che occorre tenere conto dei rischi che derivano dall'uso di strumenti di intelligenza artificiale e chiedersi se rientrino, o meno, nell'ambito dei rischi presupposti, e adeguatamente affrontati, dal regime normativo esistente. In particolare, si è proposto di tenere conto di cinque interrogativi: *1) What rights are you trying to protect? 2) What are the risks that AI poses to these rights? 3) How well does current legislation mitigate those risks? 4) What risks would the application of current legislation to AI cause? 5) What costs and trade-offs does current legislation impose? 1) What rights are you trying to protect?* [99].

Nonostante il quadro giuridico vigente in materia di sicurezza dei prodotti contempli un concetto ampio di sicurezza, non si può dire che siano coperti tutti rischi derivanti dalle tecnologie digitali emergenti [100]. Basti pensare che la normativa generale dell'UE in

---

p. 753. Con riferimento a virtù e limiti delle innovazioni tecnologiche cfr. R. PARDOLESI, A. DAVOLA, «Smart contract»: *lusinghe ed equivoci dell'innovazione purchessia,* in F. CAPRIGLIONE (a cura di), *Liber amicorum Guido Alpa,* Milano, 2019, p. 297 e ss.

[98] Si veda, per esempio, COM (2020) 65 final, cit., p. 19, ove si specifica che occorre valutare gli interessi in gioco e considerare se il settore interessato e l'uso previsto per tale applicazione implichino rischi significativi, in particolare per quanto concerne la protezione della sicurezza, dei diritti dei consumatori e dei diritti fondamentali. Più specificamente, un'applicazione di IA dovrebbe essere considerata ad alto rischio se soddisfa due criteri cumulativi: 1) l'applicazione di IA è utilizzata in un settore in cui, date le caratteristiche delle attività abitualmente svolte, si possono prevedere rischi significativi; 2) l'applicazione dell'IA nel settore in questione è inoltre utilizzata in modo tale da poter generare rischi significativi. Questo secondo criterio riconosce il fatto che non tutti gli usi dell'A nei settori selezionati comportano necessariamente rischi significativi. Quanto premesso non esclude che si tenga in considerazione che possono esistere anche casi eccezionali in cui, a causa dei rischi in gioco, l'uso di applicazioni di IA per determinati scopi deve essere considerato ad alto rischio di per sé, ossia indipendentemente dal settore interessato. Sul punto si veda, altresì, la Risoluzione del Parlamento europeo sui processi decisionali automatizzati, cit., ove si pone l'accento sulla necessità di un approccio alla regolamentazione basato sul rischio, alla luce dell'eterogeneità e della complessità delle sfide poste dalle diverse tipologie e applicazioni dell'IA e dei sistemi decisionali automatizzati e si invita la Commissione a elaborare un modello di valutazione dei rischi per l'IA e dei processi decisionali automatizzati, al fine di garantire un approccio coerente all'applicazione della normativa sulla sicurezza dei prodotti nel mercato interno.

[99] I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 776.

[100] Il dato emerge chiaramente dalla proposta di Risoluzione del Parlamento europeo sui processi decisionali automatizzati, cit., ove si legge che «occorre esaminare l'attuale quadro giuridico dell'UE, compresi l'*acquis* in materia di diritto dei consumatori, la legislazione in materia di protezione dei dati e la normativa sulla sicurezza dei prodotti e la vigilanza del mercato, nell'ottica di verificare se è in grado di rispondere all'emergere dell'IA e dei processi decisionali automatizzati e di fornire un elevato livello di protezione dei consumatori, come richiesto dall'articolo 38 della Carta dei diritti fondamentali dell'Unione europea». Sul ruolo del regime europeo della responsabilità da prodotto difettoso nella sfida dell'innovazione tecnologica si veda F.P. PATTI, *The european road to autonomousvehicles,* in *Fordham Int. Law J.,* 2019, p. 125 e ss.

materia di sicurezza, per esempio, si applica ai prodotti ma non ai servizi, per cui in linea di principio non si applica nemmeno ai servizi basati sulla tecnologia di IA (ad esempio servizi sanitari, finanziari e di trasporto), anche se non manca a livello europeo una dettagliata normativa in materia di servizi [101] arricchita, di recente, dal Regolamento 2019/1150 relativa ai servizi di intermediazione *online* [102].

Inoltre, se conformemente alla legislazione dell'UE in materia di sicurezza dei prodotti, il *software*, quando è parte del prodotto finale, deve rispettare le norme pertinenti in materia di sicurezza dei prodotto, rimane da stabilire se il *software* indipendente (*stand-alone*) rientri nell'ambito di applicazione della normativa dell'UE in materia di sicurezza dei prodotti, al di là di alcuni settori in cui esistono norme esplicite al riguardo [103]. L'integrazione del *software,* infatti, può modificare il funzionamento dei sistemi di IA durante il loro ciclo di vita e richiedere frequenti aggiornamenti del *software* [104].

---

[101] Nella proposta di Risoluzione del Parlamento europeo sui processi decisionali automatizzati, cit., si evidenzia che il quadro normativo europeo in materia di servizi dovrebbe applicarsi sia ai servizi tradizionali che a quelli che integrano processi decisionali automatizzati e si sottolinea che, sebbene i processi decisionali automatizzati possano migliorare l'efficienza e la precisione dei servizi, gli esseri umani devono sempre essere responsabili, in ultima istanza, delle decisioni prese nell'ambito di servizi professionali quali le professioni mediche, forensi e contabili, e nel settore bancario, nonché essere in grado di revocare tali decisioni, sottolineando l'importanza del controllo o della vigilanza indipendente da parte di professionisti qualificati nel caso di processi decisionali automatizzati dove sono in gioco legittimi interessi pubblici.

[102] Cfr. il *Report of the European Law Institute Model Rules on Online Platforms* dell'*European Law Institute,* in www.europeanlawistitute.eu.

[103] Il tema dei limiti della disciplina europea in tema di prodotti difettosi ad affrontare le problematiche poste dalle nuove tecnologie sono approfonditi in SWD (2018) 137 final, cit., p. 2, ove si legge che «the context of theemerging digital technologies, it may be difficult to identify whether the damage hasbeen caused by the product itself or by other elements interconnected to it in a digital ecosystem. In this respect, it will be necessary to provide for adequate safety levels for all types of products, taking also account of any new risks that may be posed regarding the emerging digital technologies». Cfr. su queste problematiche C. TRIBERTI, M. CASTEKKANI, *L'intelligenza artificiale oltre le quattro leggi della robotica. Riflessioni anche alla luce della pandemia da Covid-19,* Firenze, 2020. Sul punto si veda anche COM (2020) 61 final, p. 16, ove si precisa che nonostante l'ampiezza della definizione di prodotto fornita dalla direttiva sulla responsabilità per danno da prodotti difettosi, il suo ambito di applicazione andrebbe ulteriormente chiarito per rispecchiare meglio la complessità delle tecnologie emergenti e garantire che il risarcimento sia sempre possibile per i danni causati da prodotti difettosi a causa del *software* o di altre caratteristiche digitali. Si consentirebbe in tal modo agli operatori economici, quali gli sviluppatori di *software*, di valutare se possono essere considerati produttori ai sensi della direttiva sulla responsabilità per danno da prodotti difettosi. Sul *robot* come prodotto si veda G. CAPILLI, *La responsabilità per la produzione di robot*, in G. ALPA (a cura di), *La responsabilità del produttore,* Miano, p. 625 e ss.

[104] Cfr. A. MATTHIAS, *The responsibility gap: Ascribing responsibility for the actions of learning automata*, in *Ethics and Information Technology*, 2004, p 182. Sul *Robot* come prodotto cfr. U. RUFFOLO, *Per un fondamento di una robotica* self-learning; *dalla* machinery *produttiva all'auto* driverless: verso una "responsabilità da algoritmo"?, in U. RUFFOLO (a cura di), *Intelligenza artificiale e responsabilità*, Milano, 2017, p. 21 e ss.

L'autonomia dei sistemi di IA è una delle caratteristiche principali dell'intelligenza artificiale; caratteristica che introduce un rischio specifico e mette alla prova la legislazione europea in materia di sicurezza dei prodotti. Il dato è evidenziato chiaramente da COM (2020) 64 final, ossia dalla Relazione della Commissione del 19.2.2020 sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità [105]. *Ivi* si propone, in considerazione del fatto che i rischi da "autoaprendimento" non sono disciplinati dalla legislazione vigente [106], di introdurre *de iure condendo*, oltre alla valutazione del rischio effettuata prima dell'immissione sul mercato del prodotto, una ulteriore valutazione del rischio da effettuarsi in ipotesi di modifiche importanti intervenute durante il ciclo di vita del prodotto e tali da implicare, per esempio, una funzione diversa dello stesso e non prevista dal fabbricante nella valutazione iniziale del rischio [107].

La precisione di un algoritmo, inoltre, dipende sia dalla programmazione, sia dai dati. Le decisioni prese dagli algoritmi potrebbero, come evidenziato anche prima, basarsi su dati incompleti e quindi non affidabili, manomessi a seguito di attacchi informatici, inficiati da condizionamenti o semplicemente errati [108].

Il concetto tradizionale di sicurezza deve confrontarsi, inevitabilmente, con un'altra caratteristica fondamentale di un sempre crescente numero di prodotti e servizi che si avvalgono dell'intelligenza artificiale e cioè con la connettività. Quest'ultima può compromettere la sicurezza del prodotto non solo indirettamente, e cioè in caso di attacco informatico che rappresenti una minaccia alla *security* e crei un pregiudizio per la sicurezza degli utilizzatori, ma anche direttamente. Gli esempi potrebbero essere tanti, ma basti

---

[105] COM (2020) 64 final. Sulla rappresentata da danni provocati da processi decisionali autonomi, si sofferma anche Proposta di Risoluzione del Parlamento europeo sui processi decisionali automatizzati, cit., ove si invita la Commissione a valutare la possibilità di adattare, ai nuovi sviluppi tecnologici, i concetti come "prodotto", "danno" e "difetto" che si rinvengono nella legislazione vigente.

[106] Cfr. COM(2020) 65 final, cit., p. 16.

[107] COM (2020) 64 final, p. 8, ove si afferma che nella misura in cui il "comportamento" futuro dei prodotti basati sull'intelligenza artificiale può essere determinato in anticipo mediante la valutazione del rischio effettuata dal fabbricante prima dell'immissione sul mercato, il quadro dell'Unione in materia di sicurezza dei prodotti può essere adeguato, in quanto è già previsto l'obbligo per i produttori di tener conto, nella valutazione dei rischi, dell'uso dei prodotti per tutto il loro ciclo di vita Non si riscontra la stessa adeguatezza della legislazione vigente, nei casi in cui l'uso previsto e inizialmente progettato dal fabbricante è modificato dal comportamento autonomo del prodotto capace di autoapprendimento; comportamento che può comprometterne la conformità ai requisiti di sicurezza.

[108] Cfr. la *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità,* allegata a COM(2020) 65 final.

pensare ad un allarme antincendio che, a causa della perdita di connettività, non avverta l'utilizzatore in caso di incendio [109].

Si aggiunga che i sistemi di IA funzionano in un «technological ecosystem» in quanto «one cannot speak of AI without taking into account all of the other associated technologies and the ways in which they all interact» [110]. Tale complessità dei prodotti, che deriva dal loro essere connessi e dal poter interagire, per effetto della connessione, con altri prodotti, è un altro dei rischi specifici nuovi rispetto a quelli tradizionali [111].

L'individuazione dei rischi specifici dei sistemi di IA presuppone una adeguata conoscenza dei sistemi stessi e del loro modo di operare, così che la loro opacità è uno dei problemi da superare. In questa direzione sembra muovere, per esempio, il Regolamento (UE) 2017/589 sulle norme tecniche in tema di requisiti organizzativi delle imprese di investimento che effettuano la negoziazione algoritmica, ove si attribuisce un ruolo strategico al *self-assessment,* cioè all'autovalutazione che dovrebbe essere effettuata con regolarità e dovrebbe consentire all'impresa di investimento di comprendere appieno i sistemi, gli algoritmi di negoziazione da essa utilizzati e i rischi ad essa connessi [112].

Anche sull'opacità, come rischio specifico relativo ai sistemi di IA e tale da mettere in discussione il tradizionale concetto di sicurezza dei prodotti, si sofferma COM (2020) 64 final, ove emerge che, se non è necessario che gli esseri umani comprendano ogni

---

[109] COM (2020) 64 final, p. 6.

[110] I. GIUFFRIDA, F. LEDERER e N. VERMERYS, *A Legal Perspective on the Trials and Tribulations of AI,* cit., p. 756. Sulle questioni relative alla crittografia dei dati e alla sicurezza informatica con l'aumentare dell'automazione e dell'interconnettività, con particolare riferimento alle funzioni di guida, si rimanda a S. LANDINI, *Ethical Issues, Cybersecurity and Automated Vehicles. In: Kyriaki Noussia – Pierpaolo Marano. InsurTech: A Legal and Regulatory View*, Cham, 2020, p. 291 e ss.

[111] Cfr. COM 2020 64 final ove si legge, p. 2, che «l'intelligenza artificiale, *l'Internet* delle cose e la robotica hanno molte caratteristiche in comune. Consentono di combinare connettività, autonomia e dipendenza dai dati per svolgere compiti con un livello minimo o nullo di controllo o supervisione umani. I sistemi dotati di intelligenza artificiale possono inoltre migliorare le proprie prestazioni apprendendo dall'esperienza. La loro complessità si riflette sia nella pluralità degli operatori economici partecipanti alla catena di approvvigionamento che nella molteplicità di componenti, parti, software, sistemi o servizi, che insieme formano i nuovi ecosistemi tecnologici. A ciò si aggiunge l'apertura agli aggiornamenti e ai miglioramenti dopo l'immissione sul mercato. La grande quantità di dati necessari, la dipendenza da algoritmi e l'opacità del processo decisionale dell'intelligenza artificiale rendono più difficile prevedere il comportamento dei prodotti basati sull'intelligenza artificiale e comprendere le possibili cause di un danno. Infine, la connettività e l'apertura possono anche esporre i prodotti basati sull'intelligenza artificiale e sull'*Internet* delle cose a minacce informatiche».

[112] Si veda il considerando 8) del Regolamento delegato (UE) 2017/589 sulle norme tecniche di regolamentazione per specificare i requisiti organizzativi delle imprese di investimento che effettuano la negoziazione algoritmica. Sul punto cfr. F.J. ZUIDERVEEN BORGESIUS, *Strengthening legal protection against discrimination by algorithms,* cit., p. 12.

singola fase del processo decisionale, è fondamentale, però, che possano capire come il sistema abbia preso le decisioni algoritmiche. Questo aspetto è particolarmente importante per il meccanismo *ex post* di controllo del rispetto delle norme «in quanto consente alle autorità preposte di risalire alla responsabilità dei comportamenti e delle scelte dei sistemi di intelligenza artificiale» [113].

L'opacità dei sistemi di IA pone un altro aspetto problematico, legato alla difficoltà di risalire alle decisioni adottate da sistemi di IA [114]. Conseguentemente, coloro che hanno sofferto danni potrebbero non avere accesso effettivo agli elementi probatori necessari per giustificare un'azione in giudizio, per cui le loro possibilità di ottenere un'effettiva riparazione del danno potrebbero essere inferiori rispetto alle situazioni in cui il pregiudizio è causato da tecnologie tradizionali [115]. Ne consegue la possibilità che i costi per le vittime aumentino in maniera significativa, la potenziale difficoltà di avviare azioni per responsabilità nei confronti di soggetti diversi dal produttore e di ottenere elementi di prova a sostegno di tali azioni.

Un approccio utile a risolvere i problemi evidenziati potrebbe essere quello che utilizza il criterio dell'*accountability,* prescindendo dagli elementi soggettivi della responsabilità e financo dall'individuazione dell'errore, mutuando, e adattando, il modello di responsabilità presente nel GDPR. *Ivi*, come noto, si impone al titolare del trattamento di adottare misure giuridiche, organizzative e tecniche a protezione dei dati personali e di dimostrarne, su richiesta, l'efficacia e l'effettiva attuazione [116].

Si tratta, cioè, di «superare il paradigma basato sull'errore e sulla colpa e, invece, af-

---

[113] Cfr. COM (2020) 64 final, p. 10.

[114] In COM (2020) 65 final, cit., p. 26 si precisa che un primo problema è ripartire gli obblighi tra gli operatori economici coinvolti. Il ciclo di vita di un sistema di IA coinvolge numerosi operatori, tra cui lo sviluppatore, il soggetto che lo applica (il "*deployer*", ossia la persona che utilizza un prodotto o servizio dotato di IA) e potenzialmente altri soggetti (il produttore, il distributore o l'importatore, i prestatori di servizi, gli utenti professionali o privati). Ciascun obbligo deve essere stabilito a carico dell'operatore o degli operatori che si trovano nella posizione migliore per affrontare eventuali rischi potenziali. Ad esempio, se da un lato gli sviluppatori dell'IA sono i più qualificati per affrontare i rischi derivanti dalla fase di sviluppo, dall'altro la loro capacità di controllare i rischi durante la fase di utilizzo può essere più limitata. In tal caso il pertinente obbligo dovrebbe essere stabilito a carico del soggetto che applica l'IA. In secondo luogo va affrontata la questione dell'ambito di applicazione geografico dell'intervento legislativo, in quanto è fondamentale che le prescrizioni siano applicabili a tutti gli operatori economici interessati che forniscono prodotti o servizi basati sull'IA nell'UE, indipendentemente dal fatto che siano stabiliti o meno nell'Unione; in caso contrario, gli obiettivi dell'intervento legislativo potrebbero non essere pienamente raggiunti.

[115] Cfr. COM(2020) 65 final, cit., p. 14.

[116] In questi termini G. FINOCCHIARO, *Intelligenza artificiale e diritto – intelligenza artificiale e protezione dei dati personali,* in *Giur. it.*, 2019, p. 1676.

frontare il problema sotto il profilo dell'allocazione del rischio» [117]. Il diritto davanti al rischio non può che mostrare i suoi limiti, in quanto può intervenire o prima dell'azione, impedendo che sia compiuta, ma prima di sapere se l'eventualità del danno si sarebbe verificata oppure quando ormai l'evento si è verificato e, quindi non c'è più il rischio. Dunque, ciò che si può fare non è evitare il rischio ma, soltanto, canalizzare le conseguenze dannose. Il rischio è un concetto legato alla temporalità e al futuro ed è correlato al non sapere del futuro e alla necessità di scegliere. Si tratta, cioè, di un modo per costruire vincoli per il futuro, una strategia per imputare un danno futuro ad una decisione [118]. Se è vero che il rischio è correlato al non sapere, è vero anche che non può non es-

[117] G. FINOCCHIARO, *op. loc. cit.* Come noto ad aprire la strada del percorso che porta a ridimensionare il ruolo della colpa nel sistema della responsabilità extracontrattuale sono stati P. Trimarchi, nel 1960, con *Rischio e responsabilità oggettiva*, e S. Rodotà, nel 1964, con *Il problema della responsabilità civile.* Anche gli anni '70 sono stati fiorenti di studi sul punto. Si pensi, sull'imputazione del rischio all'impresa, con tecniche del tutto svincolate da ogni valutazione in chiave soggettiva, ad G. ALPA, *Responsabilità dell'impresa e tutela del consumatore*, Milano 1975, spec. p. 375 e ss.; quanto alle problematiche sul punto, con particolare riferimento al rapporto tra produttore e consumatore, si rimanda a C. CASTRONOVO, *Problema e sistema nel danno da prodotti,* Milano 1979, spec. p. 774 e ss. C'è da chiedersi se i problemi legati all'intelligenza artificiale daranno un maggiore impulso alla responsabilità oggettiva, considerata, invero, un pò in declino. Sulla "salute" della responsabilità oggettiva oggi Cfr. V. ROPPO, *La responsabilità civile di Pietro Trimarchi*, in *Juscivile*, 2017, p. 698. L'A. afferma che la sua impressione è che la responsabilità oggettiva abbia conosciuto un certo declino o appannamento, e sia entrata un po' in un cono d'ombra. Il dato è messo in evidenza anche fa F.P. PATTI, *Il declino della responsabilità oggettiva (a margine dell'art. 2051 c.c.),* in *Riv. dir. civ*., 2019, p. 978 e ss. All'A.si rimanda anche per l'accurata analisi del contesto europeo nel quale si è manifestato «un certo disfavore per modelli di responsabilità oggettiva che non riguardino attività pericolose o cose difettose».

[118] Si rinvia sul punto all'insegnamento di N. LUHMANN, *Soziologie des Risikos,* Berlino, 1991, trad. it., *Sociologia del rischio,* Milano, 1996. L'A. afferma, p. 14 e ss., che se si cercano definizioni del concetto di rischio ci si trova subito nella nebbia fitta e si ha l'impressione di non poter vedere al di là del proprio naso. Basti pensare che, con riferimento al tema del rischio, si pone il problema di quale idea di razionalità, di decisione, di tecnica, di futuro o semplicemente di tempo sia presupposta. Di aiuto può essere la distinzione tra rischio e pericolo che presuppone «l'incertezza in riferimento a dei danni futuri. Ci sono allora 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». Di sicuro interesse è anche il superamento di una definizione del rischio come «misura» da determinare opponendola a quella di sicurezza, sulla base di un ragionamento che analizza «in maniera ragionieristica l'aspirazione alla sicurezza e la misura di ciò che può essere ragionevolmente raggiunto». Sul punto, l'A., p. 15, afferma che «se si tratta soltanto di un problema di misura, non si capisce proprio il perché di tanto rumore. I problemi di misurazione sono problemi di convenzione e in ogni caso i rischi della misurazione sono qualcosa d'altro rispetto a ciò che viene misurato come rischio». L' impossibilità di enunciare principi generali in termini di rischio è stata evidenziata anche con riferimento a singoli ambiti. Si pensi nello specifico a quello contrattuale, in quanto ogni tipo di contratto reca in sé criteri specifici di ripartizione che obbediscono a ragioni di giustizia distributiva. Si consideri il rischio dell'inadempimento, legato alla circostanza che una delle prestazioni non venga eseguita oppure al rischio di diminuita soddisfazione economica dell'affare per

sere correlato al sapere. Infatti, quanto più si estende il sapere, tanto più si estende il non sapere [119]. Possiamo averne una prova proprio pensando ai sistemi di intelligenza artificiale che palesano che quanto più si estende la conoscenza, più si estende il non sapere delle conseguenze.

Di fronte ai rischi dei sistemi di intelligenza artificiale l'*accountability* è un modo per rassicurare i potenziali utilizzatori sul fatto che, a prescindere dagli esiti di una costosa ricerca sull'errore, otterranno un risarcimento, affidando al soggetto che trae vantaggio dall'operazione economica l'onere di individuare le misure più adeguate alla fattispecie concreta [120].

Il concetto di *accountability* è suscettibile di diverse declinazione e, sebbene, in italiano è traducibile con la parola responsabilità, non è sovrapponibile alla responsabilità in senso tecnico alla quale siamo abituati a pensare [121]. Se la responsabilità è un *post*, l'*accountability* si riferisce ad un atteggiamento proattivo rispetto ad un'azione futura, sulla quale si deve essere poi in grado di rendicontare una volta che si è verificata [122].

---

la preesistenza o sopravvenienza di circostanze che non comportano inadempimento in senso tecnico ma sconvolgono l'economia originaria dell'affare. Sul punto cfr. G. ALPA, *Rischio*, *Enciclopedia del diritto,* 1989, vol. 40, p. 1144 e ss.; ID., *Rischio contrattuale*, in *Noviss. Dig*., appendice, Torino, 1986, vol. VI, pp. 863 e ss.; M. BESSONE, *Adempimento e rischio contrattuale*, Milano, 1975; ALPA-BESSONE-ROPPO, *Rischio contrattuale e autonomia priva*ta, Napoli, 1982; R. NICOLÒ, *Alea, Enciclopedia del diritto*, Milano, 1958, I, pp. 1024 e ss.; nel senso che il rischio inerisca più alle obbligazioni che al contratto è G. GORLA, *Del rischio e pericolo nelle obbligazioni*, Padova, 1934, pp. 19 e ss. In diversa prospettiva G. PACCHIONI, *Obbligazioni*, Milano, 1898, p. 344.

[119] Cfr. M BORRELLO, *La rappresentazione del rischio e il non sapere saputo,* in R. DE GIORGI (a cura di), *Limiti del diritto,* Lecce, p. 499 e ss.

[120] G. FINOCCHIARO, *op. loc. cit.*

[121] Con riferimento al concetto di *accountability*, si tratta di saper individuare misure adeguate al rischio calcolato e saper anticipare il verificarsi di situazioni critiche, l'accadimento di eventi rischiosi possibili o molto probabili e trovare soluzioni che, *ex ante* in concreto, forniscano un certo margine di sicurezza. In questi termini, S. ATERNO, *Principio di* accountability *nel* Gdpr*, significato e applicazione*, in *https://www.agendadigitale.eu. Ivi* si precisa che «nella maggior parte delle altre lingue europee, principalmente a causa delle differenze tra i sistemi giuridici, il termine "*accountability*" non è facilmente traducibile. Da un punto di vista lessicale il termine in questione è una parola composta. Il verbo *to accoun*t è traducibile in italiano come "dar conto". Il sostantivo *ability* significa "essere in grado di" o "avere attitudine a". Il problema è che il termine è vago (anche Responsabilizzazione lo è) per capire cosa significa bisogna intendersi bene. Lo si potrebbe tradurre con "rendicontabilità" ma anche su questo termine potrebbero sorgere ipotesi e teorie interpretative».

[122] Cfr. M. BASSINI, L. LIGUORI, O. POLLICINO, *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso un nuovo paradigma?*, in F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, p. 334. Si vedano, sul punto, F. DOSHI-VELEZ, M. KORTZ, R. BUDISH, C. BAVITZ, S. GERSHMAN, D. O'BRIEN, K. SCOTT, S. SHIEBER, J. WALDO, D. WEINBERGER, A. WELLER, A. WOOD, *Accountability of AI Under the Law: The Role of Explana-*

6. – L'intelligenza artificiale è una sfida che ancora non si è affrontata in tutta la sua complessità e, per di più, lo sviluppo tecnologico è così rapido che, quando si pensa di essere pronti, ci si accorge di disporre di strumenti ormai vecchi.

Capiremo via via come affrontarla, traendo insegnamento dagli errori per non ripeterli; ciò che, intanto, è opportuno fare è limitare al massimo il rischio di compierli e, a questo fine è necessario, prima di tutto, rifiutare una visione dell'intelligenza artificiale basata sull'idea che i sistemi di IA possano essere "umanizzati", trattandosi di sistemi che possono, semmai, sostituire le decisioni umane sul piano quantitativo, ma non qualitativo [123].

Va pure scartato un approccio antropocentrico all'intelligenza artificiale che concepisca i sistemi di IA quale strumento ad esclusivo servizio della persona umana. La pericolosità di una tale prospettiva è sotto gli occhi di tutti. L'IA deve essere posta a servizio di uno sviluppo sostenibile nell'interesse dell'ecosistema nel suo complesso, in una logica di coesistenza tra parte (uomo) e tutto (sistema ecologico). Le risorse naturali e la tutela anche degli esseri non umani non può essere intesa, da parte dei singoli, come «altro da sé», al contrario, deve prevalere una sorta di «immedesimazione» del soggetto con le risorse dalle quali dipende la qualità della sua vita e la sua stessa sopravvivenza [124].

La centralità dell'uomo, allora, viene meno? Deve essere superata? Direi di no, c'è un aspetto in relazione al quale la prospettiva non può che essere antropocentrica, nel senso letterale della parola. L'uomo deve essere posto al centro di quell'attività che lui solo può fare, date le funzioni cerebrali che lo rendono diverso da ogni intelligenza artificiale, dagli altri essere senzienti e anche dalla natura. Il riferimento è all'attività di "sorveglianza" su ciò che accade o ciò che egli stesso crea, ossia, nel nostro caso, i sistemi di

---

tion, in *https://arxiv.org/abs/1711.01134*, ove si afferma che «we posit that demanding explanations from AI systems is reasonable, and that we should start by asking of our AI systems what we ask of humans. Doing so avoids AI systems from getting a "free pass" to avoid the kinds of scrutiny that may come to humans, and also avoids asking so much of AI systems that it would hamper innovation and progress. Even this modest step will have its challenges, and as they are resolved, we will gain a better sense of whether and where demands for explanation should be different between AI systems and humans. As we have little data to determine the actual costs of requiring AI systems to generate explanations, the role of explanation in ensuring accountability must also be re-evaluated from time to time, to adapt with the ever-changing technology landscape».

[123] Si pensi a quanto ci insegna Blaise Pascal ne *I Pensieri*, ove si afferma che *l'esprit de géométrie* non basta per comprendere la realtà dell'uomo, ma occorre *l'esprit de finesse.*

[124] S. TOMMASI*, I beni comuni nel dialogo tra ecologia e diritto*, in M. MONTEDURO, S. TOMMASI, *Paradigmi giuridici di realizzazione del benessere umano in sistemi ecologici ad esistenza indisponibile e ad appartenenza necessaria,* in *Benessere e regole dei rapporti civili. Lo sviluppo oltre la crisi,* Atti del 9° Convegno Nazionale SISDIC in ricordo di Giovanni Gabrielli, Esi, Napoli, 2015, p. 167 e ss.

IA, per orientarne lo sviluppo in una direzione qualitativamente elevata. Si tratta, dunque, di una responsabilità importante che ha benefiche ricadute non solo sugli umani e sul loro futuro, ma anche sulle altre componenti del pianeta. Ciò è sufficiente per rendersi conto della necessità di una prospettiva che trascenda la propria "individualità", anche al fine di dare anche alla propria individualità le migliori *chances* possibili.

Occorre, altresì, non perdersi in declamazioni o mere affermazioni o, come spesso avviene a livello europeo, in una fitta mole di documenti ripetitivi e pieni di "media della comunicazione simbolicamente generalizzati", ossia media che forniscono alla comunicazione la possibilità di venire accettata [125]. Questo per dire che non basta affermare che si deve «assicurare che l'IA sia sviluppata e applicata in un quadro adeguato che promuova l'innovazione e rispetti i valori dell'Unione e i diritti fondamentali, oltre ai principi etici come la responsabilità e la trasparenza» [126]. Chi può non essere d'accordo sulla necessità di un'IA sicura, affidabile o rispettosa della dignità? Il punto è andare oltre i generici proponimenti e pensare a come rendere concretamente possibile tutto questo, scongiurando i rischi che intanto si profilano per la vita delle persone, vittime di profilazioni e discriminazioni spesso invisibili e silenziose, e per la dignità degli esseri viventi, appartenenti e non al genere umano.

[125] Sui "media della comunicazione simbolicamente generalizzati" si rimanda a R. DE GIORGI, N. LUHMANN, *Teoria della società,* cit., p. 103.

[126] Questo è quanto si legge in COM (2018) 237 final, cit., p. 3, ma espressioni simili sono generalizzate un po' in tutti i documenti europei relativi all'IA.