DIRITTO
delle *TECNOLOGIE INFORMATICHE*,
dei *MEDIA* e delle *TELECOMUNICAZIONI*

INTELLIGENZA ARTIFICIALE AFFIDABILE TRA ETICA E PRIVACY

1. INTELLIGENZA ARTIFICIALE: VANTAGGI E SVANTAGGI DI UN MONDO IN EVOLUZIONE

"L'intelligenza artificiale non è fantascienza: fa già parte delle nostre vite. Che si tratti di utilizzare un assistente personale virtuale per organizzare la nostra giornata lavorativa, viaggiare in un veicolo a guida autonoma o avere un telefono che ci suggerisce le canzoni o i ristoranti che potrebbero piacerci, l'IA è una realtà"[1].

Si apre così la comunicazione sull'intelligenza artificiale ("AI") della Commissione europea al Parlamento europeo, al Consiglio e al Comitato economico e sociale europeo e al Comitato delle Regioni.

L'intelligenza artificiale è una tecnologia che rivoluzionerà (e, in parte, lo ha già fatto) la società così come la conosciamo, tanto che già nel 2018 il CEO di Google aveva affermato che: *"l'intelligenza artificiale è una delle cose più importanti su cui l'umanità sta lavorando. È più profondo di, non so, l'elettricità o il fuoco"* [2].

Per intelligenza artificiale si intendono quei *"sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi"*[3]. In altre parole, l'intelligenza artificiale è la simulazione dei processi di intelligenza umana da parte dei sistemi informatici.

In passato - prima che raggiungesse i livelli di evoluzione tecnologica odierni – la tipologia principale di AI era il c.d. sistema esperto, che agisce sulla base delle istruzioni impartite dal suo programmatore. Dopo essere stata adeguatamente istruita, infatti, l'AI era in grado di dedurre informazioni (***output***) da un insieme di dati di partenza (***input***), grazie al motore inferenziale e all'applicazione delle regole ***IF*** (condizione) e ***THEN*** (azione). Pertanto, volendo fare un esempio, dagli IF mal di gola e IF febbre il sistema esperto elaborava il THEN influenza.

L'AI, rispetto al passato, si è sviluppata al punto tale da apprendere autonomamente - quindi senza che gli vengano impartite ulteriori istruzioni da parte dei suoi programmatori – dai dati direttamente acquisiti nel corso del suo funzionamento. In particolare, si parla di ***machine learning*** (o apprendimento automatico***)*** quando l'evoluzione si basa su algoritmi e percorsi predefiniti in sede di programmazione, mentre si fa riferimento al c.d. ***deep learning***[4] (o apprendimento approfondito), inteso come uno degli approcci dell'apprendimento automatico, quando l'AI si evolve elaborando nuovi percorsi di apprendimento e strutturando gli algoritmi in modo da

1. Commissione europea nella Comunicazione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato dell regioni, "L'intelligenza artificiale per l'Europa", COM(2018) 237, 25 aprile 2018, p. 1.

2. https://www.cnbc.com/2018/02/01/google-ceo-sundar-pichai-ai-is-more-important-than-fire-electricity.html.

3. Sarzana di S. Ippolito F., Nicotra M., "Diritto della blockchain, intelligenza artificiale e IOT", IPSOA, 2019, p. 193.

4. Battelli E., necessità di un umanesimo tecnologico: sistemi di intelligenza artificiale e diritti della persona, in "Diritto di Famiglia e delle Persone (II), fasc. 3, 1° settembre 2022, p. 3.

generare una rete neurale artificiale, proprio perché l'apprendimento è ispirato alla struttura del cervello umano. ***Machine learning*** e ***deep learning*** non sono sinonimi di AI, ma il primo è semplicemente un modo per raggiungere l'intelligenza artificiale, mentre il secondo è uno dei molteplici approcci relativi all'apprendimento automatico.

In generale, oggi i sistemi di intelligenza artificiale funzionano ingerendo grandi quantità di dati di addestramento (c.d. ***big data***) e, tramite algoritmi, sono in grado di scoprire correlazioni tra tali dati sulla base di funzioni statistiche e di probabilità che per lo più traducono in matrici di numeri i c.d. ***patterns***, ossia modelli numerici corrispondenti a risposte presumibilmente corrette rispetto ad un determinato input. In altre parole, i sistemi di intelligenza artificiale imparano a fare previsioni dopo essere stati addestrati sulla base di ampie serie di dati. In questo modo, ad esempio, un ***chatbot*** alimentato da esempi di chat di testo può imparare a produrre scambi realistici con le persone.

L'intelligenza artificiale, dunque, possiede tre caratteristiche fondamentali: (i) l'apprendimento, ossia l'acquisizione di informazioni e la creazione di regole relative all'utilizzo di dette informazioni; (ii) il ragionamento, consistente nel saper individuare la regola corretta per giungere al risultato desiderato e (iii) l'autocorrezione, al fine di perfezionare le proprie prestazioni per giungere a risultati sempre più precisi[5].

I sistemi di AI presentano, dunque, il grande vantaggio di saper elaborare una enorme quantità di dati, con una velocità e un livello di precisione e accuratezza tale da non poter essere equiparato alla prestazione umana. Per tali motivi, tale tecnologia è destinata ad apportare numerosi benefici alla società: si pensi, ad esempio, a medici che possono fornire diagnosi migliori e terapie personalizzate e mirate, alle imprese che possono anticipare le tendenze del mercato e prendere decisioni più efficienti, ai consumatori che possono fare scelte più informate e ottenere servizi personalizzati e alle autorità pubbliche che possono ottimizzare la gestione dei beni pubblici[6].

Tuttavia, la diffusione di tecnologie caratterizzate dall'impiego di sistemi di AI, alla cui base vi sono algoritmi progettati per replicare il comportamento e la capacità di apprendimento umani, ha aperto un dibattito etico, sociale e giuridico ancora attuale.

La vera domanda è: "Una macchina guidata da algoritmi può comportarsi in modo etico?" E ancora: "Ci si può fidare dell'intelligenza artificiale?"

L'intelligenza artificiale presenta, infatti, una serie di rischi potenziali, che in parte giustificano la diffidenza con cui il legislatore e il regolatore si sono finora posti rispetto all'impiego di tali strumenti, culminata per altro nel recente provvedimento del Garante italiano che ha bloccato ChatGPT. Il potere dell'AI può essere utilizzato anche per perseguire interessi economici in modi dannosi per gli individui e la società: utenti, consumatori e lavoratori possono essere soggetti a una sorveglianza pervasiva, controllati nel loro accesso alle informazioni e alle opportunità, manipolati nelle loro scelte.

I fattori di rischio principali che riguardano l'intelligenza artificiale sono, in particolare - oltre alle intrusioni nella vita privata – meccanismi di decisione opachi, la sicurezza dei sistemi, l'affidabilità dei dati che alimentano gli algoritmi e i ***bias*** che possono avere tali sistemi nel caso in cui l'algoritmo sia stato addestrato (o si sia addestrato autonomamente) sulla base di schemi affetti da pregiudizi, con possibili effetti discriminatori.

Ulteriori criticità derivano, sul versante della protezione dei dati personali, dall'impossibilità di stabilire a priori le finalità del trattamento, in quanto la macchina, nel corso del processo di apprendimento, è in grado di evolvere in modo non sempre prevedibile le proprie condotte e quindi, di discostarsi dalle finalità in principio stabilite.

5. Grillo L., "Intelligenza Artificiale e GDPR: un binomio possibile", in Ius in Itinere, 22 aprile 2021. Disponibile online: https://www.iusinitinere.it/intelligenza-artificiale-e-gdpr-un-binomio-possibile-38043#_ftnref7.

6. European Parliamentary Research Service (EPRS), "The impact of the General Data Protection Regulation (GDPR) on artificial intelligence", giugno 2020, p. 26. Disponibile online: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf#page=16&zoom=100,93,317.

Rischi e benefici dell'AI devono, dunque, essere adeguatamente bilanciati, nel perseguimento di un "***umanesimo tecnologico***" che sappia coniugare intelligenza artificiale, etica e tutela dei diritti e delle libertà fondamentali delle persone[7], ove la domanda principale non è tanto "come funziona?", ma "come si comporta?" l'AI.

2. UN UTILIZZO ETICO DELL'INTELLIGENZA ARTIFICIALE

Le considerazioni che precedono portano ad una constatazione di carattere etico: la macchina non può sostituirsi all'agire umano. L'AI non deve risolversi in una delega in bianco in favore dell'algoritmo, tale da neutralizzare i giudizi, anche di valore, che sono propri esclusivamente dell'uomo[8].

Al fine di poter effettivamente accogliere tutti i benefici e le potenzialità dell'intelligenza artificiale è necessaria l'instaurazione di un meccanismo di fiducia tra tale tecnologia e la società, al fine di scongiurare il timore di una "***tirannia dell'algoritmo***"[9] sui diritti civili, sociali e politici degli individui.

Tale obiettivo è raggiungibile mediante la creazione di regole preventive che rendano l'AI affidabile e, quindi, una tecnologia al servizio della intelligenza umana senza che si sostituisca a questa[10]: ***"un'IA affidabile può migliorare la prosperità individuale e il benessere collettivo generando agiatezza, creando valore e massimizzando la ricchezza. Può contribuire alla realizzazione di una società equa, in quanto può aiutare ad accrescere la salute e il benessere dei cittadini secondo modalità che promuovono l'uguaglianza nella distribuzione delle opportunità economiche, sociali e politiche"***[11].

Il Gruppo di esperti ad alto livello sull'intelligenza artificiale, istituito nel 2018 dalla Commissione europea, ha pubblicato delle linee guida sugli orientamenti etici per un'AI affidabile. Secondo il gruppo di esperti, una AI affidabile dovrebbe agire nel pieno rispetto dei diritti fondamentali e dei principi etici di seguito elencati[12]:

(i) rispetto dell'autonomia umana: i sistemi di AI non devono subordinare, costringere, ingannare, manipolare, condizionare o aggregare in modo ingiustificato gli esseri umani, ma, al contrario, devono essere progettati per aumentare, integrare e potenziare le abilità cognitive, sociali e culturali umane. Deve essere consentito, quindi, agli esseri umani che interagiscono con i sistemi di AI di poter mantenere la propria autodeterminazione.

(ii) prevenzione dei danni: i sistemi di AI, oltre a non dover influenzare negativamente gli esseri umani, non devono causare danni. Pertanto, i sistemi di AI e gli ambienti in cui operano devono essere sicuri, protetti, tecnicamente robusti e si deve garantire che non siano esposti ad usi malevoli. Particolare attenzione è da porre nelle situazioni in cui i sistemi di AI possono causare o aggravare quegli effetti negativi dovuti ad asimmetrie di potere o di informazione, come ad esempio tra datori di lavoro e dipendenti, imprese e consumatori, governi e cittadini.

(iii) equità: i sistemi di IA devono rispettare sia la dimensione sostanziale che procedurale dell'equità. Quella sostanziale implica un impegno a garantire una distribuzione giusta e equa di costi e benefici, al fine di garantire che gli individui e i gruppi siano liberi di distorsioni inique, discriminazioni e stigmatizzazioni. L'AI, piuttosto, deve diventare un mezzo per promuovere le pari opportunità in termini di accesso all'istruzione, ai beni, ai servizi e alla tecnologia. La dimensione procedurale, invece, implica la possibilità tanto di contestare le decisioni elaborate dai

7. Battelli E., "Necessità di un umanesimo tecnologico: sistemi di intelligenza artificiale e diritti della persona, cit., p. 2.

8. Cfr. Ing. Fabio Lazzini, Sogei S.p.A., Il ruolo dei dati e dell'Intelligenza Artificiale nella sanità digitale: sfide ed opportunità a beneficio dei cittadini - Convegno "Amministrare con gli algoritmi" – Torino, Università di Torino, 13 ottobre 2022

9. Rodotà S., "Il mondo nella rete. Quali diritti, quali vincoli", Editori Laterza, Bari, 2014, p. 37.

10. Battelli E., "Necessità di un umanesimo tecnologico: sistemi di intelligenza artificiale e diritti della persona", cit., p. 6.

11. Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale, "Orientamenti etici per un'IA affidabile", 8 aprile 2019, p. 10. Disponibile online: https://op.europa.eu/it/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1.

12. Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale, "Orientamenti etici per un'IA affidabile", 8 aprile 2019, cit.

sistemi di AI e dagli esseri umani che li gestiscono che di presentare un ricorso efficace contro di esse. A tal fine, i processi decisionali devono essere spiegabili.

(iv) esplicabilità: i processi algoritmici devono essere trasparenti, le capacità e gli scopi dei sistemi di AI devono essere comunicati apertamente e le decisioni devono essere spiegate a coloro che sono coinvolti direttamente e indirettamente. Senza tali informazioni, infatti, una decisione non può essere debitamente contestata.

Tali principi, inoltre, necessitano di essere tradotti in requisiti concreti. Pertanto, la realizzazione di una AI affidabile presuppone il rispetto, da parte dei diversi portatori di interessi che partecipano al ciclo di vita dei sistemi (quali, quindi, sviluppatori, distributori e utenti finali), dei seguenti requisiti: (i) intervento e sorveglianza umani; (ii) robustezza tecnica e sicurezza; (iii) riservatezza e governance dei dati; (iv) trasparenza; (v) diversità, non discriminazione ed equità; (vi) benessere sociale e ambientale e (vii) ***accountability***[13].

La Commissione europea con la Proposta di regolamento europeo sull'utilizzo dell'intelligenza artificiale del marzo 2021 (***Artificial Intelligence Act*** - AI Act)[14] è intervenuta con lo scopo di fornire regole comuni per lanciare nuovi prodotti e servizi sul mercato europeo in un contesto di fiducia da parte dei cittadini dell'Unione. L'obiettivo è di assicurare che i sistemi di AI immessi sul mercato europeo siano sicuri ed etici e, secondo un approccio basato sul rischio, rispettino la normativa vigente in materia di diritti fondamentali dell'individuo e i valori dell'Unione e contrastino la cd opacità dell'algoritmo.

L'AI Act riconosce proprio nella sorveglianza umana il punto di equilibrio tra i rischi intrinseci di tali tecnologie e la perdita di opportunità in termini di crescita e benessere sociale che il loro utilizzo comporta. In tal senso, vengono previsti in capo ai fornitori di sistemi di AI una serie di obblighi di monitoraggio e segnalazione di eventuali incidenti e malfunzionamenti dei prodotti immessi sul mercato, affidando alle autorità settoriali, esistenti o di nuova costituzione – e ci si chiede se tale competenza verrà affidata al Garante Privacy - la vigilanza sul rispetto dei predetti obblighi da parte degli operatori di mercato.

In questo scenario, si sente - sia a livello europeo che nazionale - la necessità dell'intervento della trasparenza quale soluzione al nodo dell'opacità tecnologica dell'algoritmo dell'AI. La trasparenza è vista come l'unico elemento che possa consentire la realizzazione di soluzione tecnologiche in grado di creare fiducia e benessere sociale.[15]

L'obiettivo di una piena trasparenza algoritmica - intesa come totale ripercorribilità e intelligibilità dei processi digitali – si compie nel momento in cui l'algoritmo che conduce l'AI alla decisione automatizzata sia conoscibile al cittadino e realizzi la conoscibilità algoritmica. E ancora, si esprime nel diritto del cittadino a conoscere la cd logica algoritmica, ossia l'iter logico sulla base del quale è stata presa quella decisione. Il principio di conoscibilità della logica algoritmica è, per altro, presente anche nel GDPR che, con riferimento alle decisioni automatizzate, prevede il diritto dell'interessato di conoscere ed avere informazioni in merito alla logica cui risponde qualsiasi trattamento automatizzato di dati (Art. 13 e Considerando 63 del GDPR).

13. Ivi, p. 16.

14. Cfr Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione, 21 aprile 2021. Disponibile online: https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206&from=EN.

15. Lo Stesso Consiglio di Stato si è espresso per la trasparenza intesa quale "conoscibilità dell'algoritmo" che "deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti" Cfr. in tal senso Cons. Stato, sez. V, 4 febbraio 2020 n. 881.

3. INTELLIGENZA ARTIFICIALE E PROTEZIONE DEI DATI PERSONALI: LE DECISIONI AUTOMATIZZATE

Un sistema di intelligenza artificiale che generi un senso di fiducia in individui, imprese, organizzazioni e in tutta la società non solo deve essere affidabile e etico, ma deve anche essere rispettoso delle leggi.

Sono molte le intersecazioni esistenti tra AI e diritto: si pensi, ad esempio, alla branca della proprietà intellettuale, laddove esistono sistemi che sono in grado di generare essi stessi opere d'arte, o ancora alla responsabilità civile e penale dei danni causati dalla macchina intelligente.

Primo fra tutti, però, è il legame indissolubile tra l'intelligenza artificiale e la disciplina sulla protezione dei dati personali.

Come anticipato, infatti, l'AI si nutre di dati per alimentare se stessa, oltre che per analizzare, prevedere e influenzare il comportamento umano. I dati necessari a questo scopo sono presenti, in enormi quantità, nell'infosfera (ossia l'insieme dei mezzi di comunicazione e delle informazioni che da tali mezzi vengono prodotte), non solo perché conferiti dagli stessi utenti di internet nell'ambito della loro navigazione online, ma anche tramite il ricorso sempre più frequente a dispositivi elettronici connessi alla rete che comunicano informazioni senza intervento umano (si pensi all'***Internet of Things* - *IoT***). Questo utilizzo dell'IoT e dell'AI ha determinato l'importanza e il valore dei c.d. big data e anche le informazioni che non venivano raccolte o che venivano scartate come "***data exhaust***[16]", senza valore - ad esempio, le tracce delle attività online - sono ora diventate una risorsa preziosa[17]. Il ricorso all'intelligenza artificiale rende inoltre tutti i dati che essa lavora quali personali, ciò in quanto gli algoritmi consentono la coordinazione dei dati non personali con quelli personali portando, così, a una ricostruzione dell'identità della persona, che se non era identificata diventa così identificabile[18].

Diviene inevitabile, dunque, trovare un giusto bilanciamento tra i duplici obiettivi perseguiti anche a livello europeo[19]: (i) creare un "ecosistema di eccellenza" al fine di diventare ***leader*** del settore dell'AI e, dunque, mettere a disposizione dei sistemi tutti i dati necessari al suo funzionamento e (ii) creare un "ecosistema di fiducia" e, quindi, garantire il pieno rispetto dei diritti fondamentali e delle norme dell'UE, compresa la principale fonte normativa in materia di protezione dei dati personali, quale il Regolamento n. 679/2016, meglio noto come GDPR.

Il GDPR non menziona espressamente l'intelligenza artificiale ma, come detto, tratta in maniera specifica la profilazione e il processo decisionale automatizzato relativo alle persone fisiche.

L'AI è infatti spesso utilizzata per porre in essere tecniche di profilazione o per l'assunzione di decisioni automatizzate anche in ambiti che richiedono scelte complesse, basate su molteplici fattori e criteri non predefiniti. Per di più, spesso, le scelte assunte dai sistemi di intelligenza artificiale risultano essere più precise e imparziali rispetto a quelle degli esseri umani. A tal proposito, infatti, molte decisioni vengono assunte grazie all'ausilio dell'intelligenza artificiale: si pensi, ad esempio, allo ***screening*** dei ***curricola*** in fase di selezione del personale, o ancora, ai meccanismi di ***credit scoring*** per valutare l'affidabilità finanziaria nell'ambito di procedure di finanziamento.

L'utilizzo dell'AI, tuttavia, non porta a decisioni perfette. Il problema è che anche la macchina è fallibile e può dare luogo a decisioni discriminatorie a causa di ***bias*** decisionali: accade sovente che i c.d. dati di addestramento (ossia i dati a cui ricorrono gli algoritmi di intelligenza artificiale per assumere decisioni)

16. I dati esauriti sono i dati generati come sottoprodotto delle azioni e delle scelte online delle persone. Tali dati sono costituiti dai vari file generati dai browser web e dai loro plug-in, come i cookie, i file di log, i file temporanei di Internet, ecc.

17. European Parliamentary Research Service (EPRS), "The impact of the General Data Protection Regulation (GDPR) on artificial intelligence", cit., p. 32.

18. Grasselli F., "Il "diritto alla spiegazione" ex art. 22 GDPR e il Brexit case", 30 novembre 2020. Disponibile online: https://www.irpa.eu/il-diritto-alla-spiegazione-ex-art-22-gdpr-e-il-brexit-case/.

19. Così come dichiarato dalla Commissione europea nel Libro Bianco sull'intelligenza artificiale – un approccio europeo all'eccellenza e alla fiducia, 19 febbraio 2020. Disponibile online: https://op.europa.eu/it/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1.

siano essi stessi discriminatori[20]. I sistemi basati sull'apprendimento, infatti, possono essere addestrati o addestrarsi sulla base di giudizi umani e possono riprodurre i punti di forza e di debolezza degli esseri umani che hanno espresso tali giudizi, compresa la loro propensione all'errore.

Il trattamento dei dati personali tramite sistemi di intelligenza artificiale deve avvenire quindi nel rispetto dell'interesse dei cittadini a un trattamento algoritmico equo, ossia del loro interesse a non essere soggetti a pregiudizi ingiustificati derivanti dal trattamento automatizzato.

Si è resa, dunque, necessaria l'adozione di garanzie adeguate al fine di evitare che si possano instaurare forme di controllo e monitoraggio degli individui e, nei casi peggiori, ipotesi di esclusione sociale.

L'art. 22 del GDPR stabilisce, a tal fine, un divieto generale di adozione di decisioni totalmente automatizzate: *"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"*. Sono esempi di una decisione completamente automatizzata in grado di produrre effetti significativi nei confronti dell'interessato, ai sensi del considerando n. 71 del GDPR, il rifiuto automatico di una domanda di credito ***online*** o pratiche di assunzione elettronica senza interventi umani.

Le decisioni totalmente automatizzate, tuttavia, non sono vietate ***tout court***. Tale divieto incontra delle eccezioni nei casi in cui la decisione sia: a) necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento; b) 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 o c) si basi sul consenso esplicito dell'interessato. Laddove si applichi una di queste eccezioni, *"il titolare del trattamento deve attuare misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato"*.

Tra le misure che il titolare deve adottare nel caso in cui ponga in essere un trattamento di dati personali che dia luogo a una decisione totalmente automatizzata rientra il su menzionato diritto alla spiegazione a favore dell'interessato, quale principio etico a cui i sistemi dovrebbero attenersi per risultare affidabili.

4. DIRITTO ALLA SPIEGAZIONE: TRA NORMA GIURIDICA E OPACITÀ DELLE DECISIONI

Il diritto alla spiegazione per l'interessato soggetto alla decisione assunta nei suoi confronti da processi automatizzati si desume in diverse norme del GDPR: viene, infatti, espressamente richiamato dal considerando n. 71, che prescrive che il trattamento automatizzato di dati personali dovrebbe essere subordinato a garanzie adeguate, fra le quali rientra il diritto dell'interessato di ottenere una spiegazione della decisione conseguita dopo tale valutazione e di contestare la stessa; e si rinviene negli artt. 13, par. 1, lett. f), 14, par. 2, lett. g) e 15, par. 1, lett. h) che – con formula identica – sanciscono il diritto dell'interessato di ottenere *"informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato"*. Si precisa che le informazioni devono essere rese all'interessato al momento della raccolta dei dati presso l'interessato (art. 13) o presso soggetto diverso (art. 14), nonché nell'esercizio del diritto di accesso da parte dell'interessato (art. 15).

Tuttavia, mentre un sistema esperto è sempre in grado di esibire i passaggi logici che sottendono le decisioni prese, nel caso di AI basata sul ***machine learning*** o ***deep learning*** non è sempre possibile spiegare perché un modello abbia generato un particolare risultato o decisione e quale combinazione di fattori di ***input*** vi abbia contribuito, a causa della c.d. opacità degli algoritmi che è alla base dell'intelligenza artificiale.

Il concetto di opacità fonda il c.d. modello *"**black box**"*, nel quale l'intelligenza artificiale registra ed utilizza

20. Pellecchia E., "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 "Le nuove leggi civili commentate", fasc. 5, 2018, p. 3.

informazioni personali senza che sia possibile, tuttavia, risalire al processo induttivo e deduttivo che ha portato il sistema a un determinato risultato[21]: "***we can observe its inputs and outputs, but we cannot tell how one becomes the other***"[22]. Con il termine black box si intende, infatti, l'impossibilità di "guardare dentro" il meccanismo di funzionamento del sistema da parte degli stessi programmatori.

In particolare, l'opacità può essere declinata in (i) opacità intenzionale, consistente nella mancata possibilità di spiegare i processi seguiti dall'intelligenza artificiale nell'ottica del segreto industriale e del vantaggio concorrenziale, (ii) opacità tecnica, ossia il grado di conoscenza tecnica necessaria alla decifrazione dell'algoritmo, possibile solo agli esperti e (iii) opacità propria dell'algoritmo stesso, che deriva dalle scelte da questo compiute nel processo di decisione e che sono sconosciute anche ai suoi sviluppatori[23].

Neanche la comunicazione del codice sorgente dell'AI potrebbe efficacemente risolvere il problema. Ciò sia in quanto lo stesso risulta incodificabile per i non esperti, sia poiché il codice sorgente espone solo il metodo di apprendimento automatico utilizzato e non il percorso delle scelte compiute dall'AI che, come sopra evidenziato, è in alcuni casi sconosciuto[24].

La problematica relativa alla trasparenza dei sistemi AI è stata affrontata nella proposta di Regolamento sull'Intelligenza Artificiale (AI ACT) del Parlamento europeo.

L'art. 52 della proposta, rubricato "Obblighi di trasparenza per determinati sistemi di IA", impone ai fornitori di sistemi di AI destinati ad interagire con persone fisiche di garantire che tali sistemi siano progettati e sviluppati in modo tale che gli utenti siano informati del fatto di stare interagendo con un sistema di AI. Tale garanzia non trova applicazione (i) nel caso in cui l'interazione risulti evidente dalle circostanze e dal contesto di utilizzo e (ii) nei confronti di sistemi di AI autorizzati dalla legge per accertare, prevenire, indagare e perseguire reati.

L'art. 13, invece, prescrive degli obblighi di trasparenza per i sistemi considerati ad alto rischio. In particolare, la disposizione stabilisce che tali sistemi debbano essere progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'***output*** del sistema e utilizzarlo adeguatamente. Per di più, i sistemi di AI ad alto rischio devono essere sempre accompagnati da istruzioni per l'uso, che comprendano informazioni concise, complete, corrette e chiare, che siano pertinenti, accessibili e comprensibili per gli utenti. Tali informazioni devono specificare, tra l'altro e ove opportuno, le specifiche per i dati di ***input*** o qualsiasi altra informazione pertinente in termini di set di dati di addestramento, convalida e prova, tenendo conto della finalità prevista del sistema di AI.

5. CONCLUSIONI

L'ampia diffusione dell'intelligenza artificiale nella società odierna pone sfide e solleva interrogativi sotto molteplici punti di vista.

Come ogni nuova tecnologia, l'AI offre benefici e rischi. La combinazione di intelligenza artificiale e ***big data*** concede grandi opportunità per la ricerca scientifica, il benessere e l'amministrazione, ma comporta anche gravi rischi per gli individui e la società: intensificazione della sorveglianza, del controllo, della manipolazione e della discriminazione.

L'AI si relaziona con il diritto a diversi livelli. In quanto tecnologia pervasiva e multiforme, l'AI può migliorare

21. Pellecchia E., "Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation", cit., p. 3.

22. Pasquale F., "The Black Box Society. The secret Algorithms That Control Money and Information", Harvard University Press, Cambridge-London, 2015, p. 3.

23. Pellecchia E., "Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation", cit., pp. 7-8.

24. Ivi, p. 8.

o compromettere l'esercizio di molteplici diritti fondamentali, tra tutti la protezione dei dati personali.

È stato osservato, infatti, che i dati personali possono essere utilizzati per prevedere il comportamento umano, per apprendere le propensioni e gli atteggiamenti degli individui, per esercitare un'influenza sulla psiche umana. La possibilità di utilizzare l'AI stimola la raccolta di vasti insiemi di dati personali e la disponibilità di grandi insiemi di dati, a sua volta, stimola nuove applicazioni dell'AI.

Per realizzare una AI che sia allo stesso tempo etica e rispettosa del diritto è augurabile che la stessa sia dotata della caratteristica dell'esplicabilità, la quale si declina - nel caso di assunzione di una decisione automatizzata - sia nel diritto dell'interessato di essere informato di essere sottoposto ad un trattamento automatizzato, sia nel diritto alla spiegazione, consistente nel ricevere informazioni significative sulla logica dell'intelligenza artificiale utilizzata e sulle conseguenze previste.

Come analizzato, tuttavia, il diritto alla spiegazione non può sempre essere soddisfatto: a causa della ***black box***, che contraddistingue l'intelligenza artificiale basata su ***machine learning*** e ***deep learning***, non è possibile conoscere i percorsi logici che hanno portato l'AI ad assumere una determinata scelta.

Tale problematica è stata affrontata dalle istituzioni europee nell'ambito della proposta dell'***Artificial Intelligence Act*** che, assumendo un approccio basato sul rischio, richiede un determinato grado di trasparenza dei sistemi di intelligenza artificiale sin dal momento della loro progettazione e, dunque, ***by design***.

L'obiettivo sarebbe quello di permettere all'interessato sottoposto ad una decisione automatizzata di conoscere, anche in maniera non necessariamente complessa, i criteri presi in considerazione per la decisione (come, ad esempio, nel caso della concessione di un'assicurazione sulla vita, in cui possono rilevare le condizioni di salute del soggetto) e il loro peso nel processo decisionale.

In conclusione, sarebbe auspicabile la realizzazione di un modello di AI etica ed affidabile nel quale si innestano i principi di trasparenza e conoscibilità algoritmica, quali strumenti posti a tutela dei diritti fondamentali degli individui e a garanzia della ***data protection*** e ***data security***.

Luca Tufarelli ltufarelli@ristuf.it
Maria Lilia La Porta mlaporta@ristuf.it