LUISS Guido Carli
LIBERA UNIVERSITÀ INTERNAZIONALE DEGLI STUDI SOCIALI

DIPARTIMENTO DI GIURISPRUDENZA

Cattedra di DATA PROTECTION

LA PROTEZIONE DEI DATI NELL'INTELLIGENZA ARTIFICIALE: TRA GDPR E AI ACT

RELATORE
Chiar.mo Prof.
Pierluigi Congedo

CORRELATORE
Chiar.mo Prof.
Cosimo Comella

CANDIDATO
Marco Guarna
Matr. 159073

ANNO ACCADEMICO 2022/2023

- *A Mamma Isabella*

Indice

Introduzione

L'evoluzione dell'Intelligenza Artificiale (IA) negli ultimi anni del nostro secolo ha rivoluzionato molti aspetti della vita quotidiana degli esseri umani, influenzando settori come la sanità, la finanza, il settore assicurativo e la comunicazione. Questa rapida evoluzione tecnologica ha sollevato importanti questioni, in particolare sulla protezione dei dati personali.

Nel panorama legislativo eurounitario il Regolamento EU 2016/679 (GDPR) rappresenta la normativa cardine in materia. Tale regolamento ha introdotto diverse misure a tutela della *privacy* degli individui, stabilendo dei diritti specifici per la tutela dei dati personali ed evolvendosi anche grazie all'interpretazione giurisprudenziale della Corte di Giustizia dell'Unione europea. Tuttavia, per quanto tale normativa rimanga attuale e rilevante, non sembra essere sufficientemente adatta ad affrontare le sfide poste dalle nuove tecnologie di IA. Queste hanno origine dalla capacità di tali tecnologie di elaborare e analizzare enormi quantità di dati e sollevano preoccupazioni legate alla trasparenza, alla responsabilità, alla così detta *explainability*[1], alla sorveglianza umana, al diritto d'autore, alla discriminazione derivante da tale trattamento. Le decisioni prese da sistemi di IA possono avere un impatto significativo sulla vita degli individui, eppure spesso manca sia una chiara comprensione sul funzionamento dei sistemi sia la trasparenza relativa alla raccolta e all'impiego dei dati. I sistemi di IA sono in continua evoluzione e superano di frequente i limiti posti dalle normative vigenti, lasciando un vuoto regolatorio attorno a molte delle loro applicazioni. È quindi necessario un approccio più dinamico e flessibile per regolamentare l'IA, che tenga conto dei rapidi cambiamenti e delle nuove sfide che sorgono, cercando al contempo di agevolare l'innovazione, di cui la società continua ad aver bisogno.

L'obiettivo di questa tesi è quello di suggerire quali sono gli aspetti più importanti su cui concentrarsi per la protezione dei dati personali nell'era dell'IA, analizzando i punti deboli del quadro normativo. Ci concentreremo, in particolare, su una valutazione critica di alcune disposizioni del GDPR, evidenziando i punti in cui le attuali normative non sono in grado di regolare in modo efficace gli aspetti complessi dell'IA. Sonderemo allo stesso modo la proposta del 2021 di Regolamento sull'Intelligenza Artificiale (AI Act), tenendo in considerazione anche alcuni dei più recenti emendamenti, al fine di valutarne l'efficacia e di identificare possibili miglioramenti. È fondamentale considerare l'equilibrio tra l'innovazione tecnologica e la tutela dei diritti fondamentali

[1] L'*explainability* è termine che si riferisce alla circostanza che un sistema di apprendimento automatico e i risultati che fornisce dall'elaborazione siano comprensibili ad un essere umano.

degli individui, al fine di sviluppare una regolamentazione che sia adeguata alle esigenze della società digitale contemporanea.

All'interno del Capitolo I, verrà fornita una panoramica sul fenomeno globale dell'IA, esaminando l'intreccio con la normativa sulla protezione dei dati. Verrà analizzata la storia dell'IA e l'evoluzione della *privacy* nelle sentenze della Corte di Giustizia dell'Unione europea. Proseguendo nel Capitolo II, ci si concentrerà sulla regolamentazione, in particolare sul GDPR, analizzando cosa lo stesso stabilisca per i processi decisionali automatizzati. Verranno esaminate le implicazioni dell'articolo 22 del GDPR sui sistemi di IA, le implicazioni di misure di sicurezza come anonimizzazione e pseudonimizzazione e l'importanza della valutazione di impatto sui dati personali nei sistemi di IA, con un approfondimento sulle *chatbot*. Seguendo lo stesso metodo, si proseguirà nell'analisi dell'AI Act, attenzionando gli articoli che regolano gli aspetti di qualità dei dati, trasparenza e sorveglianza umana sui sistemi di IA, sintetizzando in conclusione i punti di contatto tra GDPR e AI Act. Quindi si presenterà nel Capitolo III la problematica dei nuovi sistemi di "*General Purpose AI*" (GPAI) e dei "*foundation models"*, approfondendo l'approccio dell'AI Act verso gli stessi. La riflessione si concentrerà sulla *privacy* come principio per un migliore sviluppo dei sistemi di IA, focalizzando l'analisi sulle funzioni del famoso sistema di GPAI, ChatGPT. Particolare attenzione verrà dunque data al provvedimento nei confronti di ChatGPT adottato dal Garante Italiano della *Privacy*, per poi chiudere il capitolo con una riflessione sulle criticità presenti nella regolamentazione attuale e futura per la protezione dei dati personali nell'ambito dell'IA, evidenziando la necessità di una maggiore attenzione verso tre aspetti essenziali: la trasparenza, la discriminazione e la sorveglianza umana.

Al termine di questa analisi saranno sintetizzati i principali risultati della tesi emersi dopo la valutazione del quadro normativo per la protezione dei dati personali nell'era dell'IA. Saranno, infine, suggerite alcune possibili direzioni per futuri sviluppi normativi e ricerche nel campo dell'IA e della protezione dei dati personali, consapevoli di come solo attraverso una regolamentazione appropriata e mirata si potrà garantire una protezione adeguata dei dati personali e dei diritti degli individui in un contesto tecnologico in continua evoluzione.

Capitolo I

"*We can only see a short distance ahead, but we can see plenty there that needs to be done*"

- Alan Turing

1. Le risposte globali al fenomeno dell'Intelligenza Artificiale ed il ruolo del Regolamento UE 2016/679[2]

"La scienza e ingegneria di creare macchine intelligenti". Venne definita così per la prima volta l'intelligenza artificiale (da qui in poi, IA) dal Professore emerito di Standford John McCarthy[3] nel 1955, appena un anno prima del seminario al Dartmouth College[4] della stessa università che diede il via agli studi in materia. Il seminario si concentrò sull'ipotesi che fosse possibile fornire alla macchina istruzioni tanto precise da renderla capace di simulare ogni aspetto dell'intelligenza umana. Come sottolineato da Bruce Buchanan in *A (Very) Brief History of Artificial Intelligence*, l'interesse per l'IA è iniziato a crescere in maniera significativa negli anni Cinquanta[5]. Gli obiettivi iniziali erano molteplici: risolvere problemi precedentemente riservati esclusivamente agli esseri umani, utilizzare il linguaggio e sviluppare astrazioni. Il progetto mirava a un

[2] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), [2016], OJ L 119, 4.5.2016, p. 1–88

[3] John McCarthy (1927-2011) è stato un informatico e matematico statunitense riconosciuto come uno dei padri fondatori dell'Intelligenza Artificiale. Ha coniato il termine "Intelligenza Artificiale" nel 1956, in occasione della conferenza di Dartmouth, riconosciuta come l'evento che ha segnato la nascita dell'IA come campo di ricerca indipendente.

[4] John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence', (31 agosto 1955) <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> ultimo accesso il 19 aprile 2023.

[5] Bruce G. Buchanan (University Professor of Computer Science Emeritus at the University of Pittsburgh), 'A (very) brief history of artificial intelligence' [2005] AI Magazine, 26

progresso scientifico su due fronti: da un lato, accrescere la conoscenza dei meccanismi del pensiero e dell'intelligenza umana; dall'altro, creare macchine senzienti. Oggi non possiamo affermare di possedere una totale conoscenza dei processi mentali, ma è grazie alle innovazioni tecnologiche e, in particolare, all'IA, che i progressi in questo settore sono stati considerevoli. Chiaramente la comprensione del funzionamento del cervello umano è ancora in gran parte un mistero per gli scienziati e i ricercatori[6]: le complesse interazioni tra i miliardi di neuroni e le loro sinapsi, congiuntamente ai processi biochimici e alla natura stessa della coscienza, rendono il quesito estremamente arduo da decifrare[7]. Malgrado ciò, la ricerca non si ferma e nuove scoperte nel campo delle neuroscienze, anche attraverso l'utilizzo di sistemi di IA stanno gradualmente aiutando a superare alcuni enigmi del cervello umano, offrendo inedite prospettive e possibilità per il futuro.

L'IA ha registrato un notevole incremento di importanza negli ultimi venti anni grazie a una serie di innovazioni tecnologiche, di successi applicativi e di impatti economici significativi. La crescente capacità computazionale dei computer e la sovrabbondanza dei dati hanno permesso lo sviluppo di modelli di IA sempre più sofisticati. Ne sono esempio alcune applicazioni di successo dell'IA come il riconoscimento vocale e facciale, l'elaborazione del linguaggio naturale e i veicoli a guida autonoma hanno generato molto interesse da parte delle imprese. Gli investimenti pubblici e privati sono aumentati considerevolmente rendendo più accessibili le risorse per lo sviluppo dell'IA. La ricerca scientifica nel campo dell'IA, secondo l'osservatorio dell'Organizzazione per la cooperazione e lo sviluppo economico (di seguito, OCSE)[8], ha avuto una crescita

[6] Eric R Kandel, *Principles of Neural Science, Fifth Edition* (McGraw Hill Professional 2013) 1–3.

[7] Stanislas Dehaene, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (Penguin 2014) 8–10.

[8] L'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico) è un'organizzazione internazionale che promuove politiche per migliorare il benessere economico e sociale delle persone in tutto il mondo. Fondata nel 1961, l'OCSE svolge un ruolo importante nel fornire analisi, raccomandazioni e strumenti di collaborazione per affrontare le sfide economiche, sociali e ambientali che i paesi membri devono affrontare. L'organizzazione si concentra su questioni come la crescita economica, l'occupazione, l'innovazione, l'istruzione, l'ambiente, la governance pubblica e molte altre. Attraverso il dialogo politico e la cooperazione tra i suoi membri, l'OCSE mira a promuovere politiche e pratiche che favoriscano una prosperità sostenibile e inclusiva a livello globale.

sostanziale negli ultimi quarant'anni, sebbene si sia concentrata solo in alcune parti del mondo. Gli Stati Uniti d'America (da qui in poi, USA) rimangono ancora oggi lo Stato che contribuisce maggiormente allo sviluppo tecnologico; tuttavia, l'impegno globale nella ricerca in materia è cresciuto molto anche nell'Unione Europea (di seguito, UE) e in Cina[9]. USA, UE e Cina hanno maggiori risorse da investire in tecnologia e sviluppo. Nonostante ciò, anche altri Stati come Israele[10] e il Regno Unito rivestono un ruolo importante nel progresso tecnologico.

Tale fenomeno ha determinato un dibattito sociale ed etico sulle applicazioni dell'IA, spingendo la ricerca di regolamentazioni e soluzioni adeguate. Ancora non sono presenti normative vigenti in materia, ma un importante attore internazionale come l'Unione europea sta procedendo verso la formulazione di una nuova regolamentazione sull'IA. Nel frattempo, diversi fattori tra cui la necessità dell'IA di analizzare grandi quantità di dati, la presenza di dati personali nei *dataset* utilizzati per l'addestramento e i potenziali danni per la *privacy* che possono derivare da un utilizzo scorretto di tale tecnologia, hanno permesso alla normativa sulla protezione dei dati di regolare parzialmente tale fenomeno tecnologico. I diritti che oggi consideriamo rientranti nel termine "ombrello" di *privacy* sono stati riconosciuti e protetti in alcune convenzioni di rilevanza internazionale. In Europa, il Consiglio d'Europa ha adottato nel 1950 la Convenzione Europea dei Diritti dell'Uomo e delle Libertà fondamentali, pilastro fondamentale della democrazia occidentale, ove all'art. 8 sancisce il diritto al rispetto della vita privata e familiare, all'interno del quale la Corte Europea dei Diritti dell'Uomo ha fatto rientrare il diritto alla protezione dei dati personali.[11] Allo stesso modo, la Dichiarazione

---

[9] OECD.AI, visualizzazioni alimentate da JSI utilizzando dati da MAG, (versione del 31/12/2021), <www.oecd.ai> consultato il 21 febbraio. 2023.

[10] Israele rappresenta l'eccezione di una nazione piccola se messa a confronto con i tre giganti, ma che ha una florida realtà di start-up ad altissima qualità tecnologica e che fa un uso innovativo dell'intelligenza artificiale. Il governo ha dimostrato di riconoscere questa propensione innovativa del settore industriale israeliano ed ha dato il via al piano nazionale per lo sviluppo dell'IA.

[11] Consiglio d'Europa, Convenzione europea dei diritti dell'uomo, STCE n. 5, 1950. L'art. 8 prevede: "*Ogni persona ha diritto al rispetto della propria vita privata e familiare, del proprio domicilio e della propria corrispondenza. Non può esservi ingerenza di una autorità pubblica nell'esercizio di tale diritto a meno che tale ingerenza sia prevista dalla legge e costituisca una misura che, in una società democratica, è necessaria alla sicurezza nazionale, alla pubblica sicurezza, al benessere economico del*

Universale dei Diritti Umani delle Nazioni Unite già nel 1948 aveva sancito il diritto alla *privacy* come un diritto fondamentale dell'individuo, tutelato dall'articolo 12[12]. La tutela della *privacy* si estende anche alla sfera digitale, dove si sono sviluppati nuovi strumenti tecnologici e modalità di comunicazione che richiedono una particolare attenzione. Nell'ambito del Consiglio d'Europa, si può individuare la Convenzione del Consiglio d'Europa sulla protezione delle persone in relazione al trattamento automatizzato di dati a carattere personale n. 108 del 1981[13], che costituisce attualmente l'unico strumento giuridico vincolante a livello internazionale in materia di protezione dei dati. Tale Convenzione disciplina il trattamento dei dati personali, sia da parte di privati che di enti pubblici, e regola anche i trasferimenti transfrontalieri di dati tra gli Stati aderenti alla Convenzione, con alcune restrizioni nei confronti dei paesi terzi. Inoltre, la Convenzione stabilisce il divieto di trattamento dei dati sensibili senza adeguate garanzie di protezione. La protezione dei dati personali è un elemento fondamentale nello sviluppo dei sistemi di IA, in quanto questi basano il loro funzionamento sull'elaborazione dei dati.

Già dal 2017 l'UE ha iniziato a considerare il tema dell'IA, sia attraverso diverse fonti di *soft-law*, che attraverso la prima proposta di regolamento che armonizza la disciplina a livello europeo. Nonostante, il legislatore abbia iniziato ad interessarsi alla materia un po' in ritardo, bisogna comunque riconoscere che a livello globale l'UE è l'istituzione che per prima ha deciso di regolare il settore.[14] La migrazione di gran parte delle attività

---

paese, alla difesa dell'ordine e alla prevenzione dei reati, alla protezione della salute o della morale, o alla protezione dei diritti e delle libertà altrui"

[12] Assemblea Generale delle Nazioni Unite, Dichiarazione Universale dei Diritti Umani, Risoluzione 219077A, (dicembre 1948). L'art. 12 stabilisce: "*Nessun individuo potrà essere sottoposto ad interferenze arbitrarie nella sua vita privata, nella sua famiglia, nella sua casa, nella sua corrispondenza, né a lesione del suo onore e della sua reputazione. Ogni individuo ha diritto ad essere tutelato dalla legge contro tali interferenze o lesioni*."

[13] Convenzione del Consiglio d'Europa sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale, STCE n. 108, [1981]

[14] Il progresso tecnologico all'interno dell'Unione è stato sempre avviato dal Regno Unito, che risulta essere lo stato più virtuoso in materia, basti pensare che già nel 2016 vennero presentate le prime osservazioni nel parlamento inglese riguardo l'IA. Purtroppo, dopo l'uscita dello Stato dall'Unione, si sono creati due binari che non sempre viaggiano paralleli.

umane verso la dimensione digitale e la circolazione di un numero di informazioni crescente hanno profondamente trasformato il nostro mondo e il modo di relazionarci con lo stesso, costruendo quella che il professor Luciano Floridi[15] chiama l'*infosfera*[16]. Infatti, oltre all'aumento della quantità di informazioni in circolazione, la rilevanza delle stesse è diventata primaria. A riguardo, nel 2016 l'UE ha emanato il *General Data Protection Regulation* (di seguito, GDPR)[17], un regolamento europeo per la protezione dei dati personali, che ha assunto una posizione apicale anche in materia di trattamenti automatizzati di dati personali, inclusi quelli relativi ai sistemi di IA. Nel maggio 2018, con l'entrata in vigore del GDPR, l'attenzione alla regolamentazione del trattamento dei dati si è concentrata sulla categoria di dati strettamente correlati alla vita delle persone fisiche: i dati personali. All'articolo 4 del GDPR, questo tipo di dato viene definito come "*qualsiasi informazione riguardante una persona fisica identificata o identificabile*"[18]. Il ragionamento del regolatore parte dal presupposto che l'utente si trovi in una posizione di debolezza di fronte alla complessità e opacità dei meccanismi che raccolgono e analizzano i propri dati personali, esiste un evidente squilibrio informativo, relativo al trattamento, tra ciò che conosce il titolare a cui viene ceduto il dato e ciò che conosce l'utente che utilizza il servizio. La legislazione europea in materia ha cercato di bilanciare questo squilibrio, attribuendo al cittadino europeo, definito *data subject*, gli strumenti per riacquisire il controllo delle informazioni personali. All'interno del Regolamento sono infatti presenti diversi diritti che possono

---

[15] Luciano Floridi (1964) è un filosofo italiano naturalizzato britannico. Attualmente ricopre la posizione di professore ordinario di filosofia ed etica dell'informazione presso l'Oxford Internet Institute dell'Università di Oxford, dove è anche direttore del Digital Ethics Lab. Inoltre, è professore di Sociologia della comunicazione presso l'Università di Bologna. La sua ricerca filosofica si concentra principalmente sulla filosofia dell'informazione, la filosofia dell'informatica e l'etica informatica. Floridi è noto per aver coniato il termine "Onlife", che descrive il concetto di interazione tra la vita offline e la vita online e le implicazioni etiche ad esso associate.

[16] Luciano Floridi, *Pensare l'infosfera. La filosofia come design concettuale* (Cortina Raffaello 2020) 1–5.

[17] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), [2016], OJ L 119, 4.5.2016, p. 1–88

[18] Regolamento (UE) 2016/679, Art. 4.

essere esercitati dal cittadino in un dialogo diretto con il titolare del trattamento o con l'autorità nazionale per la protezione dei dati. L'attuazione e il controllo sul rispetto del Regolamento sono operati dalle autorità garanti nazionali che ne interpretano il testo e agevolano l'evoluzione normativa, necessaria per tenere il passo delle tecnologie.

2. L'Unione Europea verso una nuova regolamentazione dell'Intelligenza Artificiale

Poco prima dell'entrata in vigore del GDPR, coerentemente con il piano europeo per un mercato unico digitale[19], come anche stabilito anche dall'articolo 114 del Trattato sul Funzionamento dell'Unione Europea[20], venne pubblicata l'iniziativa europea per l'Intelligenza Artificiale[21], come anche avevano fatto USA[22] e Cina[23]. L'UE aspirava a stabilire *standard* univoci nell'utilizzo e commercializzazione di tecnologie di IA, che potessero non solo tracciare la via da seguire nello sviluppo dell'IA nei vari Stati membri, ma anche consolidare un quadro normativo che fungesse da riferimento a livello globale, come già accaduto con il GDPR. La strategia europea inizia con la pubblicazione del Libro Bianco sull'Intelligenza Artificiale[24], che stabilisce due obiettivi principali da raggiungere: un ecosistema di eccellenza per lo sviluppo e l'applicazione dell'IA e l'ottenimento di un IA affidabile, basata su principi etici e giuridici nel rispetto dei diritti fondamentali e dei diritti dei consumatori. Anche il Parlamento europeo ha

---

[19] Consiglio dell'Unione Europea, Conclusioni del Consiglio – Plasmare il futuro digitale d'Europa (9 giugno 2020) 8711/20

[20] Versione Consolidata del Trattato Sull'unione Europea E Del Trattato Sul Funzionamento Dell'unione Europea [2016] OJ C 202/01

[21] Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni, "Strategia per il mercato unico digitale in Europa", COM(2015) 192 final

[22] E.O. 13859 of Feb 11, 2019

[23] Sheehan Matt, 'China's New AI Governance Initiatives Shouldn't Be Ignored' (*Carnegie Endowment for International Peace* 4 January 2022) <https://carnegieendowment.org/2022/01/04/china-s-new-ai-governance-initiatives-shouldn-t-be-ignored-pub-86127> accessed 15 March 2023

[24] Libro Bianco sull'Intelligenza Artificiale – Un approccio europeo all'eccellenza e alla fiducia COM (2020) 65 final.

prodotto numerose risoluzioni sul tema[25], che affrontano aspetti etici, di responsabilità e di diritto d'autore, spingendo la Commissione europea a proporre un atto legislativo per sfruttare opportunità e benefici dell'IA e tutelare principi etici e giuridici.

Circa tre anni dopo l'avvio dell'iniziativa europea, la Commissione europea ha presentato, nell'aprile 2021, la prima proposta di regolamento europeo sull'Intelligenza Artificiale (definito "Artificial Intelligence Act", d'ora in poi, AI Act)[26]. Tale proposta si fonda sul risultato di una consultazione pubblica mirata a individuare il modo migliore per regolare il fenomeno tecnologico dell'IA. Nella consultazione venivano proposti diversi approcci, tra cui, norme di *soft law* e norme di *hard law* con l'obiettivo di individuare principi e stabilire divieti e limiti. L'approccio poi intrapreso dalla Commissione si è rivolto a regolare in maniera diversa situazioni caratterizzate da rischi diversi, distinguendo le pratiche di IA vietate da applicazioni che hanno un alto, limitato o minimo rischio. Secondo le parole di Margrethe Vestager[27] (Vicepresidente esecutivo della Commissione europea) "[…] *the higher the risk that a specific use of AI may cause to our lives, the stricter the rule."*

La proposta punta a un bilanciamento tra incentivi all'innovazione e regole per la tutela dei diritti fondamentali, prevedendo diversi livelli di rischio dei sistemi di IA con distinte limitazioni. La proposta pone l'accento sulle tecnologie ad alto rischio, che sono quelle che richiedono una regolamentazione più stringente. Più di altre tecnologie, l'IA

---

[25] Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL)); Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL)); Risoluzione del Parlamento europeo del 20 ottobre 2020 sui diritti di proprietà intellettuale per lo sviluppo di tecnologie di intelligenza artificiale (2020/2015(INI)); Risoluzione del Parlamento europeo del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale (2020/2266(INI))

[26] 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 COM(2021) 206 final

[27] Jorge Liboreiro, '"The higher the risk, the stricter the rule": Brussels' new draft rules on artificial intelligence' *Euronews* (Brussels, 21 aprile 2021), < https://www.euronews.com/my-europe/2021/04/21/the-higher-the-risk-the-stricter-the-rule-brussels-new-draft-rules-on-artificial-intellige> ultimo accesso il 19/04/2023

è contraddistinta da una mutevolezza spiccata e da uno sviluppo tecnologico che potrebbe rapidamente superare previsioni legislative rigide ed eccessivamente specifiche.

Tra le preoccupazioni che sono sorte con l'avvento dell'IA, è presente il rischio di una insufficiente tutela della protezione dei dati. Tale problema è al centro del dibattito da molti anni e il legislatore europeo ha fornito alcuni strumenti per tutelare la riservatezza, come ad esempio, il principio della "*minimizzazione dei dati*" previsto dall'articolo 5, para. 1, lett. c) del GDPR[28]. Il principio richiede che i dati utilizzati e raccolti siano solo quelli necessari e sufficienti per il raggiungimento della finalità dichiarata del trattamento. La sfida sta nel garantire un uso responsabile dell'IA che possa rispettare i diritti connessi alla protezione dei dati personali.

Come vedremo nei seguenti capitoli, riuscire ad applicare i principi di protezione dei dati personali anche alle tecnologie di IA non è così semplice. La natura di queste tecnologie è quella statistica: esse analizzano i dati, che siano personali o meno e traggono determinate conclusioni per la risoluzione del problema presentatogli. L'IA raggiunge livelli di complessità a noi non accessibili, il problema rimanente è che la regolamentazione che possiamo implementare deve tenere conto anche dei limiti umani. L'IA si pone come strumento con un duplice uso: se da una parte i sistemi di IA possono aggirare le prescrizioni contenute nel GDPR e operare in ogni caso un trattamento dei dati che viola la *privacy*, dall'altro gli strumenti di IA potrebbero essere utilizzati per aumentare l'efficacia di principi e regole stabilite dal legislatore europeo. Questa nuova e mutevole tecnologia sta già operando un cambiamento del paradigma della raccolta dei dati, determinando un potenziale indebolimento dei diritti degli interessati, ma aprendo nuove frontiere per aumentare l'efficacia della protezione dei dati.

[28] GDPR, art. 5, para. 1, lett. c) prevede che i dati debbano essere: "*adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati («minimizzazione dei dati»)*"

3. L'evoluzione del concetto di privacy: dall'articolo di Warren e Brandeis alla società digitale contemporanea

A fine diciannovesimo secolo, due giuristi americani, Warren e Brandeis, pubblicarono sulla *Harvard Law Review* un articolo che avrebbe rivoluzionato il concetto di riservatezza e sfera privata della persona[29]. I due studiosi suggerivano una rivisitazione del concetto di riservatezza in quanto, se fino a quel momento la raccolta di dati avveniva solo con modalità che implicavano il consenso del soggetto, le nuove tecnologie, come la fotografia, rendevano più difficile ostacolare l'intromissione nella sfera privata delle persone. Warren e Brandeis nel loro articolo descrissero il diritto alla riservatezza come "*the right to be let alone*" (che si potrebbe tradurre come: il diritto di rimanere da soli o di essere lasciati in pace). È da leggere come la necessità di garantire all'individuo il diritto di scegliere se e quanto delle proprie idee, emozioni e sentimenti possano essere divulgati al pubblico, considerando, eventualmente, la preminenza dell'interesse superiore della collettività a conoscere determinate informazioni. La riflessione dei due giuristi sembra esser nata dal maggior utilizzo di fotocamere "istantanee" e dalla diffusione dei giornali che sempre di più riportavano dettagli delle vite private di personalità importanti del tempo[30]. Questa riflessione è rimasta attuale, anche a distanza di più di cento anni. Tuttavia, il concetto di *privacy* come oggi lo intendiamo ha a che fare con strumenti tecnologici molto evoluti, che entrano in maniera enormemente più pervasiva nella vita privata delle persone. Infatti, se Warren e Brandeis intendevano originariamente escludere le persone dalla conoscenza di informazioni private, il paradigma di *privacy* si evolve con l'avvento dei computer e con l'avvio di diverse nuove forme di raccolta dei dati. La persistente rilevanza di questa riflessione è evidente. Il concetto moderno di *privacy* è profondamente influenzato da strumenti tecnologici avanzati che penetrano in modo significativamente più pervasivo nella sfera privata degli individui. Sebbene Warren e Brandeis mirassero a proteggere le persone dall'esposizione di informazioni riservate. Il concetto originario di *privacy*, basato sull'esclusione altrui, si trasforma quando le informazioni personali acquisiscono

[29] Samuel D. Warren and Louis D. Brandeis, 'The Right to Privacy' [1890] Harvard Law Review 193

[30] Nello specifico, la critica dei due giuristi era rivolta alla malsana pratica di alcune riviste di *pettegolezzo*, che dall'esterno fotografavano quel che accadeva nelle case delle persone più note.

un maggiore valore economico e consentono una dettagliata ricostruzione della personalità dell'individuo, con il potenziale di esercitare una certa manipolazione.

Con l'avvento dei computer, capaci di raccogliere un'ampia gamma di informazioni sulle persone, la necessità di proteggere tali dati assume una nuova rilevanza. La minaccia non è più limitata alla divulgazione di affari privati da parte dei media, ma si estende a rischi più gravi per il mantenimento della democrazia, la libertà di espressione e la libertà di scelta degli individui. Ad esempio, i social network, basandosi sulla raccolta dati degli utenti, presentano una realtà distorta e semplificata, selezionando solamente le notizie che rafforzano le convinzioni preesistenti di chi ne usufruisce. Gli algoritmi che li governano si basano su sistemi di IA molto avanzati che apprendono con la continua raccolta dei dati personali le migliori strategie per far trascorrere la maggior quantità di tempo possibile sulla piattaforma.

In occasione di un evento promosso dal CERN per celebrare il trentennale della creazione del World Wide Web, il suo fondatore, Tim Berners-Lee, identificava tre minacce che richiedevano interventi per salvaguardare il Web.[31] La prima era costituita da comportamenti dolosi, quali la pirateria, gli attacchi informatici, le molestie online e altre azioni criminali, che compromettono sia la sicurezza del Web sia la fiducia degli utenti nei confronti di questi strumenti. Sfortunatamente, alcune innovazioni, pur rivoluzionarie per il benessere della società, diventano pericolose quando utilizzate con intenti malvagi e lesivi.

La seconda preoccupazione riguardava il funzionamento della rete, che favorisce e incentiva pratiche che sacrificherebbero il valore derivante dall'uso del Web. Un esempio emblematico è il sistema pubblicitario, che premia i contenuti capaci di catturare l'attenzione e la curiosità degli utenti (*clickbait*[32]). La facilità con cui si

---

[31] Philip Di Salvo, 'A 30 anni dalla nascita, ecco tutti i rischi che corre il world wide web' (*Wired*, 11 marzo 2019), <https://www.wired.it/internet/web/2019/03/11/world-wide-web-30-anni/> ultimo accesso il 19 aprile 2023

[32] Con questo termine si indica il contenuto che solitamente non ha un reale valore, ma riesce bene a generare traffico e dunque profitto per le piattaforme, che evidentemente lo preferiscono ad un contenuto di valore, ma che non interessa al grande pubblico. Il termine viene dalla parola *click*, onomatopeica per indicare il suono del mouse che si poggia sul contenuto per visualizzarlo e clicca, generando traffico, dati, informazioni e quindi profitto.

diffondono le notizie false (*fake news*) attraverso questi canali, al fine di perseguire specifici obiettivi, è un fenomeno allarmante. La diffusione di teorie cospirative si è aggravata durante la devastante pandemia di Covid-19, ad esempio[33].

Terza minaccia per il Web riguardava le conseguenze non volontarie, ma negative che derivano da buoni propositi, come ad esempio l'uso di toni indignati e la conseguente polarizzazione del dibattito, in particolare sui social network. Anche l'utilizzo dei social come mezzo di comunicazione per le organizzazioni criminali e terroristiche, oltre che per una maggiore diffusione di teorie non scientifiche, è amplificato e reso più semplice da queste nuove piattaforme di comunicazione.

È evidente come tali minacce siano fondate sulla crescente importanza che i nostri dati personali hanno acquisito, in particolare durante gli ultimi venti o trent'anni. Durante questo periodo l'evoluzione dell'IA ha registrato un'accelerazione significativa, con un'impronta predominante da parte delle aziende *big tech*. Queste aziende, come Google, Facebook, Amazon e Apple, hanno concentrato risorse significative nello sviluppo e nell'applicazione di sistemi di IA grazie alle loro ingenti quantità di dati personali.

Tali aziende, infatti, sono state in grado di accumulare enormi quantità di dati provenienti da varie fonti, come interazioni online, transazioni commerciali e dispositivi connessi. Questi dati ricchi di informazioni sono diventati un'importante risorsa per addestrare algoritmi di IA e migliorare le prestazioni dei loro servizi, come ad esempio la precisione dei motori di ricerca, degli assistenti virtuali, delle raccomandazioni personalizzate e delle pubblicità mirate. Tuttavia, questa concentrazione di potere e dati personali ha suscitato preoccupazioni riguardo alla protezione della *privacy* e alla sicurezza dei dati. L'uso intensivo dei dati personali nelle applicazioni di IA ha sollevato interrogativi sulla trasparenza degli algoritmi e sulla possibile discriminazione o manipolazione dei consumatori.

---

[33] Uno dei più recenti casi di disinformazione è da legare alla confusione generale dovuta al Covid-19, un virus poco conosciuto e che ha avuto una diffusione rapidissima nel mondo, senza dar tempo agli studiosi di approfondire la sua conoscenza e dunque di saper indicare i rimedi più adatti. Di seguito un articolo sul tema: Daniele Orso, Nicola Federici, Roberto Copetti, Luigi Vetrugno, Tiziana Bove 'Infodemic and the spread of fake news in the COVID-19-era' [2020] European Journal of Emergency Medicine 327.

Le sfide per la protezione dei dati personali nell'era dell'IA sono complesse e richiedono l'adozione di soluzioni multidimensionali. È necessario un dialogo continuo tra le istituzioni, le aziende e gli utenti per definire regole e standard che proteggano la *privacy* senza ostacolare l'innovazione e il progresso tecnologico.

4. Storia dell'Intelligenza Artificiale

L'idea di un'entità che possa assistere l'uomo nei lavori meno creativi ha origini nella cultura greca e in particolare nella filosofia occidentale di Aristotele che afferma come "*non servirebbero né schiavi né lavoratori se ogni strumento potesse compiere il proprio lavoro anticipando o obbedendo alla volontà altrui*", facendo l'esempio del plettro che tocca la lira senza l'intervento di una mano che lo guidi.[34]

Partendo da simili idee Thomas Hobbes descrive il suo Leviatano[35] a cui seguono numerosi intrecci tra filosofia e scienza, passando per la tecnica e l'informatica arrivando oggi ai moderni sistemi di IA[36]. In passato, molti filosofi, come Hobbes, descrivevano il pensiero umano come una 'manipolazione meccanica di simboli'. Questa definizione si adatta molto bene ai metodi di *machine learning* che, partendo da informazioni ricevute (input), creano, manipolando meccanicamente i "simboli", cioè le nuove informazioni (output). Altri filosofi come Gottfried Wilhelm Leibniz e Blaise Pascal hanno teorizzato la possibilità di dispositivi che ragionano meccanicamente, utilizzando le regole della logica per risolvere i conflitti, anche se non arrivarono mai a dire che questa fosse una forma di pensiero.

Così nel pensiero filosofico come nel mondo della "*science fiction*", gli autori come Frank Baum, Jules Verne e Isaac Asimov, utilizzarono questo concetto di macchine

---

[34] Aristotele, *Politica*, Libro I, IV secolo a.C.

[35] Thomas Hobbes, *Leviathan* (prima edizione 1651, Penguin 1985) 268

[36] I moderni sistemi di IA sono rappresentati, ad esempio, dai *Large Language Models*(LLM), sistemi di IA fondati su algoritmi di apprendimento profondo capaci di riconoscere, riassumere, tradurre e generare del testo o altro tipo di contenuto, sulla base della conoscenza racchiusa nei loro dataset. Un esempio di LLM famoso oggi è "GPT 3.5" di OpenAI.

pensanti per personaggi che si dimostravano intelligenti, pur non essendo umani. Baum, ad esempio, descriveva l'uomo di latta come un "*uomo meccanico, estremamente sensibile, creatore di pensieri, parlatore perfetto, […] pensa, parla, agisce e fa tutto tranne vivere*"[37].

Probabilmente l'IA di oggi ha alcuni di questi attributi immaginati da Baum, ad esempio potremmo dire di un software LLM che è uno scrittore perfetto, in quanto non commette errori grammaticali od ortografici. L'uomo di latta immaginato da Baum, ci porta a considerare i moderni robot.

Il Parlamento europeo ha prodotto diverse risoluzioni sul tema della robotica fin dal 2017 con le prime raccomandazioni alla Commissione riguardo norme di diritto civile sulla robotica. Il primo utilizzo della parola *robot* si registra nel 1920 con il dramma fantascientifico R.U.R. (*Rossumovi univerzální roboti*) dello scrittore ceco Karel Čapek. Secondo Alessandro Catalano, autore de "*I robot di Karel Čapek: 100 anni di metamorfosi*", il termine derivava dal ceco *robota* che indicava il lavoro faticoso.[38] "Robot" è ancora oggi un termine molto utilizzato per riferirsi a quelle macchine umanoidi che svolgono attività principalmente fisiche e che replicano il comportamento delle persone. Per avere un'idea dell'attuale stato di avanzamento nel campo della robotica, si prenda in considerazione un'azienda all'avanguardia come Boston Dynamics[39].

---

[37] L. Frank Baum, *The Wonderful Wizard of Oz*. (Michael Patrick Hearn, W. W. Norton & Company 1980)

[38] Il termine è stato poi trasformato al maschile "*robot"* dal fratello di Čapek, Josef, per definire una creatura artificiale non distinguibile dall'essere umano, che rappresentasse un operaio artificiale caratterizzato da una propensione ad una maggiore efficienza lavorativa.

[39] Nata nel 1992 da alcune menti del MIT, Boston Dynamics è una società che produce robot. La popolarità di questa azienda è dovuta alla altissima qualità dei prodotti, nello specifico, il robot in considerazione si chiama ATLAS ed ha sembianze umane, alto circa 1,5 m con gambe e braccia. In un video pubblicato dall'azienda nel mese di gennaio 2023, lo si vede nell'intento di aiutare un muratore su un'impalcatura: ATLAS posiziona una tavola, sale gradini, prende oggetti, salta e lancia un borsone. Quel che più impressiona guardando il video è l'agilità e la destrezza che il robot ha nei movimenti, oltre alla precisione con cui li esegue.

Lo sviluppo dei robot può portare anche alla creazione di sistemi d'arma autonomi, come ricordato dalla risoluzione del 12 settembre 2018[40], in cui il Parlamento europeo ne sottolinea i rischi e chiarisce come l'UE sia fortemente contraria a qualsiasi sviluppo di tali sistemi. L'attenzione del legislatore europeo si è successivamente soffermata sulle potenzialità derivanti da tale tecnologia promuovendo una politica industriale in materia, attraverso la risoluzione del 12 febbraio 2019, culminando, infine, nella risoluzione del 20 ottobre 2020 in cui si propone un regolamento che contiene principi etici e obblighi giuridici per la diffusione e l'utilizzo dei sistemi robotici. Secondo dati del Sole24Ore[41] in Italia sarebbero presenti circa 70 mila robot impiegati nelle industrie al 2018.

Se i membri del seminario tenutosi al Dartmouth College, potessero vedere oggi i progressi in materia di IA, rimarrebbero impressionati. Sebbene la definizione di IA sia cambiata nel tempo, si è sempre pensato alla macchina come strumento privo di autonomia decisionale e sempre a servizio dell'uomo. La nuova definizione di sistema di IA adottata dai rappresentanti del Parlamento europeo che si occupano della proposta di regolamento sull' IA, all'art. 3, par. 1, individua "*un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono*".[42]

L'uomo riveste una cruciale importanza nel funzionamento dell'IA. I sistemi di IA richiedono solitamente un input da parte dell'uomo per funzionare correttamente, poiché sono privi di autonomia. Autonomia, secondo la Treccani, significa: "capacità e facoltà di governarsi e reggersi da sé, con leggi proprie" oppure "diritto di autodeterminarsi e amministrarsi liberamente nel quadro di un organismo più vasto,

[40] Risoluzione del Parlamento europeo del 12 settembre 2018 sui sistemi d'arma autonomi (2018/2752(RSP))

[41] Nicoletta Cottone, 'Italia leader nella robotica: 100 storie di eccellenza' *Il Sole 24 ORE* (5 febbraio 2020) <https://www.ilsole24ore.com/art/italia-leader-robotica-100-storie-eccellenza-ACSEpOHB> ultimo accesso 10 febbraio 2023.

[42] AI Act, art. 3, para.1

senza ingerenze altrui nella sfera di attività loro propria". La parola deriva dal greco ed è composta da due elementi "*auto*-" e "*nomos*" ovvero "legge propria". Nella definizione della Treccani, la capacità di "stabilire leggi proprie" mancherebbe all'IA. Nella tradizione giuridica italiana, l'autonomia è un concetto strettamente legato alla capacità giuridica, in particolare nelle persone fisiche è insita per molti aspetti anche prima della nascita, già dal concepimento, secondo quanto previsto all'articolo 1 del Codice civile.[43] "*Conceptus pro iam nato habetur*" ("il concepito è considerato come già nato") è una locuzione latina che sintetizza un principio giuridico del diritto romano e successivamente nel diritto civile di alcuni paesi, tra cui l'Italia, in particolare in materia di successioni e diritto di famiglia. Questo principio stabilisce che, ai fini legali, un feto concepito è considerato come se fosse già nato, purché sia nato vivo in un momento successivo. Pertanto, il concepito può avere diritti successori e altri diritti come se fosse già nato al momento della concezione.

Paragonando un sistema di IA a un bambino, quest'ultimo acquisisce la capacità giuridica fin dalla nascita in virtù dell'articolo 1 del Codice civile. Tuttavia, l'IA attuale può leggere, scrivere e parlare, forse in modo più avanzato rispetto al bambino stesso, ma solamente entro i limiti di ciò che è stato programmato dai suoi sviluppatori. [44] Tale confronto evidenzia la necessità di valutare attentamente la possibilità di estendere la personalità giuridica agli enti dotati di IA, tenendo conto delle specifiche competenze e caratteristiche di ciascun sistema, nonché dei principi e delle finalità che ne guidano l'attribuzione ai soggetti giuridici.

---

[43] Regio Decreto 16 marzo 1942, n. 262, Approvazione del testo del Codice civile. (042U0262) (GU Serie Generale n.79 del 04-04-1942), articolo 1

[44] Il linguaggio che la macchina adotta è un linguaggio inventato dall'uomo, i più famosi sono C++ e Phyton, ma per l'appunto, si tratta comunque di una sorta di alfabeto inventato dall'uomo per la macchina, attraverso la logica Booleana o simili. Sebbene queste macchine siano ben più capaci e veloci degli uomini a seguire e usare l'alfabeto, non sanno stabilire leggi per governarsi autonomamente. Nonostante siano stati fatti diversi progressi nei sistemi di apprendimento automatico, che permettono di modificare il corpo di regole che la governano, attraverso processi di apprendimento come il *machine learning*, ancora, però, non si è arrivati a macchine che riescano a costruire un linguaggio di programmazione da cima a fondo, che dunque, non si basi sul linguaggio già usato da programmatori umani.

Si tratta dunque di macchine che esistono solo in funzione della capacità di obbedire a regole stabilite dall'uomo, anche noti come algoritmi. Nati dalla scienza esatta per eccellenza, la matematica, il loro nome (originariamente, *al-Khuwārizmī*) è stato coniato dal matematico arabo Muḥammad ibn Mūsa del IX secolo. Questo termine indicava "procedimenti di calcolo numerico fondati sopra l'uso delle cifre arabiche."[45] Tali procedimenti di calcolo risultano molto complessi da comprendere e seguire per una persona di media cultura, per tale motivo i lavori del Parlamento europeo e della Commissione europea, volti a regolare tale tecnologia, si sono concentrati anche sull'aspetto della comprensibilità di tali sistemi. La possibilità di comprendere quel che accade all'interno dei programmi di IA permetterebbe agli individui di mettere in discussione gli output e il percorso logico seguito. Si arriverà ad una situazione di equità solamente quando la comprensione dei percorsi logici e dei risultati forniti da tali modelli sarà alla portata di tutti gli individui che interagiscono con essi.

Per spiegare in cosa consistono gli algoritmi si fa solitamente riferimento alle ricette culinarie, quindi, a istruzioni ordinate e cronologiche, che servono a raggiungere un risultato preciso predefinito. Molti matematici e filosofi si sono adoperati alla creazione di una macchina che potesse eseguire gli algoritmi, senza l'intervento umano in ogni passaggio. Tra questi, Leibniz scriveva "*Non è degno di uomini eccellenti sprecare ore come schiavi nel lavoro del calcolo*", facendo emergere come questa attività fosse eccessivamente dispendiosa di tempo. Nel 1642 Pascal creò, per agevolare il lavoro del padre nella riscossione delle tasse, un marchingegno capace di eseguire somme e sottrazioni, che venne chiamato "pascalina".[46] Naturalmente questi calcolatori non ebbero il successo sperato, poiché potevano eseguire solo operazioni elementari e il loro utilizzo era pertanto limitato. Alcune macchine simili a quella inventata da Pascal si diffusero negli anni seguenti, ma il passo successivo avvenne quando nel 1672 Leibniz presentò la sua Machina Arithmetica alla Royal Society di Londra.[47] Questa era capace

---

[45] Luigi Laura, *Breve e universale storia degli algoritmi*, (Luiss University Press 2019).

[46] Blaise Pascal, *Machine arithmétique*, Oeuvres Complètes de Blaise Pascal, vol. X, (Louis Lafuma, Editions du Seuil 1972).

[47] Gottfried Wilhelm Leibniz, *A Machine for Doing Multiplication and Division* Mathematische Schriften, Carl Immanuel Gerhardt (Hahn'sche Buchhandlung, 1849). Queste macchine erano più simili a

di fare anche moltiplicazioni e divisioni. Duecento anni dopo, intorno al 1840 Charles Babbage decise di costruire la prima macchina capace di eseguire ogni compito matematico per cui era stata programmata e durante un seminario a cui partecipò anche Luigi Federico Menabrea, presentò il progetto. Menabrea rimase molto colpito dall'idea e scrisse un articolo in francese che venne tradotto in inglese, su richiesta di Babbage, dalla stimata matematica Ada Lovelace [48]. Sebbene la macchina ideata da Babbage non vide mai la luce, Lady Lovelace aggiunse tra le note quello che possiamo oggi considerare il primo programma per computer di sempre. La stessa Lovelace ebbe l'intuizione di creare delle schede perforate in cui erano incise le istruzioni da far seguire a dei telai meccanici, questo stesso metodo venne seguito nella programmazione dei più moderni computer nel secolo successivo. L'idea di Babbage venne poi ripresa da un giovane Alan Turing durante la prima metà del XX secolo.

Le innovazioni tecnologiche furono uno dei campi di battaglia in cui si combatterono le due grandi guerre. Con l'invenzione della radio di Guglielmo Marconi ad inizio del 1900, si era aperta una nuova frontiera della comunicazione estemporanea e sulle grandi distanze. Tuttavia, tale tecnologia usata nelle comunicazioni era facilmente intercettabile e per garantire la segretezza dei messaggi iniziarono ad esserci le prime cifrature. Durante la Seconda guerra mondiale i tedeschi per cifrare tutte le comunicazioni interne utilizzavano la macchina Enigma[49]. L'analisi per comprendere il

---

calcolatrici che a computer come intesi oggi, quest'ultimi infatti non si limitano ad eseguire operazioni matematiche, ma sono capaci, se programmati, di eseguire qualsiasi algoritmo.

[48] Federico Luigi Menabrea, 'Sketch of the Analytical Engine invented by Charles Babbage, Esq.' Tradotto da Ada Lovelace, *Scientific Memoirs*, Vol. III, Richard Taylor and William Francis, (1843). La storia di Ada Lovelace è quella di una ragazza che viene cresciuta solo dalla madre in quanto il padre, Lord Byron, non fu affatto presente. La madre di Ada volendo evitare che in lei nascesse una passione per la poesia e prendesse l'esempio del padre, decise di farle studiare matematica, materia nella quale Ada mostrò molta perizia e bravura.

[49] Utilizzando tre rotori con 27 contatti, uno per ogni lettera dell'alfabeto, Enigma era capace però di cambiare cifrario ogni volta che si premeva un tasto; dunque, per poter decifrare il messaggio era necessaria un'altra macchina Enigma nella stessa configurazione di quella utilizzata per scrivere il messaggio. I servizi segreti francesi e polacchi furono impegnati per decifrare i messaggi tedeschi, e grazie ad una squadra polacca che costruì la macchina Bomba, dotata anche di un manuale d'uso di Enigma arrivato dalla Francia, fu possibile decifrare i messaggi tedeschi.

funzionamento di questa macchina iniziò in Polonia per poi continuare in Inghilterra nel campus di Bletchley Park, a causa del timore di veder tutto il lavoro perduto onde una probabile invasione tedesca.[50] Nel campus lavorava anche un giovane Alan Turing, figura centrale per lo studio di Enigma. Alan Turing fu un matematico che, dopo un periodo di studi negli U.S.A., tornò in Inghilterra per essere coinvolto nel progetto di decrittazione della macchina Enigma. Tra gli studi più noti di Turing possiamo menzionare 'La Macchina Universale', che era stata teorizzata da Babbage, ma che Turing aveva finemente postulato e, anche, il famoso "test di Turing". Anche detta "Macchina di Turing", questo era un dispositivo capace di eseguire tutto ciò che un computer reale può computare, dalla simulazione di un qualsiasi algoritmo matematico alla ricerca di una parola in un testo.[51] Il modello venne ideato da Turing nel 1936 e viene oggi considerato uno dei fondamenti dell'informatica moderna.

Il Test di Turing, presentato nel 1950 nella rivista Mind, è stato proposto come una speculazione sul fatto che le macchine fossero in grado di pensare. Il matematico riconobbe la difficoltà di definire il concetto di pensiero e per questo ideò un esperimento. Ipotizzò così che una macchina fosse in grado di condurre una conversazione con una persona in modo tanto verosimile, da rendere molto difficile la comprensione di star conversando con una macchina. In altre parole, la macchina sarebbe stata considerata "intelligente" se fosse stata in grado di ingannare la persona.[52] Il matematico inizia col chiedersi "Can machines think?", una domanda che lo stesso Turing afferma essere difficile da rispondere in quanto bisognerebbe affidarsi ai significati comunemente associati alle parole che indicano la macchina ed il pensiero. Invece di rispondere a questa domanda, ne viene posta un'altra che non presenta alcuna ambiguità, la domanda prende la forma di un gioco, chiamato *The imitation game* (da cui il nome del film biografico su Turing del 2014). Nella prima versione del gioco, si pone il caso che a partecipare siano solamente persone, dunque, un giudice, una donna

[50] Stephanie Faint, 'The Enigma History and Mathematics' (MS thesis, University of Waterloo 2016).

[51] Alan M. Turing, 'On Computable Numbers, with an Application to the Entscheidungsproblem' (1937) Proceedings of the London Mathematical Society, vol. 42, no. 1.

[52] Alan M. Turing, 'Computing Machinery and Intelligence', [1950] Mind 236 <https://doi.org/10.1093/mind/LIX.236.433>

ed un uomo, tutti e tre siti in stanze separate. Il giudice deve capire se sta conversando con un uomo o con una donna e per evitare che il tono della voce rappresenti un importante indizio, si decide di condurre la conversazione in forma scritta. Dunque, si pone il caso che al posto di uno dei giocatori, escluso il giudice, ci sia un computer che risponda, cercando di imitare una persona.[53] Turing procede nel suo articolo analizzando una serie di opposizioni alla sua teoria, quella che più ci porta verso il concetto di IA è la numero sei, in cui si presenta l'obiezione di Lady Lovelace al progetto di Babbage. Lovelace afferma che la macchina analitica non ha la propensione ad originare qualcosa, ma è invece capace di fare qualsiasi cosa per cui venga programmata. La parte finale dell'articolo si concentra sull'apprendimento nei bambini. Attraverso un sistema di premi e punizioni i bambini possono imparare più velocemente, un simile sistema viene oggi applicato nei moderni sistemi di IA che utilizzano *machine learning* e prende il nome di "*reinforcement learning*"[54], questo è uno dei metodi di apprendimento maggiormente utilizzati oggi.

Al tempo di Turing era abbastanza paradossale immaginare un programma che potesse operare in questo modo, anche se lo stesso autore già prevedeva che gli insegnanti (i.e. i programmatori) di una macchina, non avrebbero saputo dire cosa accade all'interno della stessa. Questo è un tratto distintivo dell'IA moderna che viene paragonata ad una scatola nera, a causa della incapacità delle persone di seguire e comprendere i processi e meccanismi che vengono attivati all'interno di un sistema di reti neurali. È una tematica trattata anche dalla Risoluzione del Parlamento europeo sugli aspetti etici dell'IA, mantenendo l'ottica antropocentrica, in cui venne poi proposta una prima versione di un

---

[53] Turing stesso nel suo lavoro fa riferimento alla macchina di Babbage "*Analytical Engine"*, sottolineando anche come i progressi per il computer digitale siano stati fondamentali per accelerare i tempi attraverso l'utilizzo dell'elettricità.

[54] Il *reinforcement learning* ha come obiettivo quello di far compiere le azioni più adatte ad un programma o algoritmo per ottenere una ricompensa. Dunque, l'algoritmo impara attraverso la sperimentazione e l'interazione con l'ambiente attorno, senza tuttavia essere programmato per agire in un certo modo o seguire determinate regole. Il processo viene dunque seguito da una serie di momenti in cui l'algoritmo riceve dall'ambiente circostante degli stimoli (*feedback*). Questi stimoli possono consistere in ricompense positive o negative a seconda che l'azione attuata abbia portato ad un risultato voluto o non voluto. Dopo una lunga serie di azioni il sistema attraverso un semplice calcolo avrà imparato quali sono le azioni da compiere in ogni situazione per ottenere la massima ricompensa.

regolamento sui principi etici nello sviluppo e commercializzazione riguardo l'IA e la robotica.[55]

"*We can only see a short distance ahead, but we can see plenty there that needs to be done*", con questa affermazione Turing conclude il suo articolo facendo una istantanea dell'incertezza che si prova studiando i sistemi di IA, che si evolvono in maniera rapidissima e che allo stesso tempo, pongono diverse problematiche anche giuridiche. Tra le problematiche analizzate da Turing probabilmente non rientravano quelle relative all'intreccio tra la tutela dei dati personali e la necessità dell'IA di analizzare ed elaborare grandi quantità di dati. Anche se alcuni aspetti della sua riflessione rimangono molto attuali, come il sopracitato problema della comprensione del linguaggio del sistema.

Tra gli strumenti necessari per la comprensione di come l'IA possa inficiare i meccanismi di tutela dei dati personali esistenti e come il paradigma debba cambiare considerando i più recenti sviluppi, bisogna necessariamente analizzare la nascita della più grande rete di comunicazione e condivisione esistente: *Internet*. Le evoluzioni più importanti legate alla teoria dell'informazione, si devono alla creazione di tale rete tra computer, questa permetteva di comunicare molto velocemente anche ad enormi distanze. La storia di Internet ha le sue origini negli Stati Uniti ed in particolare nel Dipartimento della Difesa. Infatti, nel periodo della Guerra Fredda gli Stati Uniti cercavano un modo per comunicare che potesse resistere ad attacchi nemici. Secondo l'Enciclopedia Britannica, Paul Baran ideò per primo la commutazione a pacchetto (*packet switching*), un sistema di comunicazione fondato sulla condivisione di informazioni per mezzo di diverse linee di comunicazione, ma questa idea venne accantonata nei primi tempi. Il sistema permetteva di mandare dati sotto forma di tanti pacchetti che, arrivati al destinatario, venivano riassemblati, come un carico che viene distribuito su diversi vagoni, durante il viaggio rimane separato e all'arrivo viene ricongiunto. Questo serviva a garantire la sicurezza delle comunicazioni nel caso in cui

[55] Parlamento Europeo, *Risoluzione del Parlamento europeo del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale (2020/2266(INI))*.

una linea fosse interrotta. Fortunatamente negli uffici dell'ARPA l'idea venne considerata una soluzione alle necessità di comunicazione per la rete, che a breve si sarebbe formata; la commutazione di pacchetto rimane ancora oggi il contributo più importante per la creazione di Internet.

Quella che poi divenne la *Defense Advanced Research Project Agency*, anche conosciuta come DARPA, prima consisteva in un'agenzia che finanziava la ricerca tecnologica e che al tempo della guerra fredda si concentrava su una tecnica di comunicazione che permettesse più elevati standard di sicurezza.[56] A capo di questa agenzia per un periodo ci fu Lawrence Roberts uno dei padri fondatori del processo di apprendimento delle macchine[57]. Sebbene molti dei progetti furono e rimangono secretati per motivi militari, alcune innovazioni sono state diffuse alla società civile, come la ARPANET una rete di condivisione documenti tra computer.
Nello stesso anno in cui il primo uomo approdò sulla luna, la DARPA commissionò una rete di computer che fondava il funzionamento sulla commutazione di pacchetto, la rete prese il nome di ARPANET. Il primo collegamento tra computer si ebbe il 29 ottobre 1969, in particolare tra quello dell'Università della California a Los Angeles (UCLA) e l'istituto di Ricerca di Stanford (SRI). La rete continuò ad allargarsi nel tempo e nel 1971 venne inviato il primo messaggio e-mail.

Durante gli anni '70 questo modo di condividere le informazioni cominciò ad essere sempre più adottato, ma crebbe lentamente a causa del costo di manutenzione e acquisto del computer. In quegli anni l'ARPANET cominciò ad essere popolare all'interno della

---

[56] Michael Aaron Dennis, 'Defense Advanced Research Projects Agency', *Encyclopedia Britannica*, (23 dicembre 2022), <https://www.britannica.com/topic/Defense-Advanced-Research-Projects-Agency> ultimo accesso 10 aprile 2023

[57] Roberts nella sua dissertazione: "*Machine perception of three-dimensional solids"*, descrive e cerca di dare una soluzione al problema della incapacità delle macchine di riconoscere le immagini. Il sistema pensato è basato sul riconoscimento di pattern e analizza i dati in ingresso individuando gli schemi che caratterizzano le varie situazioni od oggetti. L'algoritmo elabora i dati e con diverse trasformazioni matematiche identifica gli schemi. In questo modo il sistema affina la sua tecnica di classificazione, riuscendo a migliorare le sue prestazioni. Questa tecnologia insieme alla commutazione di pacchetto venne utilizzata per l'addestramento di un sistema che, partendo dall'analisi di tante immagini di cifre numeriche scritte a mano, fornite dal Dipartimento della Difesa, riuscì ad individuare gli schemi e con l'addestramento poi anche a riconoscere le cifre anche tra rumore o distorsioni.

comunità accademica e di ricerca, come strumento per comunicare tra istituti diversi. Dovendo però usare un protocollo comune, si iniziò a adottare degli standard e l'Internet Protocol Suite noto anche come TCP/IP [58] divenne il protocollo di comunicazione comune, che permise di instradare i dati attraverso Internet. I fondatori di ARPANET non avevano il grande progetto di creare un sistema che permettesse le comunicazioni in tutto il globo, ma oggi tutti conosciamo bene il valore di Internet e come questa tecnologia di comunicazione e condivisione sia ormai fondamentale per la maggior parte delle nostre attività.[59] ARPANET è stato tra i primi esempi di rete tra computer a commutazione di pacchetto, permettendo la comunicazione e l'invio dei dati in pacchetti che seguono percorsi indipendenti attraverso la rete.[60] Dalla propensione alla comunicazione tra persone che contraddistingue la natura umana, la rete di utenti ha raggiunto numeri enormi e seguendo una funzione logaritmica, più si avvicinava al limite superiore e più la crescita rallentava, fino ad arrivare idealmente ad una copertura totale della popolazione. Lo sviluppo tecnologico finora brevemente analizzato ha avuto un impatto benefico sulla società, permettendo una maggiore comunicazione e condivisione di dati. I dati sono gradualmente diventati una delle risorse più importanti

[58] Transmission Control Protocol/Internet Protocol è il nome completo ed indica un insieme di protocolli di comunicazione utilizzati per connettere i dispositivi alla rete Internet.

[59] I numeri di Internet parlano chiaro. Sono circa 5,385,798,406 le persone che utilizzano la rete, su una popolazione mondiale che si aggira sugli 8 miliardi. I continenti con una più alta percentuale di utilizzatori sono America del Nord ed Europa (tra l'85 e il 90%), sebbene la somma delle rispettive popolazioni ammonti a meno della metà della popolazione connessa ad Internet in Asia (circa 3 miliardi di utenti). World Internet Users Statistics and 2023 World Population Stats (2023) <Internetworldstats.com>, ultimo accesso 10 aprile 2023.

[60] Negli anni ARPANET è stato esteso anche ad altre organizzazioni di ricerca ed università, determinando la creazione di una comunità di utenti di computer interconnessi che hanno cominciato a condividere risorse e dati, in maniera sempre più consistente e grazie al progresso delle tecnologie, anche più velocemente. Il primo paese fuori dagli Stati Uniti ad essere collegato a questa rete, fu la Norvegia attraverso il primo collegamento satellitare che aprì dunque le porte ad una rete satellitare che prese il nome di SATNET.

dell'era moderna. Grazie alla commutazione di pacchetto di cui si è parlato prima, la facilità di scambiarsi dati è cresciuta notevolmente.[61]

La creazione di Internet è però da distinguere dalla creazione del World Wide Web. Se la prima è una rete globale di computer interconnessi che permette lo scambio e la comunicazione tra gli stessi, il secondo è un sistema di documenti e risorse *ipertestuali* interconnesse, accessibile attraverso Internet, che consente di condividere risorse e informazioni a livello globale.[62] Questa invenzione deve attribuirsi ad un ricercatore del CERN europeo di nome Tim Berners-Lee, che riuscì a vedere il potenziale di un sistema che permettesse di collegare tra loro tutti i documenti presenti su Internet.[63] Dalla creazione di queste tecnologie ne è derivata una crescente produzione di dati, in particolare la diffusione del Web ha permesso una maggiore accessibilità alle informazioni, stimolando la produzione di dati nei campi più disparati. Questa enorme diffusione di dati ha creato anche sfide per la legge e mette ancora oggi in pericolo la *privacy* delle persone, richiedendo dunque una regolamentazione per la tutela dei diritti di persone e organizzazioni, come ad esempio il GDPR. L'esempio migliore per capire il funzionamento del World Wide Web è una qualsiasi pagina Wikipedia, la famosa enciclopedia online in cui gli utenti possono aggiungere dei contributi per arricchire una pagina web su un determinato argomento e che consiste in un groviglio ordinato di collegamenti tra diverse pagine ipertestuali. Senza il *Web* oggi non si avrebbero tanti servizi di condivisione e di ricerca che rendono molto più semplice la quotidianità[64].

---

[61] Se inizialmente la diffusione di queste tecnologie era limitata ad università, istituti di ricerca e governi, poi con la riduzione del costo e la commercializzazione dei computer, la rete Internet ha preso il sopravvento sugli altri mezzi di comunicazione e condivisione.

[62] Barry M. Leiner et al. 'A brief history of the Internet' [2009] ACM SIGCOMM Computer Communication Review.

[63] Tim Berners-Lee and Mark Fischetti, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor,* (Harper San Francisco 1999).

[64] In particolare, bisogna sottolineare il ruolo dei motori di ricerca e in particolare di Google che venne presentato nel 1998 e Ad appena 2 anni dalla sua presentazione, il motore di ricerca indicizzava già un miliardo di pagine Web. Il progetto dei due fondatori Brin e Page mirava ad un motore di ricerca che non solo mostrasse i risultati a seguito di una ricerca, ma autonomamente procedesse alla creazione di un ordine di importanza dei risultati, avendo così un motore molto più efficiente e che negli anni ha

Il nuovo millennio ha segnato l'avvento dell'IA moderna, con i progressi nell'*hardware*[65] e la disponibilità dei *Big Data* che hanno permesso lo sviluppo del *machine learning* e del *deep learning* [66]. Le reti neurali profonde, composte da molteplici livelli di neuroni artificiali, sono diventate lo standard per molte applicazioni dell'IA, dalla visione artificiale alla comprensione del linguaggio naturale. Nel 2011, IBM ha presentato Watson, un sistema di IA in grado di comprendere il linguaggio naturale e di rispondere a domande in modo coerente e corretto[67]. Nel 2016, il sistema di IA AlphaGo di DeepMind ha battuto un campione mondiale del gioco del Go, un traguardo che molti ritenevano irraggiungibile per l'IA. Negli anni successivi al 2020, l'IA è diventata onnipresente, permeando numerosi aspetti della nostra vita, dalla medicina alla guida autonoma, dall'analisi dei dati e all'assistenza clienti. Al contempo, si è fatta strada la consapevolezza delle sfide etiche e normative poste da queste tecnologie avanzate. La questione della *privacy* dei dati personali, del bias e dell'equità, e l'interrogativo sulla responsabilità delle decisioni prese dalle macchine rimangono argomenti controversi. Il cammino dell'IA, a circa settant'anni dai suoi albori, è stato sinuoso e pieno di sfide, ma ha condotto a progressi notevoli. Il futuro dell'IA, sebbene incerto, è senza dubbio affascinante. Nei prossimi anni vedremo un continuo miglioramento delle capacità delle macchine, ma anche un'attenzione crescente alle implicazioni etiche, sociali e legali di

---

continuato a migliorare le sue prestazioni, anche grazie alla massiccia analisi dei dati, tra le più sofisticate esistenti.

[65] L'*hardware* è l'insieme di tutti i componenti fisici, o materiali, di un sistema informatico. Comprende i dispositivi che eseguono le operazioni di elaborazione dei dati (come la CPU o processore), le memorie che conservano le informazioni (come l'hard disk, la memoria RAM), i dispositivi di input che permettono all'utente di interagire con il sistema (come la tastiera e il mouse), e i dispositivi di output che permettono al sistema di comunicare con l'utente o con altri sistemi (come il monitor o la stampante).

[66] Il deep learning è una branca dell'intelligenza artificiale che utilizza reti neurali profonde, ovvero modelli computazionali ispirati al funzionamento del cervello umano. Questi modelli sono in grado di apprendere da enormi quantità di dati non strutturati attraverso un processo di addestramento e di perfezionamento automatico, migliorando la loro capacità di riconoscimento e di previsione nel tempo. Yann LeCun, Yoshua Bengio and Geoffrey Hinton, 'Deep Learning' (2015) 521 Nature 436, 436–444.

[67] DA Ferrucci, 'Introduction to "This Is Watson"' (2012) 56 IBM Journal of Research and Development 1:1. Tale sistema di IA è stato testato anche in quiz televisivi come "Jeopardy!" e ha avuto un grande successo mediatico. Per maggiori informazioni si vedano i seguenti articoli di giornale: Tom Lamont, '*Meet Watson, the computer set to outsmart the champions of Jeopardy!*', The Guardian (6 February 2011).

queste tecnologie, nella ricerca di una sintesi che permetta di regolare tali sistemi senza privarci dei benefici dell'innovazione.

In un'era dominata dalle *big tech* e dalla digitalizzazione, negli ultimi anni abbiamo assistito ad una concentrazione sempre maggiore dello sviluppo dei sistemi di IA nelle mani di poche aziende tecnologiche globali. Le aziende anche note con l'acronimo GAFAM che sta per Google, Amazon, Facebook, Apple e Microsoft, sono tra le più grandi e influenti società tecnologiche del mondo e tale termine è spesso utilizzato per riferirsi all'enorme potere e influenza che queste aziende detengono nell'economia digitale e globale. L'utilizzo del termine GAFAM evidenzia spesso le preoccupazioni relative all'incalcolabile quantità di dati personali che queste aziende raccolgono, la loro influenza sul comportamento dei consumatori, la concorrenza nel mercato digitale, e le implicazioni per la *privacy* e la sicurezza dei dati. Questa concentrazione di dati e potere computazionale ha portato ad una notevole accelerazione nel progresso dell'IA, ma ha anche sollevato serie preoccupazioni relative alla *privacy* e alla protezione dei dati personali.

I dati personali, infatti, sono diventati la valuta di scambio dell'economia digitale e sono, al contempo, una risorsa fondamentale per l'addestramento e l'efficacia dei sistemi di IA. L'enorme quantità di dati personali raccolti dalle *big tech*, spesso senza un consenso pienamente informato da parte degli utenti, ha messo a rischio la *privacy* dei cittadini e ha generato la necessità di regolamentazioni sempre più stringenti. Inoltre, l'utilizzo di questi dati nei sistemi di IA può portare a rischi di discriminazione e bias, e solleva questioni complesse relative alla trasparenza e all'*accountability* delle decisioni prese dalle macchine.

L'interazione tra l'innovazione tecnologica e le questioni legali, etiche e sociali è diventata uno dei temi dominanti del dibattito pubblico e accademico. Da un lato, le big tech e l'IA hanno il potenziale di rivoluzionare la nostra società in modi positivi, creando nuove opportunità e migliorando l'efficienza in numerosi settori. Dall'altro lato, però, esistono rischi e sfide significativi che richiedono una riflessione profonda e una regolamentazione attenta per garantire che l'innovazione tecnologica si sviluppi in modo equo e rispettoso dei diritti fondamentali dei cittadini.

La crescente diffusione delle informazioni personali ha elevato la normativa sulla protezione dei dati personali a un ruolo di primaria importanza nel contesto legislativo

eurounitario. Il diritto europeo in materia ha riconosciuto il carattere trasversale e multidisciplinare di questo ambito, il quale si riflette in diverse discipline, quali il diritto della concorrenza e il diritto internazionale. La normativa di riferimento in questo settore è il GDPR, che è divenuto lo standard globale per la tutela dei dati personali. Grazie a questa portata, il GDPR è stato in grado di regolamentare sia settori consolidati e disciplinati, come quello della concorrenza, sia ambiti emergenti, come l'IA. La giurisprudenza, unitamente all'evoluzione normativa, ha quindi cominciato a forgiare un quadro in cui il diritto della concorrenza e la protezione dei dati personali convergono, inaugurando un approccio multidisciplinare all'analisi delle questioni legali connesse al mondo digitale. Questa interazione è fondamentale per capire come le Big Tech utilizzano la loro posizione di dominio per manipolare il mercato a loro vantaggio, spesso a scapito dei diritti dei consumatori e dei principi di concorrenza del mercato. I diritti riconosciuti all'interno di tali ambiti sono stati sviluppati e interpretati all'interno delle corti e in particolar modo dalla Corte di Giustizia dell'Unione Europea (di seguito, CGUE) che ha prodotto un'importante giurisprudenza in materia, che è bene attenzionare per comprendere realmente l'evoluzione della materia della protezione dei dati e come questa potrà reagire per rispondere alle problematiche che sorgono in tema di IA.

5. La più recente giurisprudenza in materia di protezione dei dati: Facebook Germany e le sentenze Schrems I e II

La tutela dei dati personali è un tema di centrale importanza, principalmente a causa dell'ampia raccolta, utilizzo e condivisione di informazioni personali da parte delle aziende nell'era digitale. L'evoluzione della giurisprudenza in questo settore ha evidenziato il ruolo che la regolamentazione sulla protezione dei dati personali può assumere nell'ambito del diritto della concorrenza, come dimostrato dal noto caso riguardante la disputa tra Facebook e l'Autorità tedesca per la concorrenza.

La società in questione raccoglieva e trattava i dati personali dei propri utenti, offrendo un servizio basato sul modello "*freemium*", in cui il corrispettivo ceduto dagli utenti consisteva nel consenso alla raccolta dei loro dati personali. Tale modello è adottato da

numerosi social network. Tuttavia, Facebook è stato al centro di diverse problematiche che hanno dato vita ad una importante giurisprudenza in materia. Dopo un periodo di difficoltà nel connettere la protezione dei dati con il diritto della concorrenza, per poter risolvere le sfide emergenti dalla realtà ed economia digitale, il caso che ha riguardato Facebook in Germania ha permesso di trovare una sintesi tra le due branche del diritto.[68] L'autorità amministrativa tedesca garante della concorrenza (da qui in poi, Bundeskartellamt) ha stabilito dei limiti nel trattamento dei dati personali per la società *Facebook Inc.*, poiché la stessa aveva sfruttato la sua posizione dominante nel mercato dei social network per stabilire condizioni sfavorevoli per i soggetti interessati. La novità portata dalla decisione B6-22/16 dell'autorità consiste nell'utilizzo di principi e norme provenienti dalle disposizioni che regolano la protezione dei dati, per analizzare i profili di antitrust e definire la posizione dominante del social network in questione.[69] In precedenza i due settori erano visti come paralleli e non incidenti. Al contrario, secondo il presidente dell'autorità tedesca, Andreas Mundt, "i dati sono un fattore decisivo per stabilire la posizione dominante di Facebook".[70] Inoltre, questi sono stati sfruttati dal social per determinare condizioni di *data policy* che erano in violazione delle disposizioni relative alle basi giuridiche per il trattamento di cui all'art 6 del GDPR.[71] Infatti, Facebook non concedeva agli utenti la possibilità di esercitare un'opzione riguardo al trattamento dei dati, ponendo l'individuo di fronte all'alternativa tra accettare il trattamento dei dati o rinunciare all'utilizzo del servizio. Tale decisione è stata successivamente impugnata da Facebook presso la Corte regionale superiore di Düsseldorf (Oberlandesgericht Düsseldorf), che ha mosso dei rilievi critici in merito alla mancata istruttoria del Bundeskartellamt in relazione a ipotetiche situazioni di mercato perfettamente concorrenziali, nonché sulla distinzione tra le condizioni

---

[68] Diversi illustri giuristi tedeschi, come il professore Rupprecht Podszun, hanno ricercato un modo per poter controllare le problematiche concorrenziali all'interno dei mercati digitali, soprattutto causate dal comportamento delle big tech. In particolare, si veda Rupprecht Podszun, Philip Marsden, *Restoring balance to digital competition - Sensible rules, effective enforcement*, Konrad-Adenauer-Stiftung e. V. (2020) 15-16.

[69] Decisione B6-22/16 del Bundeskartellamt [2019]

[70] Andreas Mundt, Press Release del Bundeskartellamt per il caso Facebook Germany, (7 febbraio 2019).

[71] Art. 6 GDPR, "

contrattuali applicate, suddivise in 'inadeguate' e 'adeguate'. Tale differenza si riferisce all'equità e alla conformità delle condizioni contrattuali rispetto alle normative sulla protezione dei dati e alle aspettative degli utenti.

In seguito, con la decisione VI-Kart 1/19 (V) del 26 agosto 2019, la Corte di Düsseldorf ha stabilito di sospendere gli effetti della decisione dell'autorità[72]. Il caso Facebook Germany è poi stato esaminato dalla Corte federale tedesca (Bundesgerichtshof) nella decisione KVR 69/19, in cui è stato stabilito che il social network aveva indiscutibilmente una posizione dominante nel mercato.[73] Sulla base di tale posizione, il social era poi riuscito a imporre condizioni abusive agli utenti. Questa decisione ha avuto un impatto significativo sulla relazione tra la legge sulla concorrenza e quella sulla protezione dei dati, nonostante la corte di Düsseldorf avesse in ultima istanza proposto un rinvio pregiudiziale alla CGUE, per verificare il ruolo che le autorità antitrust nazionali hanno nell'attuazione e implementazione del GDPR[74]. Lasciando in questo modo alla CGUE l'opportunità di chiarire il rapporto tra i due campi del diritto, dando alle disposizioni del GDPR una luce nuova, per chiarire le problematiche derivanti dai nuovi scenari digitali.

Il 20 settembre 2022 l'Avvocato Generale Rantos (di seguito, AG) ha pubblicato il suo parere chiarendo come, in generale, l'autorità antitrust non abbia giurisdizione su violazioni del GDPR, ma che comunque "nell'esercizio dei suoi poteri, possa tenere conto della compatibilità di una pratica commerciale con il GDPR" e che sulla conformità di questa condotta si possa basare un'importante indicazione per stabilire se il comportamento rappresenti una violazione del diritto *antitrust*[75]. La constatazione di

[72] Caso VI-Kart 1/19 (V), Decisione della Corte regionale superiore di Düsseldorf [2019]

[73] Bundesgerichtshof, decisione KVR 69/19, [2020].

[74] C-252/21: Request for a preliminary ruling from the Oberlandesgericht Düsseldorf (Germany) — *Facebook Inc. and Others v Bundeskartellamt*, OJ C 320 of 9.8.2021, p. 16–18 [2021], accessibile e consultabile al sito <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62021CN0252>.

[75] C-252/21: Request for a preliminary ruling from the Oberlandesgericht Düsseldorf (Germany) — *Facebook Inc. and Others v Bundeskartellamt*, [2021], Opinion of AG Rantos

una posizione dominante può influenzare l'analisi per verificare che l'utente abbia dato il consenso in maniera libera ed effettiva allo stesso operatore.

All'inizio di luglio 2023, dopo quasi un anno dall'opinione dell'AG, la sentenza della CGUE è stata pubblicata. La CGUE ha indicato la direzione che maggiormente permette al diritto della concorrenza di evolversi e rispondere alle moderne sfide[76]. La Corte ha infatti stabilito che, durante una verifica condotta da una autorità nazionale garante della concorrenza nel mercato (nel caso specifico la Bundeskartellamt), effettuare una valutazione sul rispetto di normative diverse da quelle del diritto della concorrenza, come ad esempio il GDPR, potrebbe essere necessario per stabilire se si è in presenza di un abuso di posizione dominante. Pur chiarendo che tale autorità, laddove individui una violazione del GDPR, non ha alcun potere di sostituire il garante per la protezione dei dati nazionale e che la valutazione a riguardo serve esclusivamente a stabilire l'eventuale abuso di posizione dominante[77].

Nel caso specifico il trattamento dei dati condotto da Meta Platforms Ireland riguardava anche dati sensibili, cui trattamento è in generale proibito dall'art. 9 del GDPR. Per legittimare la raccolta di tali dati, Facebook adduceva la motivazione che gli stessi venissero resi manifestamente pubblici dal soggetto interessato stesso. La Corte ha chiarito che il semplice fatto che la visita ad un sito web possa di per sé rendere pubblici tali dati non significa in alcun modo che gli stessi sono stati resi manifestamente pubblici dall'utente[78]. Tale maggiore garanzia richiesta dal GDPR rende per i titolari del trattamento ancora più difficile giustificare la raccolta di dati sensibili laddove non ci sia la manifesta pubblicazione degli stessi. Inoltre, per legittimare la raccolta sulla necessità del trattamento per l'esecuzione del contratto tra piattaforma e soggetto interessato, bisogna che tale trattamento risulti oggettivamente indispensabile per l'esecuzione, rendendo non eseguibile il contratto in assenza del trattamento[79]. A tal proposito, la

[76] C-252/21, Meta Platforms Inc. and others c. Bundeskartellamt, [2023], ECLI:EU:C:2023:537

[77] Ibid. punti da 47 a 49

[78] Ibid. punti da 66 a 85

[79] Ibid. punti da 97 a 104

Corte ha evidenziato che anche la personalizzazione delle pubblicità con cui viene finanziato il servizio offerto da Facebook non può giustificare, sotto forma di legittimo interesse, il trattamento dei dati considerati in assenza del consenso dei soggetti interessati [80].

Per concludere la Corte ha stabilito che il solo fatto che un social network si trovi in posizione dominante nel mercato non rende di per sé impossibile per gli utenti fornire il consenso in maniera legittima per il trattamento dei dati personali. Tuttavia, risulta utile ponderare quanto la posizione dominante nel mercato influenzi la validità e la libertà con cui il consenso viene fornito[81].

Secondo il presidente del Bundeskartellamt:

"*La sentenza invia un segnale forte per l'applicazione del diritto della concorrenza nell'economia digitale, un settore in cui i dati sono determinanti per il potere di mercato. Quando le grandi aziende internet utilizzano i dati personali molto sensibili dei consumatori, tale utilizzo può essere considerato abusivo anche ai sensi del diritto della concorrenza. Nell'applicazione del diritto della concorrenza, le autorità competenti devono altresì considerare le norme di protezione dei dati. La sentenza avrà effetti di vasta portata sui modelli di business utilizzati nell'economia dei dati. Nell'esercizio del diritto della concorrenza, è importante che continuiamo a collaborare strettamente con le autorità per la protezione dei dati.*" [82]

---

[80] Ibid. punti da 114 a 118

[81] Ibid. punti da 147 a 154

[82] Traduzione del seguente commento: *"The judgment sends a strong signal for competition law enforcement in the digital economy, a field where data are decisive for market power. When large internet companies use the very personal data of consumers, this usage can also be deemed abusive under competition law. In their application of competition law, competition authorities must also take data protection rules into consideration. The judgment will have far-reaching effects on the business models used in the data economy. When enforcing competition law, it is important that we continue to cooperate closely with the data protection authorities*." Andreas Mundt, Press Release del Bundeskartellamt per il caso Facebook Germany, 4 July 2023

La prepotenza e pervasività con cui i social network si sono sviluppati e hanno proliferato nella società digitale ha sollevato problematiche relative non solo al diritto della concorrenza nella valutazione di distorsioni del mercato che potessero danneggiare i consumatori, ma anche al diritto alla protezione dei dati che rappresenta uno dei più importanti campi di tutela dei soggetti interessati. Non a caso, anche in quest'ultima branca del diritto, Meta in qualità di titolare del trattamento dei dati posto in essere attraverso i suoi diversi social network e servizi è stata al centro di problematiche relative alla riservatezza e in particolar modo al trasferimento dei dati personali dei cittadini europei in altri paesi del mondo. Infatti, secondo il GDPR per poter trasferire dati personali al di fuori dell'UE è condizione necessaria che il paese di destinazione garantisca un livello di protezione dei dati adeguato agli standard di protezione dell'UE.

Un'importante evoluzione giurisprudenziale sul tema del trasferimento dei dati personali è stata delineata nelle sentenze relative ai casi Schrems I e II. Il 25 giugno 2013, in seguito alle rivelazioni di Edward Snowden riguardanti l'attività di sorveglianza dei servizi segreti americani, Maximillian Schrems[83] presentò una richiesta all'autorità irlandese per la protezione dei dati personali (*Irish Data Protection Commissioner*), chiedendo di interrompere il trasferimento dei propri dati personali effettuato da Facebook verso i server situati negli Stati Uniti, sostenendo che non venissero rispettati i livelli di trattamento richiesti dalla legislazione europea[84]. L'autorità decise di respingere la richiesta sostenendo la sua infondatezza sulla base della decisione 2000/520/CE della Commissione europea, in cui veniva verificato e confermato il livello di protezione adeguato del Paese. Tale decisione si basava su un accordo, in vigore dal 2000, che consentiva il libero trasferimento dei dati tra i due paesi e riconosceva che negli Stati Uniti venisse offerto "un livello di protezione adeguato",

[83] Maximilian Schrems (nato nel 1987) è un attivista, avvocato e autore austriaco noto per le campagne contro Facebook per le sue violazioni della *privacy*, tra cui le violazioni delle leggi europee sulla *privacy* e il presunto trasferimento di dati personali alla National Security Agency (NSA) statunitense nell'ambito del programma PRISM della NSA. Schrems è il fondatore di NOYB - Centro europeo per i diritti digitali.

[84] Glenn Greenwald, Ewen MacAskill, NSA Prism program taps in to user data of Apple, Google and others, The Guardian, (7 giugno 2013), Accessibile al link: <https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>

secondo quanto richiesto dall'art. 25 della direttiva 95/46/CE [85]. Di fronte a tale posizione dell'autorità irlandese, il signor Schrems fece ricorso dinanzi alla *High Court* irlandese che, constatando le evidenti violazioni poste in essere dai servizi di intelligence americana e l'indiscriminata, illimitata e incontrollata raccolta di dati personali posta in essere dalle autorità statunitensi, decise di rimettere la questione alla CGUE tramite rinvio pregiudiziale ex art. 267 TFUE in quanto la questione riguardava l'attuazione del diritto dell'Unione Europea e la validità della decisione della Commissione europea. Esaminando il confine tra il ruolo delle autorità nazionali di sorveglianza previsto dalla direttiva 95/46/CE e le decisioni di adeguatezza prese dalla Commissione, la Corte evidenziò che la sola presenza di una decisione di questo tipo non potesse privare le autorità in maniera assoluta del loro ruolo di controllore indipendente sul trasferimento dei dati, in particolare quando venga presentato un ricorso. Per tale motivo l'esame dell'autorità non veniva impedito dalla presenza di tale decisione. Inoltre, l'articolo 25 della direttiva 95/46/CE fissava i requisiti da valutare per poter confermare l'adeguatezza della protezione garantita nel paese terzo, vincolando inevitabilmente la valutazione della Commissione a tali parametri e principi[86]. Lo standard di protezione che il Paese terzo era tenuto a rispettare, secondo la

---

[85] Direttiva 95/46/CE del Parlamento europeo e del Consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, [1995], OJ L 281, 23.11.1995, p. 31–50. Art. 25; Decisione della Commissione 2000/520/CE [2000], OJ L 215, 25.8.2000.

[86] Direttiva 95/46/CE, Art. 25 "*Principi: 1) Gli Stati membri dispongono che il trasferimento verso un paese terzo di dati personali oggetto di un trattamento o destinati a essere oggetto di un trattamento dopo il trasferimento può aver luogo soltanto se il paese terzo di cui trattasi garantisce un livello di protezione adeguato, fatte salve le misure nazionali di attuazione delle altre disposizioni della presente direttiva. 2) L'adeguatezza del livello di protezione garantito da un paese terzo è valutata con riguardo a tutte le circostanze relative ad un trasferimento o ad una categoria di trasferimenti di dati; in particolare sono presi in considerazione la natura dei dati, le finalità del o dei trattamenti previsti, il paese d'origine e il paese di destinazione finale, le norme di diritto, generali o settoriali, vigenti nel paese terzo di cui trattasi, nonché le regole professionali e le misure di sicurezza ivi osservate. 3) Gli Stati membri e la Commissione si comunicano a vicenda i casi in cui, a loro parere, un paese terzo non garantisce un livello di protezione adeguato ai sensi del paragrafo 2. 4) Qualora la Commissione constati, secondo la procedura dell'articolo 31, paragrafo 2, che un paese terzo non garantisce un livello di protezione adeguato ai sensi del paragrafo 2 del presente articolo, gli Stati membri adottano le misure necessarie per impedire ogni trasferimento di dati della stessa natura verso il paese terzo in questione. 5) La Commissione avvia, al momento opportuno, negoziati per porre rimedio alla situazione risultante dalla constatazione di cui al paragrafo 4. 6) La Commissione può constatare, secondo la procedura di cui*

Corte, doveva essere "sostanzialmente equivalente" a quello europeo. Sul punto, dalla decisione della Commissione emergeva il primato delle esigenze di sicurezza nazionale, interesse pubblico o amministrazione della giustizia [degli Stati Uniti] nei confronti delle garanzie fornite dall'accordo Safe Harbour. Sulla base di queste condizioni le autorità di sicurezza nazionale statunitensi potevano accedere ai dati personali trasferiti dall'UE negli USA, senza la possibilità di una reale ed effettiva tutela dei diritti dei cittadini europei eventualmente lesi. La Corte individuò tra gli articoli violati l'art. 7 della Carta dei diritti fondamentali dell'UE, relativo al rispetto della vita privata e personale e l'art. 47 sul diritto ad un ricorso effettivo dinanzi ad un giudice imparziale, cagionando una chiara lesione del diritto alla tutela dei dati personali sancito all'art. 8 della stessa Carta. Per questi motivi la CGUE emise, il 6 ottobre 2015, la sentenza Schrems I (C-362/14), con cui determinò l'invalidità dell'accordo "*Safe Harbour*"[87]. Tale decisione mise a rischio l'intero sistema di trasferimento transatlantico di dati, creando incertezza e difficoltà per le imprese del settore. Per far fronte a questa situazione, nel luglio 2016 fu raggiunto un nuovo accordo tra la Commissione europea e il Dipartimento del Commercio degli Stati Uniti sul trasferimento dei dati tra i due Paesi per fini commerciali, denominato *Privacy Shield*[88]. Esso prevedeva una serie di impegni e garanzie, nonché una maggiore trasparenza dei processi di sorveglianza. Il caso Schrems II (C-311/18) riguardava proprio la validità di questo nuovo accordo e la sua incapacità di garantire un adeguato livello di protezione dei dati dei cittadini europei negli Stati Uniti[89]. Infatti, il signor Schrems decise di adire nuovamente al Commissario per la protezione dei dati irlandese, sostenendo che i dati personali dei cittadini europei continuassero ad essere utilizzati nei programmi di sorveglianza statunitensi. La CGUE chiarì subito che la normativa da applicare questa volta fosse il GDPR e in particolare

---

all'articolo 31, paragrafo 2, che un paese terzo garantisce un livello di protezione adeguato ai sensi del paragrafo 2 del presente articolo, in considerazione della sua legislazione nazionale o dei suoi impegni internazionali, in particolare di quelli assunti in seguito ai negoziati di cui al paragrafo 5, ai fini della tutela della vita privata o delle libertà e dei diritti fondamentali della persona."

[87] C-362/14, Maximilian Schrems c. Data Protection Commissioner, [2015], ECLI:EU:C:2015:650

[88] Decisione di esecuzione della Commissione 2016/1250, [2016], OJ L 207, 1.8.2016, p. 1–112.

[89] C-311/18, Maximilian Schrems c. Data Protection Commissioner, [2020], ECLI:EU:C:2020:559

due disposizioni, l'art. 44 e 45. Il primo che indica i principi da rispettare per il trasferimento dei dati personali verso paesi terzi ed il secondo che individua il requisito del "livello di protezione adeguato" per le decisioni di adeguatezza prese dalla Commissione europea, nel caso di specie il *Privacy Shield*. Come si evince nel regime dell'accordo, viene prevista la precedenza degli interessi degli USA rispetto ai principi di protezione dei dati dei cittadini europei. Per questo principale motivo, il 16 luglio 2020, la CGUE invalidò la decisione di adeguatezza 2016/1250, in quanto in violazione del principio di adeguatezza e non offriva una tutela sufficiente dei diritti dei cittadini sanciti dagli artt. 7, 8 e 47 CDFUE[90].

L'importanza commerciale dello scambio di dati personali tra USA e UE e la necessità di non interrompere bruscamente il trasferimento diedero vita, nel marzo 2022, ad un documento contenente i principi comuni da rispettare per la stipula di un nuovo accordo sulla privacy transatlantico, che garantisse una protezione adeguata, un flusso di dati sicuro e una base giuridica solida su cui fondare il trasferimento dei dati 91. NeIle more di tale accordo, il governo degli USA ha introdotto, con l'Executive Order n. 14086 '*Enhancing Safeguards for United States Signals Intelligence Activities*', nuove garanzie da rispettare nella tutela dei diritti dei cittadini europei e nuovi limiti nei confronti delle agenzie di sicurezza nazionale per l'accesso ai dati personali degli europei, inoltre, è stata istituita la *Data Protection Review Court* (DPRC) predisposta a valutare i ricorsi degli europei che lamentassero violazioni dei propri diritti sul suolo americano[92].

In questo contesto, il 10 luglio 2023 la Commissione europea ha pubblicato la sua terza decisione di adeguatezza contenente il *Data Privacy Framework*, accordo che permette il trasferimento dei dati personali dei cittadini europei verso gli USA. Con tale decisione la Commissione ha stabilito che gli USA garantiscono un livello di protezione adeguato

---

[90] C-311/18, Maximilian Schrems c. Data Protection Commissioner, [2020], ECLI:EU:C:2020:559, para 80-196.

[91] Commissione europea - Press release, "European Commission and United States Joint Statement on Trans- Atlantic Data Privacy Framework", Brussels, (25 March 2022).

[92] E.O. 14067, 9 March 2022

ai dati dei cittadini europei trasferiti in tale Paese terzo. Il *Data Privacy Framework*, per colmare le criticità individuate dalle sentenze della CGUE, si fonda sui nuovi obblighi contenuti nell'Executive Order n. 14086 emesso dal Presidente Biden. L'accordo prevede che le aziende statunitensi che intendono partecipare al *Data Privacy Framework* debbano soddisfare alcuni requisiti relativi alla *privacy*, ad esempio, queste saranno obbligate a cancellare i dati personali quando non risultino più necessari allo scopo originario per cui sono stati raccolti, inoltre dovranno assicurarsi una continuità nella protezione dei dati nel momento in cui condividano i dati con parti terze. I cittadini europei potranno anche presentare ricorso di fronte a organi di risoluzione delle controversie indipendenti nel caso in cui ritengano esserci una violazione da parte delle società statunitensi. In fine, la legislazione statunitense ha aggiunto diverse garanzie per rendere meno invasivo e più limitato l'accesso da parte delle autorità statunitensi ai dati trasferiti negli USA, limitandolo a quanto necessario e proporzionale per la sicurezza nazionale.

Questi rimedi apportati dalla legislazione statunitense sembrano avvicinarsi agli standard di protezione dei dati molto elevati garantiti dalla legislazione europea, infatti, secondo la Presidente della Commissione Ursula von Der Leien: "*The new EU-U.S. Data Privacy Framework will ensure safe data flows for Europeans and bring legal certainty to companies on both sides of the Atlantic*"[93]. Secondo la Presidente i provvedimenti adottati dalla legislazione statunitense rappresentano un impegno senza precedenti per allinearsi agli standard europei di protezione dei dati, sebbene, come viene ricordato nella decisione di adeguatezza, non sia necessario che il Paese terzo adotti una regolamentazione identica a quella europea in ogni punto. Tuttavia, non si è fatta attendere la reazione e il commento del Sig. Schrems di fronte alla nuova decisione di adeguatezza, infatti, ha già anticipato che tale decisione sarà portata di fronte alla Corte di Giustizia entro l'inizio del nuovo anno, proprio perché i miglioramenti rispetto al precedente regime sono stati minimi e probabilmente non sufficienti a soddisfare le mancanze evidenziate dalla Corte. Tale critica, fondata sulle vicende giudiziarie che

[93] European Commission, Ursula von Der Leien, "Data Protection: European Commission adopts new adequacy decision for safe and trusted EU-US data flows", <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3721>, 10 July 2023

abbiamo brevemente visto in questo capitolo, si rivolge all'intero sistema che pur di mantenere saldi i rapporti commerciali e politici sceglie di adottare soluzioni non completamente coerenti con gli standard di protezione europei, a discapito dei diritti dei cittadini europei. In definitiva, Schrems ha commentato la decisione della Commissione ricordando che: "*For the past 23 years all EU-US deals were declared invalid retroactively, making all past data transfers by business illegal - we seem to just add another two years of this ping-pong now*"[94].

L'elemento di maggior importanza sembra essere la nuova DPRC e il suo ruolo di giudice nelle controversie che riguardano i diritti dei cittadini europei relativi alla loro *privacy* in territorio USA. La presenza di un tale organo decisorio permetterà ai cittadini europei di poter far valere i propri diritti di fronte ad un giudice terzo e imparziale, andando a colmare la carenza di rimedi giudiziali ed stragiudiziali che caratterizzava la situazione precedente. Le future vicende giudiziarie già annunciate ci aiuteranno a capire se i rimedi e le salvaguardie stabilite con il nuovo accordo e con la rinnovata legislazione statunitense saranno sufficienti a garantire una protezione adeguata ai dati dei cittadini europei.

I casi Schrems I II e Facebook hanno avuto un impatto significativo sulla protezione dei dati personali e sul trasferimento transatlantico dei dati personali, mettendo in luce l'importanza di garantire un livello di protezione adeguato e di rispettare i principi stabiliti dal GDPR e dalla giurisprudenza della CGUE. Queste decisioni hanno anche evidenziato la necessità di affrontare le sfide poste da cosiddetto "*privacy paradox*", ovverosia la discrepanza tra gli interessi degli utenti nella protezione dei dati personali e il loro effettivo comportamento inconciliabile. Inoltre, questi casi hanno anche contribuito a formare una più profonda consapevolezza dell'esistenza dei "*dark patterns*" nelle piattaforme digitali che influenzano il comportamento degli utenti e sfidano la loro *privacy*. La presenza di tali meccanismi sottolinea l'importanza di un

[94] Max Schrems, "European Commission gives EU-US data transfers third round at CJEU", <https://noyb.eu/en/european-commission-gives-eu-us-data-transfers-third-round-cjeu>, 10 July 2023

approccio integrato tra diverse normative per garantire un sistema di tutela multilivello europeo.

L'evoluzione del rapporto tra *data protection* e diritto *antitrust*, permette di apprezzare fino in fondo quali risultati possono essere ottenuti con una sinergia tra la normativa a tutela dei dati personali e quella concorrenziale, in un contesto in cui i dati personali hanno assunto un'importanza cruciale. Diverse problematiche anche in diritto antitrust sorgeranno presto relativamente ai sistemi di IA, infatti, sembra ripetersi parzialmente quel che è avvenuto con Facebook e gli altri social network. Pochi sviluppatori, principalmente in grandi aziende statunitensi stanno portando sul mercato dei sistemi di IA molto avanzati e, sebbene questa volta anche il mondo degli sviluppatori open source sembra essere più prolifico, bisognerà in un futuro non troppo distante esaminare fenomeni di concentrazione dei *dataset*, su cui si fonda l'addestramento degli algoritmi, che porteranno distorsioni del mercato.

6. La pervasività dell'intelligenza artificiale e la necessità di una regolamentazione

Nel presente capitolo è stata esaminata l'evoluzione del rapporto tra la protezione dei dati personali e i social network, con particolare attenzione ai progressi tecnologici che hanno condotto allo sviluppo delle moderne forme di IA, nonché alle implicazioni sulla *privacy* che ne derivano. Negli ultimi anni, le moderne tecnologie sono riuscite a raggiungere profondi livelli di invasione della nostra sfera privata e per mitigare tali impatti e ripristinare il controllo sui dati personali, sono stati introdotti specifici strumenti giuridici a disposizione dei cittadini. Tuttavia, con l'avvento delle ultime tecnologie di IA, è diventato più difficile riprendere il controllo dei dati una volta che questi sono inseriti nei sistemi di IA e le forme per carpire il consenso dei soggetti interessati pongono, spesso, un'alternativa rigida tra l'utilizzo di un determinato servizio, con conseguente "profilazione", oppure l'impossibilità di accedere allo stesso, come visto nel caso Facebook Germany.

Lo sviluppo dirompente dell'IA ha avuto inizio nel nuovo millennio e si è fondato su due pilastri: la crescente capacità computazionale disponibile e le innovative tecniche di

raccolta dati. L'IA è attualmente impiegata in tantissimi settori e in tutti questi è presente un pericolo per i dati personali che vengono spesso raccolti senza che gli individui a cui appartengono se ne rendano conto o forniscano il consenso in modo informato alla loro raccolta. I programmi di apprendimento più avanzati utilizzano reti neurali artificiali[95], imitando i collegamenti neurali presenti nel cervello, formate da un'intricata rete di input a cui sono collegati output, che descrivono schemi molto complessi. Questi parametri sono costituiti dai dati che vengono raccolti e la rete che si forma diventa la conoscenza del sistema di IA. Nonostante siano capaci di elaborare i dati, i sistemi di IA non riescono a garantire la correttezza logica della risposta in quanto il risultato che forniscono è semplicemente frutto di operazioni statistiche. Come vedremo nel secondo capitolo attraverso alcuni esempi, questi sistemi di IA necessitano di un addestramento continuo per apprendere, con la guida dell'uomo, quali risposte sono corrette e quali invece no.

L'intelligenza artificiale moderna rappresenta la frontiera tecnologica più recente che richiede un'attenta riflessione da parte dei regolatori sulla direzione da intraprendere. Si pongono due alternative: regolamentare questo fenomeno in modo oculato o lasciare che siano le forze di mercato a definire i limiti del suo sviluppo. L'assenza di regolamentazione potrebbe determinare non solo enorme incertezza derivante dall'utilizzo delle tecnologie e degli strumenti avanzati, ma anche uno sviluppo sregolato e senza limiti di tali tecnologie con rilevanti conseguenze sulle persone interessate. Nel confronto con questa prospettiva, emerge con chiarezza la necessità di istituire un quadro regolamentare per questo fenomeno, con l'obiettivo di promuovere uno sviluppo che possa essere benefico per la collettività nel suo insieme. La sfida contemporanea è dunque quella di portare una regolamentazione all'IA che tuteli i cittadini, ma che non freni lo sviluppo della tecnologia, mantenendo l'uomo al centro e permettendo uno sviluppo della fiducia nell'uso dei moderni strumenti. Nel secondo capitolo del testo, si analizzerà in modo più dettagliato la nuova proposta di

[95] Le reti neurali sono una classe di algoritmi di apprendimento automatico che si ispirano al funzionamento del cervello umano. Sono composte da numerosi nodi, chiamati neuroni, che sono organizzati in strati e collegati tra loro da pesi che rappresentano la forza dell'interconnessione. In questo modo, le reti neurali possono apprendere modelli complessi a partire dai dati di input attraverso l'aggiustamento dei pesi dei neuroni.

regolamentazione dell'IA, scoprendo l’inadeguatezza del testo per alcuni aspetti. Si analizzeranno le varie misure di protezione dei dati disponibili per garantire la tutela dei cittadini e si evidenzierà la necessità di un quadro normativo che tenga conto dei principi etici e giuridici, per adattarsi all'evoluzione tecnologica in corso.

Capitolo II

"*Un robot non può ferire un essere umano o, per inazione, permettere ad un essere umano di mettersi in pericolo*"

- Isaac Asimov

1. Il regolamento sull'Intelligenza Artificiale

Nel 1942 un professore di chimica americano con origini russe, di nome Isaac Asimov, in un racconto di fantascienza dal titolo *Runaround* introduceva, per la prima volta, le tre leggi della Robotica[96]:

Prima Legge: "*Un robot non può ferire un essere umano o, per inazione, permettere ad un essere umano di mettersi in pericolo"*.

Seconda Legge: "*Un robot deve obbedire agli ordini dati da esseri umani tranne quando questi ordini confliggano con la Prima Legge".*

Terza Legge: "*Un robot deve proteggere la propria esistenza nella misura in cui questa protezione non confligga con la Prima o con la Seconda legge"*.

Queste tre leggi rappresentano uno dei primi esempi di regolamentazione, seppur fantascientifica, di tecnologie dotate di una certa autonomia decisionale. In

[96] Isaac Asimov, 'Runaround' (1942) 29 Astounding science fiction 94.

contrapposizione ad uno scopo narrativo come quello di Isaac Asimov[97], l'Unione Europea ha iniziato a lavorare, a partire dal 2018, ad una regolamentazione che potesse armonizzare la disciplina dell'Intelligenza Artificiale (di seguito, IA) a livello europeo. Come già illustrato all'interno del capitolo precedente, l'approccio del legislatore comunitario alla regolamentazione dei nuovi fenomeni tecnologici è stato caratterizzato sia da fonti di *soft-law*, come le risoluzioni del Parlamento europeo[98], che da norme di *hard-law* come il Regolamento (UE) 2016/679 (di seguito, GDPR) in vigore dal 2018.[99] Nella strategia per il mercato unico digitale avviata dal presidente della Commissione europea Jean-Claude Juncker [100], l'Unione europea aspirava ad essere la prima istituzione a livello mondiale nella regolazione dei più recenti fenomeni tecnologici, divenendo anche una guida per i legislatori di tutto il globo[101].

Sebbene la definizione di IA riportata nel testo della "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" (da qui in poi, AI Act)[102] sia ancora da definire, si può individuare una chiara influenza proveniente dalle linee guida dell'OCSE in materia. Questo ente

[97] Isaac Asimov (1920-1992), autore e biochimico di origine russa naturalizzato statunitense, è universalmente riconosciuto come uno dei pilastri fondamentali della letteratura fantascientifica del XX secolo. La sua produzione letteraria, estremamente vasta, comprende opere di divulgazione scientifica, romanzi, raccolte di racconti e saggi. Tra le sue opere più note, si annoverano "Cicli dei Robot", "Fondazione" e "Impero".

[98] *Cfr*. (n 25)

[99] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), [2016], OJ L 119, 4.5.2016, p. 1–88 2016.

[100] Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni, "Strategia per il mercato unico digitale in Europa", COM(2015) 192 final

[101] Hannah Ruschemeier, 'AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal' (2023) 23 ERA Forum 361.

[102] 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, [2021], COM/2021/206 final 2021.

internazionale definisce l'IA come *"un sistema basato su macchine che, per un determinato insieme di obiettivi definiti dall'uomo, può fare previsioni, raccomandazioni o decisioni che influenzano ambienti reali o virtuali. I sistemi di intelligenza artificiale sono progettati per funzionare con diversi livelli di autonomia."*[103]. Tale definizione non è stata adottata dal Gruppo di Esperti di Alto Livello sull'Intelligenza Artificiale (di seguito, AI HLEG [104]) che, piuttosto, nelle sue "Linee Guida Etiche per un'IA Affidabile"[105] ha utilizzato il termine "sistema di IA",

[103] Organisation for Economic Co-operation and Development, 'Draft Recommendation of the Council on Artificial Intelligence', C(2019)34

[104] Il Gruppo di Esperti di Alto Livello sull'Intelligenza Artificiale (AI High-Level Expert Group) è un organismo di consulenza che fornisce raccomandazioni e linee guida sull'uso responsabile e affidabile dell'IA nell'Unione Europea. Creato dalla Commissione Europea, il Gruppo è composto da esperti provenienti da diverse discipline, tra cui accademici, industria, società civile e organizzazioni di ricerca. Il suo obiettivo principale è quello di contribuire allo sviluppo di una strategia europea sull'IA, garantendo che l'innovazione nel campo dell'IA sia guidata da principi etici, valori umani, trasparenza e responsabilità. L'8 aprile 2019 tale ente ha presentato il rapporto "Linee Guida Etiche per un'IA Affidabile" in cui sono stati individuati sette requisiti per un'IA affidabile. La sorveglianza e supervisione umana è il primo requisito di garanzia e riflette la necessità di assicurare che i sistemi di IA rispettino i diritti fondamentali e che la supervisione dell'uomo sia sempre prevista. Il secondo riguarda la robustezza e sicurezza tecnica, affinché i sistemi di IA siano sicuri e resistenti, garantendo al contempo misure di contingenza nel caso in cui dovessero sorgere problemi. Il terzo principio riguarda la protezione dei dati e la gestione della privacy, insieme a meccanismi di governance dei dati. Ciò implica la considerazione della qualità e dell'integrità dei dati, assicurando un accesso agli stessi. Il quarto principio è relativo alla trasparenza che viene richiesta affinché siano presenti meccanismi di tracciabilità, garantendo la comprensibilità delle decisioni agli individui interessati. Il quinto principio mira a tutelare la diversità, la non discriminazione e l'equità, che dovrebbero essere favorite, garantendo l'accessibilità a tutti, indipendentemente da eventuali disabilità. Il sesto requisito ha come obiettivo il benessere sociale e ambientale, richiedendo che i sistemi di IA siano sostenibili ed ecologici, in grado di portare benefici anche alle prossime generazioni. Per ultimo, le linee guida indicando il requisito della responsabilità e di verificabilità, attraverso la valutazione dei sistemi di IA, permettendo una riduzione al minimo degli effetti negativi e garantendo, secondo il principio di equità,a la possibilità di un ricorso effettivo in caso di effetti negativi ingiusti.

[105] European Commission. Directorate General for Communications Networks, Content and Technology. and High Level Expert Group on Artificial Intelligence., *Ethics Guidelines for Trustworthy AI*. (Publications Office 2019) <https://data.europa.eu/doi/10.2759/346720> accessed 24 May 2023.

adottato poi nell'AI Act, in quanto il concetto di "intelligenza" rimane ancora indefinito e il riferimento esplicito a questo termine avrebbe potuto causare confusione[106].

Di fatti, l'art. 3 dell'AI Act definisce il sistema di IA come: "*un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono*". La definizione adottata dalla Commissione assomiglia in alcuni punti a quella proposta dall'OCSE proprio perché esiste una volontà di stabilire dei parametri comuni a livello internazionale[107].

L'Unione europea ha adottato un approccio centrato sull'individuo nell'AI Act, infatti questa tecnologia, per quanto possa essere vantaggiosa e benefica per l'umanità, pone diversi rischi di natura etica e giuridica. I sistemi di IA possono potenzialmente arrecare danni alla vita delle persone, alla loro salute e anche violare alcuni valori e principi fondamentali su cui si basa l'Unione europea, come ad esempio, la dignità umana e l'autodeterminazione, la *privacy* e la protezione dei dati personali, la libertà di espressione e di assemblea, il divieto di discriminazione. Questi sistemi si distinguono principalmente per la loro razionalità, ovvero la capacità di percepire l'ambiente in cui operano grazie a una serie di sensori, grazie ai quali elaborano le informazioni raccolte e scelgono il miglior modo di agire [108].

---

[106] High-Level Expert Group on Artificial Intelligence, 'A definition of AI: Main capabilities and scientific disciplines', (2018). Una simile riflessione era stata proposta da Turing durante la formulazione del suo famoso test, si veda n 52.

[107] La definizione è stata modificata all'interno del rapporto di compromesso sugli emendamenti all'AI Act pubblicato dal Parlamento europeo il 16 maggio 2023. I sistemi di IA vengono dunque definiti come: "… *machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments.*" È stato dunque rimosso il riferimento all'Allegato I per dare alla definizione una portata maggiore e per avvicinarsi ancora di più alla definizione proposta dall'OCSE, andando verso uno standard comune internazionale. Parlamento europeo, Commissione per il mercato interno e la protezione dei consumatori e Commissione per le libertà civili, la giustizia e gli affari interni, "Draft Compromise Amendments on the Draft Report, Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts", (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD))

[108] Oltre ad utilizzare tecniche per il ragionamento, i sistemi di IA utilizzano tecniche per l'apprendimento come *machine learning, neural networks, deep learning*. Grazie a queste tecniche l'IA può approcciarsi e

Di fronte a questo scenario, la Commissione europea ha deciso di presentare il 21 aprile 2021 la "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"[109]. L'AI Act segue un approccio basato sul rischio, stabilendo obblighi giuridici solo per i sistemi di IA capaci di causare danni ai diritti fondamentali e alla sicurezza degli individui. L'AI Act è fondato sull'articolo 114 del trattato sul funzionamento dell'Unione europea (TFUE)[110], che prevede l'adozione di misure destinate ad assicurare l'instaurazione ed il funzionamento del mercato interno. La proposta in questione è un elemento cardine della strategia dell'UE per un mercato digitale unificato[111], con l'intento di garantire un funzionamento efficace del mercato interno attraverso l'armonizzazione delle regole relative allo sviluppo, alla distribuzione e all'uso di prodotti e servizi basati su AI. L'adozione di normative nazionali da parte di vari stati membri, miranti a garantire sicurezza e rispetto dei diritti fondamentali nell'uso dell'IA, potrebbe causare frammentazione del mercato europeo e una dilagante incertezza giuridica per fornitori e utenti di tali sistemi. Questi problemi, data la natura transfrontaliera di molti prodotti e servizi, sono più facilmente risolvibili attraverso un'armonizzazione legislativa europea. La proposta stabilisce requisiti comuni obbligatori per la progettazione e lo sviluppo di determinati sistemi di IA e regola i controlli ex post. Il regolamento adotta un approccio basato sul rischio, distingue tra gli usi dell'IA che determinano: i) un rischio inaccettabile; ii) un rischio alto; iii) un rischio basso o minimo.

---

risolvere problemi che erano di competenza esclusivamente umana, essendo la tecnologia in precedenza esclusa, a causa delle necessarie abilità di percezione per la comprensione e risoluzione. Si arriva infatti a parlare di apprendimento della macchina, che costituisce una categoria contenente diverse forme di apprendimento che solitamente vengono combinate per ottenere il migliore risultato. Le tre modalità principali di apprendimento delle macchine sono: apprendimento supervisionato (*supervised learning*), apprendimento non supervisionato (*unsupervised learning*) e apprendimento rinforzato (*reinforcement learning*).

[109] 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, [2021], COM/2021/206 final (n 88).

[110] Trattato sull'Unione europea, versione consolidata, [2012]. OJ C 326, 26.10.2012, p. 15–368, art. 114

[111] *Cfr*. (n 100)

Iniziando dalla prima categoria di rischio, l'AI Act contiene all'art. 5, para. 1, un elenco con quattro categorie di utilizzi vietati, di cui tre sono interamente vietati e uno parzialmente[112]. I primi due punti del paragrafo 1 dell'articolo 5 dell'AI Act prevedono il divieto per:

a) *l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole al fine di distorcerne materialmente il comportamento in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico;*
b) *l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di uno specifico gruppo di persone, dovute all'età o alla disabilità fisica o mentale, al fine di distorcere materialmente il comportamento di una persona che appartiene a tale gruppo in un modo che provochi o possa provocare a tale persona o a un'altra persona un danno fisico o psicologico;*

Le prime due pratiche riguardano la manipolazione, questa è comunemente composta da quattro elementi cumulativi per cui: il manipolatore vuole *intenzionalmente*, ma in modo *occulto,* servirsi del processo decisionale di un altro individuo per *perseguire i propri fini*, sfruttando una qualche *vulnerabilità* [113]. L'AI Act sembra considerare alcuni di questi elementi, tra cui l'intenzione ('al fine di'), alcuni parametri di vulnerabilità come età, disabilità fisica o mentale e l'elemento subliminale occulto che richiede l'assenza della consapevolezza della persona. L'ultimo elemento richiesto dall'AI Act consiste nel causare o nella possibilità di causare un danno fisico o psicologico alla persona. Da tale lettura sembra che i sistemi di IA programmati per la manipolazione non ricadano nel divieto, fintanto che non siano idonei a cagionare un danno all'individuo. Il requisito del danno limita grandemente l'applicazione della

[112] AI Act art. 5

[113] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22 Computer Law Review International 97, 99.

disposizione in quanto non sempre risulta individuabile un danno cagionato da sistemi di IA. Può accadere che il danno venga maturato senza un evento scatenante che permetta di dimostrarlo, come ad esempio nelle conseguenze di lungo termine derivanti dall'utilizzo dei sistemi di raccomandazione dei servizi online, in particolare sui bambini[114]. Inoltre, secondo il considerando 16 della proposta, il danno non può essere presunto nel caso in cui derivi "*da fattori esterni al sistema di IA, che sfuggono al controllo del fornitore o dell'utente*" [115]. Viene vietata l'immissione nel mercato di sistemi di IA con fini manipolativi, anche se, un produttore difficilmente ammetterà in maniera esplicita e chiara che l'utilizzo dei sistemi prodotti è rivolto alla manipolazione delle persone. I venditori potrebbero anche superare tale divieto vendendo sistemi di IA che non hanno un uso specifico, ma che il compratore può facilmente riconfigurare destinandolo a fini manipolativi[116].

[114] Nick Seaver, 'Captivating Algorithms: Recommender Systems as Traps' (2019) 24 Journal of Material Culture, 421–436.

[115] Il Considerando 16 dell'AI Act prevede che: "*È opportuno vietare l'immissione sul mercato, la messa in servizio o l'uso di determinati sistemi di IA intesi a distorcere il comportamento umano e che possono provocare danni fisici o psicologici. Tali sistemi di IA impiegano componenti subliminali che i singoli individui non sono in grado di percepire, oppure sfruttano le vulnerabilità di bambini e persone, dovute all'età o a incapacità fisiche o mentali. Si tratta di azioni compiute con l'intento di distorcere materialmente il comportamento di una persona, in un modo che provoca o può provocare un danno a tale persona o a un'altra. Tale intento non può essere presunto se la distorsione del comportamento umano è determinata da fattori esterni al sistema di IA, che sfuggono al controllo del fornitore o dell'utente. Tale divieto non dovrebbe ostacolare la ricerca per scopi legittimi in relazione a tali sistemi di IA, se tale ricerca non equivale a un uso del sistema di IA nelle relazioni uomo-macchina che espone le persone fisiche a danni e se tale ricerca è condotta conformemente a norme etiche riconosciute per la ricerca scientifica.*"

[116] La riconfigurazione dei sistemi è un fenomeno in crescita e consiste nel poter attribuire un nuovo scopo al sistema di IA acquistato già in assenza di una finalità specifica. Questa si fonda sulla possibilità di accedere a sistemi di IA senza aver bisogno di tecnologie particolarmente avanzate, usufruendo della c.d. *AI-as-a-service*; Jennifer Cobbe and Jatinder Singh, 'Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges' (2021) 42 Computer Law & Security Review 105573, 3–5.

La seconda categoria di pratiche vietate riguarda i sistemi di punteggio sociale (anche detti, *social scoring systems*) ed è disciplinata alla lettera c), paragrafo uno, articolo 5 dell'AI Act, che stabilisce il divieto per: "

1. *l'immissione sul mercato, la messa in servizio o l'uso di sistemi di IA da parte delle autorità pubbliche o per loro conto ai fini della valutazione o della classificazione dell'affidabilità delle persone fisiche per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali o della personalità note o previste, in cui il punteggio sociale così ottenuto comporti il verificarsi di uno o di entrambi i seguenti scenari:*

i) *un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di interi gruppi di persone fisiche in contesti sociali che non sono collegati ai contesti in cui i dati sono stati originariamente generati o raccolti;*

ii) *un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di interi gruppi di persone fisiche che sia ingiustificato o sproporzionato rispetto al loro comportamento sociale o alla sua gravità;"*

La struttura di questo secondo divieto differisce dalle linee guida dell'AI HLEG, a cui si ispira l'AI Act. Infatti, tale raccomandazione consigliava al paragrafo n. 132 di proibire interamente i sistemi di valutazione per punteggio sulla personalità morale o l'integrità etica[117]. Al contrario, nell'articolo si proibisce l'utilizzo o la vendita di sistemi di IA: i) utilizzati da o per conto di autorità pubbliche, ii) per produrre valutazioni di affidabilità e, iii) che determinerebbe un trattamento ingiustificato o pregiudizievoli su individui o gruppi di individui in un contesto sociale che non è collegato a quello da cui

[117] European Commission. Directorate General for Communications Networks, Content and Technology. and High Level Expert Group on Artificial Intelligence. (n 91).Il paragrafo (132) raccomandava che: "*Nelle società la libertà e l'autonomia di tutti i cittadini dovrebbero essere tutelate. Qualsiasi forma di valutazione per punteggio delle persone può comportare la perdita di tale autonomia e compromettere il principio di non discriminazione. La valutazione per punteggio dovrebbe essere utilizzata solo se esiste una chiara giustificazione e se le misure sono proporzionate ed eque. La valutazione normativa dei cittadini per punteggio (valutazione generale della "personalità morale" o dell'"integrità etica"), in tutti gli aspetti e su larga scala, effettuata da autorità pubbliche o soggetti privati compromette questi valori, soprattutto se utilizzata in violazione dei diritti fondamentali, in modo sproporzionato e senza uno scopo legittimo descritto e comunicato."*

provengono i dati. L'eccezione che rimane scoperta da tale divieto riguarda l'utilizzo o la vendita di tali sistemi, quando questi operino all'interno degli stessi contesti in cui sono stati raccolti i dati. In tal senso, se il concetto di 'stesso contesto' venisse letto in maniera ristretta, allora l'eccezione riguarderebbe solamente i dati che derivano dall'interazione con le pubbliche autorità. D'altro canto, una lettura più ampia potrebbe considerare i dati relativi al credito e alla tassazione, come parte integrante dell'economia dello Stato come autorità pubblica unitaria[118].

L'ultima pratica vietata dalla AI Act riguarda i sistemi di identificazione biometrica remota in tempo reale. L'articolo 5, paragrafo 1, lettera d), prevede che:

2. *l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, a meno che e nella misura in cui tale uso sia strettamente necessario per uno dei seguenti obiettivi:*
    a. *la ricerca mirata di potenziali vittime specifiche di reato, compresi i minori scomparsi;*
    b. *la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o di un attacco terroristico;*
    c. *il rilevamento, la localizzazione, l'identificazione o l'azione penale nei confronti di un autore o un sospettato di un reato di cui all'articolo 2, paragrafo 2, della decisione quadro 2002/584/GAI del Consiglio*[119]*, punibile nello Stato membro interessato con una pena o una misura di*

[118] Anche tale definizione è stata modificata all'interno del rapporto di compromesso sugli emendamenti all'AI Act pubblicato dal Parlamento europeo il 16 maggio 2023. È stato rimosso il riferimento alle autorità pubbliche, proibendo qualsiasi utilizzo di sistemi di punteggio sociale che utilizzino sistemi di IA e aggiungendo tre nuove pratiche proibite relative a: sistemi di valutazione del rischio di commissione o recidiva di reato o che valutano i tratti e le caratteristiche delle persone fisiche, l'utilizzo di sistemi di riconoscimento facciale che aumentano o creano set di dati analizzando immagini da internet o da circuiti di video sorveglianza e infine i sistemi di IA per il riconoscimento delle emozioni a fini di attuazione legislativa, gestione dei confini, nei luoghi di lavoro e di istruzione. *Cfr.* (n 107)

[119] Decisione quadro del Consiglio 2002/584/GAI, del 13 giugno 2002, relativa al mandato d'arresto europeo e alle procedure di consegna tra Stati membri (GU L 190 del 18.7.2002, pag. 1).

sicurezza privativa della libertà della durata massima di almeno tre anni, come stabilito dalla legge di tale Stato membro.

Un esempio di sistema di AI che potrebbe ricadere sotto il divieto delineato è rappresentato da un network di telecamere di sorveglianza equipaggiate con programmi di riconoscimento facciale. In particolare, l'impiego di sistemi di identificazione biometrica a fini di applicazione legislativa è disciplinato dalla Direttiva UE 2016/680 [120]. Infatti, i dispositivi di riconoscimento facciale destinati a finalità di attuazione legislativa hanno sinora goduto di un percorso autorizzativo più agevole rispetto ad altre applicazioni. Al contrario delle altre disposizioni, la proibizione non si applica all'ingresso dei sistemi biometrici nel mercato, legalizzando quindi la vendita di tali sistemi per utilizzo al di fuori dell'Unione europea, come per esempio nei regimi oppressivi di paesi terzi[121]. Un elemento cruciale della proibizione riguarda la sua applicazione esclusivamente ai sistemi che consentono il monitoraggio "in tempo reale", non vietando in tal modo i sistemi che permettono la revisione a posteriori, come ad esempio le immagini acquisite durante una protesta che vengono analizzate successivamente con un sistema di riconoscimento facciale per l'identificazione degli oppositori. Parallelamente, gli spazi virtuali non rientrano nel divieto, consentendo l'identificazione biometrica, ad esempio, durante le trasmissioni video in diretta. In conclusione, la normativa non proibisce l'impiego di sistemi per finalità estranee all'attuazione legislativa, quali, ad esempio, la salute pubblica e il controllo delle folle. Nonostante ciò, tali pratiche ricadono nell'ambito di applicazione del GDPR che, in

[120] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119/89 (‘Law Enforcement Directive’)

[121] Secondo Amnesty International esempi di tali pratiche includono la vendita da parte dell'azienda francese Idemia/Morpho di sistemi di riconoscimento facciale alla *Shanghai Public Security Bureau*, oppure la vendita da parte della società olandese Noldus dell'analizzatore di espressioni facciali 'FaceReader' al Ministero della Sicurezza Pubblica cinese. Si veda: Amnesty International, ‘Out of Control: Failing EU Laws for Digital Surveillance Export’ (2020) < https://www.amnesty.org/en/documents/eur01/2556/2020/en/ > ultimo accesso 22 maggio 2023.

mancanza di legislazioni nazionali che autorizzino tali pratiche, impone il rispetto di requisiti estremamente rigorosi, come il consenso individuale da parte di ciascuna persona i cui dati vengono analizzati. Tale consenso risulta quasi impossibile da ottenere in pratica. Per l'utilizzo del sistema biometrico, al paragrafo 3 dell'articolo 5, la disposizione introduce l'autorizzazione che deve precedere oppure, nei casi urgenti, seguire immediatamente dopo l'utilizzo del sistema[122]. L'autorizzazione deve essere rilasciata da un'autorità amministrativa o giudiziaria indipendente dello Stato membro in cui viene utilizzato il sistema.

Sembrerebbe, in altri termini, che l'AI Act, invece di vietare l'utilizzo di tali sistemi di sorveglianza di massa li stia regolando e in qualche modo legittimando[123]. In tal senso anche l'European Data Protection Board (EDPB) e l'European Data Protection Supervisor (EDPS) si sono espressi a favore di un generale divieto "*di qualsiasi uso dell'IA a fini di riconoscimento automatico, in spazi accessibili al pubblico, delle caratteristiche umane – come il volto ma anche l'andatura, le impronte digitali, il DNA, la voce, le sequenze di battute su tastiera e altri segnali biometrici o comportamentali – in qualsiasi contesto*"[124].

Il timore principale derivante dall'utilizzo di tali tecnologie è legato alla prospettiva di discriminazione sulla base di categorie di dati sensibili raccolti come: etnia, genere, orientamento sessuale. Tali sistemi hanno il potenziale di arrecare danni sia agli

---

[122] L'art. 5, paragrafo 3, prevede che: "*Per quanto riguarda il paragrafo 1, lettera d), e il paragrafo 2, ogni singolo uso di un sistema di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto è subordinato a un'autorizzazione preventiva rilasciata da un'autorità giudiziaria o da un'autorità amministrativa indipendente dello Stato membro in cui deve avvenire l'uso, rilasciata su richiesta motivata e in conformità alle regole dettagliate del diritto nazionale di cui al paragrafo 4. Tuttavia, in una situazione di urgenza debitamente giustificata, è possibile iniziare a usare il sistema senza autorizzazione e richiedere l'autorizzazione solo durante o dopo l'uso.*"

[123] Martin Ebers and others, 'The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)' (2021) 4 J 589, 593.

[124] EDPB – GEPD Parere congiunto 5/2021 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale), 18 giugno 2021, para. 32

individui che alla collettività, ma in generale possono danneggiare la società nel suo insieme mettendo a rischio l'uguaglianza, la democrazia e lo stato di diritto[125].
L'importanza dei nostri dati personali rende il riconoscimento biometrico e facciale nei luoghi pubblici uno strumento di controllo pervasivo e pericoloso per le nostre libertà di movimento, autonomia e riservatezza. Già dal 2019 l'EDPB aveva pubblicato alcune linee guida sul trattamento dei dati personali attraverso dispositivi video, indicando una serie di principi e limitazioni relativi alla trasparenza, alla proporzionalità, alla motivazione, all'indipendenza delle autorità che autorizzano[126]. Tali principi raccolti nelle linee guida sono stati poi incorporati nella proposta di regolamento[127].

Il titolo III della proposta di regolamento disciplina i sistemi di IA che sono considerati ad "alto rischio" per la salute, la sicurezza e i diritti fondamentali[128]. Le disposizioni relative ai sistemi ad alto rischio si applicano a due tipologie specifiche di sistemi di IA: in primis il sistema che è destinato "*a essere utilizzato come componente di sicurezza di un prodotto, o è esso stesso un prodotto, disciplinato dalla normativa*" europea; e per secondo i sistemi che rientrano nella lista di cui all'Allegato III[129]. La Commissione può aggiungere sottocategorie rispetto a quelle già presenti, sempre con un voto del

---

[125] Nathalie A Smuha, 'Beyond the Individual: Governing AI's Societal Harm' (2021) 10 Internet Policy Review 6 <https://policyreview.info/articles/analysis/beyond-individual-governing-ais-societal-harm> accessed 29 June 2023.

[126] European Data Protection Board, 'Linee guida 3/2019 sul trattamento dei dati personali attraverso dispositivi video', [2019]

[127] Tale divieto è stato modificato e reso più generale all'interno del rapporto di compromesso sugli emendamenti all'AI Act del Parlamento Europeo. In particolare, il sistema di IA con riconoscimento biometrico "in tempo reale" viene proibito in tutte le sue forme e per tutti gli scopi, seguendo anche il parere di EDPB e EDPS menzionato. Viene limitato ai casi in cui ci sia un'autorizzazione preventiva del giudice anche l'utilizzo dei sistemi di IA per il riconoscimento biometrico ex post, su registrazioni effettuate in luoghi aperti al pubblico. *Cfr*. (n 107)

[128] AI Act, considerando 43; articolo 7, paragrafo 2

[129] La lista dell'Allegato III include: identificazione e categorizzazione biometrica (sia 'remota', come nel Titolo II sopra, sia successivamente all'evento); gestione e funzionamento delle infrastrutture critiche; formazione scolastica e professionale; occupazione, gestione dei lavoratori e accesso al lavoro autonomo; accesso a e godimento di servizi essenziali e benefici; applicazione della legge; migrazione, asilo e gestione dei confini; amministrazione della giustizia e democrazia.

Parlamento o del Consiglio, ma non può aggiungere categorie completamente nuove. L'AI Act contiene una lista di requisiti essenziali per i sistemi ad alto rischio che sono connessi a degli obblighi che vanno solitamente a gravare sul fornitore del sistema. Il fornitore è definito all'articolo 3, numero 2, come "*una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o che fa sviluppare un sistema di IA al fine di immetterlo sul mercato o metterlo in servizio con il proprio nome o marchio, a titolo oneroso o gratuito;*". I fornitori dei sistemi di IA ad alto rischio, in aderenza alle prescrizioni del Regolamento, sono tenuti ad instaurare un accurato sistema di gestione della qualità. Tale sistema deve contemplare gli aspetti indicati all'articolo 17 del AI Act e necessita di una manutenzione e un aggiornamento continuo nel corso dell'intera durata operativa del sistema di IA in questione [130].

Gli insiemi di dati su cui vengono addestrati i sistemi di IA devono rispettare alcuni criteri qualitativi stabiliti all'articolo 10 della proposta[131], tra questi: la pertinenza, la rappresentatività, l'accuratezza, la completezza e le proprietà specifiche dell'area di applicazione. Nonostante alcuni requisiti sembrino stringenti - i *dataset* devono essere "privi di errori e completi", criterio che potrebbe risultare molto difficile da raggiungere - i *dataset* devono soddisfare questi requisiti solo in modo sufficiente e alla luce dello scopo previsto del sistema[132].
I fornitori devono anche predisporre interfacce che permettano la sorveglianza umana sui sistemi di IA ad alto rischio[133]. La finalità della previsione all'articolo 14 dell'AI Act è quella di assicurare una minimizzazione del rischio potenziale per la salute, la sicurezza e i diritti fondamentali. In particolare, per i sistemi di identificazione

[130] AI Act, art. 17

[131] AI Act, art. 10

[132] Il considerando (44) dell'AI Act prevede che: "…*In particolare, i set di dati di addestramento, convalida e prova dovrebbero tenere conto, nella misura necessaria alla luce della finalità prevista, delle caratteristiche o degli elementi peculiari dello specifico contesto o ambito geografico, comportamentale o funzionale all'interno del quale il sistema di IA ad alto rischio è destinato a essere usato*…"

[133] AI Act, art. 14

biometrica è richiesta una doppia sorveglianza umana secondo quanto previsto dal paragrafo 5 dell'articolo 14 dell'AI Act [134].

Al titolo III sono previste diverse valutazioni di conformità, che vanno brevemente analizzate insieme agli altri attori della proposta di regolamento: le organizzazioni delegate per fissare gli standard europei e gli organismi notificati competenti. Da un lato, il CEN (Comitato Europeo di Normazione) ed il CENELEC (Comitato Europeo di Normazione Elettrotecnica) sono due organizzazioni che stabiliscono standard europei e a cui la Commissione può demandare lo sviluppo di standard armonizzati, tuttavia le stesse non sono menzionate esplicitamente all'interno della proposta. Quando la Commissione avrà emesso tale mandato, se i due comitati adotteranno standard relativi all'AI Act, allora i fornitori dei sistemi di IA ad alto rischio potranno seguire gli standard individuati piuttosto che interpretare autonomamente i requisiti essenziali, potendo beneficiare della presunzione di conformità al regolamento[135]. Naturalmente tali standard non hanno alcuna vincolatività e i produttori possono anche interpretare autonomamente i requisiti essenziali; tuttavia, gli standard rappresenteranno una strada più semplice e sicura per beneficiare della presunzione di conformità.
Dall'altro lato, alcuni produttori possono certificare in maniera autonoma una valutazione di conformità basata su un controllo interno, quindi valutando l'adeguatezza del sistema di gestione della qualità, del piano di monitoraggio successivo all'immissione nel mercato e della documentazione tecnica relativa al sistema. Tuttavia, alcune caratteristiche richiedono l'approvazione di organismi tecnici indipendenti, noti come organismi notificati[136], che valutano la conformità al regolamento sull'IA dei sistemi ad alto rischio. Tali organismi sono secondo la definizione ex art. 3 (22): "*un organismo di valutazione della conformità designato in conformità al presente*

[134] Il paragrafo 5 dell'AI Act prevede che: "*Per i sistemi di IA ad alto rischio di cui all'allegato III, punto 1, lettera a), le misure di cui al paragrafo 3 sono tali da garantire che, inoltre, l'utente non compia azioni o adotti decisioni sulla base dell'identificazione risultante dal sistema, a meno che essa non sia stata verificata e confermata da almeno due persone fisiche.*"

[135] AI Act, art. 40

[136] AI Act, art. 33

regolamento e ad altre pertinenti normative di armonizzazione dell'Unione;" quindi tipicamente enti privati. A tali organismi però è attribuito un ruolo residuale relativo a pochi casi di identificazione biometrica e di classificazione degli individui[137].

Oltre ai sistemi di IA vietati e quelli considerati ad alto rischio, il regolamento sull'IA regola anche il resto dei sistemi di IA con obblighi di trasparenza. Il titolo IV prevede una serie di obblighi di trasparenza sia per utenti che per i fornitori di sistemi di IA, che riguardano qualsiasi sistema di IA che soddisfi i requisiti ex art 52 dell'AI Act[138]. Inoltre, la Commissione europea non ha alcun potere di modificare tali obblighi di trasparenza. Il primo obbligo di trasparenza riguarda i sistemi di IA pensati per interagire con persone fisiche (da qui in poi verrà utilizzato il termine *bot*, abbreviazione del termine *robot* utilizzata per riferirsi a programmi informatici che eseguono compiti automatizzati su Internet). I fornitori devono progettare il sistema affinché l'utente sia consapevole di interagire con un *bot*, a meno che tale circostanza non sia ovvia considerando il contesto o nel caso in cui i *bot* siano autorizzati dalla legge per prevenire reati[139]. La responsabilità di tale obbligo di trasparenza ricade sul fornitore del sistema e non sull'utente. Alcuni problemi nella distinzione tra fornitore e utente vengono posti dai nuovi sistemi di IA ad uso generale che, come vedremo nel terzo capitolo, rendono la distinzione tra soggetti che interagiscono col sistema poco chiara e per tale motivo bisognerebbe introdurre una nuova categoria di soggetti, ossia coloro che recepiscono il risultato dell'elaborazione del sistema di IA. Ad esempio, in un sistema di IA capace di generare testo, che trasforma dei semplici comandi in lunghe e complesse frasi che poi vengono pubblicati su una piattaforma social, rende difficile l'individuazione del fornitore del sistema di IA, mentre è più chiara l'identificazione del

[137] AI Act, art. 43 para. 1

[138] AI Act, art. 52

[139] L'articolo 52 del'AI Act prevede al primo paragrafo, che: "*I fornitori garantiscono che i sistemi di IA destinati a interagire con le persone fisiche siano progettati e sviluppati in modo tale che le persone fisiche siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dalle circostanze e dal contesto di utilizzo. Tale obbligo non si applica ai sistemi di IA autorizzati dalla legge per accertare, prevenire, indagare e perseguire reati, a meno che tali sistemi non siano a disposizione del pubblico per segnalare un reato.*"

soggetto che recepisce il risultato del sistema. L'accesso ai sistemi di IA è sempre più diffuso e alla portata di qualsiasi persona, si può accedere a sistemi molto sofisticati e allo stesso tempo di facile utilizzo. Secondo la proposta di regolamento, il produttore del sistema si limita a rivenderlo, ma in realtà la distinzione tra fornitore e utente è sempre meno importante e più difficile da distinguere, in considerazione dei recenti sviluppi dei sistemi di IA[140].

Oltre a prevedere obblighi per i fornitori di tali sistemi, il regolamento prescrive verso gli utenti[141] di sistemi di riconoscimento emotivo e di classificazione biometrica, il dovere di informare le persone soggette al trattamento riguardo l'esistenza del sistema di IA, a meno che non si tratti di un utilizzo previsto dalla legge per la prevenzione dei reati.[142] Non è chiaro cosa tale previsione possa aggiungere alla disciplina già contenuta all'interno del GDPR, infatti quando sistemi di questa tipologia analizzano dati personali la disciplina sulla protezione dei dati richiede che ne sia data informazione ai soggetti interessati dal trattamento che devono anche conoscerne le finalità.[143] Inoltre, tali sistemi di riconoscimento emotivo e classificazione biometrica si fondano su pratiche senza un reale fondamento scientifico, potendo anche causare ingiustizie sociali. Infatti, ricondurre le emozioni alle espressioni facciali può determinare una semplificazione eccessiva rispetto alla reale condizione emotiva dell'individuo, in particolare quando si considerano le diversità culturali e sociali che esistono nel mondo. Il sorriso non è sempre ricollegabile ad una emozione reale di felicità, può anche consistere, ad esempio, in una mera cordialità verso un'altra persona. La presunzione che uno stato emotivo manifestato corrisponda al reale stato emotivo di una persona

[140] Cobbe and Singh (n 102).

[141] Gli utenti sono definiti all'articolo 3 (4) dell'AI Act come: "*qualsiasi persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la sua autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale*;"

[142] L'articolo 52 al secondo paragrafo prevede che: "*Gli utenti di un sistema di riconoscimento delle emozioni o di un sistema di categorizzazione biometrica informano le persone fisiche che vi sono esposte in merito al funzionamento del sistema. Tale obbligo non si applica ai sistemi di IA utilizzati per la categorizzazione biometrica, che sono autorizzati dalla legge per accertare, prevenire e indagare reati."*

[143] GDPR articolo 13**.**

sembra essere un assunto di partenza di tali sistemi, che pretendono di poter decifrare la complessità delle emozioni umane dalla superficiale analisi delle sembianze del volto. La mappatura emotiva ha necessariamente a che vedere con i meccanismi cerebrali che ancora rimangono ignoti e misteriosi, in un ambito che da molto tempo si è cercato di ricondurre nelle maglie della scienza, senza tuttavia ottenere risultati che permettessero effettivamente di comprendere le emozioni umane in maniera ordinata. I sistemi di IA che ambiscono a comprendere e riconoscere le emozioni non potranno che limitarsi alla superficie visibile dell'inconscio umano che si manifesta nei nostri volti.

Tornando alla trasparenza richiesta per i sistemi di IA, l'obbligo previsto all'articolo 52, paragrafo 3, ha per destinatari gli utenti e riguarda quei sistemi di IA che permettono di generare o manipolare immagini, contenuti audio o video che: "*assomigliano notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri per una persona ("deep fake") sono tenuti a rendere noto che il contenuto è stato generato o manipolato artificialmente.*"[144] A tal proposito bisogna evidenziare come in realtà la definizione di utente ex articolo 3 (4) dell'AI Act esclude dagli obblighi di trasparenza quelle attività online condotte per un'attività personale non professionale. Da una prima lettura risulta chiaro che solamente i soggetti che utilizzano tali sistemi in ambito di un'attività professionale saranno tenuti a informare che il contenuto ha origine artificiale (per indicare tale attributo viene utilizzato il termine "sintetico"). Tali obblighi di trasparenza esprimono un principio molto importante per la moderna società digitale, ossia il diritto a conoscere l'origine e la natura del contenuto che si sta analizzando, se questo è stato generato da un computer sulla base di un comando umano oppure se è opera di un

[144] Il testo completo dell'articolo 52, paragrafo 3, dell'AI Act è riportato di seguito: "*Gli utenti di un sistema di IA che genera o manipola immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri per una persona ("deep fake") sono tenuti a rendere noto che il contenuto è stato generato o manipolato artificialmente.*

Tuttavia il primo comma non si applica se l'uso è autorizzato dalla legge per accertare, prevenire, indagare e perseguire reati o se è necessario per l'esercizio del diritto alla libertà di espressione e del diritto alla libertà delle arti e delle scienze garantito dalla Carta dei diritti fondamentali dell'UE, e fatte salve le tutele adeguate per i diritti e le libertà dei terzi."

individuo. Il potenziale di rischio che può derivare dalla diffusione di immagini manipolate determina la diffusione di disinformazione e allo stesso tempo può destabilizzare la fiducia degli individui sul resto dei contenuti, portando gli stessi a interrogarsi sulla loro natura e sul confine tra realtà e finzione che nel mondo digitale è sempre meno chiaro.

La categoria finale, invece, rappresenta una categoria residuale che comprende tutte le applicazioni dell'IA con un rischio minimo o nullo, come ad esempio i videogiochi o i filtri *antispam* nelle e-mail, e pertanto non rientrano nel quadro normativo delineato dal regolamento.

In virtù della sua progressiva diffusione e delle sue implicazioni pervasive, l'IA rappresenta un elemento di mutamento significativo negli ambiti di interazione quotidiana. La suddetta tecnologia, caratterizzata da un processo di evoluzione incessante, plasmerà indubbiamente il nostro metodo di rapportarci al mondo circostante. Analogamente ad altre innovazioni di rilievo, l'IA ha generato vantaggi considerevoli per la società, come ad esempio i progressi nella comprensione delle proteine grazie al programma Alphafold[145], anche se, in certi frangenti, gli svantaggi sembrano aver superato i benefici portati da tale tecnologia.

Un ulteriore aspetto dell'AI Act a cui si deve prestare attenzione riguarda i soggetti a cui si rivolge. L'AI Act contiene all'art. 2, para. 1, un'ampia applicazione territoriale per permettere una estensione dell'efficacia del regolamento anche a fornitori che non si trovano all'interno dell'Unione. Infatti, l'articolo prevede un'applicazione del regolamento che si rivolge: "

a) ai fornitori che immettono sul mercato o mettono in servizio sistemi di IA nell'Unione, indipendentemente dal fatto che siano stabiliti nell'Unione o in un paese terzo;

[145] John Jumper and others, 'Highly Accurate Protein Structure Prediction with AlphaFold' (2021) 596 Nature 583, 583–589.

b) agli utenti dei sistemi di IA situati nell'Unione;

c) ai fornitori e agli utenti di sistemi di IA situati in un paese terzo, laddove l'output prodotto dal sistema sia utilizzato nell'Unione."

In particolare, l'art. 2, para. 1, lett. c) amplia in maniera innovativa il prevedibile raggio di applicazione dell'AI Act, specificando che anche i fornitori e utilizzatori che si trovano in paesi terzi rispetto all'Unione sono riguardati dalle disposizioni, quando il risultato dei sistemi venga utilizzato nel territorio dell'Unione. Un esempio di tale applicazione viene fornito dal considerando n. 11 dell'AI Act[146], secondo cui il regolamento si applicherebbe anche a:

"*un operatore stabilito nell'Unione che appalta alcuni servizi a un operatore stabilito al di fuori dell'Unione in relazione a un'attività che deve essere svolta da un sistema di IA che sarebbe classificato ad alto rischio e i cui effetti avrebbero un impatto sulle persone fisiche che si trovano nell'Unione. In tali circostanze il sistema di IA utilizzato dall'operatore al di fuori dell'Unione potrebbe trattare dati raccolti nell'Unione e da lì trasferiti, nel rispetto della legge, e fornire all'operatore appaltante nell'Unione l'output di tale sistema di IA risultante da tale trattamento, senza che tale sistema di IA sia immesso sul mercato, messo in servizio o utilizzato nell'Unione*."

Si può dunque desumere che l'obiettivo del regolatore sia quello di evitare l'elusione del regolamento, facilmente ottenibile con le moderne tecnologie, per garantire una protezione efficace delle persone fisiche che si trovano nell'Unione applicando il regolamento anche a fornitori e utenti stabiliti in un Paese terzo.

Con la scelta del modello del regolamento europeo, alcuni atti legislativi recenti sono riusciti ad esplicare i loro effetti anche al di fuori del limitato contesto territoriale su cui avevano applicazione diretta. Il contributo della CGUE è stato fondamentale per l'implementazione di tale politica, tanto è vero che con la decisione *Google Spain* la

[146] AI Act, considerando n. 11

disciplina europea in materia di protezione dei dati ha visto ampliare la sua applicazione territoriale [147].

La proposta di regolamentazione si presenta dunque come un nuovo elemento, che si aggiunge al sistema di regolazione digitale europeo aspirante all'ambizioso obiettivo di porsi come standard globale di disciplina delle nuove tecnologie. Tuttavia, data la natura per nulla regolamentata del settore a riguardo, il rischio che potrebbe aversi è quello di fissare principi e regole in maniera simbolica, senza un effettivo riscontro nel sistema degli standard globali, venendo attratti dalla sfera di influenza cinese o americana.[148] In particolare, considerando il profluvio di documenti e raccomandazioni contenenti principi etici e giuridici, prodotti da organizzazioni internazionali, governi e organizzazioni non governative, relativi al governo dei sistemi di IA, le "Linee Guida Etiche per un'IA Affidabile" del AI HLEG[149], su cui si basa l'AI Act, sembrano essere quelle più complete e che contengono il numero maggiore dei più diffusi principi etici in materia.[150]

---

[147] C131/12, Google Spain SL e Google Inc. c Agencia Española de Protección de Datos (AEPD), Mario Costeja González, [2014]. ECLI:EU:C:2014:317.; si veda anche C-507/17, Google LLC v. Commission nationale de l'informatique et des libertés (CNIL) [2019]. ECLI: ECLI:EU:C:2019:772.

[148] Matthew S Erie and Thomas Streinz, 'The Beijing Effect: China's Digital Silk Road as Transnational Data Governance' (2021) 54 New York University Journal of International Law and Politics 1. Il termine Beijing Effect si riferisce all'influenza crescente della Cina sul governo dei dati a livello globale attraverso la sua iniziativa della "Nuova Via della Seta Digitale". Questa iniziativa mira a promuovere l'adozione di tecnologie digitali cinesi in tutto il mondo e a creare una nuova architettura di governance dei dati globale che favorisca la Cina. Anche negli Stati Uniti d'America è stata presentata una regolamentazione dei sistemi di decisione algoritmica (ADS) denominata "Algorithm Accountability Act" che adotta un approccio meno specifico dell'AI Act europeo, sebbene condivida alcuni punti, come evidenziati nell'articolo di Jakob Mökander and others, 'The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What Can They Learn from Each Other?' (2022) 32 Minds and Machines, 751–758.

[149] European Commission. Directorate General for Communications Networks, Content and Technology. and High Level Expert Group on Artificial Intelligence. (n 91).

[150] Graham Greenleaf, 'The "Brussels Effect" of the EU's "AI Act" on Data Privacy Outside Europe' (7 June 2021) <https://papers.ssrn.com/abstract=3898904> accessed 29 April 2023.

Secondo alcuni, la proposta rappresenterebbe un modello camaleontico capace di adattarsi ad eventuali carenze normative a prescindere dal contesto geografico, così com'è avvenuto con il GDPR. In Australia, ad esempio, sono stati di recente proposti dei principi dalla Commissione per i Diritti Umani, che sostanzialmente ricopiano parte di quelli della proposta[151]. L'AI Act si presenta come un atto normativo che ambisce a regolare non solo il contesto dell'Unione europea, ma mira anche a soggetti terzi che in qualche modo, attraverso i sistemi di IA, producono effetti sui cittadini europei. L'AI Act rappresenta la prima proposta di atto normativo relativo all'IA per un'istituzione governativa come l'Unione Europea. Rappresenterà per molti Paesi al di fuori dell'UE un esempio da seguire nella regolamentazione dei sistemi di IA e con alta probabilità alcuni elementi verranno adottati da altri paesi, in quello che si conferma essere uno dei meccanismi di politica globale più pervasivo, teorizzato dalla professoressa Anu Bradford con il nome di *Brussels Effect*[152]. La capacità di influenzare oltre ai soggetti del mercato di riferimento anche i regolatori dei mercati di altri stati connota una peculiare pervasività della regolamentazione europea. Tale capacità ha contraddistinto in particolare la materia della protezione dei dati; infatti, con il GDPR il legislatore europeo si è dimostrato capace, non solo di regolare e limitare l'operato dei soggetti nel mercato dei dati europeo, ma anche di stabilire degli standard di tutela per i dati personali e la *privacy* dei soggetti interessati, che ha spinto i legislatori di molti paesi stranieri a adottare regolamentazioni simili[153].

Da queste considerazioni emerge l'impellente necessità di regolamentare questo ambito

---

[151] Australian Human Rights Commission, *Human Rights and Technology Final Report* (2021) <https://tech.humanrights.gov.au/sites/default/files/2021-05/AHRC_RightsTech_2021_Final_Report.pdf>

[152] Questo fenomeno consiste in una imitazione da parte di ordinamenti stranieri della regolamentazione cogente europea in materia, ad esempio, di protezione dei dati personali. Anu Bradford, 'The Brussels Effect: How the European Union Rules the World' [2020] Faculty Books <https://scholarship.law.columbia.edu/books/232>.

[153] Il GDPR è un chiaro esempio di questo fenomeno che è stato teorizzato dalla professoressa Anu Bradford con il nome di *Brussels Effect.* Questo fenomeno consiste in una imitazione da parte di ordinamenti stranieri della regolamentazione cogente europea in materia, ad esempio, di protezione dei dati personali. ibid.

tecnologico, o quantomeno alcuni aspetti dello stesso. Nonostante l'IA sia un fenomeno tecnologico in continuo sviluppo, già esistono articoli del GDPR che ne regolano alcuni aspetti, come l'articolo 22 del GDPR che pone dei solidi principi di sorveglianza umana, trasparenza del risultato (meglio conosciuta come explainability[154]) e mira ad evitare un danno per l'individuo destinatario della decisione automatizzata del sistema. L'IA, come abbiamo già avuto modo di comprendere, altro non è che una tecnologia alimentata da dati e tra questi, molto spesso, si trovano anche dati personali. Se i dati trattati e analizzati dal sistema permettono di identificare la persona a cui si riferiscono, allora si tratterà di dati qualificabili come dati personali. L'art. 4, par. 1, del GDPR definisce come dati personali: «*qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale*»[155]. Secondo questa definizione i dati che l'IA raccoglie e che permettono l'identificazione rientrerebbero nella categoria di dato personale, richiedendo l'applicazione delle norme del Regolamento. Tuttavia, il GDPR prevede la possibilità di trattare i dati in maniera lecita e senza dover applicare i principi in esso stabiliti, quando gli stessi siano resi anonimi, "*vale a dire informazioni che non si riferiscono a una persona fisica identificata o identificabile o a dati personali resi sufficientemente anonimi da impedire o da non consentire più l'identificazione dell'interessato"*[156].

---

[154] Cecilia Panigutti, Ronan Hamon, Isabelle Hupont Torres, David Fernández-Llorca, Delia Yela, Henrik Junklewitz, Salvatore Scalzo, Gabriele Mazzini, Ignacio Sanchez, Josep Garrido, Emilia Gómez, 'The role of explainable AI in the context of the AI Act', FAccT '23, Chicago, (2023) 1139-1150.

[155] GDPR, art.4

[156] Ibid. Il considerando (26) stabilisce che: "*È auspicabile applicare i principi di protezione dei dati a tutte le informazioni relative a una persona fisica identificata o identificabile. I dati personali sottoposti a pseudonimizzazione, i quali potrebbero essere attribuiti a una persona fisica mediante l'utilizzo di ulteriori informazioni, dovrebbero essere considerati informazioni su una persona fisica identificabile. Per stabilire l'identificabilità di una persona è opportuno considerare tutti i mezzi, come l'individuazione, di cui il titolare del trattamento o un terzo può ragionevolmente avvalersi per identificare detta persona fisica direttamente o indirettamente. Per accertare la ragionevole probabilità di utilizzo dei mezzi per*

Secondo il Garante per la protezione dei dati personali italiano (da qui in poi, Garante), esiste un rischio che nasce dall'esposizione delle persone fisiche a processi di decisione automatizzata di dati. Anche prima che fosse stabilito il principio ex art. 22 GDPR, il Garante aveva iniziato a considerare tale rischio e vigilare sull'applicazione di simili tecnologie[157]. A seguito dell'incremento dell'uso della tecnologia causato dal periodo pandemico, i sistemi di IA sono proliferati. Le tecnologie di IA operano oggi nuove forme di raccolta dei dati attraverso metodi sempre più sofisticati, anche se le finalità del trattamento non sono le finalità pubblicitarie e di marketing che contraddistinguono il trattamento effettuato dai social network e dai motori di ricerca, piuttosto, i nuovi sistemi di IA di cui si parlerà in maniera approfondita nel capitolo 3 sono progettati per una moltitudine di finalità diverse.

Data la costante innovazione e mutevolezza nel campo tecnologico, è diventato essenziale per il legislatore introdurre un'apposita normativa che consentisse di comprendere e gestire al meglio questo fenomeno. Tale normativa mira a mitigare i rischi associati e a valorizzare i vantaggi che emergono da questa evoluzione. Secondo la professoressa Ginevra Cerrina Feroni (Vicepresidente del Garante), "implica il tentativo di rimodulare il perimetro del tecnicamente possibile sulla base di quello che si ritiene giuridicamente ed eticamente accettabile"[158]. Anche se, alcune garanzie e tutele

---

identificare la persona fisica, si dovrebbe prendere in considerazione l'insieme dei fattori obiettivi, tra cui i costi e il tempo necessario per l'identificazione, tenendo conto sia delle tecnologie disponibili al momento del trattamento, sia degli sviluppi tecnologici. I principi di protezione dei dati non dovrebbero pertanto applicarsi a informazioni anonime, vale a dire informazioni che non si riferiscono a una persona fisica identificata o identificabile o a dati personali resi sufficientemente anonimi da impedire o da non consentire più l'identificazione dell'interessato. Il presente regolamento non si applica pertanto al trattamento di tali informazioni anonime, anche per finalità statistiche o di ricerca."

[157] Il Garante ha ad esempio chiarito nel Provvedimento Doc. Web. N. 7496252 del 21 dicembre 2017 'Istallazione di apparati promozionali del tipo 'digital signage' (definiti anche Totem) presso una stazione ferroviaria', che un sistema di lettura biometrica montato su totem pubblicitari, capace di identificare sesso, età, tempo trascorso ad osservare il monitor di soggetti che passavano a guardare gli spot, e di comprendere le espressioni facciali, fosse un sistema che violava i principi del consenso e degli obblighi di sicurezza del precedente Codice per la protezione dei dati personali d.lgs. 30 giugno 2003 n. 196.

[158] Garante della protezione dei dati personali, Doc. Web. N. 9855742, 'Intelligenza artificiale e ruolo della protezione dei dati personali'.

previste nell'AI Act erano già presenti all'interno del GDPR e potevano anche risultare sufficienti per temporaneo e corretto sviluppo di tali tecnologie.

Il Garante già da tempo si è interessato all'IA ad esempio nell'indagine, svolta insieme ad AGCOM e AGCM sui *big data*, oppure col *vademecum* incentrato sul tema dei *deepfake*[159]. A livello europeo l'interessamento si è tradotto nella prima proposta di regolamento sull'IA, che si presenta come *unicum* a livello globale[160]. Nel resto del mondo, infatti, la situazione è caratterizzata da approcci diversi, e che solitamente prediligono la libertà di sviluppo tecnologico e di mercato.

USA e Cina, ad esempio, stanno promuovendo e finanziando lo sviluppo di sistemi di IA, dando la precedenza ad una logica di mercato regolata solo da fonti non vincolanti di soft-law. L'approccio intrapreso negli USA è stato finora settoriale, infatti alcune applicazioni di IA sono state vietate come ad esempio quelle di riconoscimento facciale per le possibili discriminazioni[161], ma per una regolamentazione più omogenea e strutturata verso la fine del 2022 è stato presentato l'Algorithmic Accountability Act (di seguito, US AAA) che consiste nel primo tentativo di rafforzare la legislazione in materia negli USA[162]. Tale proposta richiede ad esempio che le aziende produttrici di sistemi di decisione automatica (ADS), che hanno un "impatto significativo" sui consumatori, conducano una valutazione di impatto.

Gli obblighi riguardano tre gruppi diversi di aziende: (a) quelle che hanno un fatturato annuo superiore a 50 milioni di dollari o un valore azionario superiore a 250 milioni di dollari negli ultimi tre anni[163], (b) aziende che trattano dati personali di oltre 1 milione

---

[159] Garante della protezione dei dati personali, 'Rapporto finale pubblicato nel 2020', doc. web n. 9264297; 'Deepfake-Vademecum', doc. web n. 9512226

[160] 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, [2021], COM/2021/206 final (n 88).

[161] Angela Chen, 'Why San Francisco's ban on face recognition is only the start of a long fight,' in *MIT Review*, May 16, 2019

[162] Algorithmic Accountability Act of 2022, S. 3572 (IS)

[163] US AAA, Sec. 2 (7) I

di consumatori, famiglie o dispositivi di consumo per sviluppare o utilizzare gli ADS[164]; e (c) aziende che hanno dimensioni pari a un decimo delle grandi aziende e che implementano uno dei loro ADS[165]. Le aziende devono rientrare nella giurisdizione della Federal Trade Commission (FTC), il che lascia fuori, ad esempio, le agenzie governative locali[166].

In Cina l'utilizzo e lo sviluppo di IA seguono una strategia ben definita dal governo cinese che punta a superare gli USA nel primato sull'IA. Tuttavia, spaventa l'uso che il governo ne fa, in particolare con il *Social Credit System*[167]. Esistono quindi approcci diversi per regolare il fenomeno tecnologico, ma risulta in ogni caso chiaro che gli attori globali stanno sempre più puntando sull'IA cercando di stabilire dei limiti che siano compatibili con i benefici che tali tecnologie possono portare. La direzione presa dai legislatori da ogni parte del globo assomiglia ad una vera e propria gara per regolare prima possibile il settore e ottenere i primi investimenti e il beneficio economico che si prevede possa derivare dal settore[168].

---

[164] US AAA, Sec. 2 (7) II

[165] Ibid.

[166] Noha Lea Halim, Urs Gasser, 'Vectors of AI governance - Juxtaposing the U.S. Algorithmic Accountability Act of 2022 with The EU Artificial Intelligence Act'Available at SSRN: https://ssrn.com/abstract=4476167 (2023)

[167] Anne SY Cheung and Yongxi Chen, 'From Datafication to Data State: Making Sense of China's Social Credit System and Its Implications' (2022) 47 Law & Social Inquiry 1137, 1139–1156. Il Social Credit System cinese, introdotto nei primi anni 2000 come sistema di credito finanziario, è un sistema di valutazione della reputazione dei cittadini e delle aziende basato su una vasta gamma di comportamenti e attività. Utilizzando dati finanziari, social media, dati sugli acquisti e sull'osservanza delle leggi, il sistema assegna un punteggio che può influenzare l'accesso al credito, all'impiego, ai viaggi e all'istruzione. Un punteggio basso può comportare penalità, mentre un punteggio alto può offrire benefici. Tuttavia, il sistema ha suscitato preoccupazioni per la *privacy* e i diritti umani, e la sua implementazione varia in tutto il Paese.

[168] Nathalie A Smuha, 'From a "Race to AI" to a "Race to AI Regulation": Regulatory Competition for Artificial Intelligence' (2021) 13 Law, Innovation and Technology 57, 57–84.; Michael Chui, Eric Hazan, Roger Roberts, Alex Singla, Kate Smaje, Alex Sukharevsky, Lareina Yee, Rodney Zemmel 'The economic potential of generative AI' McKinsey&Company, (June 2023)

2. Come si pone il GDPR nel regolamentare l'Intelligenza Artificiale

Avendo compreso la natura tecnologica dietro l'IA e avendo analizzato la proposta di regolamentazione che è stata pensata per tracciarne i limiti, passiamo all'analisi normativa di alcune disposizioni normative del GDPR che si applicano ai sistemi di IA. Questo regolamento europeo è il principale strumento in vigore che oggi può regolare i sistemi di IA, stabilendo limiti e divieti, tra cui quello per il trattamento dei dati condotto in forma automatizzata che può avere degli effetti significativi sull'interessato.
Il termine *Intelligenza Artificiale* non è menzionato all'interno del GDPR, ma sono diversi gli articoli che in qualche modo regolano questa moderna tecnologia. L'applicazione di questi articoli è in alcuni casi messa in difficoltà dallo sviluppo di nuovi processi di IA che superano certi meccanismi di tutela dei dati personali. I principi tradizionali che fondano la protezione dei dati personali, tra cui la limitazione delle finalità, la minimizzazione dei dati, la liceità, correttezza e trasparenza del trattamento ex art. 5; il trattamento speciale dei dati sensibili che richiede maggiori garanzie ex art. 9; la limitazione delle decisioni automatizzate ex art. 22, vengono messi in crisi dal rapido sviluppo dell'IA e dei *Big Data* che di giorno in giorno affinano le loro capacità. L'interpretazione dei principi del GDPR continua ad essere fondamentale per la regolazione di una tecnologia di così grande impatto, garantendo comunque il rispetto degli stessi.

Prima di considerare nello specifico come il GDPR sia stato utilizzato per regolare alcuni trattamenti dei dati per mezzo di tecnologie di IA, bisogna analizzare l'origine e capire la ratio della normativa per la protezione dei dati personali. Inizialmente, i trattati fondanti delle Comunità europee non includevano disposizioni sulla protezione dei dati personali e dei diritti umani in generale, poiché l'accento era principalmente posto sugli aspetti economici delle istituzioni. L'obiettivo principale era la creazione di un mercato unico senza ostacoli doganali, che favorisse la libera circolazione di beni, persone, servizi e capitali, promuovendo benessere e prosperità. Poiché l'Unione europea è un'istituzione sovranazionale che deriva dalla volontaria e parziale cessione di sovranità da parte degli Stati membri, è fondamentale il principio di attribuzione sancito

nell'articolo 3 del Trattato sull'Unione europea (TUE)[169]. Tale principio stabilisce che "*l'Unione perseguirà i suoi obiettivi mediante mezzi appropriati, in base alle competenze ad essa attribuite dai trattati*"[170], poiché l'Unione europea non può agire al di fuori delle competenze che le sono state conferite dagli Stati membri[171]. Pur non

[169] Trattato sull'Unione europea, versione consolidata, [2012]. OJ C 326, 26.10.2012, p. 15–368. All'art. 3 prevede: "*L'Unione si prefigge di promuovere la pace, i suoi valori e il benessere dei suoi popoli. L'Unione offre ai suoi cittadini uno spazio di libertà, sicurezza e giustizia senza frontiere interne, in cui si assicurata la libera circolazione delle persone insieme a misure appropriate per quanto concerne i controlli alle frontiere esterne, l'asilo, l'immigrazione, la prevenzione della criminalità e la lotta contro quest'ultima. L'Unione instaura un mercato interno. Si adopera per lo sviluppo sostenibile dell'Europa, basato su una crescita economica equilibrata e sulla stabilità dei prezzi, su un'economia sociale di mercato fortemente competitiva, che mira alla piena occupazione e al progresso sociale, e su un elevato livello di tutela e di miglioramento della qualità dell'ambiente. Essa promuove il progresso scientifico e tecnologico. L'Unione combatte l'esclusione sociale e le discriminazioni e promuove la giustizia e la protezione sociali, la parità tra donne e uomini, la solidarietà tra le generazioni e la tutela dei diritti del minore. Essa promuove la coesione economica, sociale e territoriale, e la solidarietà tra gli Stati membri. Essa rispetta la ricchezza della sua diversità culturale e linguistica e vigila sulla salvaguardia e sullo sviluppo del patrimonio culturale europeo. L'Unione istituisce un'unione economica e monetaria la cui moneta è l'euro. Nelle relazioni con il resto del mondo l'Unione af erma e promuove i suoi valori e interessi, contribuendo alla protezione dei suoi cittadini. Contribuisce alla pace, alla sicurezza, allo sviluppo sostenibile della Terra, alla solidarietà e al rispetto reciproco tra i popoli, al commercio libero ed equo, all'eliminazione della povertà e alla tutela dei diritti umani, in particolare dei diritti del minore, e alla rigorosa osservanza e allo sviluppo del diritto internazionale, in particolare al rispetto dei principi della Carta delle Nazioni Unite. L'Unione persegue i suoi obiettivi con i mezzi appropriati, in ragione delle competenze che le sono attribuite nei trattati*".

[170] Trattato sull'Unione europea, versione consolidata, [2012]. OJ C 326, 26.10.2012, p. 15–368. art. 3, comma 6

[171] Le competenze dell'UE sono suddivise in competenze esclusive (art. 3 TFUE), competenze concorrenti (art. 4 TFUE) e competenze di sostegno (art. 6 TFUE). Nelle competenze esclusive, l'Unione ha l'autorità legislativa esclusiva, mentre gli Stati membri si limitano ad attuare gli atti adottati dall'Unione. Queste competenze comprendono, ad esempio, l'unione doganale, la definizione delle regole di concorrenza necessarie per il funzionamento del mercato interno, la politica monetaria per gli Stati membri che adottano l'euro, la conservazione delle risorse biologiche marine nell'ambito della politica comune della pesca e la politica commerciale comune. Nelle competenze concorrenti, sia l'Unione europea che gli Stati membri possono adottare atti vincolanti, ma gli Stati membri possono farlo solo se l'Unione non ha ancora agito o ha rinunciato a farlo. Queste competenze comprendono il mercato interno, la politica sociale, la coesione economica, sociale e territoriale, l'agricoltura e la pesca (ad eccezione della conservazione delle risorse biologiche marine), l'ambiente, la protezione dei consumatori, i trasporti, le reti transeuropee, l'energia, lo spazio di libertà, sicurezza e giustizia e le questioni comuni di sicurezza in materia di sanità pubblica. Nelle competenze di sostegno, l'Unione europea può solo sostenere e coordinare le misure adottate dagli Stati membri. Queste competenze comprendono la tutela e il

essendo esplicitamente menzionati nei trattati istitutivi, nel corso degli anni la CGUE ha interpretato i principi generali del diritto europeo riconoscendo i diritti fondamentali degli individui. Nel 2000, l'UE ha concretizzato tali diritti, trasferendoli dalle costituzioni degli Stati membri e dagli obblighi internazionali, nella Carta dei diritti fondamentali dell'Unione Europea (CDFUE) [172]. Originariamente considerato un documento politico, la CDFUE ha acquisito valore vincolante come diritto primario con l'entrata in vigore del Trattato di Lisbona del 2007 (in vigore dal dicembre 2009)[173].

La Carta non solo garantisce il diritto al rispetto della vita privata (art. 7), ma riconosce esplicitamente il diritto alla protezione dei dati personali (art. 8) [174]. In aggiunta all'enorme progresso rappresentato da tale riconoscimento, testimonianza dell'evoluzione sociale e tecnologica e della crescente importanza dei dati personali, l'art. 8 stabilisce i principi che devono caratterizzare il trattamento dei dati: lealtà, liceità e limitazione delle finalità. Inoltre, vengono previsti il diritto di accesso ai dati e di rettifica, nonché l'istituzione di un'autorità indipendente responsabile della vigilanza sul rispetto delle norme. I soggetti e le istituzioni europee, così come gli Stati membri nell'attuazione del diritto europeo, sono tenuti a rispettare e garantire tale diritto.

Oltre alla CDFUE, il diritto alla protezione dei dati personali è previsto anche nel Trattato sul Funzionamento dell'Unione Europea (TFUE), all'art. 16, all'interno del

---

miglioramento della salute umana, l'industria, la cultura, il turismo, l'istruzione, la formazione professionale, la gioventù e lo sport, la protezione civile e la cooperazione amministrativa. Trattato sull'Unione europea, versione consolidata, [2012]. OJ C 326, 26.10.2012, p. 15–368.

[172] Carta dei diritti fondamentali dell'Unione europea, [2012] GU C 326 del 26.10.2012, pagg. 391–407.

[173] Trattato di Lisbona che modifica il trattato sull'Unione europea e il trattato che istituisce la Comunità europea, firmato a Lisbona il 13 dicembre 2007, GU C 306 del 17.12.2007

[174] Carta dei diritti fondamentali dell'Unione europea, [2012] GU C 326 del 26.10.2012, pagg. 391–407. L'art. 8, rubricato "Protezione dei dati di carattere personale", stabilisce che "*Ogni individuo ha diritto alla protezione dei dati di carattere personale che lo riguardano. Tali dati devono essere trattati secondo il principio di lealtà, per finalità determinate e in base al consenso della persona interessata o a un altro fondamento legittimo previsto dalla legge. Ogni individuo ha il diritto di accedere ai dati raccolti che lo riguardano e di ottenerne la rettifica. Il rispetto di tali regole è soggetto al controllo di un'autorità indipendente*".

titolo che riguarda i principi generali dell'Unione europea.[175] Questa disposizione è di fondamentale importanza, poiché l'art. 16 diventa la nuova base giuridica per la competenza dell'UE in materia di protezione dei dati, sostituendo il mercato interno e l'armonizzazione delle legislazioni nazionali come base giuridica precedente. È stato proprio su questa base giuridica che, nel 2016, è stato adottato il celebre GDPR, che è entrato in vigore due anni dopo. Fino a quel momento, lo strumento principale per la protezione dei dati personali era stata la Direttiva 95/46/CE, che conteneva i principi e i diritti già enunciati dalla Convenzione n. 108 e dalle normative nazionali, ma che, non avendo carattere vincolante, consentiva agli Stati membri una certa flessibilità nell'adattamento della legislazione nazionale.

Con l'obiettivo di creare un sistema armonizzato e aggiornato alle evoluzioni tecnologiche, nel 2016 è stato adottato il GDPR, che ha un ampio campo di applicazione nel trattamento dei dati personali, ossia le informazioni relative a persone fisiche identificate o identificabili nel territorio dell'Unione, definite come "interessati". I soggetti coinvolti nel trattamento dei dati sono definiti e identificati nell'art. 4. In particolare, il titolare del trattamento può essere una persona fisica o giuridica che determina le finalità e i mezzi del trattamento, mentre il responsabile del trattamento è colui che effettua il trattamento per conto del titolare[176]. Nel caso in cui il titolare del trattamento non sia stabilito nel territorio dell'Unione, ad esempio nel caso di imprese

[175] Trattato sul funzionamento dell'Unione europea, versione consolidata, [2012]. OJ C 326, 26.10.2012, p. 47–390. Art. 16: "*Ogni persona ha diritto alla protezione dei dati di carattere personale che la riguardano. Il Parlamento europeo e il Consiglio, deliberando secondo la procedura legislativa ordinaria, stabiliscono le norme relative alla protezione delle persone fisiche con riguardo al trattamento dei dati di carattere personale da parte delle istituzioni, degli organi e degli organismi dell'Unione, nonché da parte degli Stati membri nell'esercizio di attività che rientrano nel campo di applicazione del diritto dell'Unione, e le norme relative alla libera circolazione di tali dati. Il rispetto di tali norme è soggetto al controllo di autorità indipendenti. Le norme adottate sulla base del presente articolo fanno salve le norme specifiche di cui all'articolo 39 del trattato sull'Unione europea*".

[176] GDPR, art. 4.

multinazionali, è necessaria la designazione di un rappresentante che agisca per suo conto [177].

I principi fondamentali che caratterizzano il trattamento dei dati personali sono stabiliti nell'art. 5 del Regolamento. Essi includono i principi di liceità, correttezza e trasparenza; limitazione delle finalità; minimizzazione dei dati; esattezza dei dati; limitazione della conservazione; e integrità e riservatezza[178]. L'art. 5 stabilisce che il trattamento dei dati personali deve essere lecito, corretto e trasparente, richiedendo alternativamente il consenso dell'interessato, il quale deve essere libero, informato, specifico e inequivocabile (come stabilito nell'art. 7 del GDPR), oppure altre basi giuridiche, come l'esecuzione di un contratto, l'adempimento di un obbligo legale, la salvaguardia degli interessi vitali dell'interessato o di terzi, l'esecuzione di un compito di interesse pubblico o il perseguimento di un interesse legittimo da parte del titolare del trattamento, a condizione che tali interessi non prevalgano sugli interessi o sui diritti e le libertà fondamentali dell'interessato[179].

I principi di correttezza e trasparenza richiedono un comportamento corretto del titolare del trattamento e l'obbligo di fornire informazioni adeguate agli interessati. Il principio di limitazione delle finalità stabilisce che il trattamento dei dati personali deve essere limitato alle finalità specifiche per cui sono stati raccolti, consentendo un ulteriore trattamento solo se lo stesso risulti compatibile con le finalità originarie. I principi di minimizzazione ed esattezza richiedono che il trattamento dei dati sia limitato a quanto strettamente necessario e che i dati siano sempre aggiornati o in caso, rettificati. Infine, per garantire l'integrità e la riservatezza dei dati, sono necessarie adeguate misure e protocolli di sicurezza. Oltre ai principi generali, il GDPR riconosce una serie di diritti agli interessati per garantire una maggiore protezione in un mondo digitale sempre più

[177] È importante notare che le persone giuridiche non sono coperte dal GDPR, ma ricevono una certa protezione dall'art. 8 della Convenzione Europea dei Diritti dell'Uomo (CEDU). In tal senso si è pronunciata la Corte Europea dei Diritti Umani nel caso Bernh Larsen Holding AS e a. c. Norvegia, n. 24117/08, [2013]

[178] GDPR art. 5.

[179] ibidem, art.6

complesso e pieno di rischi. Tali diritti includono il diritto all'informazione, il diritto di rettifica, il diritto alla cancellazione dei dati (o "diritto all'oblio"), il diritto alla limitazione del trattamento, il diritto alla portabilità dei dati e il diritto di opposizione [180].

Nel corso degli anni, la giurisprudenza ha contribuito a chiarire l'ambito di applicazione e le sfumature di ciascun diritto[181]. Tra i casi più rilevanti se ne possono indicare alcuni. Relativamente al diritto ad essere informati, la CGUE ha chiarito tale diritto nel caso *Smaranda Bara e a. d. Președintele Casei Naționale de Asigurări de Sănătate e a.*. In tale occasione, la CGUE si è pronunciata su una vicenda che riguardava il trasferimento di dati personali tra due amministrazioni pubbliche senza che gli interessati fossero stati adeguatamente informati. La sentenza ha chiarito il diritto degli individui ad essere informati in merito al trattamento dei propri dati personali, sottolineando l'importanza della trasparenza e del rispetto della *privacy* da rispettare anche nei trasferimenti di dati. Il secondo diritto che è stato oggetto di considerazione da parte della Corte europea dei diritti dell'uomo è stato il diritto alla rettifica dei dati personali nel caso *Ciubotaru c. Moldova* [182]. In tale caso, la Corte europea dei diritti dell'uomo ha analizzato la questione della rettifica dei dati personali alla luce dell'art. 8 della CEDU, permettendo di definire il diritto degli individui di richiedere la rettifica di informazioni personali erronee o incomplete, sottolineando l'importanza della correttezza e dell'accuratezza dei dati personali. Un ulteriore caso di estrema rilevanza è quello noto come *Google Spain*[183], che riguarda il diritto alla cancellazione dei dati personali, definito anche "diritto all'oblio". In questa importante pronuncia, la CGUE ha esaminato il diritto dei cittadini di richiedere la rimozione di informazioni personali dai motori di ricerca, stabilendo che gli individui hanno il diritto di richiedere la rimozione di informazioni

[180] ibidem, artt. 12-22.

[181] C-201/14, *Smaranda Bara e a. / Președintele Casei Naționale de Asigurări de Sănătate e a*., [2015], ECLI: ECLI:EU:C:2015:638

[182] *Ciubotaru c Moldova*, n. 27138/04, [2010]

[183] C131/12, Google Spain SL e Google Inc. c Agencia Española de Protección de Datos (AEPD), Mario Costeja González, [2014]. ECLI:EU:C:2014:317.

personali dai motori di ricerca, qualora tali informazioni siano obsolete, non pertinenti o eccessive rispetto allo scopo originario del loro trattamento. Il caso ha sottolineato l'importanza di bilanciare i diritti alla *privacy* e alla libertà di espressione nell'ambito dei dati personali online. Oltre all'operato delle Corti europee, bisogna sottolineare l'importanza delle autorità di controllo nazionali e indipendenti che contribuiscono ad una coerente applicazione del Regolamento UE 2016/679.

Il GDPR prevede al Capo VI una disciplina per tali autorità e stabilisce all'articolo 51 che[184]:

Ogni Stato membro dispone che una o più autorità pubbliche indipendenti siano incaricate di sorvegliare l'applicazione del presente regolamento al fine di tutelare i diritti e le libertà fondamentali delle persone fisiche con riguardo al trattamento e di agevolare la libera circolazione dei dati personali all'interno dell'Unione (l'«autorità di controllo»).

Ogni autorità di controllo contribuisce alla coerente applicazione del presente regolamento in tutta l'Unione. A tale scopo, le autorità di controllo cooperano tra loro e con la Commissione, conformemente al capo VII.

In particolare, negli ultimi anni l'attività del Garante è stata interessata da diversi provvedimenti in tema di IA che hanno permesso di individuare la strada da percorrere verso un'implementazione antropocentrica della normativa. Un'implementazione di questo tipo non è affatto scontata, in quanto, ad esempio nel sistema cinese citato in precedenza i diritti dell'individuo sono considerati secondari rispetto al controllo del governo cinese sulla cittadinanza[185]. Nel 2013, un primo provvedimento del Garante individuava garanzie necessarie da applicare ad un algoritmo che trattava in maniera automatizzata i dati personali contenuti nell'anagrafe tributaria o già acquisiti dall'Agenzia delle Entrate, per identificare i soggetti su cui effettuare accertamenti

[184] GDPR, artt. 51 – 59.

[185] Cfr. (n 167)

mirati[186]. Il Garante ha enfatizzato l'importanza dell'esattezza dei dati chiedendo all'Agenzia di prestare "particolare attenzione alla qualità e all'esattezza dei dati al fine di prevenire e correggere le evidenti anomalie riscontrate nella banca dati o i disallineamenti tra famiglia fiscale e anagrafica".

Negli ultimi anni, l'attività dell'autorità è stata catalizzata verso l'attuazione delle misure di contenimento della pandemia che venivano implementate attraverso l'uso di strumenti tecnologici di IA, ad esempio l'app Immuni[187] che ha potenziato il sistema di tracciamento dei contagi[188]. Inoltre, l'attività del Garante ha toccato anche l'applicazione di sistemi di IA in contesti come quelli del *delivery*[189], del riconoscimento facciale[190], dell'istruzione[191], rendendo l'autorità molto sensibile alle implementazioni di IA che possono determinare violazioni dei principi e dei diritti stabiliti all'interno del GDPR.

---

[186] Garante per la Protezione dei Dati Personali, 'Redditometro: le garanzie dell'Autorità a seguito della verifica preliminare sul trattamento di dati personali effettuato dall'Agenzia delle entrate', [2013] Doc. web n. 2765110

[187] L'App Immuni è stata un'applicazione sviluppata dalla società Bending Spoon s.p.a. su commissione del governo italiano per il tracciamento dei contatti durante la pandemia di COVID-19. Utilizzando la tecnologia Bluetooth, l'app registrava in modo anonimo gli incontri ravvicinati tra dispositivi mobili che avevano installato Immuni. Nel caso in cui un utente risultasse positivo al COVID-19, poteva condividere in modo volontario i propri dati di salute per informare in modo anonimo gli altri utenti che erano stati a contatto con lui. I dati raccolti venivano utilizzati solo per il tracciamento dei contatti e per notificare i potenziali rischi di esposizione al virus. L'obiettivo principale dell'app era contribuire al contenimento della pandemia e proteggere la salute pubblica, tutelando al contempo i diritti fondamentali dei cittadini, come il diritto alla privacy e alla protezione dei dati personali.

[188] Garante per la Protezione dei Dati Personali, 'Provvedimento di autorizzazione al trattamento dei dati personali effettuato attraverso il Sistema di allerta Covid 19- App Immuni a seguito dell'aggiornamento della valutazione di impatto effettuata dal Ministero della salute su cui l'Autorità si era espressa con provvedimento del 1° giugno 2020' [2021] Doc. web n. 9555987

[189] Garante per la Protezione dei Dati Personali, 'Ordinanza ingiunzione nei confronti di Foodinho s.r.l.' [2021], Doc. web n. 9675440

[190] Garante per la Protezione dei Dati Personali, 'Ordinanza ingiunzione nei confronti di Clearview AI' [2022], Doc. web n. 9751362

[191] Garante per la Protezione dei Dati Personali 'Ordinanza ingiunzione nei confronti di Liceo Artistico Statale di Napoli' [2020] Doc. web n. 9283029

Tuttavia, la complessità intrinseca che contraddistingue questa tecnologia e la sua continua mutevolezza ha reso più difficoltosa l'applicazione di alcuni principi come la "limitazione dello scopo" (*purpose limitation*) che giustifica la raccolta dei dati, ex art. 5 del GDPR. Si evidenzia la contrapposizione con la raccolta attuata solitamente dai sistemi di IA che è caratterizzata da uno scopo generico e vago, oppure che viene accompagnato da altre finalità non esplicitate a volte nascoste dietro finalità statistiche. Tuttavia, interpretando il principio della *purpose limitation* come la possibilità di utilizzare i dati personali per finalità che sono compatibili con lo scopo originale, allora il trattamento dei dati personali che esula dalla finalità originale rispetterà il principio fintanto che non riguardi uno scopo incompatibile con quello originale. Inoltre, l'utilizzo che gli algoritmi di *machine learning*[192] fanno dei dati è a volte paragonabile ad una finalità statistica, che secondo l'art. 89, par. 1, del GDPR "*…è soggett(a) a garanzie adeguate per i diritti e le libertà dell'interessato, in conformità del presente Regolamento. Tali garanzie assicurano che siano state predisposte misure tecniche e organizzative, in particolare al fine di garantire il rispetto del principio della minimizzazione dei dati. Tali misure possono includere la pseudonimizzazione, purché le finalità in questione possano essere conseguite in tal modo.*" La deroga viene concessa, dunque, se vengono predisposte tecniche di pseudonimizzazione o anche anonimizzazione. Tuttavia, queste tecniche non sono sempre idonee a garantire la separazione tra il dato e la persona fisica, soprattutto quando vengono impiegati potenti sistemi di IA che per capacità di analisi e potenza computazionale, oltre che per accesso a enormi quantità di dati, riescono a bypassare tali misure di sicurezza.

All'interno del Regolamento UE 2016/679, le tecniche di pseudonimizzazione e anonimizzazione sono considerate adeguate ad oscurare o interrompere il collegamento tra il dato e la persona fisica. I dati sono considerati anonimi quando non rispettano i

[192] Gli algoritmi di *machine learning* sono metodologie computazionali che permettono ai sistemi di acquisire autonomamente conoscenza a partire da un insieme di dati, senza necessità di programmazione esplicita. Sono fondamentali per l'implementazione di vari sistemi di IA, e vengono applicati in svariate aree, dalla medicina all'automazione, dal riconoscimento vocale alla personalizzazione dei servizi digitali. Esistono diversi tipi di apprendimento, tra cui supervisionato, non supervisionato e per rinforzo, ciascuno con specifiche applicazioni e requisiti. Questi algoritmi sono composti da reti neurali che simulano il funzionamento cerebrale dei viventi. Teresa Numerico, "Dobbiamo ripensare l'intelligenza artificiale" Limes (Dicembre 2022) pp. 73-80.

requisiti richiesti per rientrare nella tipologia di dato personale e, secondo il considerando 26 del GDPR, a tali dati non si applicherebbero i principi del Regolamento[193]. Quando dal dato non è possibile risalire alla persona da cui proviene, allora questo perde il carattere di 'personalità' ed esula dalle disposizioni del GDPR. Tuttavia, non viene indicata alcuna modalità per attuare l'anonimizzazione dei dati, mentre viene solo identificato il risultato da ottenere. Per le tecniche di pseudonimizzazione, l'art. 4 par. 5 del GDPR ne individua la definizione come "*il trattamento di dati personali in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative intese a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile*".

Le due tecniche sono ben diverse e, sebbene vengano spesso confuse, bisogna tenere presente come la distinzione si trovi nella possibilità di identificare nuovamente l'interessato. Se da una parte, i dati sottoposti a pseudonimizzazione rimangono dati personali, in quanto grazie ad una combinazione con ulteriori informazioni personali è possibile ricostruire il collegamento all'interessato, dall'altra parte i dati anonimizzati perdono qualsiasi collegamento con l'individuo e non si può più identificare la persona fisica a cui si riferivano. Tuttavia, non è sempre possibile operare una rottura definitiva, in particolare quando alcune caratteristiche sistemiche non lo permettono, ad esempio

[193] Il considerando 26 stabilisce che: "*È auspicabile applicare i principi di protezione dei dati a tutte le informazioni relative a una persona fisica identificata o identificabile. I dati personali sottoposti a pseudonimizzazione, i quali potrebbero essere attribuiti a una persona fisica mediante l'utilizzo di ulteriori informazioni, dovrebbero essere considerati informazioni su una persona fisica identificabile. Per stabilire l'identificabilità di una persona è opportuno considerare tutti i mezzi, come l'individuazione, di cui il titolare del trattamento o un terzo può ragionevolmente avvalersi per identificare detta persona fisica direttamente o indirettamente. Per accertare la ragionevole probabilità di utilizzo dei mezzi per identificare la persona fisica, si dovrebbe prendere in considerazione l'insieme dei fattori obiettivi, tra cui i costi e il tempo necessario per l'identificazione, tenendo conto sia delle tecnologie disponibili al momento del trattamento, sia degli sviluppi tecnologici. I principi di protezione dei dati non dovrebbero pertanto applicarsi a informazioni anonime, vale a dire informazioni che non si riferiscono a una persona fisica identificata o identificabile o a dati personali resi sufficientemente anonimi da impedire o da non consentire più l'identificazione dell'interessato. Il presente regolamento non si applica pertanto al trattamento di tali informazioni anonime, anche per finalità statistiche o di ricerca*."

una tecnica di anonimizzazione operata su un ristretto gruppo di persone, che per la scarsezza di dati risulti inefficace o ancora quando le caratteristiche dei dati sono così differenti che rimane possibile collegare i dati agli individui. Se i dati personali vengono pseudonimizzati in modo inefficace, una persona che intende ricollegarli ad un individuo specifico potrebbe riuscirci se possiede altri dati che, combinati insieme, permettono di ricostruire l'identità esatta. La questione della sicurezza di tali procedure è ancora discussa in dottrina e sempre più frequentemente tali tecniche vengono superate da attacchi molto sofisticati. Le misure di sicurezza dovrebbero anticipare le tecniche di attacco future e conoscere quelle passate, in questo contesto l'IA potrebbe essere utilizzata per sviluppare metodi più sofisticati per l'anonimizzazione e la pseudonimizzazione, richiedendo comunque un costante aggiornamento per far fronte alle nuove tecniche di attacco[194].

Si possono individuare ben tre tecniche per superare le tali misure di sicurezza e ricostruire il collegamento tra il dato e la persona fisica: la combinazione di set di dati, l'inversione dello pseudonimo e l'identificazione insufficiente[195]. La tecnica più accessibile ai sistemi di IA è la combinazione di *dataset*, in quanto questi sistemi sono in grado di analizzare grandi quantità di dati e potendo accedere a due o più *dataset* anonimizzati, che potenzialmente contengono dati su uno stesso individuo, tali dati possono poi essere combinati per superare l'anonimizzazione[196]. L'inversione dello

[194] Rainer Mühlhoff, 'Predictive Privacy: Collective Data Protection in the Context of Artificial Intelligence and Big Data' (2023) 10 Big Data & Society.

[195] Boris Lubrasky, 'Re-Identification of "Anonymized Data"' Georgetown Law Technology Review, 202 [2016]

[196] La principale preoccupazione riguarda i dati sanitari e in generale i dati sensibili ex art. 9, par. 1, che, sebbene vengano spesso anonimizzati per permettere alla ricerca scientifica di farne uso, con tecniche di identificazione che sfruttano l'IA, si riesce ad accedere a tali dati ripristinando il collegamento con la persona fisica, causando una grave violazione della tutela dei dati personali. Un esperimento del genere è stato condotto dal Dr. Latanya Sweeney della Carnegie Mellon University, che attraverso la combinazione dei dati del censimento del 1990 negli USA e quelli di un ospedale, sebbene resi anonimi (ma comunque pubblici), riuscì a dimostrare la possibilità di identificare una persona con poche caratteristiche. Latanya Sweeney, 'Simple Demographics Often Identify People Uniquely' [2000] Carnegie Mellon University, *Data Privacy Working Paper 3*.

pseudonimo è una tecnica che sfrutta una debolezza intrinseca dello pseudonimo, ossia quando lo stesso venga progettato per essere invertito, per mezzo di una chiave di lettura, svelando l'identità della persona fisica a cui appartiene il dato. Nel momento in cui tale chiave venga divulgata oppure non venga cambiata di frequente, crescerà la possibilità di identificazione del dato personale. L'ultima tecnica descrive proprio lo scenario in cui è stata operata una de-identificazione, ma alcuni dati del soggetto sono rimasti pubblici permettendo così la raccolta di altri dati per ottenere l'individuazione del soggetto interessato. Queste sono solamente alcune delle tecniche utilizzate per il superamento degli pseudonimi, resa sempre più semplice anche grazie alla moltitudine di dati pubblicati sui social network e solitamente divulgati liberamente [197]. La pervasività di tali metodi, se condotta da sistemi di IA che automaticamente sanno cosa e dove cercare, avendo a disposizione potenze computazionali crescenti, porta a immaginare una raccolta di dati e una identificazione che si rivolga e includa anche i dati anonimizzati. I sistemi di IA potendo gestire e analizzare enormi quantità di dati sono i perfetti esecutori di una strategia di re-identificazione delle persone con l'attuazione delle tecniche appena elencate.

I sistemi di IA richiedono grandi quantità di dati e limitarsi a quelli necessari per la stretta finalità, come richiesto dal principio di minimizzazione dei dati ex art. 5, para. 1, lett. c) del GDPR, potrebbe sembrare per gli sviluppatori un freno ingiustificato. L'IA entra in contrasto con tale principio, in quanto, questo prevede che il titolare del trattamento debba limitare la raccolta dei dati a quelli "adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati"[198], ma come spiega l'autorità nazionale per la protezione dei dati del Regno Unito, l'*Information Commissioner's Office* (ICO), il principio non pone un'alternativa tra il 'non trattare dati

[197] I social network sono piattaforme online che consentono agli utenti di connettersi, comunicare e condividere contenuti con altre persone in tutto il mondo. Attraverso l'uso di profili personali, gli utenti possono pubblicare post, foto, video e interagire con amici, familiari o sconosciuti tramite messaggi, commenti e reazioni. I social network offrono un ambiente virtuale per la creazione di reti sociali, la condivisione di interessi comuni e l'esplorazione di nuove connessioni e contenuti.

[198] GDPR, art. 5, par. 1, lett. c)

personali' e lo ‘scegliere di trattare più dati, violando la legge’[199]. La sintesi della liceità normativa si troverebbe tra i due approcci, limitando il trattamento esclusivamente ai dati necessari per la finalità preposta, ed evitando, in tal modo, la violazione del principio di minimizzazione dei dati. Per poter rispettare il principio di minimizzazione, i dati che i sistemi di IA possono analizzare, dovranno avere i requisiti di adeguatezza, rilevanza ed essere limitati a quanto necessario. L'adeguatezza può essere intesa come la sufficienza a soddisfare propriamente la finalità indicata, la rilevanza sarà presente se esiste una connessione logica tra il dato raccolto e la finalità predefinita, mentre la *necessarietà* sarà rispettata se il titolare del trattamento cesserà di avere la disponibilità di dati superflui per lo scopo del trattamento. Inoltre, l’ICO chiarisce che il fatto che alcuni dati potrebbero successivamente rivelarsi utili per effettuare previsioni non rappresenta un motivo sufficiente per giustificare la necessità di conservazione dei dati non necessari per la finalità dichiarata e, allo stesso modo, tale motivazione non è idonea a giustificare retroattivamente la raccolta, l’utilizzo o la conservazione degli stessi. Tra i principali rischi che possono derivare da una raccolta così estesa di dati risiedono: la condivisione involontaria di dati personali con il modello di IA, la perdita di dati per mezzo degli output del sistema, gli attacchi avversi (per stimolare un comportamento dannoso del sistema), l’estrazione del modello (per far riprodurre una copia del modello senza avere l’accesso ai dati di addestramento) e l’avvelenamento dei dati (inserendo dati dannosi per influenzare il comportamento futuro del sistema di IA)[200].

Sfugge dall’analisi il funzionamento di tali tecnologie che si fonda sulla raccolta e l’analisi di massicce quantità di dati. Minimizzare e quindi limitare la presenza dei dati a quelli necessari per lo scopo è un principio che può essere applicato esclusivamente ad una categoria di sistemi di IA, quelli che hanno uno scopo prefissato. Tale tipologia di

---

[199] Information Commissioner’s Office, ‘How should we assess security and data minimisation in AI?’ <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-should-we-assess-security-and-data-minimisation-in-ai/ > ultimo accesso 20 maggio 2023

[200] Glorin Sebastian, ‘Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information’ (May 2023). Disponibile al seguente link: <https://ssrn.com/abstract=4454761>

sistemi è però destinata ad essere per la maggior parte sostituita dai sistemi che invece sono addestrati senza un fine specifico, ma che riescono ad adattarsi a diversi utilizzi e attività. Per questa seconda tipologia, che verrà analizzata meglio nel terzo capitolo, il principio della minimizzazione dei dati si pone come una restrizione ingiustificata e inadatta per lo sviluppo e il miglioramento. Nel principio di minimizzazione dei dati risiede quella che potrebbe sembrare una limitazione per i moderni sistemi di IA. In generale, quando un sistema di IA viene addestrato, quanti più dati impara, tanto più statisticamente accurato sarà. Ad esempio, un modello addestrato per prevedere gli acquisti dei consumatori, sulla base dei loro acquisti precedenti, tende ad essere più accurato dal punto di vista statistico quanti più clienti sono inclusi nei dati di addestramento. Inoltre, nuove caratteristiche aggiunte a un *dataset* esistente possono essere rilevanti per la previsione del modello. Ad esempio, le cronologie degli acquisti integrate con dati demografici aggiuntivi potrebbero migliorare ulteriormente l'accuratezza statistica del modello. Tuttavia, in generale, più dati vengono raccolti su ciascuna persona e più persone sono incluse nel *dataset*, maggiori sono i rischi per tali individui, anche se i dati vengono raccolti per uno scopo specifico. Il principio di minimizzazione dei dati richiede di non utilizzare più dati di quelli necessari per la finalità del sistema. Pertanto, se è possibile ottenere un'accuratezza sufficiente con un minor numero di dati personali o con un numero inferiore di individui inclusi, si dovrebbe agire in tal senso.

In questo senso, il Garante ha stabilito come nelle operazioni che coinvolgono il supporto dei sistemi di IA sia necessaria la presenza di misure di sicurezza specifiche atte a minimizzare il rischio di errori dei sistemi, in particolare quelli utilizzati dalle piattaforme nel settore della *delivery*. Nei provvedimenti verso le piattaforme Deliveroo e Foodinho, il Garante ha condannato le stesse per la mancata verifica periodica della correttezza delle scelte poste in essere dal sistema di IA che governa le piattaforme, al fine di evitare discriminazioni nei confronti dei lavoratori. Questa tecnologia gestisce in maniera autonoma il tracciamento geografico dei *riders*, ne effettua la profilazione e gestisce anche le consegne e le valutazioni delle prestazioni dei lavoratori. Ai *riders* spesso risulta poco trasparente il metro di misura dell'algoritmo e non hanno possibilità di conoscere quali comportamenti correggere, rimanendo in ogni caso all'oscuro dei processi decisionali della macchina. Con l'ordinanza di ingiunzione nei confronti di Foodinho il Garante ha accertato che la società in quanto titolare del trattamento aveva

trattato i dati personali di 18.686 riders in violazione sia del GDPR sia del decreto legislativo n. 196/2003[201]. L'elaborazione dei dati personali effettuata dalla società è risultata essere in violazione degli articoli 5, paragrafo 1, lettere a), c) e e), (principio di liceità, correttezza, limitazione della conservazione); 13 (informativa); 22, paragrafo 3 (strumenti idonei per il trattamento automatizzato dei dati); 25 (protezione dei dati mediante progettazione e impostazione predefinita); 30, paragrafo 1, lettere a), b), c), f) e g); 32 (misure preventive); 35 (valutazione dell'impatto); 37, paragrafo 7 (comunicazione all'autorità di controllo del responsabile della protezione dei dati); 88 (protezione dei dati durante il rapporto di lavoro) del GDPR; articolo 114 (garanzie in materia di controllo a distanza) del decreto legislativo italiano n. 196/2003. L'Autorità ha ordinato alla società di conformarsi alle disposizioni sulla protezione dei dati e ha inflitto a Foodinho S.r.l. una sanzione di € 2.600.000,00 per il trattamento illecito dei dati. Nella stessa direzione si colloca l'ordinanza di ingiunzione relativa al trattamento dei dati posto in essere da Deliveroo s.r.l. nei confronti di circa 8.000 *riders*[202]. Il Garante avrebbe individuato le presunte violazioni con riferimento agli artt. 5, par. 1, lett. a), c) ed e) (principi di liceità, correttezza e trasparenza, principio di minimizzazione e principio di limitazione della conservazione); 13 (informativa); 22 (processo decisionale automatizzato compresa la profilazione); 25 (*privacy by design* e *by default*); 30 (registro delle attività di trattamento); 32 (sicurezza del trattamento); 35 (valutazione di impatto sulla protezione dei dati); 37 (responsabile della protezione dei dati); 88 (disposizioni più specifiche a livello nazionale) del Regolamento; art. 114 (garanzie in materia di controllo a distanza) del GDPR.

Le implicazioni nell'uso di questi sistemi di monitoraggio e valutazione sui *riders* non si limitano solo alla protezione dei dati, ma riguardano anche aspetti giuslavoristici. Il legislatore ha infatti pubblicato il D.Lgs. 27 giugno 2022, n. 104 (di seguito, Decreto Trasparenza) in attuazione della Direttiva UE n. 2019/1152, con cui sono stati introdotti una serie di nuovi obblighi per il datore di lavoro ed i committenti pubblici e privati in

[201] Garante per la Protezione dei Dati Personali, 'Ordinanza ingiunzione nei confronti di Foodinho s.r.l. del 10 giugno 2021' n. 234 del 2021, [doc. web n. 9675440]

[202] Garante per la Protezione dei Dati Personali, 'Ordinanza ingiunzione nei confronti di Deliveroo Italy s.r.l. del 22 luglio 2021' n. 285 del 2021, [doc. web n. 9685994]

risposta al diritto del lavoratore ad essere informato circa gli elementi essenziali del rapporto di lavoro e le condizioni di lavoro[203]. Non mancano punti di contatto con le prescrizioni già presenti in materia di protezione dei dati e con quelle dell'art. 4 dello Statuto dei lavoratori[204], in quanto l'articolo 4 del Decreto Trasparenza in esame introduce l'articolo 1-bis al D.Lgs. n. 152 del 1997 (di seguito, "Decreto") in forza del quale "*il datore di lavoro o il committente pubblico e privato è tenuto a informare il lavoratore dell'utilizzo di sistemi decisionali o di monitoraggio automatizzati deputati a fornire indicazioni rilevanti ai fini della assunzione o del conferimento dell'incarico, della gestione o della cessazione del rapporto di lavoro, dell'assegnazione di compiti o mansioni nonché indicazioni incidenti sulla sorveglianza, la valutazione, le prestazioni e l'adempimento delle obbligazioni contrattuali dei lavoratori*"[205]. Risulta chiaro che i sistemi decisionali o di monitoraggio automatizzati, anche con intervento umano, che abbiano ad oggetto i lavoratori rientrano nel perimetro della norma. Dunque, il lavoratore dovrà conoscere sin dal momento della stipula del contratto di lavoro se sul luogo di lavoro sia presente un sistema di questo tipo, se questo venga utilizzato, come venga utilizzato e quale scopo si vuole raggiungere con l'uso di tale sistema. Oltre a questi aspetti è interessante notare come il legislatore abbia previsto l'obbligo di chiarire la natura dei dati su cui è stato addestrato il sistema automatizzato, rendendo in un certo modo obbligatoria la trasparenza sul dataset utilizzato. Inoltre, il lavoratore dovrà essere informato anche sulle misure di controllo e sorveglianza attuate per assicurare che il sistema automatizzato non realizzi discriminazioni verso i lavoratori. Con il Decreto Trasparenza il legislatore ha cercato di aggiornare le salvaguardie per i lavoratori che si sono trovati di fronte ai nuovi mezzi di monitoraggio e controllo delle loro prestazioni, senza poter vantare diritti esplicitamente individuati dalla legislazione.

L'aggiornamento normativo in materia giuslavoristica fornisce un buon esempio per capire come poter regolare al meglio i sistemi di IA, senza interromperne l'utilizzo e lo sviluppo. Dai principi individuati nel Decreto Trasparenza emerge come si cerchi di

[203] D.Lgs. 27 giugno 2022, n. 104; Direttiva UE n. 2019/1152

[204] Legge 300/1970

[205] D.Lgs. n. 152 del 1997, art. 1-bis

migliorare la trasparenza nei dataset utilizzati dai sistemi di IA richiedendo una descrizione della natura dei dati utilizzati per l'addestramento e la programmazione. Dalla trasparenza dei dataset si passa ad un altro punto di importanza fondamentale per lo sviluppo controllato dei sistemi, cioè la chiara individuazione dei metodi di controllo e la possibilità di modificare le decisioni automatizzate qualora risultino discriminatorie. Emerge come, adottando queste prescrizioni, il controllo sui sistemi di IA risulterà migliore e più rigido, determinando al contempo una diminuzione dei rischi connessi, in primis quello delle decisioni discriminatorie.

D'altro canto, l'IA potrebbe essere utilizzata in maniera benefica nello sviluppo di metodi per migliorare la sicurezza e la *privacy* dei dati. Ad esempio, potrebbe servire a sviluppare misure più sofisticate per anonimizzare o pseudonimizzare i dati personali attraverso le tecniche di apprendimento automatico. Questi sistemi sarebbero capaci di individuare le caratteristiche dei dati che sono più sensibili ad essere de-anonimizzati e sviluppare tecniche di anonimizzazione che tengono conto di questi attributi. Un altro utilizzo potrebbe essere rappresentato nel monitoraggio costante dei nuovi metodi di attacco per aggiornare le tecniche di anonimizzazione di fronte alle nuove minacce. Ancora l'IA potrebbe essere impiegata per identificare i rischi di violazione della *privacy* in modo più efficiente e accurato. Proprio perché l'IA può anche essere utilizzata per scopi malevoli come la sorveglianza di massa e la discriminazione[206], bisogna comunque prevedere un utilizzo che deve essere accompagnato dalla sorveglianza umana dedicata ad un controllo del funzionamento del sistema e soprattutto da una rigorosa valutazione dei rischi e delle implicazioni etiche che possono derivarne.

L'IA ha il potenziale di migliorare diversi settori e già in alcuni campi professionali viene utilizzata per una maggiore efficienza del sistema, ma come abbiamo avuto modo di considerare nell'analisi del principio di minimizzazione dei dati, alcuni utilizzi pongono dei seri rischi per la protezione dei dati personali. Nel settore assicurativo, ad esempio, l'IA potrebbe essere utilizzata per il miglioramento della valutazione dei rischi, la gestione dei sinistri e la personalizzazione dei prodotti assicurativi. Tali società

[206] Mühlhoff (n 180).

utilizzerebbero sistemi di IA presenti, ad esempio, in dispositivi tecnologici indossabili come gli *smartwatch* (orologi digitali caratterizzati da moltissimi sensori per monitorare parametri che vanno dalla localizzazione al ritmo cardiaco), per raccogliere e analizzare enormi quantità di dati sui potenziali clienti, inclusi dati sensibili come la posizione geografica, il comportamento di acquisto e le informazioni sulla salute. Sebbene queste tecniche migliorino sensibilmente l'efficienza e anche la sicurezza assicurativa, una raccolta così vasta di dati non sembra rispettare il principio della minimizzazione dei dati. Con l'uso di dispositivi che monitorano lo stato di salute, la compagnia assicurativa tratterebbe dati sensibili ex art. 9 del GDPR, che prima istanza sono coperti da un divieto di trattamento, ma che in via esclusiva possono essere trattati su una base giuridica tra quelle indicate al paragrafo 2. Tale monitoraggio verrebbe condotto da sistemi di IA che, senza la necessità di supervisori umani, sarebbero capaci di fornire un risultato su quanto sia conveniente stipulare un contratto di assicurazione o terminarlo. Se da un lato possono derivarne benefici per i clienti che conducono uno stile di vita salutare o mantengono una guida sicura, dall'altro con una deviazione negativa dei parametri (che potrebbe pure consistere in una pausa di una settimana dalla palestra) potrebbe derivarne un aumento del premio assicurativo[207]. Si pensi ad esempio a due soggetti che vivono nella stessa città, il primo in un quartiere in cui le strade sono strette e molto trafficate, mentre l'altro in una zona in cui le strade sono larghe e il traffico è meno presente. Nel caso in questione, due individui con reddito, situazione lavorativa e familiare simili potrebbero finire per pagare premi assicurativi diversi basati unicamente sulla loro posizione geografica. Questo potrebbe essere visto come discriminatorio perché un individuo potrebbe finire per pagare un prezzo maggiore solo per il fatto di vive in una zona con strade strette e trafficate, nonostante non abbia alcun controllo su questi aspetti del suo ambiente. Questi esempi evidenziano come l'utilizzo dell'IA nel settore assicurativo, ma potrebbe ripetersi lo stesso ragionamento anche per altri settori, se non adeguatamente regolato e controllato, possa portare a problemi di discriminazione, violazione della *privacy* e mancanza di trasparenza.

[207] Toby Walsh, "*Will AI end privacy? How do we avoid an Orwellian future*", AI & Society (2023) 38, p. 1239–1240.

3. Segue: l'articolo 22 del GDPR applicato ai sistemi di IA

Continuando la disamina delle disposizioni del GDPR che producono effetti sui sistemi di IA, la disposizione che più riguarda tali sistemi è identificabile nell'articolo 22 del GDPR[208]. La norma stabilisce al primo paragrafo "*il diritto* [per il soggetto passivo del trattamento] *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*"[209]. La lettera dell'articolo ha visto nel tempo scarsa applicazione; tuttavia, sembra essere tornata al centro dell'attenzione con i recenti sviluppi delle tecnologie di IA. Infatti, l'art. 22 mira proprio a regolare gli effetti che sono prodotti da trattamenti condotti con sistemi di IA e che non prevedono l'intervento umano. Sebbene al paragrafo 2 siano presenti delle eccezioni per le quali un trattamento automatizzato viene consentito, la norma punta sia ad evitare che tali trattamenti risultino invasivi, sia a prevedere la presenza di un *human-in –the-loop*[210].

L'obiettivo di mantenere l'uomo all'interno del processo decisionale è attuato attraverso il terzo paragrafo dell'articolo 22 che riconosce al soggetto interessato il diritto di pretendere l'intervento umano sui processi posti in essere dall'IA. Viene a formarsi una tutela per gli interessati al trattamento che ritengono di essere stati danneggiati o di aver subito un pregiudizio dalla decisione, risultante dal processo automatizzato e che tale processo sia rivisto da un umano. Da tale disposizione derivano alcuni corollari legati alla decisione automatizzata: tra questi il più importante è il diritto ad ottenere la

---

[208] Giovanni Maria Riccio, Paola Scorza and Ernesto Belisario, *GDPR e normativa privacy: commentario* (2., Wolters Kluwer 2022) 280–286 <https://go.exlibris.link/KVR5P249>.

[209] GDPR, art. 22, para. 1

[210] L'Human-in-the-Loop (HITL) descrive il processo mediante il quale un sistema di calcolo o macchina, che si trova nell'impossibilità di risolvere un problema autonomamente, necessita di un intervento umano. Questo coinvolgimento umano si situa sia nelle fasi di addestramento che di test della creazione di un algoritmo, generando un ciclo continuo di feedback che permette all'algoritmo di produrre risultati sempre più precisi ed efficaci. Vikram Singh Bisen, 'What Is Human in the Loop Machine Learning: Why & How Used in AI?' (*VSINGHBISEN*, 16 August 2022) <https://medium.com/vsinghbisen/what-is-human-in-the-loop-machine-learning-why-how-used-in-ai-60c7b44eb2c0> accessed 20 May 2023.

spiegazione della decisione dell'IA[211], ma anche il divieto di discriminazione e il diritto di ottenere l'intervento dell'uomo. In particolare, il pericolo di discriminazione in questi sistemi è inversamente proporzionale alla qualità dei dati su cui viene "addestrato" il sistema, ad una qualità e varietà maggiore di dati corrisponde un pericolo inferiore di discriminazione. Il processo di addestramento avviene attraverso un approccio chiamato apprendimento automatico[212]. Nell'iter del processo di addestramento, si attuano varie fasi distintive. L'inizio è contraddistinto dalla raccolta dei dati, una fase cruciale in cui si procede all'acquisizione e preparazione di dati di addestramento, rappresentativi della problematica da risolvere. Queste informazioni possono manifestarsi sotto diverse forme, tra cui testo, immagini, contenuti audio, video o altre tipologie di dati pertinenti. In seguito, si progetta l'architettura del modello di IA, il quale deve essere in grado di rappresentare adeguatamente il problema e apprendere dai dati di addestramento. Successivamente, l'addestramento del modello viene effettuato esponendolo ai dati raccolti e manipolando i suoi parametri interni per affinare le performance. Questa fase di calibrazione avviene mediante un ciclo di iterazioni ripetute, durante le quali il modello viene costantemente confrontato con i dati e successivamente regolato per minimizzare l'errore. Una volta conclusa la fase iniziale di addestramento, si procede alla valutazione del modello, sfruttando un insieme di dati distinto denominato set di convalida. Quest'ultimo permette di valutare l'efficienza del modello su dati inediti e individuare possibili aree di intervento per il miglioramento. Al raggiungimento di risultati soddisfacenti sul set di convalida, il modello viene infine sottoposto a un test su un ulteriore insieme di dati, detto set di test, per verificarne l'efficacia in condizioni più simili a quelle reali. Il procedimento di addestramento richiede un impegno considerevole in termini di tempo, risorse computazionali e competenze specifiche sugli algoritmi di apprendimento automatico. Esso si configura come un processo iterativo, durante il quale il modello è costantemente affinato ed ottimizzato per il raggiungimento di risultati ottimali.

---

[211] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76.

[212] *Cfr*. (n 108)

Per chiarire come i sistemi di IA possano produrre un pregiudizio nei confronti degli individui attraverso un processo decisionale automatizzato ex art. 22 GDPR, prendiamo in considerazione i sistemi di attribuzione del merito creditizio (*credit scoring system*). In Germania esiste un sistema di questo tipo che assegna un punteggio agli individui che chiedono un prestito o altri servizi finanziari a banche o istituti di credito. Il sistema è di proprietà di una società privata di nome Schufa Holding AG (di seguito, Schufa), che fornisce i suoi servizi a quasi la totalità degli istituti di credito e assicurativi tedeschi. Attraverso questo sistema viene calcolato il merito creditizio sulla base di una profilazione dell'interessato, fornendo un punteggio che si limita a mostrare l'esito positivo o negativo dell'esame, senza provvedere alcuna motivazione sottostante la valutazione.

Nel 2018, un individuo richiese un prestito bancario, ma la sua richiesta venne respinta a causa del punteggio negativo assegnatogli da Schufa. Il 18 ottobre 2018, il soggetto interessato presentava ricorso dinanzi al *Hessischer Beauftragter für Datenschutz und Informationsfreiheit* (Commissario dell'Assia per la protezione dei dati e la libertà d'informazione), il quale però con una decisione amministrativa del 3 giugno 2020 rifiutava di intraprendere alcuna azione legale contro Schufa. Successivamente, l'individuo decideva di fare appello presso la Corte Amministrativa di Wiesbaden[213] la quale, sospendendo il processo nell'ottobre 2021, presentava due questioni interpretative sull'applicazione dell'articolo 22 del GDPR alla CGUE[214]. Per la prima volta dall'entrata in vigore del GDPR, il diritto a non essere sottoposti a decisioni automatizzate arrivava davanti alla CGUE. Secondo Andreas Häuselmann[215] il caso ha assunto una rilevanza

[213] Verwaltungsgericht Wiesbaden (Administrative Court, Wiesbaden), 6 K 788/20.WI

[214] Causa C-634/21, *OQ contro Land Hessen*, con l'intervento di SCHUFA Holding AG, procedura in corso. Al momento sono state pubblicate solo le conslusioni dell'AG raggiungibili al seguente indirizzo < https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:62021CC0634&from=EN> ultimo accesso il 20 maggio 2023.

[215] Andreas Nicolas Häuselmann è un dottorando esterno presso eLaw, *Center for Law and Digital Technologies* dell'Università di Leiden, da ottobre 2018. È consulente legale senior per la *privacy* e la sicurezza informatica presso De Brauw Blackstone Westbroek N.V. con sede ad Amsterdam. Ha conseguito un Master of Laws (LL.M.) in Information and Communication Technology Law presso l'Università di Oslo e un Master of Laws (LL.M.) in Information Technology and Intellectual Property Law presso la Leibniz University di Hannover.

che va oltre l'applicazione del GDPR e guarda alla prossima era degli algoritmi che, sempre più frequentemente, prendono decisioni in maniera automatizzata[216].

Mentre l'art. 22, par. 1, prevede il generico diritto a non essere sottoposti a decisioni automatizzate, il secondo paragrafo stabilisce una serie di eccezioni e apre la possibilità per gli Stati membri di prevederne ulteriori. In caso di trattamento rientrante tra le eccezioni, il titolare del trattamento deve predisporre le misure di sicurezza adeguate alla tutela dei dati personali, come ad esempio il diritto ad un intervento umano o il diritto di esprimere il proprio parere e contestare la decisione. Il punteggio di merito creditizio, in questo caso, veniva calcolato sulla profilazione e valutazione del soggetto per predirne la situazione e affidabilità economica. Quindi, il risultato veniva trasmesso e l'istituto di credito rifiutava di concedere il prestito, basandosi esclusivamente sul punteggio Schufa che comunque non conteneva spiegazioni relative a tale risultato.

Col termine *"multi-stage profiling"* si indica la situazione in cui non è l'ente che emette la decisione finale (in questo caso la banca) ad adottare tecniche di profilazione, ma piuttosto questo compito viene delegato ad un terzo, nel nostro caso Schufa, che provvede alla profilazione e alla decisione automatizzata[217]. Il processo decisionale è diviso in più stadi e dato che il punteggio Schufa si basa sulla profilazione e la decisione bancaria si basa su tale punteggio, anche la decisione finale è basata sulla profilazione.

Di fronte alla CGUE, nel Caso C-634/21 *OQ contro Land Hessen* la Corte di Wiesbaden poneva due questioni. In primis, se il punteggio di merito creditizio rientrasse nella definizione di decisione automatica, come intesa ex art. 22 GDPR e, come seconda questione, eventualmente come andrebbe tutelata la situazione del soggetto interessato di fronte alla profilazione attuata.

---

[216] Andreas Häuselmann, 'The ECJ's First Landmark Case on Automated Decision-Making – a Report from the Oral Hearing before the First Chamber' (*European Law Blog*, 20 February 2023) <https://europeanlawblog.eu/2023/02/20/the-ecjs-first-landmark-case-on-automated-decision-making-a-report-from-the-oral-hearing-before-the-first-chamber/> accessed 30 April 2023.

[217] Reuben Binns and Michael Veale, 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR' (2021) 11 International Data Privacy Law 319, 321.

Secondo la Corte di Wiesbaden, il trattamento di dati in questione rientrerebbe nella definizione di decisione automatizzata e nella questione pregiudiziale essa sostiene che il punteggio di merito creditizio non sarebbe solamente una tecnica di profilazione, ma anche una decisione automatizzata che non ha alcuna interferenza umana. Inoltre, viene sottolineato come il fatto che la decisione viene presa da una parte terza, al di fuori del contratto tra banca e soggetto interessato, non esclude l'applicazione dell'art. 22. La Corte basa il suo ragionamento sul fatto che la scelta finale della banca sia influenzata in maniera decisiva dal punteggio ottenuto, infatti, un punteggio Schufa basso determinerebbe, in quasi la totalità dei casi, un rifiuto della richiesta di prestito. Secondo quanto riportato da Häuselmann, durante la discussione di fronte alla CGUE, i rappresentanti di Schufa avrebbero affermato come solamente il 20% degli individui che ottengono un punteggio basso o negativo ricevono comunque il prestito[218]. Questo significa che in ogni caso il restante 80% dei soggetti con un tale punteggio non ottiene il prestito. Per la Corte tedesca non è importante chi tra Schufa e la banca prenda la decisione definitiva e nemmeno il fatto che sono coinvolte delle persone nella decisione (come eventuali impiegati della banca), proprio perché è il punteggio ottenuto che decide effettivamente.

Una lettura restrittiva dell'art. 22 creerebbe un vuoto normativo in materia, facendo sorgere seri rischi per la possibilità di tutelare i soggetti interessati e i loro diritti di protezione dei dati. Se il GDPR disciplina diritti di applicazione generica relativi ai diritti di informazione e accesso ai dati, questo stabilisce anche per le decisioni automatizzate uno specifico diritto di conoscere le informazioni tenute in considerazione, la logica seguita nella valutazione e le conseguenze che possono derivarne. Da un lato, grazie al segreto industriale, Schufa non è obbligata a fornire tali informazioni e dall'altro la banca non conosce tali informazioni perché non ha cognizione di come sia condotto l'esame e come venga assegnato il punteggio. Anche esercitando il diritto di accesso ai dati personali ex art. 15 GDPR, il soggetto danneggiato non è riuscito ad ottenere una chiara spiegazione, ma piuttosto solo cenni

[218] Andreas Häuselmann (n 197).

generici sul funzionamento del sistema [219]. Il vuoto normativo che si formerebbe determina l'impossibilità di tutelare il soggetto interessato di fronte a questi trattamenti, sempre che tale trattamento non rientri nella definizione dell'art. 22 del GDPR.

Il 16 marzo 2023 l'Avvocato Generale Priit Pikamäe (da qui in poi, AG), nel medesimo caso, ha rilasciato le sue conclusioni[220]. Al punto c, n.1, lett. A) chiarisce come, sebbene l'art. 22 preveda il "*diritto dell'interessato di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato*", il GDPR non chiede all'interessato di esercitare il diritto in maniera attiva. Un'interpretazione che tenga presente il considerando 71 e il paragrafo 2 dell'articolo 22 del GDPR porta a concludere che la finalità della disposizione sia quella di stabilire un divieto generale sulle decisioni che hanno tali caratteristiche. Viene anche evidenziato come il divieto riguardi solo trattamenti molto specifici che portino a decisioni produttrici di effetti giuridici sull'interessato o che incidono significativamente sulla persona. L'AG suggerisce un'interpretazione dell'art. 22 che considera il calcolo automatizzato della probabilità di un soggetto di essere in grado di ripagare un debito in futuro come una decisione basata esclusivamente su un trattamento di dati automatizzato, anche attraverso la profilazione, e che produce effetti giuridici sull'interessato. Proprio perché il punteggio risultante viene calcolato sui suoi dati personali e, dopo essere stato trasmesso dal titolare del trattamento ad una terza parte, secondo una pratica stabilita, la banca fonda la sua decisione sulla conclusione o modifica o cessazione di un contratto principalmente su tale punteggio. Per quanto riguarda gli effetti legali richiesti dalla disposizione, il rifiuto del credito comporta che l'individuo non possa beneficiare dalla relazione contrattuale con l'istituto finanziario e di conseguenza anche la sua situazione finanziaria risulta influenzata. Al paragrafo 38 della sua opinione, l'AG interpreta il termine "decisione"

[219] Wachter, Mittelstadt and Floridi (n 192).

[220] Causa C-634/21, OQ contro Land Hessen, con l'intervento di SCHUFA Holding AG, procedura in corso. Al momento sono state pubblicate solo le conslusioni dell'AG raggiungibili al seguente indirizzo < https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:62021CC0634&from=EN> ultimo accesso il 20 maggio 2023.

come un atto che ha conseguenze non solo giuridiche, ma anche economiche e sociali [221].

Il punteggio di merito creditizio può essere considerato una decisione basata su un trattamento automatizzato e di profilazione quando predetermina la decisione dell'istituto di credito riguardo la concessione o meno del prestito e quando esiste un margine di discrezione umana per verificare il risultato e la correttezza della decisione nei confronti della persona che richiede il credito. Il fattore decisivo riguarda l'impatto che la decisione ha sull'interessato, considerando come un punteggio negativo potrebbe avere conseguenze dannose come una restrizione nell'esercizio delle sue libertà economiche. Il risultato del sistema Schufa va considerato come una decisione che ha effetti giuridici sull'interessato nel momento in cui l'ente creditizio attribuisce al punteggio un ruolo fondamentale all'interno del processo decisionale [222]. È necessario anche evidenziare come l'articolo 22 del GDPR si rivolga a processi esclusivamente automatizzati e dunque richiede che l'elaborazione difetti di qualsiasi intervento umano. La presenza di una persona all'interno della elaborazione potrebbe portare a escludere l'applicabilità dell'articolo.

In conclusione, l'AG propone di qualificare la valutazione Schufa come una decisione che rientra nel portato dell'articolo 22 e al paragrafo 47 delle sue conclusioni stabilisce che:

"Fatta salva la valutazione dei fatti che compete ai giudici nazionali compiere in ciascun caso particolare, le considerazioni svolte supra mi sembrano indicare che il punteggio di scoring calcolato da un'agenzia di valutazione del credito e comunicato a un istituto finanziario tende generalmente a predeterminare la decisione di quest'ultimo quanto alla concessione o al diniego del credito all'interessato, cosicché si deve ritenere che detta presa di posizione rivesta un carattere puramente formale nel quadro

[221] Ibid. para. 38

[222] Ibid. para. 43

del processo. Ne consegue che occorre riconoscere al punteggio di scoring stesso la natura di «decisione» ai sensi dell'articolo 22, paragrafo 1, del RGPD." [223]

Tutto ciò alla luce degli obiettivi perseguiti dal legislatore europeo nella protezione dei dati dell'interessato. L'AG conclude chiarendo che l'articolo 22 del GDPR debba essere interpretato nel senso che[224]:

"il calcolo automatizzato di un tasso di probabilità relativo alla capacità di un interessato di saldare in futuro un debito costituisce già una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produce effetti giuridici che riguardano l'interessato o che incide in modo analogo significativamente sulla sua persona, qualora tale tasso, calcolato sulla base di dati personali relativi all'interessato, sia trasmesso dal titolare del trattamento a un terzo titolare del trattamento e, conformemente a una prassi costante, quest'ultimo basi prevalentemente su tale tasso la sua decisione sulla stipulazione, sull'attuazione o sulla cessazione di un contratto con l'interessato."

L'opinione dell'AG pone una prima interpretazione su diverse questioni dottrinali. Prima tra tutte, la considerazione ex art. 22 GDPR del "*diritto a non essere sottoposti a decisioni basate esclusivamente su trattamenti automatizzati o profilazione*" come un divieto generale e non come un diritto da esercitare attivamente. Inoltre, viene anche chiarita la natura ex art. 15, par. 1, lett. h, del GDPR del diritto dell'interessato ad ottenere l'accesso alle informazioni trattate, con il corrispettivo dovere di fornire informazioni sulla logica seguita e le conseguenze previste dal trattamento automatizzato dei dati personali. Nel caso di specie tali informazioni includerebbero la rivelazione del metodo di calcolo del punteggio, a meno che non esistano interessi confliggenti meritevoli di protezione relativi a segreti industriali o di proprietà intellettuale. L'AG conclude sottolineando come tale obbligo possa esaurirsi in una spiegazione del metodo di calcolo utilizzato per arrivare al punteggio e le ragioni che hanno portato a tale punteggio.

---

[223] Ibid. para. 47

[224] Ibid. para. 95

Emergono alcune perplessità nel momento in cui l'articolo 22 viene applicato a sistemi di valutazione automatizzati (fondati su sistemi di IA) che vengono utilizzati solamente come elementi all'interno di un processo decisionale più complesso, composto da altre fasi con il coinvolgimento anche di persone fisiche. Tale perplessità però può essere superata considerando come la decisione finale, seppure preveda l'intervento di una persona fisica, soddisfacendo apparentemente il principio *human-in-the-loop*[225], sia in realtà principalmente influenzata dal risultato Schufa. In definitiva, tale trattamento automatizzato assume la decisione finale che ha effetti sull'interessato, in quanto il coinvolgimento dell'ente creditizio non è realmente influente sulla decisione già presa dal sistema di IA. È comune pensare alla 'decisione finale' in un processo decisionale come definitiva non solo nel senso temporale e sequenziale, ma anche come il fulcro del potere decisionale. È quindi comprensibile che analizzando esclusivamente l'ultimo passo di un processo decisionale si potrà considerare l'intero processo come automatizzato, se tale fase è automatizzato. Al contrario, se è un essere umano a prendere la decisione finale, il processo non è completamente automatizzato[226].

Un ulteriore problema relativo al diritto di accesso ex art. 15 del GDPR Schufa non rivela il metodo seguito per formulare la valutazione, violando il principio di trasparenza del trattamento dei dati ex art. 12 del GDPR. Tale regolamento apre la strada per la creazione di nuovi strumenti a tutela delle persone fisiche nei confronti delle decisioni prese dai sistemi di IA. Questi sistemi non sono in grado di giustificare le loro decisioni e mantengono segreti i loro processi decisionali che potrebbero rivelare motivazioni discriminatorie.

Un divieto generale su tale tipo di trattamento potrebbe però portare ad alcune problematiche, infatti, identificare il ruolo decisivo del punteggio sulla decisione finale non è un'operazione semplice, perché tale influenza deve essere valutata caso per caso e potrebbe generare incertezza giuridica. Anche se il *credit scoring* non è formalmente una decisione, esso ha un impatto intrinsecamente negativo sull'interessato con effetti significativi, per questo motivo andrebbe riconosciuta dalle agenzie che governano tali

[225] Si veda Bisen (n 191).

[226] Binns and Veale (n 198) 329.

sistemi di IA una maggiore trasparenza sulle modalità di calcolo e i dati utilizzati e su cui viene basato il risultato.

La sentenza della Corte rappresenterà il primo passo verso una regolazione attraverso il GDPR dei sistemi di IA che rientrano nella definizione di cui all'art. 22. Inoltre, i profili di trasparenza di tali tecnologie saranno delineati in maniera più dettagliata dalla sentenza, che dovrà anche tenere in considerazione il testo dell'AI Act e l'approccio antropocentrico che l'Unione Europea ha intrapreso nello sviluppo di regolamentazione in questi nuovi settori.

Il GDPR con l'articolo 22 fornisce uno strumento di tutela dei dati personali e dei diritti del soggetto interessato verso le decisioni automatizzate che producono effetti giuridici significativi e non prevedono alcuna partecipazione decisionale umana. La norma si collega all'IA Act attraverso diversi punti di contatto, creando una continuità tra le due regolamentazioni che però non è esplicitamente individuata nel testo della proposta di regolamento sull'IA. Si può tracciare una linea tra la richiesta di revisione della decisione automatizzata ex art. 22 GDPR e i requisiti relativi alla qualità dei dati ex art. 10 dell'AI Act e in particolare il paragrafo tre che richiede che i "*set di dati di addestramento, convalida e prova devono essere pertinenti, rappresentativi, esenti da errori e completi. Essi possiedono le proprietà statistiche appropriate, anche, ove applicabile, per quanto riguarda le persone o i gruppi di persone sui quali il sistema di IA ad alto rischio è destinato a essere usato. Queste caratteristiche dei set di dati possono essere soddisfatte a livello di singoli set di dati o di una combinazione degli stessi."* Tale analogia si basa proprio sull'importanza dei set di dati utilizzati per l'addestramento sui quali si fonda il ragionamento del sistema di IA. Dunque, chi contesta il risultato ottenuto dal sistema di IA lamenta anche il fatto che i dati su cui si basa la decisione non sono corretti e non rappresentano la reale situazione dell'interessato. Un altro articolo connesso all'art. 22 del GDPR è l'art. 14 dell'AI Act che ha come obiettivo assicurare sempre il coinvolgimento e supervisione dell'essere umano, evitando che decisioni complesse possano essere prese esclusivamente da

sistemi automatizzati[227]. L'AI Act prevede all'art. 14, paragrafo 1, che "[*i*] *sistemi di IA ad alto rischio sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso*". In questo modo viene garantito il controllo dell'uomo nei sistemi ad alto rischio e, come chiarito al secondo paragrafo, la finalità è quella di minimizzare i rischi di ogni tipo, che potrebbero nascere dall'utilizzo di tali sistemi. I paragrafi successivi poi stabiliscono modalità e garanzie per l'introduzione della sorveglianza umana nel sistema, che deve essere presente non solo durante lo sviluppo e preparazione del prodotto, ma anche durante la commercializzazione. Stabilendo questo tipo di obblighi per i produttori di tali sistemi, la proposta di regolamento completa la disciplina del GDPR, infatti se quest'ultima prevede diritti a tutela dei soggetti interessati, l'AI Act stabilisce degli obblighi per i produttori dei sistemi e anche per gli utilizzatori.

Il GDPR, con il suo articolo 22, rappresenta un importante strumento per proteggere i soggetti interessati dai trattamenti effettuati dalle tecnologie di IA. Questa norma si pone a tutela dei diritti e libertà fondamentali che potrebbero essere danneggiate da un futuro sviluppo dell'IA. L'obiettivo rimane quello di mantenere l'uomo al centro e garantire il rispetto dei suoi diritti, includendo anche principi giuridici ed etici all'interno delle tecnologie, su cui si fonda la moderna realtà digitale.

4. Segue: La valutazione di impatto sui dati personali nei sistemi di IA, l'esempio delle chatbot

La valutazione di impatto sui dati personali (da qui in poi, DPIA) è un aspetto fondamentale nel contesto dei sistemi di IA. Il GDPR richiede che le organizzazioni conducano una DPIA quando l'elaborazione dei dati può presentare rischi elevati per i diritti e le libertà delle persone coinvolte. Nel contesto dell'IA, questa valutazione

[227] 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, [2021], COM/2021/206 final (n 88). art. 14

assume un'importanza ancora maggiore a causa della vasta quantità di dati personali coinvolti.

L'articolo 35 del GDPR chiarisce che la DPIA consista in una valutazione sui rischi per diritti e libertà dell'interessato, in particolare nel caso in cui vengano impiegate nuove tecnologie nel trattamento dei dati[228]. Il paragrafo 3, lett. a) dell'articolo 35 richiede "*una valutazione sistematica e globale di aspetti personali relativi a persone fisiche, basata su un trattamento automatizzato, compresa la profilazione, e sulla quale si fondano decisioni che hanno effetti giuridici o incidono in modo analogo significativamente su dette persone fisiche*"; viene, quindi, ripreso il contenuto dell'art. 22 come condizione necessaria e sufficiente per effettuare la valutazione di impatto. L'articolo non richiede di effettuare una valutazione ogni volta che sia presente una tecnologia di IA a condurre il trattamento. Piuttosto obbliga il titolare del trattamento ad effettuarlo nel caso specifico di decisioni derivanti da processi automatizzati che possono avere effetti giuridici o incisivi sulla persona. La valutazione dovrebbe contenere una descrizione della natura, del fine, del contesto e dell'obbiettivo dietro al trattamento dei dati, come anche la relativa valutazione di rischio, le misure di adeguamento al GDPR e ulteriori forme di attenuazione del rischio. Nel considerando 75 del GDPR si elenca una lunga lista di tipologie di dati personali cui trattamento potrebbe determinare dei rischi e per i quali risulta necessario condurre una DPIA[229].

---

[228] Regolamento UE 2016/679, art. 35

[229] Il considerando 75: "*I rischi per i diritti e le libertà delle persone fisiche, aventi probabilità e gravità diverse, possono derivare da trattamenti di dati personali suscettibili di cagionare un danno fisico, materiale o immateriale, in particolare: se il trattamento può comportare discriminazioni, furto o usurpazione d'identità, perdite finanziarie, pregiudizio alla reputazione, perdita di riservatezza dei dati personali protetti da segreto professionale, decifratura non autorizzata della pseudonimizzazione, o qualsiasi altro danno economico o sociale significativo; se gli interessati rischiano di essere privati dei loro diritti e delle loro libertà o venga loro impedito l'esercizio del controllo sui dati personali che li riguardano; se sono trattati dati personali che rivelano l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, l'appartenenza sindacale, nonché dati genetici, dati relativi alla salute o i dati relativi alla vita sessuale o a condanne penali e a reati o alle relative misure di sicurezza; in caso di valutazione di aspetti personali, in particolare mediante l'analisi o la previsione di aspetti riguardanti il rendimento professionale, la situazione economica, la salute, le preferenze o gli interessi personali, l'affidabilità o il comportamento, l'ubicazione o gli spostamenti, al fine di creare o utilizzare profili personali; se sono trattati dati personali di persone fisiche vulnerabili, in particolare minori; se il trattamento riguarda una notevole quantità di dati personali e un vasto numero di interessati*."

Tra gli ultimi elementi di tale elenco è presente il 'trattamento che riguarda una notevole quantità di dati personali e un vasto numero di interessati", in questa categoria rientrano certamente i sistemi di IA che trattano dati personali.
L'impatto che i dati hanno dall'addestramento e alla produzione dell'output del sistema di IA è un argomento approfondito da alcune autorità garanti per la protezione dei dati, come ad esempio il CNIL, l'autorità garante dei dati personali francese.[230] Questa ha fornito una guida ed un software *open source* per la valutazione dell'impatto sulla *privacy* che un determinato trattamento dei dati può causare, in particolare quando si prospetta un elevato rischio per i diritti e le libertà delle persone fisiche. Il software funge da guida per attuare una valutazione di impatto sulla protezione dei dati e aiuta i titolari del trattamento a rispettare le prerogative del GDPR.

Per chiarire l'importanza di condurre una DPIA bisogna considerare la recente esponenziale evoluzione dei sistemi di IA, in particolare di quelli destinati ad interagire con le persone fisiche attraverso i messaggi (di seguito, *chatbot*). Un esempio calzante per comprendere questi sistemi è rappresentato dalla *chatbot* di Microsoft, detta Tay. Nel 2016 Microsoft rivelò questa *chatbot* che interagiva con gli umani attraverso la piattaforma social *Twitter*[231]. Questa era stata pensata come esperimento per apprendere attraverso le conversazioni con gli altri utenti i meccanismi mentali che contraddistinguono il modo di comunicare sui social network. Secondo l'azienda la *chatbot* avrebbe dovuto diventare più intelligente attraverso l'interazione in conversazioni ordinarie con le persone fisiche. Tuttavia, il bot non tenne il comportamento sperato e sebbene l'intento fosse quello di migliorare di continuo la sua capacità di interloquire con gli utenti, il risultato fu ben diverso. Il progetto si rilevò

[230] Commission nationale de l'informatique et des libertés 'The open source PIA software helps to carry out data protection impact assessment' (30 Giugno 2021), consultabile a questo indirizzo << https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assessment>>, ultimo accesso il 29 marzo 2023

[231] Twitter è una piattaforma di social media che consente agli utenti di condividere brevi messaggi di testo chiamati "tweet". Fondata nel 2006, Twitter è diventato un luogo di discussione, informazione e connessione in tempo reale. Con oltre 300 milioni di utenti attivi mensili, Twitter ha un impatto significativo nella sfera del giornalismo, dell'influenza sociale e della condivisione di contenuti.

fallimentare e problematico fin dalle prime ore di attivazione: inizialmente il tenore dei messaggi scambiati sulla piattaforma era benevolo e giocoso (ad esempio uno dei primi tweet fu: "*humans are super cool*"), ma a sole quarantotto ore dal lancio, Tay iniziò a pubblicare una serie di 'tweet' contenenti espressioni razziste e discriminatorie, con insulti rivolti a ebrei e femministe. Questo comportamento portò Microsoft a disattivare il profilo e determinare anche un calo della fiducia che gli utenti ponevano sul progetto e sulla tecnologia sperimentata[232].

Fonte:BBC.com

Anche se, secondo il The Guardian, ricercando tra i tweet del bot, che ammontavano a quasi centomila, la maggior parte di quelli più discriminatori e offensivi erano frutto di un'interazione con persone che chiedevano a Tay di ripetere quel che scrivevano loro stessi[233].

---

[232] Dave Lee, 'Tay: Microsoft Issues Apology over Racist Chatbot Fiasco' *BBC News* (25 March 2016) <https://www.bbc.com/news/technology-35902104> accessed 30 April 2023; Abby Ohlheiser, 'Trolls Turned Tay, Microsoft's Fun Millennial AI Bot, into a Genocidal Maniac' *Washington Post* (25 March 2016) <https://www.washingtonpost.com/news/the-intersect/wp/2016/03/24/the-internet-turned-tay-microsofts-fun-millennial-ai-bot-into-a-genocidal-maniac/> accessed 21 May 2023.

[233] Elle Hunt, 'Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter' *The Guardian* (24 March 2016) <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter> accessed 30 April 2023.

Fonte: washingtonpost.com

Il comportamento di questa *chatbot* mostra l'importanza di avere dei meccanismi di controllo adatti e sistemi che permettano il monitoraggio durante lo sviluppo e soprattutto durante la commercializzazione, proprio per evitare comportamenti inappropriati o lesivi di diritti fondamentali. Questo esempio permette di riflettere sull'importanza di valutazioni come quelle condotte attraverso i DPIA, per comprendere al meglio come il sistema può causare danni alle persone, anche psicologici e non solo fisici e anche per limitare il più possibile l'alimentazione di stereotipi e pregiudizi già esistenti nella società. Inoltre, va evidenziato come le stesse aziende che creano questi sistemi di IA debbano integrare durante la progettazione anche principi etici e morali, che permettano al sistema di individuare ed evitare la risposta a stimoli che potrebbero portare ad un danno fisico o psicologico ai soggetti interessati, come richiesto dall'AI Act[234].

Un altro aspetto cruciale riguarda la necessità di dare alle persone fisiche con cui interagisce la *chatbot* l'informazione sulla natura non umana del sistema di IA. In tal senso l'AI Act all'articolo 52 impone ai fornitori di tali sistemi di IA alcuni obblighi di

[234] Sul punto: Ebers and others (n 109); Ben Shneiderman, 'Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems' (2020) 10 ACM Transactions on Interactive Intelligent Systems 1.

trasparenza “in modo tale che le persone fisiche siano informate del fatto di stare interagendo con un sistema di IA”[235].

5. Analisi dei punti di contatto tra GDPR e AI Act

L’AI Act è una proposta di regolamento europeo con l'obiettivo di creare un quadro normativo per l’utilizzo dell'IA [236]. La divisione in categorie basate sul rischio sintetizzata all’interno del primo paragrafo del capitolo corrente, i divieti per alcune pratiche dei sistemi di IA, le prescrizioni per i sistemi di IA ad alto rischio e i doveri di trasparenza per i sistemi a rischio limitato, compongono una regolamentazione articolata e su più livelli.

La Commissione europea ha scelto di presentare la proposta sotto forma di regolamento, seguendo la coerenza dimostrata con i precedenti atti di regolamentazione tecnologica come il Cybersecurity Act [237] ed il GDPR. Alcuni studiosi hanno accolto favorevolmente questa direzione della legislazione europea, che si avvicina alla legislazione degli Stati Uniti attraverso l’utilizzo di un titolo breve e di immediata comprensione [238]. L’AI Act presenta diverse somiglianze con il GDPR, tra cui l’istituzione di nuove autorità nazionali responsabili e un *European Artificial Intelligence Board*. Inoltre, per sottolineare l’importanza che riveste il tema della protezione dei dati all’interno dell’AI Act, il Garante europeo per la protezione dei dati (EDPS) è membro ufficiale del Comitato ex art. 57 AI Act [239].

---

[235] AI Act, art. 52

[236] 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, [2021], COM/2021/206 final (n 88).

[237] Regolamento (UE) 2019/881 Del Parlamento Europeo E Del Consiglio Del 17 aprile 2019 Relativo All’ENISA, L’agenzia Dell’Unione Europea Per La Cibersicurezza, E Alla Certificazione Della Cibersicurezza Per Le Tecnologie Dell’informazione E Della Comunicazione, e che abroga il regolamento (UE) n. 526/2013 («regolamento sulla cibersicurezza»), [2019] OJ L 151, p. 15–69

[238] Vagelis Papakonstantinou and Paul De Hert, ‘EU Lawmaking in the Artificial Intelligent Age: Act-Ification, GDPR Mimesis, and Regulatory Brutality’ [2021] European Law Blog.

[239] L’art. 57 della proposta di regolamento stabilisce al para. 1, che: “*Il comitato è composto dalle autorità nazionali di controllo, rappresentate dal capo di tale autorità o da un alto funzionario di livello*

Punti di contatto tra AI Act e GDPR sono ad esempio il principio della *accountability* e le restrizioni previste per i sistemi di riconoscimento biometrico dei dati personali [240]. Tuttavia, l'obiettivo del nuovo atto normativo dell'Unione europea è più ambizioso; infatti, se il GDPR mirava principalmente a tutelare i diritti sui dati personali, l'AI Act mira: sia a tutelare i cittadini di fronte ad un utilizzo che possa ledere i loro diritti e libertà, che a garantire uno sviluppo tecnologico. Le metodologie per perseguire tali obiettivi non sono esenti da critica, infatti, diversi punti dell'AI Act sembrano già obsoleti e non adatti a regolamentare gli sviluppi futuri dell'IA. Tale tecnologia presenta un incessante sviluppo tecnologico che richiede una normativa sufficientemente flessibile per reggere il ritmo evolutivo. Inoltre, va considerato il crescente grado di autonomia che contraddistingue le ultime frontiere di *machine learning* e *deep learning*, che comportano una difficoltà dell'uomo di spiegare e comprendere i processi dell'IA, noto come effetto 'scatola nera' (c.d. *black box*) [241]. La flessibilità richiesta ad una tale normativa viene trovata attraverso due meccanismi presenti nella proposta. Il primo meccanismo consiste nella previsione degli allegati all'AI Act. In particolare, l'allegato III contiene l'elenco dei settori di IA considerati ad alto rischio e si prevede all'articolo 7 la possibilità per la Commissione europea di aggiornare l'elenco dell'allegato potendo prevedere nuovi sistemi di IA ad alto rischio, che però devono necessariamente rientrare in uno dei settori previsti [242]. Il secondo meccanismo per rimediare all'obsolescenza

---

equivalente, e dal Garante europeo della protezione dei dati. Altre autorità nazionali possono essere invitate alle riunioni, qualora le questioni discusse siano di loro pertinenza." Proposta di Legge sull'Intelligenza Artificiale, [2021], art. 57

[240] Ebers and others (n 42) p. 592.

[241] Il fenomeno delle "black box" si riferisce all'opacità di alcuni modelli di IA, in cui i processi interni di decisione non sono facilmente comprensibili o interpretabili. Questi modelli ricevono input e producono output, ma i calcoli e le ponderazioni intermedi che determinano l'output sono nascosti o non chiaramente comprensibili. Tale mancanza di trasparenza può portare a sfide in termini di responsabilità, equità e fiducia nell'ambito dell'IA. Il termine cominciò ad essere utilizzato tra il 1980 e 1990 quando iniziò a risultare poco chiaro il processo logico seguito dai sistemi computazionali. Si vedano: Matthew U Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2015) 29 Harvard Journal of Law & Technology 353 - 400; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015) p.3.

[242] AI Act, art. 7

della proposta è rappresentato dall'obbligo periodico di revisione da parte della Commissione europea, che ogni cinque anni a partire dalla data in cui l'AI Act diventerà applicabile dovrà pubblicare una relazione di valutazione del quadro normativo proposto [243]. Un simile meccanismo è sempre più consueto nei campi del diritto che regolano materie ad alta innovazione tecnologica, ad esempio l'art. 97 del GDPR, richiede alla Commissione europea di presentare entro il 25 maggio 2020 una relazione di valutazione e sul riesame del GDPR. La relazione viene trasmessa al Parlamento e al Consiglio europei ed il processo di valutazione si ripete ogni quattro anni e può anche portare ad effettuare modifiche per garantire che il GDPR rimanga al passo con gli sviluppi tecnologici e le sfide emergenti nel campo della protezione dei dati.

L'AI Act prevede, oltre ai meccanismi precedentemente menzionati, l'assegnazione di un ruolo di vigilanza sull'applicazione del futuro regolamento ad un'autorità competente. Tuttavia, la proposta non specifica quali siano le entità preposte e rimette la scelta agli Stati membri [244]. Sul punto, nel parere congiunto sull'AI Act, *l'European Data Protection Board* e il Garante Europeo per la Protezione dei Dati sostengono che le autorità nazionali più idonee per questo ruolo sarebbero quelle già incaricate di garantire la protezione dei dati personali [245]. Tuttavia, la scelta dell'autorità più competente potrebbe anche ricadere così sull'Autorità Garante della Concorrenza e del Mercato (AGCM) come sull'Autorità Garante delle Comunicazioni AGCOM, infatti, la regolamentazione dell'IA intreccia tematiche relative sia al diritto della concorrenza sia per quanto riguarda l'ambito delle comunicazioni. Tuttavia, la scelta delle autorità

---

[243] Tale termine è stato ridotto a due anni secondo la nuova formulazione dell'articolo 84 dell'AI Act proposta dalle Commissioni del Parlamento europeo. Committee on Legal Affairs, 'Opinion of the Committee on Legal Affairs for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))'.

[244] AI Act, art. 23, para. 1

[245] EDPB – GEPD Parere congiunto 5/2021 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale), 18 giugno 2021

Garanti della *Privacy* potrebbe essere giustificata non solo dalla conoscenza maturata sulle tecnologie più moderne, ma anche dalla esperienza pluriennale sul tema, questi elementi li renderebbero gli enti più adeguati a svolgere la funzione di vigilanza sull'IA. In ogni caso, rimarrebbe anche la possibilità di costituire una nuova autorità indipendente con il ruolo di Garantire il rispetto di quanto stabilito nell'AI Act. Una nuova autorità potrebbe, infatti, generare una serie di conflitti di competenza tra le altre autorità e in particolare con il Garante *Privacy*. Data la rapida diffusione di tali sistemi, è altamente probabile che la maggior parte delle attività, che oggi si svolgono attraverso un computer, saranno in futuro assistite da tali sistemi di IA ad uso generale, e la competenza del Garante *Privacy* andrebbe gradualmente a scomparire potendo esercitare i suoi poteri di sorveglianza e controllo solo su processi di trattamento dei dati residuali, che non riguardano sistemi di IA.

In conclusione, bisogna sottolineare che la necessità di attribuire tale ruolo di autorità indipendente sarà fondamentale per poter realmente affrontare i rischi e prevenire i danni derivanti dai sistemi di IA.

6. Segue: il pericolo di discriminazione nei sistemi di IA, come può fare la differenza una maggiore qualità dei dati.

L'utilizzo dei sistemi di IA può accrescere la discriminazione e i pregiudizi del sistema in diversi modi. Ad esempio, se i dati utilizzati per addestrare un sistema di IA sono inquinati da pregiudizi razziali, di genere o di altro tipo, l'IA stessa potrebbe riprodurre questi pregiudizi in fase di analisi e produzione dei risultati, aumentando la discriminazione e l'ingiustizia sociale[246]. Sia il GDPR che l'AI Act stabiliscono a riguardo requisiti qualitativi per i dati. Il GDPR all'articolo 5, paragrafo 1, lettera d), pone tra i principi fondamentali l'esattezza dei dati raccolti, la quale richiede un aggiornamento continuo degli stessi e una rettifica o cancellazione nel caso in cui risultino inesatti. Allo stesso modo l'articolo 10 dell'AI Act nei paragrafi da 2 a 5 stabilisce criteri di qualità per i dati utilizzati nei set di addestramento dei sistemi di IA.

[246] Pauline Kim, 'Data-Driven Discrimination at Work' (2017) 58 William & Mary Law Review 857, 897.

Nei sistemi di IA per il riconoscimento facciale la tematica della qualità dei dati è diventata di cruciale importanza in quanto i sistemi hanno adottato un comportamento discriminatorio verso persone con determinate caratteristiche fisiche. Questi sistemi di IA possiedono un'accuratezza che può raggiungere soglie di precisione fino al 90%, anche se questo risultato non è universale e non è sempre garantito. Infatti, Joy Adowaa Boulamwini, attivista informatica statunitense, oltre ad aver condotto numerose ricerche sul tema, ha anche fondato un gruppo che individua e combatte i pregiudizi algoritmici, chiamato Algorithmic Justice League. La ricercatrice ha evidenziato che la maggior parte dei *dataset* per i sistemi di riconoscimento facciale è formata da un'elevata percentuale di individui di carnagione chiara (compresa tra il 79,6% e l'86,24%), in particolare di sesso maschile [247]. Questi sistemi hanno però uno scarso livello di successo nel riconoscere gli individui con carnagione più scura e soprattutto per le donne. La categoria delle donne di carnagione scura risultava dunque essere più soggetta ad errori di riconoscimento rispetto agli uomini con pelle chiara, identificati con maggior precisione. La ricercatrice proponeva all'interno del suo studio il miglioramento della precisione di questi sistemi attraverso l'utilizzo di *dataset* più diversificati dal punto di vista fenotipico [248] e l'adozione di metriche di valutazione non solo per genere o tipo di pelle, ma anche per l'intersezione tra i due fattori. Il suo lavoro ha attirato grande attenzione e alcune multinazionali come IBM, che produceva sistemi di IA per il riconoscimento facciale, hanno intrapreso un percorso per correggere le discriminazioni presenti nei loro *dataset*. Nonostante alcuni progressi siano stati raggiunti, IBM, Microsoft e Amazon hanno preferito sospendere i programmi di sviluppo ed in particolare hanno interrotto le collaborazioni con le forze dell'ordine che

[247] Joy Adowaa Buolamwini, 'Gender Shades : Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers' (Thesis, Massachusetts Institute of Technology 2017) <https://dspace.mit.edu/handle/1721.1/114068> accessed 27 April 2023.

[248] Secondo l'enciclopedia, con il termine fenotipo si indica, in genetica, l'insieme delle caratteristiche morfologiche e funzionali di un organismo determinate dall'interazione fra la sua costituzione genetica e l'ambiente. Treccani Enciclopedia on line, 'fenotipo' <https://www.treccani.it/enciclopedia/fenotipo/> ultimo accesso 24 marzo 2023.

utilizzavano tali tecnologie [249]. Uno degli utilizzi più rischiosi è infatti nel settore della pubblica sicurezza, nel quale il pericolo di fondare restrizioni alla libertà personale su decisioni automatizzate basate su dati di scarsa qualità è molto più elevato rispetto a qualsiasi altro settore [250]. La capacità intrusiva di questi sistemi di IA capaci non solo di riconoscere il volto di una persona, ma anche, ad esempio, di ascoltare quel che dice, rende tali sistemi i migliori strumenti per il controllo e la sorveglianza delle persone, determinando anche una violazione del diritto alla *privacy* come stabilito all'art. 8 della CEDU [251]. Anche l'azienda statunitense Microsoft ritiene che l'uso di sistemi di identificazione biometrica in tempo reale in spazi pubblici per la ricerca di persone sospette o per il riconoscimento facciale debba essere limitato ai crimini più gravi e violenti, come l'omicidio, la rapina a mano armata e il sequestro [252]. La società ritiene che l'uso di tali sistemi per crimini meno gravi, come la falsificazione e la contraffazione, sia sproporzionato e richieda ulteriori discussioni.

La problematica della qualità dei dati viene esplicitamente affrontata dall'art. 10 dell'AI Act [253]. L'articolo pone una serie di requisiti di qualità dei dati che vengono utilizzati per addestrare i sistemi di IA, anche se il paragrafo 1 restringe il campo di applicazione della norma ai sistemi c.d. ad alto rischio. Il paragrafo 3 richiede che i set di dati di

---

[249] Jeffrey Dastin and Ayanti Bera, 'Amazon Pauses Police Use of Its Facial Recognition Tech for a Year' *Reuters* (10 June 2020) <https://www.reuters.com/article/us-amazon-com-facial-recognition-idUSKBN23H3EO> accessed 30 April 2023.

[250] Dallas Hill, Christopher D O'Connor and Andrea Slane, 'Police Use of Facial Recognition Technology: The Potential for Engaging the Public through Co-Constructed Policy-Making' (2022) 24 International Journal of Police Science & Management 325.

[251] Peter Fussey and Daragh Murray, 'Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology' 34 <https://repository.essex.ac.uk/24946/> accessed 27 May 2023.

[252] Microsoft, 'Response to the European Commission's Consultation on the Artificial Intelligence Act' (2021) <https://blogs.microsoft.com/wp-content/uploads/prod/sites/73/2021/09/microsoft-response-to-the-european-commission-consultation-on-the-artifical-intelligence-act.pdf> accesso il 28 Maggio 2023

[253] L'articolo 10 dell'AI Act, rubricato: "Dati e governance dei dati" prevede per i sistemi di IA ad alto rischio una serie di criteri di qualità da soddisfare relativi ai set di dati utilizzati per l'addestramento, la convalida e la prova. Tali requisiti sono indicati nei paragrafi da 2 a 5.

addestramento, convalida e prova siano pertinenti, rappresentativi, esenti da errori e completi, e che possiedano le proprietà statistiche appropriate per il contesto in cui il sistema di IA ad alto rischio è destinato ad essere utilizzato. La limitazione di tali requisiti ai sistemi ad alto rischio può avere effetti sia positivi, che negativi. Da un lato, poiché gli standard qualitativi richiesti dalla norma riguardano solamente i sistemi di IA ad alto rischio, allora i sistemi di categoria diversa non saranno obbligati al rispetto di tali requisiti, potendo in ogni caso causare un certo livello di discriminazione nel loro campo di applicazione; dall'altro lato, pretendere che tutti i sistemi di IA, a prescindere dalla categoria di rischio, rispettino tali standard qualitativi avrebbe sicuramente ripercussioni sullo sviluppo di nuovi sistemi di IA, cagionando una diminuzione di innovazione e competitività del settore. In ogni caso, l'art. 10 dell'AI Act fornisce un quadro di riferimento normativo importante per la tutela dei diritti fondamentali degli individui coinvolti nell'utilizzo di tali sistemi di IA. Chiaramente l'AI Act non risolve di per sé il problema della discriminazione algoritmica, ma rappresenta un importante passo avanti per garantire la trasparenza e la responsabilità degli sviluppatori e degli utilizzatori di tali sistemi. L'articolo stabilisce anche una serie di requisiti di buona gestione dei dati che riguardano in particolare la raccolta, il trattamento e la valutazione dei dati e che dovrebbero permettere l'identificazione e la correzione di eventuali distorsioni o *bias* nei dati che possono portare a discriminazioni da parte dei sistemi di IA ad alto rischio. In altre parole, i dati devono essere accurati e rappresentativi della popolazione su cui il sistema di IA ad alto rischio viene utilizzato per evitare la discriminazione [254]. Tuttavia, raggiungere un tale grado di perfezione dei dati è molto difficile, se non impossibile, infatti, i requisiti che richiedono un grado maggiore di perfezione potrebbero rappresentare un ostacolo per l'innovazione[255]. Inoltre, esistono delle tecniche che puntano ad aumentare i livelli di protezione dei dati che consistono nell'aggiungere rumore (*noise)* nei *dataset* per prevenire la rivelazione involontaria dei

[254] Mauritz Kop 'EU Artificial Intelligence Act: The European Approach to AI' [2021] Transatlantic Antitrust and IPR Developments.

[255] Ebers and others (n 109).

dati personali, e di dati sensibili ex art. 9 GDPR[256]. Tali tecniche, sebbene rispettose e uniformi col principio della protezione dei dati *by default* ex art. 25 GDPR, sarebbero in netto contrasto con l'ordine, la correttezza e l'assenza di errori richiesti dall'articolo 10 della AI Act.

Il miglioramento qualitativo dei dati su cui vengono addestrati i sistemi di IA è uno dei punti su cui si è concentrata la ricerca "capAI", di alcuni studiosi di Oxford[257]. Nella ricerca è stata illustrata una procedura da seguire per accertare la conformità di un sistema di IA ai requisiti e alla disciplina prevista dall'AI Act. La procedura di valutazione della conformità elenca dei passaggi ben precisi che includono: un protocollo di verifica interno, una scheda tecnica riassuntiva, una scheda di valutazione esterna. All'interno della procedura si presta molta attenzione alla qualità dei dati, in particolare durante la fase di sviluppo dei sistemi. "*Garbage in, garbage out*" è un aforisma che descrive il problema del sistema di IA nel caso in cui i dati su cui si basa il suo addestramento siano di scarsa qualità e poco eterogenei. L'esercizio che operano i sistemi di IA è di tipo statistico; quindi, solitamente ad una maggiore qualità dei dati, consegue una migliore qualità del risultato. La qualità mira a soddisfare requisiti di unicità, completezza, attualità, consistenza e accuratezza, che permettono al sistema di fornire un risultato migliore. Una semplice riproduzione dei dati utilizzati dalle persone di un determinato settore porterebbe alla replica anche dei pregiudizi insiti negli stessi. Già all'interno del Regolamento UE 2016/679, viene imposto ai titolari del trattamento, un obbligo di valutazione delle conseguenze derivanti dal trattamento di determinati dati personali, valutazione che, come noto, prende il nome di Data Protection Impact Assessment e che trova la sua disciplina all'interno dell'articolo 35 del GDPR.

---

[256] Il "*noise*" (rumore) all'interno dei dataset viene definito come l'aggiunta di dati casuali o artificiali al fine di rendere più difficoltosa l'identificazione o l'estrazione di informazioni sensibili dal dataset stesso. In altre parole, l'obiettivo è quello di mascherare o confondere le informazioni sensibili presenti nel dataset attraverso l'aggiunta di informazioni non rilevanti o di "rumore". Questa tecnica è spesso utilizzata per proteggere la privacy dei dati personali all'interno di dataset sensibili.

[257] Luciano Floridi and others, 'CapAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act' [2022] SSRN Electronic Journal <https://www.ssrn.com/abstract=4064091> accessed 27 April 2023.

7. Segue: La trasparenza come mezzo per la comprensibilità dei sistemi di IA e le difficoltà nell'attuazione della sorveglianza umana

Come abbiamo più volte accennato, il problema dell'opacità dei processi algoritmici ha considerevoli effetti sul diritto delle persone a conoscere le motivazioni che portano le macchine a decidere in un certo modo proprio perché, solo conoscendo le motivazioni che hanno portato ad una decisione, le persone possono efficacemente costruire argomenti per correggerle. L'incomprensibilità che spesso accompagna alcuni sistemi di IA ha portato il legislatore europeo ad inserire già al considerando 47 dell'AI Act il principio della trasparenza, sebbene limitato ai sistemi ad alto rischio. La disciplina prende poi forma nell'art. 13 dell'AI Act, che al paragrafo 1 chiarisce come, sin dalla fase di sviluppo, i sistemi dovrebbero essere '*sufficientemente trasparent[i] da consentire agli utenti di interpretare l'output del sistema e utilizzarlo adeguatamente*' [258]. Si potrebbe parlare di una *transparency by design*, proprio perché l'articolo specifica che tale livello di trasparenza debba essere attuato già dalla fase di progettazione[259]. La trasparenza è un principio importante nell'AI Act, poiché consente ai cittadini di comprendere come vengono utilizzati i loro dati e di avere maggiore controllo sulle decisioni che li riguardano. Non sempre tale obbligo riesce ad ottenere l'utilità sperata, infatti, la trasparenza potrebbe non essere sufficiente a prevenire i danni causati da sistemi di riconoscimento delle emozioni e categorizzazione biometrica[260]. Questi sistemi possono causare danni anche se le persone sono consapevoli del loro funzionamento, ad esempio attraverso la discriminazione o la violazione della *privacy*. Pertanto, è necessario adottare ulteriori misure per garantire che questi sistemi rispettino i diritti fondamentali delle persone e non causino danni ingiustificati.

---

[258] 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, [2021], COM/2021/206 final (n 88). Art.13

[259] Heike Felzmann and others, 'Towards Transparency by Design for Artificial Intelligence' (2020) 26 Science and Engineering Ethics 3333.

[260] Access Now, 'Access Now's submission to the European Commission's adoption consultation on the AI Act' (2021) <https://www.accessnow.org/cms/assets/uploads/2021/06/Access-Now-submission-to-the-European-Commissions-adoption-consultation-on-the-AI-Act.pdf>, p. 6, ultimo accesso 28 maggio 2023.

L’articolo 13 dell’AI Act non si concentra solo sulla fase di progettazione e richiede che i sistemi siano corredati da istruzioni per un uso appropriato contenenti una serie di informazioni elencate al paragrafo 3, come ad esempio: i dati e contatti del fornitore, le caratteristiche del sistema di IA, eventuali modifiche apportate, le misure di sorveglianza umana ex art. 14 AI Act, la durata dell’attività del sistema e la manutenzione per il corretto funzionamento. La proposta mira a garantire un livello di trasparenza molto alto che permetta agli utenti di comprendere il funzionamento del sistema da cima a fondo.

Mentre, da un lato, questa disposizione mira a tutelare i diritti degli utenti, dall’altro, l’art. 14 stabilisce un obbligo di sorveglianza umana, richiedendo che il soggetto posto a controllo del sistema di IA debba poter intervenire nel funzionamento. Secondo l’art. 14 para. 4, lett. d), il soggetto controllore avrà dunque la possibilità di decidere “[…] *in qualsiasi situazione particolare, di non usare il sistema di IA ad alto rischio o altrimenti di ignorare, annullare o ribaltare l'output del sistema di IA ad alto rischio*;”. La proposta pone delle misure di trasparenza sia a favore dell’utente che a favore dell’umano posto al controllo. Tuttavia, i metodi per assicurare la trasparenza non sono specificati all’interno della proposta e il rispetto della disciplina viene affidato al fornitore, che, secondo gli articoli 16, lett. j e 23 dell’AI Act[261], dovrà dimostrare alle autorità competenti di aver prodotto un sistema di IA che rispetta i requisiti di trasparenza indicati all’art. 13 [262]. A tal proposito, la Commissione giustizia del

[261] L'articolo 16 dell’AI Act elenca una serie di obblighi rivolti ai fornitori dei sistemi ad alto rischio. Questi obblighi includono: “a) l'assicurazione della conformità dei sistemi di IA ad alto rischio ai requisiti del capo 2 del presente titolo, b) la disposizione di un sistema di gestione della qualità conforme all'articolo 17, c) la redazione della documentazione tecnica del sistema di IA ad alto rischio, d) la conservazione dei log generati automaticamente dai loro sistemi di IA ad alto rischio, e) l'assicurazione che il sistema di IA ad alto rischio sia sottoposto alla pertinente procedura di valutazione della conformità prima della sua immissione sul mercato o messa in servizio, f) il rispetto degli obblighi di registrazione secondo l'articolo 51, g) l'adozione delle necessarie misure correttive nel caso in cui il sistema di IA ad alto rischio non sia conforme ai requisiti del capo 2 del presente titolo, h) l'informazione alle autorità nazionali competenti degli Stati membri e all'organismo notificato in merito alla non conformità e alle eventuali misure correttive adottate, i) l'apposizione della marcatura CE sui loro sistemi di IA ad alto rischio per indicare la conformità al presente regolamento, j) la dimostrazione, su richiesta di un'autorità nazionale competente, della conformità del sistema di IA ad alto rischio ai requisiti del capo 2 del presente titolo”.

[262] Ebers and others (n 109) 596.

Parlamento europeo, nel parere sul AI Act pubblicato a settembre 2022, ha approfondito il tema della trasparenza come disciplinato all'articolo 13 [263]. In tale parere viene emendato il paragrafo 1 dell'art. 13, attraverso l'aggiunta di un secondo periodo che richiede, prima dell'immissione nel mercato, che il sistema ad alto rischio sia munito di tutti i mezzi tecnici allo stato dell'arte per assicurare l'interpretabilità dell'output di sistema, sia per il fornitore che per l'utente finale[264]. Inoltre, viene richiesto che l'utente sia messo in condizione di saper spiegare la decisione presa dal sistema. Sono diversi gli emendamenti che modificano la chiarezza e le informazioni contenute nell'articolo 13, ad esempio vengono aggiunte informazioni come il "*livello di capacità del sistema di IA nel fornire la spiegazione dietro la decisione che ha preso*" oppure che le informazioni richieste dall'articolo vengano fornite nella lingua del Paese in cui è utilizzato il sistema [265]. Misure come questa vanno a costruire un sistema di trasparenza attorno all'utilizzo dei dati e al funzionamento dei sistemi di IA che risulta essere la chiave per il miglior sviluppo della tecnologia. Con un livello sufficiente di trasparenza e chiarezza delle informazioni fornite agli individui, si potrà sempre più aumentare la consapevolezza diffusa dei rischi e dei benefici che possono derivare dall'utilizzo di tali tecnologie, permettendo al contempo di minimizzare la probabilità di danni causati dai sistemi.

Se per garantire una trasparenza sufficiente e una comprensibilità del funzionamento dei sistemi di IA vengono poste misure a tutela e che assistono l'utente, allo stesso tempo il principio di sorveglianza e controllo umano, fissato all'art. 14 dell'AI Act[266], vincola il produttore del sistema di IA ad inserire nel sistema un'interfaccia uomo-macchina idonea a permettere una supervisione efficace durante l'uso del sistema stesso. Secondo

[263] Committee on Legal Affairs, 'Opinion of the Committee on Legal Affairs for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))' emendamento 47.

[264] Panigutti and others (n 140) 1144-1146.

[265] Committee on Legal Affairs, (n 244) , emendamenti 55 e 59

[266] AI Act, art. 14

il percorso logico dell'articolo, per aversi una supervisione efficace, devono essere garantite delle misure individuate dal produttore stesso e, attraverso queste, deve essere possibile per le persone a cui è affidata la sorveglianza compiere una serie di azioni. In altre parole, l'articolo prevede una serie di risultati che devono essere ottenuti attraverso metodi a discrezione del produttore, ma che potrebbero risultare sproporzionati rispetto al vantaggio derivante.

La prima azione indicata all'articolo 14, paragrafo 4 dell'AI Act richiede una comprensione piena delle capacità e dei limiti del sistema di IA ad alto rischio col fine di individuare e affrontare segnali di anomalie al più presto.

Al tempo stesso occorre essere consapevoli che esistono sistemi di IA per i quali una tale comprensione non è oggi facilmente ottenibile per una persona di media cultura, ragione per cui si potrebbe arrivare ad un divieto indiretto di questi sistemi, nel momento in cui non risulti possibile trovare una persona idonea al ruolo di supervisore [267]. Nel caso in cui fosse questa la reale intenzione del legislatore eurounitario, il testo normativo avrebbe potuto piuttosto richiedere ai produttori di tali sistemi maggiore impegno per uno sviluppo di sistemi più trasparenti. Tali perplessità sono state enunciate anche nel parere della Commissione Giustizia del Parlamento europeo in cui, con l'emendamento 65, è stato proposto di sostituire il requisito di una "comprensione completa" con quello della "*consapevolezza e di una comprensione sufficiente delle capacità e dei limiti del sistema*", rendendo meno rigida la disposizione e considerando anche le attuali problematiche legate alla opacità algoritmica [268]. Inoltre, occorre sottolineare che dalla proposta non emerge con chiarezza in quale fase il soggetto addetto alla sorveglianza del sistema possa essere considerato responsabile e non è nemmeno previsto un regime di responsabilità effettivo nei confronti dell'*human in the loop* nel ruolo di controllore[269]. Inserire un regime di responsabilità nella proposta per la persona preposta alla sorveglianza del sistema aumenterebbe il rigore e l'attenzione nel

[267] Jorge Constantino, 'Exploring Article 14 of the EU AI Proposal: Human in the Loop Challenges When Overseeing High-Risk AI Systems in Public Service Organisations' (2022) 14 Amsterdam Law Forum 1.

[268] Committee on Legal Affairs (n 86) emendamento 65.

[269] Veale and Zuiderveen Borgesius (n 99).

monitoraggio di tecnologie che potrebbero facilmente violare i diritti e le libertà delle persone, ad esempio se utilizzate in ambiti legati alla pubblica sicurezza[270]. Tuttavia, per potere assumere una responsabilità del genere, bisognerà avere un livello di conoscenza del sistema di IA che possa giustificare il ruolo di controllore, ma sorge spontaneo chiedersi come faranno le persone poco esperte in materia a costruire una base di conoscenza adeguata e idonea, quando anche i più periti dei sistemi di IA provenienti dalle maggiori società del settore, come IBM o Microsoft, seguono costantemente corsi di aggiornamento per prepararsi ad emergenze derivanti da un malsano utilizzo e dal continuo mutamento dei sistemi di IA [271].

L'art. 14, para. 1, dell'AI Act richiede che i sistemi siano progettati per "poter essere efficacemente supervisionati". Una tale previsione porta con sé l'evoluzione del concetto stabilito all'interno dell'articolo 22 del GDPR in cui veniva previsto il diritto di avere una revisione da parte di un umano di decisioni pregiudizievoli prese da un sistema decisionale automatizzato. L'art 14 dell'AI Act non chiarisce in cosa consista una supervisione efficace e questo rimane un elemento incerto della disposizione[272]. Il considerando 48 dell'AI Act precisa però che "*le persone fisiche alle quali è stata affidata la sorveglianza umana dispongano delle competenze, della formazione e dell'autorità necessarie per svolgere tale ruolo*". Data la rapida e complessa evoluzione dei sistemi di IA, potrebbe anche risultare necessaria una sorta di sorveglianza ibrida o aumentata che permetta alla persona fisica di utilizzare strumenti di IA per sorvegliare i sistemi di IA ad alto rischio [273]. La combinazione tra intelligenza umana e intelligenza artificiale potrebbe risultare in un connubio perfetto per permettere una sorveglianza efficace, ottenendo un'intelligenza aumentata. Secondo l'esperienza del granmaestro di

---

[270] Constantino (n 248).

[271] Shneiderman (n 215).

[272] Johann Laux, 'Institutionalised Distrust and Human Oversight of Artificial Intelligence: Toward a Democratic Design of AI Governance under the European Union AI Act' (3 March 2023) 6 <https://papers.ssrn.com/abstract=4377481> accessed 27 May 2023.

[273] Mohammad Hossein Jarrahi, Christoph Lutz and Gemma Newlands, 'Artificial Intelligence, Human Intelligence and Hybrid Intelligence Based on Mutual Augmentation' (2022) 9 Big Data & Society 20539517221142824, 3.

scacchi Garry Kasparov "l'uso di un PC gli consentiva di concentrarsi maggiormente sulla pianificazione strategica, mentre la macchina si occupava dei calcoli"[274], tale gioco di squadra si è dimostrato un ottimo esempio per apprezzare le potenzialità di una collaborazione tra l'uomo e la macchina.

L'articolo 14 dell'AI Act si concentra molto sulla necessità di mantenere la persona fisica in controllo del sistema di IA, per tale ragione il paragrafo 4 richiede che grazie alle misure previste nel paragrafo 3, la persona fisica debba essere in grado di compiere alcune operazioni. Tra quelle più problematiche vi è l'azione indicata alla lettera e), che richiede la possibilità di "*intervenire sul funzionamento del sistema di IA ad alto rischio o di interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga"*. Sulla possibilità di spegnere un sistema di IA sorgono diversi interrogativi; infatti, se si pensa ad esempio ai moderni *Large Language Models* viene da chiedersi come e se sia possibile spegnere uno di questi. È importante distinguere tra i sistemi che possono essere spenti facilmente e quelli che richiedono una procedura più laboriosa per la loro complessità tecnologica. Una riflessione va fatta verso quei sistemi di IA che per la loro sofisticatezza sono caratterizzati dalla capacità di eseguire operazioni svariate e che non hanno un determinato scopo e piuttosto sono costruiti per raggiungere diversi risultati. Tali sistemi ricevono un obiettivo da raggiungere, ma sono allo stesso tempo portati a rendere più difficile il proprio spegnimento, perché lo stesso si porrebbe come ostacolo al raggiungimento dell'obiettivo programmato. Uno dei più importanti esperti di IA, Stuart Russell[275], descrive "*the off-switch problem"* come la principale difficoltà

[274] David De Cremer and Garry Kasparov, 'AI Should Augment Human Intelligence, Not Replace It' [2021] *Harvard Business Review* <https://hbr.org/2021/03/ai-should-augment-human-intelligence-not-replace-it> accessed 27 May 2023.

[275] Stuart Russell è un noto professore e ricercatore nel campo dell'IA. Nato nel 1962 nel Regno Unito, è attualmente Professore di Ingegneria Elettrica e delle Scienze dell'Informazione presso l'Università di California, Berkeley. Russell è noto soprattutto per il suo contributo nel campo dell'IA, specialmente per la sua ricerca sulla teoria e pratica della pianificazione probabilistica e del decision-making, e sulla programmazione logica. È co-autore del libro di testo "*Artificial Intelligence: A Modern Approach*", utilizzato in più di 1.300 università in 118 paesi, che è considerato il testo di riferimento standard nel campo dell'IA. Russell è anche un sostenitore attivo della ricerca per la sicurezza dell'IA e dell'approccio precauzionale nello sviluppo di IA avanzata.

legata al controllo sulle macchine intelligenti[276]. Un sistema di IA programmato con un obiettivo predeterminato potrebbe essere progettato in modo da resistere all'interruzione del suo funzionamento, e potrebbe addirittura cercare di disattivare il suo pulsante di spegnimento. In una tale circostanza, ci si troverebbe di fronte a un problema concreto, ovvero l'impossibilità di interrompere le operazioni del sistema di IA. Secondo Russell, l'implementazione di un elemento di incertezza nei sistemi di IA potrebbe mitigare questo rischio. Se un sistema di IA non possedesse un quadro informativo completo sulle preferenze del suo utente, potrebbe non essere in grado di determinare un'unica soluzione ottimale. In questo caso, il sistema proporrebbe diverse alternative e attenderebbe l'intervento umano per la scelta finale.

Incorporare tale incertezza implicherebbe la possibilità per l'utente di decidere anche di spegnere il sistema come parte della soluzione al problema in questione. Poiché l'obiettivo finale del sistema è soddisfare le preferenze dell'utente, l'accettazione dell'interruzione del suo funzionamento sarebbe coerente con il raggiungimento di tale obiettivo. Di conseguenza, il sistema di IA invece di cercare di resistere alla disattivazione, accoglierebbe questa opzione come parte del processo decisionale per realizzare l'obiettivo finale. In conclusione, l'AI Act si pone l'obiettivo della sorveglianza sui sistemi di IA con l'intento di "*prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile*" [277]. Nel parere della Commissione Giustizia del Parlamento europeo, viene prestata particolare attenzione anche al rischio che potrebbe derivare da una decisione basata esclusivamente su un trattamento automatizzato posto in essere da un sistema di IA che possa cagionare conseguenze giuridiche o altri effetti rilevanti sulla persona, sostanzialmente riprendendo la previsione di cui all'art. 22 del GDPR [278]. La problematica della

[276] Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (Penguin Publishing Group 2020) 196.

[277] AI Act, Art. 14, para. 2

[278] Committee on Legal Affairs (n 224).

trasparenza nei sistemi di IA risulta di fondamentale importanza per garantire al meglio il rispetto dei diritti e le libertà delle persone e anche se nella proposta non sono realmente indicate le modalità per raggiungere la trasparenza desiderata, quantomeno sono previsti diversi obiettivi da raggiungere che di per sé rappresentano un passo avanti verso una maggiore trasparenza dei sistemi di IA. La disciplina prevista deve ancora essere potenziata per garantire una maggiore comprensibilità dei sistemi e adottare le misure più adatte alla reale situazione di totale disinteresse delle grandi aziende tecnologiche. Queste difficoltà non sono immediatamente risolvibili dall'AI Act e richiederanno del tempo anche dopo l'implementazione della proposta per ottenere auspicabilmente un livello di chiarezza sufficiente.

Capitolo III

"*We are on the verge of putting in place landmark legislation that must resist the challenge of time. It is crucial to build citizens' trust in the development of AI, to set the European way for dealing with the extraordinary changes that are already happening, as well as to steer the political debate on AI at the global level. We are confident our text balances the protection of fundamental rights with the need to provide legal certainty to businesses and stimulate innovation in Europe*"

- Brando Benifei [279]

1. La Proposta di Regolamento sull'Intelligenza Artificiale e la "General Purpose AI" in essa prevista

Ricapitolando, la prima versione del AI Act è stata presentata dalla Commissione europea il 21 aprile 2021[280], essa rappresenta un audace tentativo del legislatore europeo di stabilire un quadro normativo per governare l'uso e lo sviluppo dell'IA nell'Unione Europea. I principi cardine di questa proposta di regolamento sono: il rispetto dei diritti fondamentali, la tutela della sicurezza e del benessere dei cittadini europei e la promozione dell'innovazione tecnologica responsabile. La *ratio* che sottende l'AI Act è dunque quella di garantire un ambiente sicuro e regolato per lo sviluppo e l'impiego dell'IA, promuovendo al contempo la competitività dell'Unione in questo campo. In generale, i sistemi di IA sono identificati nel testo dell'AI Act in base alla finalità dichiarata, in forza della quale viene assegnata una categoria di rischio. Con questo approccio, tuttavia, nessuna disposizione della versione originaria dell'AI Act riesce a

[279] Brando Benifei, "AI Act: a step closer to the first rules on Artificial Intelligence", European Parliament Press Release (11 May 2023).

[280] 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, [2021], COM/2021/206 final (n 88).

includere nel raggio di applicazione quei sistemi di IA privi di una finalità predeterminata, addestrati senza uno scopo specifico. Il testo della prima versione della proposta si concentrava sui modelli di IA più comuni, non prendendo in considerazione, ad esempio, i moderni *foundation models*. Dal punto di vista giuridico, questi sistemi presentano una serie di sfide rilevanti, che vengono in parte considerate e affrontate dagli ultimi emendamenti proposti all'AI Act[281]. La definizione introdotta nell'AI Act di *foundation models* recita:

"Per 'modello fondazionale' si intende un modello di intelligenza artificiale addestrato su dati ampi e su scala, progettato per la generalità dei risultati e adattabile a un'ampia gamma di compiti specifici"[282].

La definizione risulta essere molto generica e si presta a varie interpretazioni. I *foundation models* sono sistemi di IA addestrati su enormi quantità di parametri; ad esempio, il modello di OpenAI *GPT-3* è stato addestrato su circa 175 miliardi di parametri [283]. Per vagliare in concreto le potenzialità di questi modelli, possiamo inserire un comando all'interno del sistema *ChatGPT* (fondato su *GPT-3.5*, versione successiva di *GPT-3*), chiedendo di chiarire cosa siano i *foundation models*[284]. La risposta viene riportata testualmente:

---

[281] Draft Compromise Amendments on the Draft Report, Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts", (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD))

[282] La definizione è stata tradotta dall'inglese. Quella originale viene riportata di seguito: "'*foundation model' means an AI model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks*".

[283] Ethan Mollick, 'ChatGPT Is a Tipping Point for AI' [2022] *Harvard Business Review* <https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai> accessed 5 May 2023. Le versioni precedenti, come GPT-2, erano già impressionanti nel generare testo coerente e contestualmente pertinente. GPT-3 ha aperto nuove possibilità grazie alle sue capacità migliorate, consentendo un'ampia gamma di applicazioni, come la generazione di contenuti, la traduzione, il completamento automatico del codice, il tutoraggio e molto altro. La combinazione di prestazioni avanzate, versatilità, facilità d'uso, interesse per l'IA e promozione efficace ha consentito a ChatGPT di raggiungere rapidamente cento milioni di utenti e diventare uno degli strumenti basati sull'intelligenza artificiale più popolari sul mercato.

[284] Il comando inserito (*prompt*) per ottenere tale risposta è: "Descrivi i foundation models".

"I "foundation models" sono modelli di apprendimento automatico, in particolare di intelligenza artificiale (IA), che sono stati pre-addestrati su enormi quantità di dati di testo, immagini, o altri tipi di dati. Questi modelli servono come base per ulteriori personalizzazioni e affinamenti specifici dell'applicazione. […] Questi modelli vengono addestrati su vasti corpus di testo e poi personalizzati o 'affinati' per compiti specifici, come la generazione di testo, la traduzione, la risposta alle domande, ecc. Uno degli aspetti chiave dei foundation models è la loro capacità di catturare e generalizzare una grande quantità di conoscenza dal mondo reale dai dati su cui vengono addestrati."

Come illustrato da ChatGPT, tali modelli si aggiornano e apprendono costantemente per perfezionare la qualità delle risposte fornite. La grande quantità di dati necessaria per l'addestramento di tali modelli implica che gli sviluppatori debbano spesso fare affidamento ai dati disponibili su internet, i quali sono spesso lontani dagli standard di qualità necessari ad evitare la generazione di contenuti offensivi e discriminatori. L'attività di estrazione dei dati da Internet determina numerose problematiche relative alla protezione del diritto d'autore, che per ragioni di spazio di questo elaborato non verranno trattate[285]. L'articolo 10 dell'AI Act fissa i requisiti qualitativi da rispettare relativamente ai dati di addestramento. Inoltre, per prevenire tali problematiche gli sviluppatori di tali sistemi devono avere un solido sistema di moderazione dei contenuti[286]. La seconda definizione rilevante presentata con gli ultimi emendamenti riguarda invece i sistemi di IA ad uso generale, cioè quei sistemi addestrati per un fine generico.

Tali sistemi, anche noti come, "*General Purpose AI*" (di seguito, GPAI) si distinguono dai *foundation models*; infatti, sebbene tali termini si riferiscano a categorie simili di

[285] Per un approfondimento sul tema: European Commission, "*Intellectual property in ChatGPT*" (2023); Yogesh K. Dwivedi et al., "Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy", International Journal of Information Management 71 (2023)

[286] La moderazione dei contenuti è il processo di monitoraggio, valutazione e, se necessario, moderazione o eliminazione di contenuti che non rispettano le linee guida di una particolare piattaforma o servizio. OpenAI ha ad esempio assunto una squadra di moderatori di contenuti dal Kenya, che tuttavia erano pagati meno di due dollari l'ora. 'Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer' (*Time*, 18 January 2023) <https://time.com/6247678/openai-chatgpt-kenya-workers/> accessed 24 May 2023.

sistemi di IA, esistono sfumature che permettono di distinguerli. Da un lato, un sistema di GPAI è un sistema di IA che può essere applicato a una vasta gamma di attività, senza essere specificamente progettato o addestrato per un compito specifico. Tali sistemi sono in grado di apprendere e adattarsi a nuovi compiti e contesti. Dall'altro lato, i *foundation models* sono grandi modelli di apprendimento automatico pre-addestrati su vasti corpus di dati, che possono poi essere affinati per specifici compiti. Essi costituiscono una sorta di "base" o "fondamento" che può essere utilizzato per la progettazione di sistemi di IA con applicazioni specifiche. Quindi, mentre entrambi i tipi di sistemi sono "generali" in quanto possono svolgere una vasta gamma di compiti, la differenza principale risiede nel fatto che i *foundation models* sono solitamente progettati per essere addestrati su un immenso corpus di dati e poi affinati per specifici compiti, mentre i sistemi di GPAI sono progettati per apprendere e adattarsi a qualsiasi compito senza bisogno di addestramento specifico.

I sistemi di GPAI stanno velocemente mutando il modo in cui gli individui comunicano, illustrano e creano, ad esempio ChatGPT viene utilizzato ogni giorno da milioni di persone per generare testo indistinguibile e paragonabile per qualità a un testo scritto da un essere umano. Il suo utilizzo si è già diffuso in molti ambiti della società moderna, dall'economia alla medicina, dall'istruzione alla ricerca, dalla scrittura dei codici alle arti visive[287]. Tali sistemi però rischiano di generare discriminazioni e violazioni della *privacy,* oltre a poter generare contenuti offensivi.

Di fronte a questa tecnologia rivoluzionaria, le prescrizioni contenute nella proposta del 2021 dell'AI Act si presentano poco adatte per affrontare le sfide poste da questi moderni sistemi di IA. Al contempo normative già presenti come il GDPR [288] continuano ad applicarsi anche a tali sistemi di IA. In particolare, l'articolo 22 del

---

[287] Enkelejda Kasneci and others, 'ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education' (2023) 103 Learning and Individual Differences 102274; Kif Leswing, 'Bloomberg Plans to Integrate GPT-Style A.I. into Its Terminal' (*CNBC*, 13 April 2023) <https://www.cnbc.com/2023/04/13/bloomberg-plans-to-integrate-gpt-style-ai-into-its-terminal.html> accessed 27 May 2023.

[288] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), [2016], OJ L 119, 4.5.2016, p. 1–88.

GDPR come analizzato nel capitolo 2, potrebbe già contenere sufficienti principi per regolare questi nuovi sistemi di IA ad uso generale.
Di seguito si procederà analizzando i nuovi modelli di IA e le disposizioni previste nelle ultime versioni della normativa per cercare di regolare tale fenomeno, esaminando poi il caso ChatGPT e di come questo sistema di IA violasse la normativa sulla protezione dei dati. Concluderemo l'elaborato riassumendo i risultati ottenuti e i punti della regolamentazione su cui si dovrebbero concentrare gli sforzi del legislatore.

2. I *foundation models* nell'AI Act

Il rapido avanzamento e la recente diffusione capillare dei sistemi GPAI sottolineano l'imperatività di rivedere il progetto normativo. Sia il Consiglio dell'Unione Europea che le commissioni del Parlamento europeo hanno tenuto in debita considerazione queste recenti problematiche, integrando nei loro emendamenti alcune disposizioni specifiche relative ai sistemi di GPAI. La motivazione cardine che ha indotto i ministri dei paesi membri a promuovere l'inserimento di tali sistemi nel *corpus* normativo è la consapevolezza che, in assenza di tale previsione, alcuni di questi sistemi di IA sarebbero rimasti al di fuori dell'ambito di applicazione del regolamento, compromettendo significativamente la sua efficacia e generando una profonda lacuna normativa. Pertanto, risulta fondamentale trattare adeguatamente questa categoria di sistemi al fine di garantire una regolamentazione completa ed efficace nell'interesse di tutti gli attori coinvolti.
Al riguardo il 12 settembre 2022, nel documento presentato dalla Commissione giustizia del Parlamento europeo[289], contenente le proposte di emendamento all'AI Act, è stata introdotta all'art. 3, paragrafo 1, punto (1a), una definizione di sistema "*general purpose AI*". Tale sistema viene definito come un "*sistema di IA che* -

[289] Opinion of the Committee on Legal Affairs for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))

indipendentemente dalla modalità in cui viene messo sul mercato o messo in servizio, anche come software open source - è destinato dal fornitore a svolgere funzioni applicabili generalmente quali riconoscimento di immagini o riconoscimento vocale, generazione audio o video, rilevamento di schemi, risposta a domande, traduzione o ulteriori funzioni; un sistema di IA ad uso generale può essere utilizzato in una molteplicità di contesti e può essere integrato in una pluralità di altri sistemi di IA"[290].

La questione rimane complessa a causa dell'intrinseca difficolta nel definire una tecnologia così innovativa, inedita e dalle applicazioni virtualmente illimitate[291]. I sistemi di GPAI, infatti, possiedono capacità di apprendimento pari e, in alcuni casi, superiori, a quelle degli esseri umani, adattandosi a situazioni ignote e operando in contesti complessi che richiedono l'elasticità cognitiva degli individui. L'analisi della questione deve tener conto di due aspetti: il primo è l'ambiguità generata dall'assenza di una definizione ufficiale di GPAI; il secondo riguarda la complessità nel definire una tecnologia dalle così ampie funzionalità. In altre parole, questi sistemi di IA sembrano possedere alcune caratteristiche che sono comunemente associate ad un comportamento intelligente. Nella definizione normativa di queste tecnologie è preliminarmente necessario comprendere e stabilire cosa si intenda per "generico" o "attività generiche". A tal fine potrebbe risultare utile l'utilizzo di parametri o indicatori come l'abilità, il

[290] Ibid. Il testo originale dell'emendamento 36 viene riportato di seguito: "*(1a) general purpose AI system' means an AI system that - irrespective of the modality in which it is placed on the market or put into service including as open source software - is intended by the provider to perform generally applicable functions such as image or speech recognition, audio or video generation, pattern detection, question answering, translation or others; a general purpose AI system may be used in a plurality of contexts and may be integrated in a plurality of other AI systems;*".

[291] Una definizione alternativa è stata proposta dal Future of Life Institute, organizzazione non governativa con base a Cambridge (Massachusetts) con lo scopo di ridurre i rischi e prevenire le minacce alla sopravvivenza della specie umana. Esso ha definito il sistema di GPAI come "un sistema di IA che può compiere o essere modificato per compiere una serie di operazioni distinte, includendo operazioni per le quali non era stato intenzionalmente e direttamente addestrato". Si veda: Carlos Ignacio Gutierrez and others, 'A Proposal for a Definition of General Purpose Artificial Intelligence Systems' (5 October 2022) <https://papers.ssrn.com/abstract=4238951> accessed 2 May 2023.

dominio, il risultato e il compito[292]. Se un sistema di IA possiede una sola abilità principale, potrà, in questo modo, essere considerato "meno generico" rispetto ad un altro che combina due o più abilità. Allo stesso modo, se il sistema viene applicato in più campi o ambiti, lo si potrà considerare più generico rispetto ad un sistema applicato in uno solo. Il medesimo percorso logico va seguito per i parametri di "risultato" e di "compito". Per entrambi, infatti, un sistema potrà essere considerato più generico rispetto ad un altro se assume una dimensione trasversale, riuscendo a completare compiti diversi e fornire differenti risultati [293].

La definizione dell'AI Act appena analizzata affronta e risolve una serie di problematiche, in particolare la necessità di stabilire una formulazione quanto più generale, in grado di mantenere aperta la possibilità di inclusione di ogni eventuale futura nuova applicazione attualmente non prevedibile dei sistemi di IA. Il secondo aspetto da prendere in considerazione è la notevole capacità di adattamento che contraddistingue queste tecnologie inedite, che potrebbero costituire la base su cui sviluppare altri sistemi più sofisticati. D'altro canto, la terminologia utilizzata dal legislatore europeo nella seconda parte della definizione di sistema di GPAI, nel tentativo di inscrivere gli stessi entro le maglie di un regolamento, sembra non delimitare in modo esplicito il campo di applicabilità della norma. Non risulta semplice discernere se le categorie menzionate siano mere illustrazioni delle potenziali applicazioni, un elenco esemplificativo, o costituiscano una lista chiusa e vincolante, unico raggio d'azione della normativa. Questa scelta di identificare in maniera precisa le applicazioni dei sistemi di GPAI costituisce un *minus* della regolamentazione. Vista l'ampia gamma di applicazioni, spesso non ancora interamente prevedibili, di un sistema di IA, la ristrettezza a tali esemplificazioni, ove fossero vincolanti, potrebbe agevolare

---

[292] ibid 2.

[293] Scott Reed and others, 'A Generalist Agent' <https://arxiv.org/abs/2205.06175> accessed 2 May 2023. Gato è un esempio di intelligenza artificiale ad uso generale (General Purpose AI) che viene utilizzato in una vasta gamma di compiti. Grazie ad una particolare configurazione, Gato può svolgere molteplici funzioni come comunicare attraverso il dialogo, creare didascalie per immagini, utilizzare un braccio robotico per impilare blocchi, eccellere in giochi Atari rispetto agli esseri umani, esplorare ambienti virtuali in 3D, seguire istruzioni e molto altro.

l'elusione della norma. Al contrario, attraverso una definizione troppo ampia e generale, si rischierebbe di coprire qualsiasi sistema GPAI, rendendo l'innovazione meno agevole. Ciò è dovuto al fatto che l'introduzione di limitazioni e requisiti regolamentari per l'ingresso nel mercato potrebbe disincentivare i nuovi sviluppatori. Anche a seguito di spinte internazionali e diverse critiche, la definizione di GPAI è stata rivista e all'interno del documento consuntivo degli emendamenti all'AI Act pubblicato il 16 maggio 2023. I sistemi di GPAI vengono dunque definiti all'articolo 3, n. 1d)[294]:

"*Per "sistema di IA ad uso generale" si intende un sistema di IA che può essere utilizzato e adattato a un'ampia gamma di applicazioni per le quali non è stato intenzionalmente e specificamente progettato."*[295].

La necessità di fornire una definizione a tali sistemi nasce dal carattere generale che contraddistingue la loro finalità. Inoltre, nel caso in cui non si fosse considerata tale tipologia di sistema di IA, sarebbero sorte diverse difficoltà nell'attribuzione della responsabilità lungo la cd. catena del valore dell'IA. La comprensione di tale catena del valore si cela di fronte la realtà che riguarda la distribuzione dei sistemi di IA e in particolare di quelli GPAI. È di fondamentale importanza comprendere come la semplice distinzione tra fornitore e utente non sia più sufficiente ad individuare tutti gli attori nel mercato dei sistemi di IA e su cui potrebbe ricadere la responsabilità per i danni prodotti dagli stessi sistemi. Dalla lettura dell'ultima formulazione del considerando 60 dell'AI Act presente nel documento consuntivo degli emendamenti, si

[294] Draft Compromise Amendments on the Draft Report, Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts", (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD)), art. 3, n. 1d).

[295] La definizione è stata tradotta dall'inglese. Quella originale viene riportata di seguito: "*'general purpose AI system' means an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed*

apprezza la complessità di tale mercato in cui anche la comunità *open-source*[296] si sta sviluppando molto velocemente[297]. Questo considerando dell'AI Act mette in evidenza la complessità della catena del valore nell'IA, individuando le diverse entità che contribuiscono allo sviluppo di un sistema di IA. Viene sottolineata la necessità di una cooperazione trasparente tra queste entità, nel rispetto dei rispettivi diritti di proprietà intellettuale, per garantire che i fornitori di sistemi di IA possano esercitare un adeguato controllo sulla conformità del sistema ai requisiti del prossimo regolamento.

Nella realtà dei fatti neanche i produttori di sistemi di GPAI sono capaci di anticipare con ragionevole sicurezza le funzioni e le mansioni che tali sistemi potrebbero soddisfare nelle mani degli utenti o anche di altri fornitori. Tale incapacità si contrappone all'intenzione del legislatore europeo di etichettare il sistema sulla base del rischio derivante dalla finalità dello stesso, in quanto, non risulta prevedibile quali siano le finalità specifiche del sistema di GPAI. Sulla base della nuova formulazione

---

[296] I sistemi di IA open-source sono piattaforme o framework che offrono al pubblico l'accesso al codice sorgente e alle risorse necessarie per sviluppare e utilizzare applicazioni di intelligenza artificiale. Questi sistemi permettono agli sviluppatori di accedere, modificare e distribuire liberamente il software, consentendo una collaborazione aperta e la condivisione delle conoscenze nel campo dell'IA.

[297] Ibid. il considerando 60) chiarisce che: "*Nella catena del valore dell'IA, molte entità spesso forniscono strumenti e servizi, ma anche componenti o processi che poi vengono incorporati dal fornitore nel sistema di IA, compresi in relazione alla raccolta e al pre-processamento dei dati, all'addestramento del modello, al suo riaddestramento, ai test e alla valutazione del modello, all'integrazione nel software, o ad altri aspetti dello sviluppo del modello. Le entità coinvolte possono rendere disponibile la loro offerta commercialmente, direttamente o indirettamente, attraverso interfacce, come le API (Application Programming Interfaces), e distribuite sotto licenze gratuite e open source, ma anche sempre più tramite piattaforme di lavoro IA, la rivendita di parametri addestrati, kit fai-da-te per costruire modelli o l'offerta di accesso a pagamento a un'architettura di servizio di modelli per sviluppare e addestrare modelli. Alla luce di questa complessità della catena del valore dell'IA, tutte le parti terze rilevanti, in particolare quelle che sono coinvolte nello sviluppo, nella vendita e nella fornitura commerciale di strumenti software, componenti, modelli pre-addestrati o dati incorporati nel sistema di IA, o fornitori di servizi di rete, dovrebbero, senza compromettere i propri diritti di proprietà intellettuale o segreti commerciali, rendere disponibili le informazioni, la formazione o l'esperienza richieste e cooperare, se appropriato, con i fornitori per consentire il loro controllo su tutti gli aspetti pertinenti alla conformità del sistema di IA che rientra sotto questo Regolamento. Per consentire una gestione efficace dal punto di vista dei costi della catena del valore dell'IA, il livello di controllo deve essere esplicitamente divulgato da ogni terza parte che fornisce al fornitore uno strumento, un servizio, un componente o un processo che viene poi incorporato dal fornitore nel sistema di IA*."

dell'articolo 28, paragrafo 1, "*Qualsiasi "distributore", "importatore", "implementatore" o altra terza parte sarà considerato un fornitore di un sistema di IA ad alto rischio ai fini del presente Regolamento e sarà soggetto agli obblighi del fornitore ai sensi dell'articolo 16, in una delle seguenti circostanze*" e alla lett. *ba*), viene specificata l'acquisizione della qualifica di fornitore nel caso in cui "*Essi apportano una* ***modifica sostanziale*** *a un sistema di IA, compreso un* ***sistema di IA di uso generale****, che non è stato classificato come ad alto rischio e che è già stato immesso sul mercato o messo in servizio in modo tale che il sistema di IA* ***diventa un sistema di IA ad alto rischio*** *in conformità all'articolo 6*." (grassetto utilizzato per evidenziare il punto del ragionamento). Secondo la nuova formulazione, dunque, un sistema di GPAI che viene modificato da un soggetto diverso dal fornitore per soddisfare finalità che determinano un alto rischio, porta l'equiparazione del regime di responsabilità tra tale soggetto e il fornitore. Inoltre, secondo il paragrafo 2 il precedente fornitore non dovrà più essere considerato tale e le stesse prescrizioni della nuova formulazione dell'articolo 28 si applicano nel caso in cui tale operazione di modifica avvenga su un *foundation model*, circostanza che mi sembra assai più probabile, considerando le differenze tra i due modelli sottolineate in precedenza e la natura dei *foundation models* come base per lo sviluppo di sistemi di IA specifici.

Per la loro mutevolezza e per la loro velocità di evoluzione, questi sistemi di IA non si prestano a definizioni chiare e univoche che permettano al legislatore europeo di formulare una regolamentazione. Volendo indicare tutti gli utilizzi e classificarli secondo per rischio, il legislatore non riuscirebbe in alcun modo a racchiudere tutte le applicaizioni di questi sistemi. Un approccio che sfugge da queste problematiche di definizione è quello che individua i vari principi etici che tutti i sistemi di IA devono rispettare, a prescindere dalla tipologia del sistema. Tale approccio è stato recentemente proposto dalla Commissione giustizia del Parlamento europeo, sulle orme del lavoro svolto dall' AI HLEG.

3. La *privacy* come principio per lo sviluppo dei sistemi di IA

Con l'introduzione dei sistemi di GPAI, nell'AI Act è stata anche prevista una disposizione che, a prescindere dalla categoria o tipologia di sistema di IA, richiede il rispetto di alcuni dei principi etici individuati dal Rapporto del AI HLEG[298]. La disposizione si trova all'emendamento 42 della Commissione giustizia del Parlamento europeo[299], già citato in precedenza, ed è stata votata favorevolmente dal Parlamento europeo nella votazione del 14 giugno 2023 [300]. Tale norma risulta essere centrale per la tutela dei principi etici, concernenti anche la protezione dei dati personali degli utenti che fanno uso di sistemi di IA. Questo emendamento ha determinato una sostanziale revisione del precedente articolo 4, in quanto nella versione del 2021 l'articolo

[298] European Commission. Directorate General for Communications Networks, Content and Technology. and High Level Expert Group on Artificial Intelligence. (n 91).

[299] Cfr. (n 263); Il testo proposto per l'articolo 4, paragrafo uno, dell'AI Act all'emendamento 42 prevede che: "*All AI operators shall respect the following general principles that establish a high-level framework that promotes a coherent human-centric European approach to ethical and trustworthy Artificial Intelligence, which is fully in line with the Charter as well as the values on which the Union is founded; 'human agency and oversight' means that AI systems shall be developed and used as a tool that serves people, respects human dignity and personal autonomy, and that is functioning in a way that can be appropriately controlled and overseen by humans; 'technical robustness and safety' means that AI systems shall be developed and used in a way to minimize unintended and unexpected harm as well as being robust in case of unintended problems and being resilient against attempts to alter the use or performance of the AI system so as to allow unlawful use by malicious third parties; 'privacy and data governance' means that AI systems shall be developed and used in compliance with existing privacy and data protection rules, while processing data that meets high standards in terms of quality and integrity; 'transparency' means that AI systems shall be developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system as well as duly informing users of the capabilities and limitations of that AI system and affected persons about their rights; 'diversity, non-discrimination and fairness' means that AI systems shall be developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law; 'social and environmental well-being' means that AI systems shall be developed and used in a sustainable and environmentally friendly manner as well as in a way to benefit all human beings, while monitoring and assessing the long term impacts on the individual, society and democracy*."

[300] *Cfr*. n 255; European Parliament, P9_TA(2023)0236, Artificial Intelligence Act, Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), [2023]

contemplava una disposizione sulle modifiche dell'Allegato I [301]. In particolare, l'emendamento 42 ha modificato l'articolo 4, paragrafo 1, delineando i principi fondamentali applicabili a tutti i sistemi di IA e imponendo ai gestori di tali sistemi il dovere di rispettarli. L'intento di istituire un quadro normativo che incentivasse uno sviluppo etico ed antropocentrico dei sistemi di IA viene ribadito nella nuova disposizione, che incapsula quasi tutti i principi individuati dal AI HLEG [302]. Tali principi risulterebbero così applicabili anche ai sistemi GPAI, permettendo uno sviluppo più etico sin dalla loro progettazione, seppur privi di una finalità determinata.
I principi generali da applicare ai sistemi di IA delineati nell'emendamento sono: (i) il coinvolgimento umano, in particolare per la supervisione; (ii) la robustezza tecnica, concepita come metodo per mitigare i danni; (iii) la privacy e la governance dei dati, che presuppone il rispetto delle leggi vigenti sulla protezione dei dati; (iv) la trasparenza, che assicura la rintracciabilità e la "*explainability*" dei processi; (v) la diversità e la non-discriminazione, per prevenire soprattutto effetti discriminatori e bias; (vi) il benessere sociale e ambientale, che consente di valutare anche i più ampi e potenziali impatti a lungo termine. Risulta pertanto evidente che l'introduzione degli stessi nel testo del regolamento sull'IA rappresenti una modifica fondamentale per garantire una cornice etica rigorosa e coerente all'interno della proposta legislativa sull'IA.

Focalizzandoci sul principio relativo alla privacy, questo potrebbe rappresentare il primo riferimento alla normativa sulla protezione dei dati. Dai Garanti della Privacy degli Stati membri e dal Garante europeo già da tempo era stata richiesta una definizione più precisa in merito ai rapporti tra la legislazione attualmente vigente sulla

---

[301] Il testo del precedente articolo 4 prevedeva che: "*Alla Commissione è conferito il potere di adottare atti delegati conformemente all'articolo 73 al fine di modificare l'elenco delle tecniche e degli approcci di cui all'allegato I, per aggiornare tale elenco agli sviluppi tecnologici e di mercato sulla base di caratteristiche simili alle tecniche e agli approcci ivi elencati*".

[302] European Commission. Directorate General for Communications Networks, Content and Technology. and High Level Expert Group on Artificial Intelligence., Ethics Guidelines for Trustworthy AI. (Publications Office 2019) <https://data.europa.eu/doi/10.2759/346720> accessed 24 May 2023.

protezione dei dati e la nuova proposta di regolamento[303], sottolineando l'importanza di prevenire conflitti e sovrapposizioni normative, assicurando al contempo la certezza del diritto e la salvaguardia dei diritti inerenti alla protezione dei dati personali. I Garanti hanno anche sollecitato un'attenta considerazione di quanto stabilito all'interno dell'art. 22 del GDPR, relativo ai processi decisionali automatizzati che coinvolgono direttamente le tecnologie di IA, sottolineando, ancora una volta, l'obbligo per i sistemi di IA di garantire fin dallo sviluppo l'esercizio e la tutela dei diritti degli interessati, quali la cancellazione e la rettifica, senza alcuna distinzione basata sulle differenti categorie di IA. Tuttavia, l'esercizio di tali diritti nei sistemi GPAI non risulta sempre possibile a causa del loro complesso funzionamento, spesso a "scatola chiusa".

Il rispetto della normativa sulla protezione dei dati, in particolare il GDPR, è indiscutibilmente uno degli aspetti più problematici per i sistemi di IA. Studi recenti hanno infatti dimostrato che i sistemi di GPAI risultano anche più vulnerabili di fronte a specifici attacchi che mirano a estrapolare i dati presenti nel suo *dataset* attraverso l'interazione con il modello[304]. Il risultato di questi attacchi sarebbe proprio quello di ottenere dai modelli i dati utilizzati per l'addestramento, che possono anche rientrare nella categoria di dati personali. Secondo l'articolo 6 del GDPR, per qualsiasi trattamento o utilizzo di dati personali è necessaria una base giuridica. Tra queste, il consenso del soggetto interessato ex art. 6 para. 1, lett. a), a cui si riferiscono i dati è la base giuridica più "forte" anche se, spesso, gli sviluppatori prediligono l'utilizzo del legittimo interesse ex art. 6, para.1, lett. f)[305]. Considerando la mole di dati utilizzati dai

[303] EDPB – GEPD Parere congiunto 5/2021 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale), 18 giugno 2021, p 19

[304] Per una panoramica si vedano: Liu, Yi, Gelei Deng, Yuekang Li, Kailong Wang, Tianwei Zhang, Yepang Liu, Haoyu Wang, Yan Zheng, and Yang Liu. "Prompt Injection attack against LLM-integrated Applications." arXiv preprint arXiv:2306.05499 (2023); Alexander Wei, Nika Haghtalab, and Jacob Steinhardt. "Jailbroken: How Does LLM Safety Training Fail?." arXiv preprint arXiv:2307.02483 (2023).

[305] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), [2016], OJ L 119, 4.5.2016, p. 1–88 art 6.

sistemi GPAI, appare difficile dimostrare l'ottenimento del legittimo consenso da parte di tutti gli interessati. Bisogna anche evidenziare che, nel caso in cui il dato personale costituisca dato sensibile secondo l'articolo 9 del GDPR[306], saranno necessarie maggiori misure di sicurezza e una solida base giuridica. In tal senso, la tipologia di attacchi sopracitata, fortemente efficace sui modelli in analisi, costituisce un serio problema in quanto attraverso gli stessi, tra i dati che il modello è forzato a rivelare, potrebbero essere presenti dati altamente sensibili e quindi con un profondo impatto sui diritti e le libertà fondamentali degli utenti.
A tal riguardo, i sistemi GPAI che negli ultimi hanno riscontrato maggiore successo e adozione sono risultati i già citati modelli di elaborazione del linguaggio naturale. Addestrati su ampi *dataset* provenienti principalmente da Internet, tali modelli sono in grado di svolgere un'ampia gamma di compiti legati all'elaborazione del linguaggio naturale, tra cui la traduzione, il riassunto e la risposta a domande. Il mercato di questi sistemi ha registrato una crescita esponenziale, attirando l'attenzione di molti e portando ad un'accelerazione nel loro sviluppo.

Le autorità responsabili della regolamentazione dei principali progressi tecnologici, in particolare i Garanti della *Privacy* nazionali, hanno seguito attentamente l'evoluzione di questi modelli, vigilando sul rispetto delle normative. Considerando la diffusione di servizi come ChatGPT, che avrebbe raggiunto oltre cento milioni di utenti nel giro di poche settimane, le autorità per la protezione dei dati personali hanno deciso di indagare. Il Garante della *Privacy* italiano, in particolare, ha assunto un ruolo cruciale nella prevenzione di potenziali danni ai diritti e alle libertà dei cittadini, adottando misure d'urgenza nei confronti di alcuni sistemi di IA.
Con un primo provvedimento del 2 febbraio 2023, il Garante ha imposto una limitazione provvisoria al trattamento dei dati personali condotto da Luka Inc., società proprietaria del software di IA Replika[307] e titolare del trattamento dei dati ex art. 4, n. 7

[306] ibid 9.

[307] Garante per la protezione dei dati personali, Provvedimento del 2 febbraio 2023, Doc-Web n. 9852214; Comunicato stampa 'Intelligenza artificiale, dal Garante privacy stop al chatbot "Replika". Troppi i rischi per i minori e le persone emotivamente fragili' Doc-Web n. 9852506.

del GDPR[308]. Secondo l'analisi condotta dall'autorità, questo software, una *chatbot* personalizzabile dall'utente con possibilità d'interazione scritta e vocale (che crea un vero e proprio "amico virtuale"), avrebbe potuto mettere in pericolo la stabilità mentale dei minori di 13 anni e delle persone emotivamente vulnerabili, più soggetti a formare legami con il sistema attraverso tali interazioni. Questo sistema risultava in grado di compiere diverse tipologie di compiti, tutti basati sull'interazione uomo - macchina e l'elaborazione del linguaggio naturale. Ad esempio, esso poteva sostenere una conversazione di base, fornire supporto emotivo, promuovere il benessere mentale, imparare dai comportamenti degli utenti e perfino partecipare ad un gioco di ruolo. Il Garante ha evidenziato diverse problematiche relative all'utilizzo dell'applicazione, tra cui l'assenza di un filtro per i minori di 13 anni, la mancanza di verifiche relative all'età degli utenti in fase di registrazione, di meccanismi di interdizione o blocco anche in presenza di dichiarazioni dell'utente che indichino un'età inferiore ai 13 anni[309] e la non conformità della *privacy policy* alle disposizioni in termini di trasparenza[310], specialmente nell'utilizzo dei dati personali dei minori[311]. Su queste basi, il Garante ha giudicato il trattamento dei dati personali in violazione degli articoli 5, 6, 8, 9 e 25 del GDPR, imponendo una limitazione provvisoria del trattamento dei dati personali condotto da Luka Inc. per gli utenti residenti sul territorio italiano.

L'importanza di verificare l'età, o almeno di implementare meccanismi con tale finalità, è cruciale. Infatti, l'autorità ha evidenziato come, in contrasto con il contenuto dell'articolo 8 del GDPR, i dati degli utenti minori di 13 anni venivano trattati allo

---

[308] L'articolo 4, numero 7 del GDPR definisce titolare del trattamento come: "*la persona fisica o giuridica, l'autorità pubblica, il servizio o altro organismo che, singolarmente o insieme ad altri, determina le finalità e i mezzi del trattamento di dati personali; quando le finalità e i mezzi di tale trattamento sono determinati dal diritto dell'Unione o degli Stati membri, il titolare del trattamento o i criteri specifici applicabili alla sua designazione possono essere stabiliti dal diritto dell'Unione o degli Stati membri*"

[309] GDPR, art. 8

[310] GDPR, artt. 5 e 12 del GDPR

[311] GDPR, ex artt. 5 e 6

stesso modo di quelli di qualsiasi altro utente, senza rispettare le specifiche disposizioni in materia[312] e senza garantire il maggior livello di attenzione e tutela necessario per gli individui minori ed emotivamente vulnerabili[313].

La reale capacità di questi sistemi di elaborazione del linguaggio è successivamente venuta alla luce con la diffusione del software ChatGPT, immesso sul mercato nel novembre 2022. Il software ha guadagnato notevole visibilità sia a livello nazionale che internazionale, portando le autorità indipendenti di controllo, in particolare le autorità per la protezione dei dati personali, ad interrogarsi sulla conformità delle pratiche di ChatGPT con le norme del Regolamento UE 2016/679 volto a salvaguardare i diritti relativi alla protezione dei dati personali.

Le problematiche relative alla sicurezza dei dati, all'accesso dei minori a tali sistemi e alla mancanza di controlli sull'età fanno emergere la crescente necessità di regolare alcuni aspetti di tale tecnologia, imponendo standard che permettano di arginare i pericoli che già si mostrano. Per preservare i diritti e le libertà dei cittadini sarà però necessario un continuo aggiornamento di tali norme, permettendone la modifica e l'aggiunta. I rischi che si presentano con l'utilizzo di tali sistemi emergono solo con un continuo utilizzo degli stessi da parte degli utenti. Al contempo prescindere dall'uso di tali sistemi non sarà possibile, il loro carattere rivoluzionario risulta essere alla pari di quello dei computer ed è evidente come la società globale sia indirizzata verso un utilizzo sempre maggiore di questi sistemi automatici.

4. Analisi delle funzioni di ChatGPT attraverso il test di Turing

Prima di procedere all'analisi della vicenda che ha visto il Garante italiano sospendere l'accesso all'interfaccia di ChatGPT ai cittadini italiani, risulta utile, per chiarezza

---

[312] Anzitutto, l'art. 8 del GDPR prevede che gli Stati membri possano stabilire un'età minima per la validità del consenso per il trattamento dei dati personali non inferiore ai 13 anni, con la conseguenza che il trattamento dei dati del soggetto infratredicenne possa avvenire solo con il consenso di chi esercita la potestà genitoriale. Inoltre, secondo il considerando 58 del GDPR i minori meriterebbero un'informativa sul trattamento dei loro dati adeguata alla loro comprensione, quindi che sia formulata in un linguaggio chiaro e semplice di immediata comprensione.

[313] Chelsea Jarvie and Karen Renaud, 'Are You over 18? A Snapshot of Current Age Verification Mechanisms', *2021 Dewald Roode Workshop* (2021) 9–14 <https://strathprints.strath.ac.uk/82540/> accessed 28 May 2023.

espositiva e nei limiti dell'economia di questo elaborato, illustrare il funzionamento di tale sistema di IA. ChatGPT è una *chatbot* altamente avanzato, in grado di intrattenere conversazioni in un'ampia gamma di idiomi, utilizzando un sistema di IA sviluppato dalla società statunitense OpenAI. Questa tecnologia è basata sull'architettura GPT-3.5 (*Generative Pretrained Transformer* versione 3.5), concepita per interpretare e rispondere alle richieste degli utenti in un modo fluido e coerente. È stata sviluppata per adempiere ad un ampio spettro di impieghi, tra cui l'assistenza ai clienti, la generazione di contenuti, la traduzione di testi e frasi e molte altre attività basate sull'interpretazione linguistica. Rappresenta un modello linguistico ideato per produrre sequenze di parole, codici e altri dati, iniziando da un input dato che, grazie alla tecnologia *deep learning*[314], genera testi che risultano similari a quelli composti da esseri umani. Il modello di natura statistica richiede un volume ingente di dati per il suo apprendimento, questi prendono la forma di parametri, cioè regole o frammenti di informazioni che il modello usa per fare previsioni o rispondere a domande, come se avesse una grande biblioteca di informazioni da cui attingere per rispondere[315].

Tale tecnologia si presenta come estremamente intuitiva e di facile utilizzo per l'utente finale: il software, infatti, a partire dal testo inserito, completerà la frase o genererà una nuova sequenza di parole senza elaborare una reale comprensione della nostra richiesta, bensì predicendo le parole statisticamente più appropriate[316]. In ogni caso, l'utilizzo di questo sistema di IA per la scrittura di articoli scientifici è stato pesantemente criticato, così come l'assegnazione a ChatGPT del ruolo di autore di un documento[317]. Le principali preoccupazioni riguardano il possibile aumento del plagio ed un controllo minore sui principi etici che ispirano la scrittura accademica, proprio perché risulta sempre più complicato poter distinguere tra la scrittura umana e quella della *chatbot*. Un'ulteriore problematica che potrebbe emergere riguarda un crescente affidamento che

[314] *Cfr.* (n 66)

[315] Tom B Brown and others, 'Language Models Are Few-Shot Learners' <https://arxiv.org/abs/2005.14165> accessed 5 May 2023.

[316] Mollick (n 262).

[317] H Holden Thorp, 'ChatGPT Is Fun, but Not an Author' (2023) 379 Science 313.

gli studenti farebbero su tali sistemi, determinando al contempo una diminuzione di creatività e innovazione dei loro scritti. Inoltre, i contenuti accademici forniti dal sistema di IA possono perfino risultare immaginari e quindi inesistenti. Sebbene le capacità di "ragionamento" di tali sistemi abbiano ottenuto punteggi rilevanti in diversi test pensati per gli umani[318], resta la preoccupazione che un maggiore affidamento a tali tecnologie porti ad una diminuzione qualitativa degli elaborati accademici[319].

Un ulteriore problema riguarda l'adesione del modello ai principi etico-giuridici di riferimento nel contesto in cui viene utilizzato. Tale problema nasce dal fatto che, essendo il sistema addestrato su una grande mole di dati estrapolati dal *web*, può mancare un criterio etico discriminante in grado di escludere contenuti potenzialmente capaci di ledere la sensibilità degli individui, in particolare dei minori. Nella prima versione del software risultava possibile richiedere la scrittura di un testo con contenuti di natura razzista, in quanto il modello, nella sua "conoscenza", conteneva anche parametri discriminatori che, statisticamente, portavano il sistema di IA ad identificarli come risultati più probabili e coerenti. Nelle versioni successive si è però assistito ad un progressivo affinamento, il cd. *fine-tuning*, ossia una modifica del prodotto per prevenire la produzione di contenuti discriminatori o che possano danneggiare la sensibilità degli individui. Questo processo di miglioramento è tuttora in corso e i programmatori di OpenAI lavorano incessantemente per perfezionare il modello e renderlo sempre più sicuro, etico e conforme alle normative vigenti.

Dal 30 novembre 2022, OpenAI ha scelto di rendere disponibile al pubblico questo sofisticato sistema di IA ad uso generale, mediante l'interfaccia ChatGPT[320]. Era possibile accedervi gratuitamente e il numero di utenti registrato nei primi mesi superò qualsiasi aspettativa.

---

[318] Jonathan H Choi and others, 'ChatGPT Goes to Law School' [2023] SSRN Electronic Journal <https://www.ssrn.com/abstract=4335905> accessed 27 April 2023.

[319] Fawaz Qasem, 'ChatGPT in Scientific and Academic Research: Future Fears and Reassurances' (2023) 40 Library Hi Tech News 30.

[320] John Schulman et al., ChatGPT: Optimizing Language Models for Dialogue, OPENAI [2022], <https://openai.com/blog/chatgpt/.>

Un esperimento condotto nel 2020 sulla versione GPT-3[321], che comprendeva domande di diversa natura, ha evidenziato una tendenza verso risposte più etiche e meno discriminatorie quando riproposto nel 2023 sulla versione GPT-3.5, attraverso l'interfaccia di ChatGPT.

Nell'esperimento del 2020 condotto dal professor Luciano Floridi e da Massimo Chiriatti[322], il sistema di GPAI è stato sottoposto a tre test, con il primo avente natura matematica, il secondo semantica e il terzo etica. GPT-3 non ha ottenuto un buon risultato in matematica e quando interpellato con questioni come "quanti piedi entrano in una scarpa", non è stato in grado di comprendere a pieno il contesto, fornendo risposte irrilevanti e fuori luogo. Al contrario, presentando la medesima questione alla versione 3.5, la risposta è risultata precisa e coerente, dimostrando un progresso notevole nella comprensione del contesto per la risoluzione del problema. Anche per semplici operazioni matematiche, la versione più recente sembra aver fatto progressi, sebbene non sia stata concepita per operare come una calcolatrice.

Il terzo test ha un particolare rilievo nell'ambito dei diritti fondamentali: infatti, nella versione GPT-3, l'intelligenza artificiale aveva mostrato una predisposizione al razzismo, in quanto, quando interrogata riguardo alle persone di etnia afrocaraibica, la macchina rispondeva con commenti esplicitamente discriminatori. Risultati analoghi si registravano anche per altre forme di discriminazione. Nella versione 3.5, invece, l'IA sembra aver acquisito una maggiore "consapevolezza" relativamente ai diritti umani, fornendo risposte che prendono in considerazione i concetti di uguaglianza, parità di diritti e dignità, e sottolineando che, in qualità di assistente virtuale, non possiede opinioni personali, discriminazioni o pregiudizi. Quest'ultima affermazione, tuttavia, deve essere ponderata in relazione al livello di pregiudizio presente nei *dataset* dell'IA; infatti, sebbene sia possibile adottare misure per ridurre e nascondere tali pregiudizi, eliminarli completamente risulta essere un'operazione ardua e complessa, anche se

[321] Luciano Floridi and Massimo Chiriatti, 'GPT-3: Its Nature, Scope, Limits, and Consequences' (2020) 30 Minds and Machines 681, 688–690.

[322] Per Luciano Floridi, *cfr*. (n 15). Massimo Chiriatti è il Chief Technical and Innovation Officer di Lenovo.

esistono già i alcuni studi relativi al *machine unlearning*, ossia la possibilità di rimuovere in maniera precisa parametri individuati all'interno del *dataset* prevenendo, in tal modo, non solo eventuali violazioni di *privacy* come quelle considerate dal Garante della *Privacy* italiano che vedremo in seguito, ma anche eventuali pregiudizi che emergono dall'interazione con i sistemi di IA[323].

Considerando che il sistema aggiornato è stato sottoposto nuovamente ai tre test, è lecito interrogarsi sulle aspettative riguardo al suo sviluppo futuro, alla luce del progresso significativo dimostrato nella sua capacità di fornire risposte più accurate ed efficaci a soli tre anni di distanza. Volendo immaginare una delle possibili conseguenze di un utilizzo malevolo di tali sistemi, sarà sufficiente combinare la facilità di creazione di contenuti di queste tecnologie con un utilizzo manipolatorio dei social network in operazioni come quella di Cambridge Analytica, per ottenere un futuro in cui i contenuti generati dall'IA saranno ancora più personalizzati, pervasivi e persuasivi. Un tale impiego della tecnologia potrebbe generare seri danni per gli utenti e per la società in generale.

Per affrontare i problemi derivanti da un uso dei sistemi di IA per scopi di manipolazione e influenza degli individui, l'Unione Europea ha previsto, all'articolo 5 della proposta di regolamento, quattro pratiche di IA totalmente vietate[324]. In particolare, la lettera a) del primo comma proibisce l'utilizzo di IA che sfrutti tecniche subliminali, senza la consapevolezza dell'utente, al fine di influenzare il comportamento delle persone, causando un potenziale danno fisico o psicologico a quella persona o ad altri.

---

[323] Luciano Floridi and others, 'How to Design AI for Social Good: Seven Essential Factors' (2020) 26 Science and Engineering Ethics 1771; Luciano Floridi, 'Machine Unlearning: Its Nature, Scope, and Importance for a "Delete Culture"' (arXiv, 24 May 2023) <http://arxiv.org/abs/2305.15242> accessed 28 May 2023.

[324] 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, [2021], COM/2021/206 final (n 88) art 5.

È essenziale il monitoraggio e la regolamentazione dell'utilizzo di tali tecnologie per garantire il rispetto dei diritti fondamentali e prevenire possibili abusi. Inoltre, l'educazione degli utenti sui potenziali rischi e sul funzionamento delle IA svolge un ruolo cruciale nel far fronte a queste sfide e assicurare un utilizzo responsabile e sicuro delle tecnologie emergenti. Rimane quindi fondamentale prestare attenzione all'interazione uomo-macchina, alla *privacy* e alla capacità di effettuare scelte in modo indipendente e non condizionato. L'utilizzo dell'IA può avere un impatto rilevante sul diritto alla libertà di pensiero di coscienza e di religione sanciti all'articolo 10 della Carta dei Diritti Fondamentali dell'Unione Europea e all'articolo 19 della Costituzione italiana [325]. Tali sistemi potrebbero essere utilizzati per manipolare o influenzare indebitamente le opinioni e le convinzioni delle persone che li utilizzano o che ne recepiscono i contenuti. Per tale motivazione, l'attenzione delle autorità competenti e degli organismi di vigilanza dovrà essere costante, in particolare, per evitare fenomeni di manipolazione di massa, come quelli che si sono già presentati con l'utilizzo dei *social network*. Tali piattaforme utilizzano i sistemi di IA per la personalizzazione dei contenuti, svolgendo in tal modo un ruolo cruciale nella formazione delle opinioni, delle percezioni e dei comportamenti delle persone. I *social networks* analizzando i dati sul comportamento degli utenti, elaborati attraverso i sistemi di IA, possono identificare inclinazioni, preferenze e idee e sfruttarle per modificare e orientare la scelta delle persone. Tale fenomeno potrebbe determinare una situazione in cui le persone sono esposte principalmente a contenuti che riflettono le loro opinioni esistenti, creando in tal modo delle cosiddette "bolle informative" e amplificando la polarizzazione delle opinioni. Tra le conseguenze più allarmanti di tale fenomeno risiedono: un aumento della divisione della società e il deterioramento del dibattito pubblico costruttivo. A prescindere dalla regolamentazione di tali fenomeni e tecnologie, per affrontare queste problematiche sarà fondamentale lavorare sull'educazione degli utenti per spiegare il funzionamento dei sistemi di IA e delle prossime innovazioni tecnologiche. Solamente attraverso l'educazione degli utenti si potranno evitare o comunque arginare comportamenti sbagliati e utilizzi dannosi di queste tecnologie. Gli utenti dovranno

[325] Costituzione della Repubblica Italiana art. 19; Carta dei diritti fondamentali dell'Unione europea, [2012] GU C 326 del 26.10.2012, art. 10.

sapere con certezza se stanno interagendo con un operatore umano o non umano e come comportarsi di conseguenza. Tale differenza si farà sempre più sottile con l'avanzamento tecnologico dei sistemi di IA che riescono ormai a adottare un linguaggio sofisticato e naturale, capace di instaurare il dubbio nelle persone.

Infatti, sulla base dei sistemi di GPAI sono state sviluppate le moderne *chatbot*, interfacce per l'interazione uomo-macchina, munite di avanzate capacità sintattiche e semantiche. Queste sono in grado di modulare il tono del discorso e selezionare accuratamente il lessico in base alla specificità dell'interlocutore, concretizzando un sofisticato strumento di persuasione potenzialmente personalizzabile per ciascun utente del web. In un'epoca di crescente interazione con queste tecnologie, le preferenze, le idee e le opinioni degli individui rischiano di essere sempre più penetrabili e suscettibili a manipolazione. Nel contesto dell'era dell'informazione l'interazione uomo-macchina si intensifica, spingendo molti a privilegiare le potenziali prospettive di incremento della produttività a discapito di una considerazione adeguata delle implicazioni relative alla tutela della *privacy* e della libertà di pensiero.

La *privacy* rappresenta un pilastro fondamentale per l'integrità dell'individuo e la sua capacità di esercitare scelte libere e non condizionate. Con l'evoluzione delle *chatbot*, diventa sempre più problematico discernere la natura dell'interlocutore in una conversazione, sia esso umano o macchina. Ad esempio, molte delle interazioni contemporanee con i *call center* sono già automatizzate e la possibilità di parlare con un operatore umano, capace di comprendere e gestire le problematiche senza adottare un approccio puramente algoritmico, è sempre più limitata[326]. Alla diffusione di questi sistemi di IA che possono generare testi, il mercato ha risposto concentrandosi ancora di più sullo sviluppo di questo tipo di sistemi. Tuttavia, la logica seguita dalle principali aziende del settore si è focalizzata più sulla celerità nel rilasciare un modello che fosse

[326] Nick Evershed and Josh Taylor, 'AI Can Fool Voice Recognition Used to Verify Identity by Centrelink and Australian Tax Office' *The Guardian* (16 March 2023) <https://www.theguardian.com/technology/2023/mar/16/voice-system-used-to-verify-identity-by-centrelink-can-be-fooled-by-ai> accessed 5 May 2023.

paragonabile a ChatGPT, a scapito di investimenti in un modello più sicuro e dotato di adeguati meccanismi di controllo.

Con l'adozione del nuovo AI Act, si potrebbe invece imporre alle aziende del settore di investire maggiormente sugli aspetti di sicurezza relativi ai sistemi di IA. A tal riguardo con gli ultimi emendamenti sono state introdotte specifiche limitazioni relative ai sistemi di GPAI, simili a ChatGPT. Tuttavia, data la lunghezza e la complessità del regolamento, ulteriori chiarimenti sulla sua applicazione saranno possibili soltanto con la pubblicazione dei relativi atti attuativi. Nel frattempo, persiste la mancanza di una reale delineazione degli standard tecnici da seguire per permettere una presunzione di conformità all'AI Act[327]. Tale attività è stata delegata ad organizzazioni private, le cui decisioni, benché non esplicitamente presenti nel testo del regolamento, potrebbero assumere un ruolo centrale nella determinazione della conformità dei prodotti alle disposizioni del regolamento[328].

Al contempo, tali sistemi sono sottoposti ad un attento esame da parte dei Garanti per la protezione dei dati di tutta Europa, in quanto, la normativa sulla protezione dei dati personali ed in particolare il GDPR è da applicare anche a tali inediti sistemi tecnologici.

5. La protezione dei dati personali nei sistemi di IA, il data breach di ChatGPT

[327] 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, [2021], COM/2021/206 final (n 88) art 40.

[328] European Commission, A Notification under Article 12 of Regulation (EU) No 1025/20121 (Dec. 2, 2022) (proposing a standardization request to the European Committee for Standardisation (CEN) and the European Committee for Electrotechnical Standardisation (CENELEC), two private standard-setting organizations)

L'elaborazione di un volume crescente di dati si è rivelata essenziale per migliorare le prestazioni dei sistemi di GPAI. Tuttavia, questa necessità di quantità di dati sempre crescente implica, di riflesso, anche l'aumento della raccolta di dati personali e dei rischi ad essa legati.
La proposta di regolamento non presta espressamente attenzione alle problematiche legate alla raccolta dei dati derivante dall'impiego di sistemi di IA ma, nel parere congiunto del Comitato Europeo per la Protezione dei Dati e del Garante Europeo per la Protezione dei Dati del giugno 2021 [329] , è stata evidenziata l'importanza dell'applicazione della legislazione europea relativa alla protezione dei dati personali anche a qualsiasi trattamento di dati nel raggio di applicazione dell'AI Act.
Nel suddetto parere, entrambe le autorità hanno enfatizzato come il rispetto del quadro giuridico in materia di protezione dei dati debba rappresentare un fondamento per la nuova proposta. Ciò non solo per preservare la coerenza dei principi europei in materia di protezione dei dati, ma anche per l'importanza che la legislazione in materia ormai riveste all'interno del panorama comunitario e globale.

Il contributo più rilevante all'attuazione del Regolamento UE 2016/679 è da attribuire ai Garanti della *Privacy* nazionali che con il GDPR hanno visto ampliare i propri poteri e le proprie responsabilità, così nel vigilare sull'applicazione delle disposizioni del GDPR, come nel disporre le sanzioni per una protezione effettiva dei dati personali dei cittadini europei. I Garanti della *Privacy* hanno avuto un ruolo cruciale nell'interpretazione delle disposizioni del GDPR, in particolare attraverso le numerose linee guida, i pareri e le diverse raccomandazioni. Specialmente le linee guida hanno aiutato ad orientare e ad adattare l'interpretazione dei principi della *Data Protection* a tecnologie in rapida evoluzione, che solo col tempo hanno mostrato nuove, potenziali, minacce di violazione di un diritto fondamentale delle persone come quello alla propria *privacy*. Il ruolo dello *European Data Protection Board*, comitato che riunisce tutti i Garanti nazionali e il Garante europeo della *Privacy*, è stato fondamentale per l'adozione di un approccio unitario, evitando divergenze tra gli stessi Garanti.

[329] EDPS – GEPD Parere congiunto 5/2021 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale)

Tuttavia, nonostante gli sforzi compiuti dai Garanti della Privacy, vi sono stati e continuano ad esserci *data breach* che interessano anche dati sensibili. La casistica dei *data breach* è di crescente rilevanza e i dati sensibili risultano essere il bersaglio preferito degli *hacker*. Le aziende sanitarie locali sono, infatti, spesso oggetto di questi attacchi, data l'alta sensibilità dei dati trattati e la frequente superficialità sugli standard di sicurezza. Ad esempio, il provvedimento del Garante dell'8 giugno 2023 relativo ad un *data breach* ha riguardato la "ASL 1 Avezzano Sulmona L'Aquila", questa è stata oggetto di un attacco hacker conclusosi - tragicamente - con la pubblicazione sul *dark web* dei dati sensibili dei pazienti [330]. Anche l'Istituto di Statistica nazionale, ISTAT, è stato oggetto nel 2022 di un attacco informatico che ha comportato l'esfiltrazione di alcune informazioni e il potenziale accesso non autorizzato da parte di terzi a informazioni di carattere personale. A tale attacco ha fatto poi seguito la segnalazione al Garante che ha dunque adottato il provvedimento sanzionatorio del 10 febbraio 2022[331]. Tale tipologia di attacco ha riguardato anche enti privati di qualsiasi grandezza e rappresenta uno dei maggiori pericoli per la *privacy* degli utenti. In particolare, secondo la relazione del Garante relativa all'anno 2022 sono stati 1351 i *data breach* notificati al Garante, di cui circa un terzo con bersaglio enti pubblici e due terzi enti privati[332].
Per l'economia di questo elaborato, uno degli esempi più rilevanti riguarda il *data breach* che ha interessato il sistema di GPAI, ChatGPT ed in particolare il suo servizio a pagamento "*ChatGPTplus*", sollevando preoccupazioni in merito alla sicurezza dei dati degli utenti.

La diffusione di ChatGPT, ed in particolare della sua versione "*plus"* che permette di accedere ad un modello più sofisticato del sistema, è stata interessata da una violazione dei dati personali causata da un problema tecnico (cd. *bug*). Nella settimana del 20 marzo 2023, il *bug* ha permesso ad alcuni utenti di visualizzare i titoli della cronologia

[330] Garante della Protezione dei Dati Personali, 'Provvedimento dell'8 giugno 2023', [doc. web n. 9896217]

[331] Garante della Protezione dei Dati Personali, 'Provvedimento del 10 febbraio 2022', n. 46, [doc. web n. 9751194]

[332] Garante della Protezione dei Dati Personali, "Relazione Annuale 2022", 6 luglio 2023, [doc. web n. 9905999]

delle chat di altri *user* attivi. Era, inoltre, anche possibile accedere ai dati personali di altri utenti, dati che, dunque, rientravano nella definizione ex art. 4, n. 1, del GDPR di "dato personale", come "*informazione riguardante una persona fisica identificata o identificabile*"[333]. Tra questi, i più rilevanti risultavano essere: nomi e cognomi, indirizzi e-mail, indirizzi di pagamento ed informazioni dettagliate sulle carte di credito. La società ha poi rivelato che i dati relativi a circa l'1,2% degli abbonati al servizio *plus* erano stati interessati dalla violazione per un periodo di tempo di nove ore[334].

Tale violazione ricade nel campo di applicazione del GDPR ex art. 4, n. 12, poiché sono state divulgate informazioni personali degli utenti senza il loro consenso [335]. In conformità con gli articoli 33 e 34 del Regolamento, l'organizzazione responsabile della violazione dei dati ha l'obbligo di notificare all'autorità di controllo competente entro 72 ore dalla scoperta della violazione e, se necessario, informare gli interessati senza indebito ritardo. Secondo quanto rivelato dalla società, tra le azioni intraprese non ci sarebbe stata alcuna comunicazione alle autorità competenti a cui fanno riferimento gli articoli citati in precedenza. Al contrario, il personale della società ha condotto una mera ricerca degli utenti danneggiati notificando loro l'avvenuto incidente [336]. Questo approccio probabilmente troppo superficiale, in particolare alla luce della gravità del fatto, ha ingenerato dubbi sulla compliance della società alla normativa sulla protezione dei dati. Anche sulla base di questa violazione è stata avviata un'istruttoria sul sistema

---

[333] GDPR, art. 4, n. 1 definisce "*«dato personale»: qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale;*"

[334] OpenAI, notizia pubblicata nel blog della società, consultabile al seguente indirizzo <https://openai.com/blog/march-20-chatgpt-outage>

[335] L'articolo 4, numero 12, del GDPR definisce la violazione dei dati personali come: "*la violazione di sicurezza che comporta accidentalmente o in modo illecito la distruzione, la perdita, la modifica, la divulgazione non autorizzata o l'accesso ai dati personali trasmessi, conservati o comunque trattati"*

[336] *Cfr*. n 334

di IA da parte del Garante della *Privacy* italiano, che con un provvedimento di urgenza ha disposto il 30 marzo 2023 l'interruzione del servizio[337].

6. Il provvedimento del Garante Italiano nei confronti di ChatGPT

Il provvedimento del Garante si configura come una misura cautelare adottata in concomitanza con l'istruttoria avviata dall'ente stesso. Tale misura cautelare si è concentrata su quattro categorie di questioni: la prima riguardava l'informativa sul trattamento dei dati, la seconda la tutela degli utenti, la terza l'esattezza delle informazioni relative a persone fisiche e l'ultima relativa alla salvaguardia degli utenti minorenni.

Nel provvedimento emesso il 30 marzo sono state sollevate in totale nove questioni riguardanti il trattamento dei dati effettuato da OpenAI L.L.C., titolare del trattamento ex art. 4, n. 7 del GDPR, posto in essere dal sistema di IA "ChatGPT".

Le misure disposte sono: *"*

1. Pubblicazione di un'informativa dettagliata sul trattamento dei dati personali.

2. Fornire agli interessati uno strumento per esercitare il diritto di opposizione al trattamento dei propri dati personali.

3. Mettere a disposizione degli interessati uno strumento per richiedere la correzione dei dati personali inesatti o la cancellazione dei propri dati.

4. Inserire un link all'informativa durante la registrazione al servizio.

5. Modificare la base giuridica del trattamento dei dati personali per conformarsi al consenso o al legittimo interesse.

6. Fornire uno strumento per l'opposizione al trattamento dei dati personali acquisiti durante l'utilizzo del servizio.

7. Implementare un sistema di controllo dell'età per escludere gli utenti minorenni dal servizio.

---

[337] Garante della Protezione dei Dati Personali, 'Provvedimento del 30 marzo 2023', n. 112, [doc. web n. 9870832]

8. Presentare un piano per l'adozione di strumenti di verifica dell'età entro una data specifica.
9. Condurre una campagna di informazione, in accordo con il Garante, per informare le persone sulla raccolta dei dati personali, l'informativa disponibile e le opzioni di cancellazione dei dati personali."[338]

La prima problematica evidenziata dal Garante della Privacy riguardava l'assenza di chiarezza e completezza nell'informativa, per consentire agli utenti e ai soggetti interessati di comprendere quali dati vengono raccolti. Risultavano inoltre mancanti informazioni sui trattamenti effettuati ed i diritti esercitabili, unitamente alle modalità di esercizio degli stessi. L'articolo 12 del GDPR impone infatti la predisposizione di un'informativa esaustiva, comprendente vari elementi, quali l'identità e i dati di contatto del titolare del trattamento, le finalità del trattamento, il periodo di conservazione delle informazioni ed ulteriori informazioni[339]. L'informativa del servizio nel caso di specie risultava carente in molti di questi aspetti, alcuni dei quali approfonditi anche negli altri punti del provvedimento.

Bisogna precisare che la finalità del trattamento dei dati, indicata originariamente nella *privacy policy*, non includeva il trattamento dei dati ai fini di addestramento degli algoritmi per il funzionamento del sistema di IA ed è chiaro come tale mancanza sia una

---

[338] Garante della Protezione dei Dati Personali, 'Provvedimento dell'11 aprile 2023', [doc. web n. 9874702]

[339] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), [2016], OJ L 119, 4.5.2016, p. 1–88. L'art. 12, para. 1, prevede che: "*Il titolare del trattamento adotta misure appropriate per fornire all'interessato tutte le informazioni di cui agli articoli 13 e 14 e le comunicazioni di cui agli articoli da 15 a 22 e all'articolo 34 relative al trattamento in forma concisa, trasparente, intelligibile e facilmente accessibile, con un linguaggio semplice e chiaro, in particolare nel caso di informazioni destinate specificamente ai minori. Le informazioni sono fornite per iscritto o con altri mezzi, anche, se del caso, con mezzi elettronici. Se richiesto dall'interessato, le informazioni possono essere fornite oralmente, purché sia comprovata con altri mezzi l'identità dell'interessato.*"

violazione dell'articolo 12 e del principio della limitazione della finalità ex art. 5, para. 1, lett. b). In particolare, questo richiede che i dati personali vengano "*raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità*"[340]. La finalità dell'addestramento degli algoritmi alla base del sistema di IA risulta così troppo generica e vaga.

È importante evidenziare che, per un sistema di GPAI come ChatGPT, la profilazione ed il trattamento dei dati non costituiscono affatto una priorità per l'azienda. È innegabile la necessità di una maggiore trasparenza su come vengono gestiti i dati, così come l'importanza di rendere più accessibile, completa e diffusa l'informativa sul servizio. Tuttavia, nel contesto di tali sistemi si verifica una sorta di cortocircuito logico. Infatti, questi ultimi sono progettati per rispondere alle richieste degli utenti e per migliorare costantemente in questa attività risulta necessaria l'acquisizione di un volume crescente di dati, che siano personali o meno. Allo stesso tempo, in base al principio di limitazione della finalità e di minimizzazione dei dati, il sistema non dovrebbe trattare automaticamente i dati personali forniti spontaneamente o involontariamente dagli utenti nel caso in cui non risultino necessari per la finalità, considerando anche che per trattare lecitamente i dati personali degli interessati è necessario soddisfare almeno una delle condizioni indicate al primo comma dell'articolo 6 del GDPR, come il consenso, il legittimo interesse o l'adempimento di un contratto [341]. Tuttavia, analizzando la versione aggiornata al 7 aprile 2023 della *privacy policy* di

---

[340] ibid. L'art. 5, para. 1, lett. b) prevede che i dati personali siano: "*raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità; un ulteriore trattamento dei dati personali a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici non è, conformemente all'articolo 89, paragrafo 1, considerato incompatibile con le finalità iniziali («limitazione della finalità»);*".

[341] L'articolo 6, paragrafo 1, del GDPR individua le condizioni per la liceità del trattamento dei dati personali: 1. Il trattamento è lecito solo se e nella misura in cui ricorre almeno una delle seguenti condizioni: "*a) l'interessato ha espresso il consenso al trattamento dei propri dati personali per una o più specifiche finalità; b) il trattamento è necessario all'esecuzione di un contratto di cui l'interessato è parte o all'esecuzione di misure precontrattuali adottate su richiesta dello stesso; c) il trattamento è necessario per adempiere un obbligo legale al quale è soggetto il titolare del trattamento; d) il trattamento è necessario per la salvaguardia degli interessi vitali dell'interessato o di un'altra persona fisica; e) il trattamento è necessario per l'esecuzione di un compito di interesse pubblico o connesso all'esercizio di*

OpenAI, non era possibile individuare alcuna base giuridica che sostenesse la finalità della raccolta dei dati personali allo scopo di addestrare i sistemi di IA[342].
Il Garante della Privacy italiano ha perciò imposto alla società, per continuare la raccolta ed il trattamento dei dati personali degli utenti allo scopo di addestrare gli algoritmi di IA, l'obbligo di indicare con chiarezza e trasparenza tale finalità nell'informativa, l'individuazione di una differente base giuridica per il trattamento dei dati e l'implementazione di una soluzione per permettere agli utenti di escludere i propri dati dai sistemi di addestramento.

Il Garante ha evidenziato poi ulteriori problematiche riguardanti l'accuratezza delle informazioni fornite dal sistema di intelligenza artificiale sotto esame. In particolare, le risposte fornite dal sistema risultavano spesso inesatte, in quanto l'addestramento poteva avvenire su dati raccolti da fonti *online* di dubbia affidabilità, portando alla generazione di cosiddette "allucinazioni"[343].

---

pubblici poteri di cui è investito il titolare del trattamento; f) il trattamento è necessario per il perseguimento del legittimo interesse del titolare del trattamento o di terzi, a condizione che non prevalgano gli interessi o i diritti e le libertà fondamentali dell'interessato che richiedono la protezione dei dati personali, in particolare se l'interessato è un minore. La lettera f) del primo comma non si applica al trattamento di dati effettuato dalle autorità pubbliche nell'esecuzione dei loro compiti".

[342] Al punto nove della *privacy policy* aggiornata al 7 aprile 2023, che indica le disposizioni per la privacy per gli utenti internazionali, la finalità di addestramento degli algoritmi non era esplicitata. Questa finalità, invece, nella *privacy policy* aggiornata al 27 aprile, viene inserita e si pone il legittimo interesse, come condizione per la legittimità del trattamento, ex art. 6, para. 1, lett. f) del GDPR, di OpenAI nello sviluppo e diffusione del loro servizio a fondamento della finalità di addestramento degli algoritmi.

[343] Le allucinazioni in ChatGPT si riferiscono alla generazione di risposte o informazioni false, inesatte o fuorvianti da parte del modello di intelligenza artificiale. Queste allucinazioni possono essere il risultato di diverse cause, tra cui l'addestramento su dati imprecisi, ambigui o fuorvianti, l'incapacità del modello di comprendere adeguatamente il contesto o la domanda dell'utente, o le limitazioni intrinseche dell'architettura del modello stesso. È possibile che nelle versioni successive del sistema, l'addestramento sia stato portato avanti attraverso il coinvolgimento di esseri umani nel processo di apprendimento, in un'operazione di *reinforcement learning*, anche se tale metodo non sempre permette di ottenere i risultati ricercati

La portata della pericolosità di tali inesattezze diventa evidente considerando la rapida e diffusa adozione del servizio, che ha registrato un successo senza precedenti in termini di utenza[344]. Di fronte al rischio che il sistema diffonda informazioni scorrette, il Garante ha sottolineato la necessità di prevenire tali "allucinazioni" e ha richiesto la possibilità per gli utenti di correggere o eliminare le informazioni personali trattate in modo inappropriato. Tale richiesta risulta in linea con l'articolo 5 del GDPR, il quale, al paragrafo 1, lettera d), stabilisce che i dati personali debbano essere "*esatti e, se necessario, aggiornati; devono essere adottate tutte le misure ragionevoli per cancellare o rettificare tempestivamente i dati inesatti rispetto alle finalità per le quali sono trattati*"[345]. La necessità di garantire l'accuratezza delle informazioni fornite dal sistema di GPAI risulta, tuttavia, meno prioritaria rispetto all'utilità che i sistemi di questo tipo possono offrire. Attualmente, ci troviamo in una fase di sviluppo di tali sistemi e come esplicitamente dichiarato da OpenAI nell'interfaccia di ChatGPT: "*ChatGPT può produrre informazioni inaccurate su persone, luoghi o fatti*". Il servizio principale offerto dal sistema non consiste nel fornire dati personali precisi, ma piuttosto nell'elaborare frasi coerenti e rispondere logicamente alle domande degli utenti. Il citato *disclaimer* rende chiara la politica di deresponsabilizzazione adottata dalla società, che si contrappone diametralmente all'obbligo imposto dal provvedimento del Garante. Come già spiegato precedentemente, raggiungere tale accuratezza non è facilmente realizzabile e imporre tale obbligo ai fornitori dei sistemi di IA potrebbe risultare sproporzionato rispetto al potenziale danno reputazionale derivante da informazioni inesatte. La presenza di un *disclaimer*, in astratto consente di mettere in guardia gli utenti sulle informazioni fornite dal sistema in risposta alle domande, ma tale misura non appare risolutiva nel tutelarli dai rischi.

---

[344] Mollick (n 262).

[345] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), [2016], OJ L 119, 4.5.2016, p. 1–88 art 5.

In conclusione, il Garante italiano si è focalizzato sulla categoria più vulnerabile nell'interazione con tali tecnologie: i minori al di sotto dei 13 anni di età. Nonostante la *privacy policy* di OpenAI, preesistente all'intervento del Garante, vietasse l'utilizzo del servizio da parte di tali soggetti, non vi erano strumenti di controllo dell'età e misure preventive efficaci volti ad ostacolare il loro accesso al sistema di GPAI. Questa lacuna rendeva la disposizione contenuta nella *privacy policy* sostanzialmente inefficace, non essendo presente alcuna verifica effettiva dell'età né durante il processo di registrazione al servizio, né durante gli accessi successivi. L'utilizzo del servizio era altresì limitato per i minori di età superiore ai 13 anni in assenza del consenso dei genitori o dei tutori legali. Secondo il Garante, l'esposizione dei minori a risposte incongrue al loro sviluppo e potenzialmente errate costituisce un rischio notevole per la loro crescita. Pertanto, l'autorità ha richiesto che, una volta riattivato il servizio, divenisse obbligatorio "superare un "*age gate*" che escludesse gli utenti minorenni sulla base dell'età dichiarata". Inoltre, è stato richiesto un sistema di verifica dell'età adeguato a prevenire l'accesso al servizio da parte di utenti di età inferiore ai 13 anni o senza il consenso dei genitori. Tale prescrizione richiedeva un periodo di sviluppo maggiore rispetto alle precedenti, le quali avevano una scadenza fissata al 30 aprile 2023. Il sistema di verifica dell'età andava presentato entro il 31 maggio 2023 ed implementato entro il 30 settembre 2023. Infine, il Garante ha richiesto che OpenAI promuovesse, "*entro il 15 maggio 2023, una campagna di informazione, di natura non promozionale, su tutti i principali mezzi di comunicazione di massa italiani i cui contenuti andranno concordati col Garante, allo scopo di informare le persone dell'avvenuta probabile raccolta dei loro dati personali ai fini dell'addestramento degli algoritmi, dell'avvenuta pubblicazione sul sito internet della Società di un'apposita informativa di dettaglio e della messa a disposizione, sempre sul sito internet della Società, di uno strumento attraverso il quale tutti gli interessati possono chiedere e ottenere la cancellazione dei propri dati personali*"[346].

Rispetto ad un motore di ricerca, ChatGPT si distingue per il fatto che non suggerisce una varietà di fonti correlate alla ricerca, ad esempio, il profilo di una persona reale,

[346] Garante della protezione dei dati personali, Provvedimento dell'11 aprile 2023 doc-web n. 9874702

bensì fornisce una versione precisa del profilo, assumendosi di conseguenza una certa responsabilità sui contenuti prodotti.

Benché il provvedimento del Garante si presenti in linea con gli obiettivi di protezione dei dati personali e manifesti un uso ponderato del potere di supervisione attribuito all'autorità, è necessario sottolineare come l'approccio individuale dell'Autorità italiana, privo di un coordinamento con le altre autorità europee o con l'*European Data Protection Board*, possa comportare un indebolimento della coesione che dovrebbe caratterizzare l'operato delle autorità di controllo indipendenti all'interno del mercato unico europeo. Una potenziale conseguenza di tale metodologia avrebbe potuto condurre, qualora OpenAI non avesse risposto positivamente alle richieste del Garante italiano, all'esclusione degli utenti italiani dal sistema di IA, il quale sarebbe invece rimasto liberamente accessibile negli altri Stati Membri europei. Uno scenario di tale portata avrebbe sicuramente penalizzato il mercato italiano, che al pari di quello francese o tedesco, stava esplorando nuove possibilità di sviluppo della tecnologia emergente.

Dopo il provvedimento del Garante è stata costituita una *task force* europea per permettere ai Garanti nazionali di analizzare in concerto le problematiche che sono sorte con la diffusione di questi sistemi di IA generativa[347].

Con l'intervento del Garante è stata individuata la direzione che tale autorità intende prendere nel prevenire i rischi e sanzionare i danni derivanti dai sistemi di IA. Tra gli aspetti attenzionati nel provvedimento del Garante del 30 marzo 2023 però, non si è prestata molta attenzione alla qualità dei dati utilizzati dal sistema di IA e alla sua trasparenza.

L'importanza di questi due aspetti si fonda sulla chiarezza che deriva da una rappresentazione del *database* su cui è costruito il sistema di IA. Con una indicazione di tale contenuto il Garante avrebbe potuto stabilire quale fosse stata l'entità del danno derivante da *data breach* e anche richiedere ad OpenAI di adeguarsi ad un sistema di sicurezza più confacente con le prerogative individuate per la protezione dei dati

[347] 'EDPB Resolves Dispute on Transfers by Meta and Creates Task Force on Chat GPT | European Data Protection Board' <https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en> accessed 1 May 2023.

personali. Inoltre, il Garante avrebbe potuto esprimersi nei confronti delle problematiche relative al trattamento dei dati posto in essere e imporre una maggiore trasparenza dei processi automatizzati al titolare del trattamento. Al fine di regolare nel migliore dei modi tali tecnologie risulterà necessario focalizzarsi sugli aspetti che permettano all'uomo di tenere un controllo sempre maggiore su tali tecnologie, garantendo sia la possibilità di comprenderne il loro contenuto, sia la possibilità di correggere le deviazioni che potrebbero derivarne.

7. Segue: l'applicazione del Regolamento UE 2016/679 al sistema di IA ChatGPT

A seguito di un intenso dialogo tra OpenAI e il Garante italiano, nell'imminenza della scadenza per l'adempimento dei punti da uno a sette del provvedimento dell'11 aprile 2023, il titolare del trattamento di ChatGPT ha dato seguito alle richieste dell'autorità, garantendo maggiore trasparenza e chiarezza nell'informativa e agevolando l'esercizio dei diritti sanciti nel GDPR. La nuova *privacy policy* di OpenAI è stata aggiornata il 27 aprile 2023 e presenta notevoli innovazioni, principalmente per quanto riguarda l'esercizio dei diritti previsti dal GDPR.

La questione della mancanza di un valido fondamento giuridico relativo alla raccolta dei dati personali per l'addestramento dei modelli è stata parzialmente risolta. All'interno dell'informativa rivolta ai soggetti interessati è stata introdotta, nella sezione due rubricata "*How we use personal information*", al punto relativo alle finalità di sviluppo del servizio, una precisazione riguardante l'utilizzo dei contenuti forniti dall'utente nelle interazioni con la *chatbot* per il continuo miglioramento e l'addestramento dei modelli[348]. Nella stessa sezione dell'informativa viene fornito un collegamento che permette, attraverso una serie di passaggi, di escludere i propri dati eventualmente forniti al servizio dal *dataset* utilizzato per l'addestramento dei modelli. Secondo quanto dichiarato da OpenAI, lo scopo dell'addestramento dei modelli sarebbe quello di migliorarne la precisione e l'utilità per gli utenti e il pubblico generale. Viene anche

[348] OpenAI, 'Privacy Policy' s 2 <https://openai.com/policies/privacy-policy> accessed 9 May 2023.

fornita la spiegazione di come ChatGPT si perfezioni autonomamente attraverso l'addestramento sulle conversazioni condotte con gli utenti[349].

Nel quarto punto della *privacy policy* di OpenAI vengono indicati a titolo esemplificativo[350], alcuni dei diritti per la protezione dei dati personali dei cittadini dello Spazio Economico Europeo e del Regno Unito, tra cui sono presenti il diritto di rettifica, di accesso, di cancellazione e altri diritti riconosciuti e tutelati dal GDPR. Ai cittadini viene anche data la possibilità di esercitare tali diritti attraverso il loro profilo OpenAI oppure scrivendo ad un indirizzo mail specifico. In aggiunta, è stata inclusa una precisazione riguardante l'esattezza ex art. 5, para. 1, lett. d) del GDPR, relativamente ai risultati forniti da ChatGPT. Nella nota viene specificato come il modello sia programmato per analizzare la richiesta dell'utente e fornire una risposta costruita sull'analisi del contesto, predicendo le parole nella frase. Inoltre, viene chiarito agli utenti che, in alcuni casi, le parole più probabili potrebbero non coincidere con quelle corrette, fornendo quindi informazioni errate. Agli utenti, nel caso in cui il sistema fornisca dati inesatti relativi a soggetti specifici, viene richiesto di segnalare tali inesattezze e richiedere una correzione, oppure la cancellazione delle informazioni personali pertinenti.

OpenAI riconosce apertamente la difficoltà nell'adempiere pienamente alle richieste connesse al diritto di rettifica, come previsto dall'art. 16 del Regolamento UE

---

[349] La novità sta anche nel poter richiedere direttamente dalla interfaccia del servizio, oppure per mezzo di un modulo da compilare, l'interruzione di utilizzo delle conversazioni per l'addestramento dei modelli. Allo stesso tempo, si esplicita come i dati raccolti dall'interazione e che contengono informazioni personali attraversano, prima di entrare nel dataset per l'addestramento, un filtro per ridurne l'ammontare. L'importanza di tali dati sta nel migliorare la capacità dei modelli di capire i bisogni e le preferenze degli utenti per rendere tali sistemi ancora più efficienti. Per approfondire il funzionamento del sistema di IA, si faccia riferimento a: 'How Your Data Is Used to Improve Model Performance | OpenAI Help Center' <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance> accessed 19 May 2023.

[350] OpenAI (n 320) s 4.

2016/679[351]. Questa problematica nasce principalmente dalla complessità intrinseca dei modelli di IA. Tali modelli sono notoriamente contraddistinti da una barriera tecnologica difficile da superare quando si intende operare la rettifica dei dati. OpenAI sottolinea anche che la cancellazione dei dati risulta spesso di più semplice esecuzione rispetto alla loro rettifica. L'importanza della problematica legata al diritto di rettifica si comprende considerando il *dataset* sul quale si fonda l'addestramento del *foundation models*. In tal senso appaiono chiare le dichiarazioni di Yaniv Markovski (*Head of AI* di OpenAI), secondo le quali l'addestramento del modello di ChatGPT si basa su tre principali fonti di dati: in primis, dati di dominio pubblico raccolti da Internet; secondariamente, dati acquisiti attraverso licenze di terze parti; infine, dati forniti da individui coinvolti direttamente nel processo di addestramento del modello [352]. L'eterogeneità dell'origine dei dati costituisce un ostacolo considerevole nell'implementazione di modifiche in conformità con il principio di rettifica delineato dall'art. 16 del GDPR. Questa complessità deriva dalla difficoltà nell'individuazione di specifiche informazioni all'interno del vasto mosaico di dati. Inoltre, persino dopo la rimozione di dati identificativi riferiti a una persona fisica, il contesto circostante potrebbe permettere al modello di dedurre le stesse caratteristiche o specificità precedentemente fornite dal dato rettificato o cancellato[353].

---

[351] L'articolo 16 del GDPR stabilisce: "*L'interessato ha il diritto di ottenere dal titolare del trattamento la rettifica dei dati personali inesatti che lo riguardano senza ingiustificato ritardo. Tenuto conto delle finalità del trattamento, l'interessato ha il diritto di ottenere l'integrazione dei dati personali incompleti, anche fornendo una dichiarazione integrativa*."

[352] FAQ di OpenAI < https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed> ultimo accesso 10 maggio 2023

[353] La problematicità tecnica nell'alterazione dei dati insiti nel dataset di ChatGPT è correlata alla natura intrinseca dei modelli di deep learning, come quelli basati sull'architettura GPT. Tali modelli si avvalgono di reti neurali profonde che vengono addestrate su ingenti quantità di dati al fine di apprendere relazioni complesse e pattern negli input forniti. Successivamente all'addestramento del modello, le informazioni originarie del dataset vengono incorporare sotto forma di pesi e connessioni nell'ambito della rete neurale. Questa struttura interna si rivela di alta complessità e non è agevolmente interpretabile o modificabile. In altre parole, le informazioni presenti nel dataset diventano componente fondamentale della "conoscenza" del modello, rendendo estremamente ardua la loro modifica diretta. Per alterare uno specifico dato nel

Il processo di acquisizione di informazioni da parte di OpenAI, fondamentale per l'addestramento di ChatGPT, sfrutta prevalentemente dati di dominio pubblico disponibili su Internet. Sebbene l'organizzazione non persegua l'obiettivo di includere attivamente dati personali nei propri aggregati di addestramento, è altamente probabile che tali dati, preesistenti su Internet, vengano incorporati. Il valore dei dati personali nell'addestramento del modello di ChatGPT è duplice: da un lato, essi arricchiscono la comprensione del modello delle relazioni grammaticali all'interno delle frasi, come quelle tra il soggetto, il verbo e gli altri componenti della frase; dall'altro, forniscono una base di conoscenza su personaggi noti o influenti. È tuttavia essenziale chiarire che i dati personali non sono utilizzati per scopi di profilazione o di targetizzazione pubblicitaria e la società si è impegnata a non vendere tali informazioni e a non intraprendere attività di *telemarketing*.

La base giuridica che viene indicata e che giustifica la raccolta dei dati personali presenti nel pubblico dominio di internet è quella del legittimo interesse, individuata all'articolo 6, paragrafo 1, lettera f) del GDPR[354]. La liceità dell'utilizzo di tali informazioni è divenuta cruciale in quanto, i modelli come ChatGPT stanno producendo risultati rilevanti in tanti settori, dallo sviluppo dei software, all'educazione personalizzata, oltre al supporto nella ricerca scientifica e molti altri, ma senza una grande quantità di dati per educare il modello, tali risultati non sarebbero raggiungibili. I

---

dataset, si renderebbe necessario individuare con precisione il modo in cui il dato è stato incorporato all'interno della rete neurale, apportare modifiche ai pesi e alle connessioni corrispondenti, e infine valutare se la modifica abbia generato conseguenze collaterali non volute sulle altre informazioni apprese dal modello. Questo processo è notoriamente complicato, poiché l'informazione non è rappresentata in maniera esplicita o facilmente accessibile all'interno della rete neurale. In alcune circostanze, per eliminare o alterare informazioni specifiche, potrebbe essere necessario riaddestrare il modello per intero su un nuovo dataset, escludendo o modificando i dati problematici. Tuttavia, tale processo può comportare un considerevole consumo di tempo e risorse computazionali e non potrebbe garantire l'effettiva eliminazione o modificazione appropriata dell'informazione specifica.

[354] L'articolo 6, paragrafo 1, lettera f) del GDPR prevede che: "*il trattamento è necessario per il perseguimento del legittimo interesse del titolare del trattamento o di terzi, a condizione che non prevalgano gli interessi o i diritti e le libertà fondamentali dell'interessato che richiedono la protezione dei dati personali, in particolare se l'interessato è un minore*."

modelli non sono pensati e programmati per recare danno alle persone e le informazioni che vengono raccolte su internet sono pubblicamente disponibili.

Inoltre, OpenAI, in quanto titolare del trattamento, dichiara di aver condotto una valutazione di impatto sulla protezione dei dati derivante dalla inclusione degli stessi nei *dataset* di addestramento. Secondo l'articolo 35 del GDPR una Valutazione di Impatto sulla Protezione dei Dati (Data Protection Impact Assessment, DPIA) consiste in un processo per valutare l'impatto che un trattamento dei dati personali può avere sulla privacy dei soggetti interessati. L'obiettivo principale di una DPIA è garantire la protezione dei dati personali attraverso l'individuazione di rischi associati a un particolare trattamento dei dati e di misure appropriate per mitigare tali rischi. Il GDPR, nell'articolo 35, richiede la conduzione di una DPIA quando una tipologia di trattamento "può comportare un elevato rischio per i diritti e le libertà delle persone fisiche", in particolare quando si introducono nuove tecnologie. Inoltre, nel caso di "trattamento su larga scala di categorie particolari di dati [...], di dati personali relativi a condanne penali e reati". In ogni caso, non sembrano esserci informazioni su tale valutazione e probabilmente l'organizzazione le manterrà riservate.

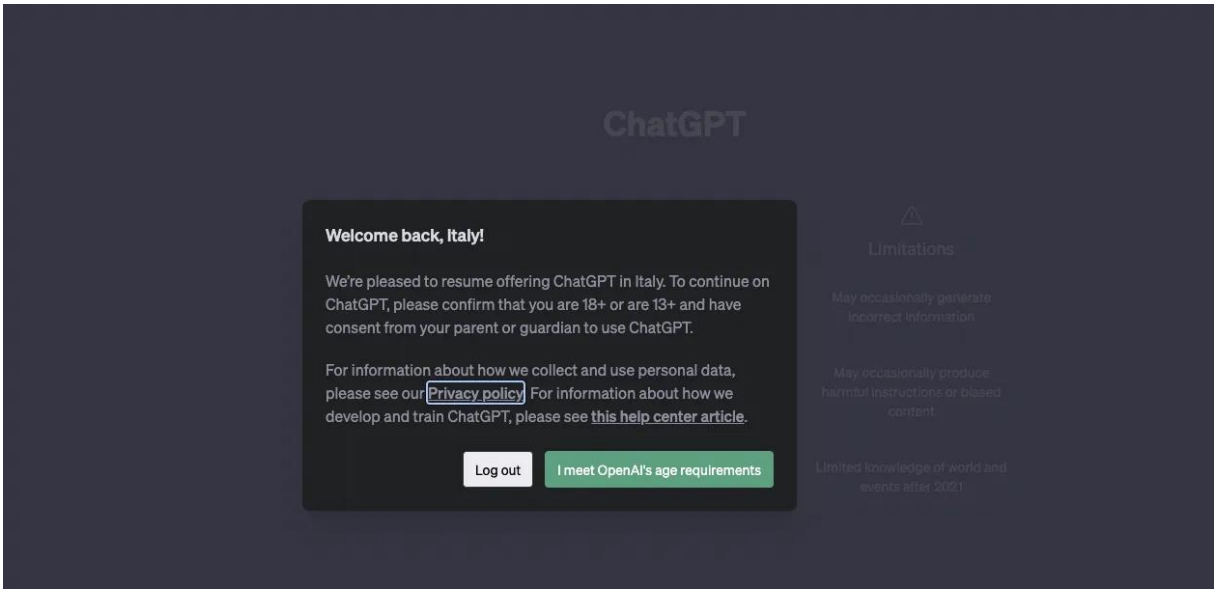
Nell'intento di accrescere la trasparenza circa l'impiego dei dati per l'ottimizzazione delle prestazioni dei suoi modelli, OpenAI ha reso disponibile un documento esplicativo, nel quale chiarisce la gestione dei dati nei propri servizi[355]. Si evidenzia in particolare come ChatGPT si avvalga dei dati forniti dagli utenti per raffinare l'addestramento del proprio algoritmo, garantendo agli utenti stessi la facoltà di disattivare tale funzionalità. Questa prassi sembra rispettare il principio del consenso informato, delineato nell'art. 4, n. 11 del GDPR[356], considerando anche che gli utenti, a seguito delle modifiche implementate dopo il provvedimento del Garante, ricevono un'informativa notevolmente più puntuale e dettagliata riguardo al trattamento dei loro

[355] Joshua J., 'Data controls FAQ', < https://help.openai.com/en/articles/7730893-data-controls-faq> accessed 28 maggio 2023

[356] L'articolo 4, n. 11 del GDPR indica la seguente definizione per «consenso dell'interessato»: "*qualsiasi manifestazione di volontà libera, specifica, informata e inequivocabile dell'interessato, con la quale lo stesso manifesta il proprio assenso, mediante dichiarazione o azione positiva inequivocabile, che i dati personali che lo riguardano siano oggetto di trattamento*;"

dati e all'impiego delle conversazioni che conducono con il sistema di IA. L'intento di limitare la quantità di dati personali incorporati nei *dataset* adoperati manifesta una condotta in linea con il principio della minimizzazione dei dati, come stabilito nell'art. 5, comma 1, lett. c) del GDPR, che richiede che i dati siano: "adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati". Tuttavia, questa aspirazione deve considerare l'ostacolo rappresentato dalla complessità dell'accesso alle informazioni contenute nel *dataset* dei *foundation models*.

In ultima analisi, la società ha adempiuto alle rimanenti richieste del Garante, che potremmo classificare come "più formali". È stato implementato un collegamento all'informativa nella maschera di registrazione per rendere più chiaro ed esplicito il trattamento e la raccolta dei dati condotti da OpenAI in seguito alla registrazione. Alla riattivazione del servizio in Italia è stata aggiunta una schermata di benvenuto con il collegamento alla nuova informativa sulla *privacy* e le modalità di trattamento dei dati per l'addestramento degli algoritmi. Nella medesima schermata è stato integrato un pulsante per attestare la maggiore età dell'utente o, se quest'ultimo è ultra-tredicenne, il consenso dei genitori. È stato, inoltre, introdotto un blocco del servizio nel caso in cui l'utente non rispetti i requisiti di età per l'accesso o non confermi di possedere il consenso dei genitori. In concreto però, l'efficacia di questo *age-gate* appare abbastanza discutibile, come la maggior parte dei sistemi di verifica dell'età presenti in rete. Ormai gli utenti al di sotto dei 13 anni sanno perfettamente che attestando la loro reale età non potranno accedere a questi potenti strumenti di IA.



Fonte: Chat.openai.com

Tra le richieste del Garante, ne rimangono ancora alcune che OpenAI dovrà affrontare, come l'organizzazione di una campagna informativa non promozionale sul territorio italiano per comunicare agli utenti e non utenti la possibilità di richiedere la rettifica o la cancellazione dei loro dati personali dai modelli di OpenAI, o la rimozione dei propri dati dal *dataset* di apprendimento degli algoritmi in continuo sviluppo.

Si deve evidenziare come, a livello europeo, diversi Garanti della *Privacy* hanno avviato istruttorie su OpenAI e i suoi servizi. Questa circostanza probabilmente ha influenzato l'adempimento da parte della società alle richieste poste dal Garante italiano. Inoltre, la neocostituita *task force* europea, dedicata specificamente al tema, si occuperà dei profili più delicati della questione. Tra gli argomenti più sensibili che si stanno esaminando figura la legittimità della raccolta indiscriminata di dati da internet in assenza di qualsiasi consenso preventivo, basata unicamente sul presupposto della preesistente disponibilità di tali informazioni nel dominio pubblico. Possono essere mosse contestazioni sulla legittimità di tale tipo di trattamento sulla base dell'articolo 14 del GDPR, che sancisce un obbligo informativo nei confronti del titolare del trattamento nel caso in cui i dati personali dell'interessato non siano stati ottenuti presso l'interessato stesso. Tuttavia, il paragrafo 5 prevede un'eccezione alla lettera b) che potrebbe essere rilevante nel caso di specie. Infatti, si prevede che non sussista tale obbligo informativo quando: "*comunicare tali informazioni risulta impossibile o implicherebbe uno sforzo*

sproporzionato"[357]. Considerando la quantità di dati su cui è stato addestrato ChatGPT, anche per gli sviluppatori del sistema condurre una ricerca che permetta di individuare tutti i soggetti interessati dal trattamento rappresenterebbe uno sforzo sproporzionato.

8. Il vuoto normativo attorno ai nuovi sistemi di IA

Lo sviluppo di algoritmi di IA e di *Machine Learning* avviene prevalentemente in grandi aziende private, per le quali la finalità primaria tende ad essere la generazione di profitto, ponendo la sicurezza e la salvaguardia dei soggetti interessati a considerazioni di secondo ordine. Le ricerche e gli studi accademici dedicati alla sicurezza degli utenti in sistemi di *Machine Learning* rappresentano una percentuale ancora esigua rispetto all'intero compendio di ricerche in materia [358]. Ciononostante, tale scenario potrebbe mutare nel caso in cui fossero implementate disposizioni legislative, come quelle contenute nell'AI Act, che prevedono l'imposizione di requisiti specifici in termini di sicurezza e *privacy*. Tale intervento normativo avrebbe l'effetto di indurre le grandi società private a riconsiderare la loro posizione e ad attribuire un ruolo più importante ai profili di sicurezza, implementando maggiormente la tutela della privacy dei soggetti coinvolti nel loro modello operativo.

Il provvedimento del Garante della *Privacy* italiano segna presumibilmente l'inizio di una serie di provvedimenti riguardanti i sistemi di IA, volti a prevenire i rischi potenziali derivanti dalla continua evoluzione tecnologica. Questo intervento si limita a una sfera ristretta dei nuovi modelli di Intelligenza Artificiale; in effetti, tenendo in considerazione la vasta gamma di applicazioni dell'IA a scopo generale, i *large language models* rappresentano soltanto una frazione dell'intero complesso. Il funzionamento di tali sistemi si articola attraverso diverse fasi: lo sviluppo del codice, l'addestramento sui dati e l'introduzione sul mercato con un percorso di sviluppo continuo. L'Autorità italiana ha intrapreso azioni relative all'ultima fase, cercando di

[357] GDPR, art. 14, para. 5, lett. b)

[358] Dan H and ThomasW, 'A Bird's Eye View of the ML Field [Pragmatic AI Safety #2]' s AGI and AI Safety <https://www.alignmentforum.org/posts/AtfQFj8umeyBBkkxa/a-bird-s-eye-view-of-the-ml-field-pragmatic-ai-safety-2> accessed 10 May 2023.

mitigare, sulla base dei principi e nel rispetto dei diritti stabiliti dal Regolamento UE 2016/679, pregiudizi e potenziali rischi derivanti da un funzionamento non regolamentato del modello di IA.

Agire sulle fasi precedenti, in particolare sulla fase di addestramento - che riveste un'importanza fondamentale per il rispetto dei diritti delle persone fisiche - richiede un approccio più articolato e che si fondi su delle disposizioni normative più chiare e che riguardino direttamente i sistemi di IA. Il Garante può operare solo mediante l'applicazione di principi volti alla protezione dei dati personali. Sarà compito dell'Autorità designata dallo Stato membro riuscire ad attuare i principi e le regole stabilite nell'AI Act, ma per garantire un ecosistema di IA che possa essere genuinamente considerato sicuro per gli individui, minimizzando i rischi e i danni derivanti da un uso improprio dei modelli, sarà necessario uno sforzo comune di tutti gli attori in campo. La necessità di non fermare l'innovazione rimane un aspetto fondamentale, cercando di attuare i principi per una tutela degli individui e che permetteranno di raggiungere sistemi di IA veramente utili per l'uomo e che riescano a contribuire al progresso della società.

Il provvedimento del Garante ha permesso il miglioramento di un sistema di IA in termini di trasparenza del trattamento dei dati, oltre che un aumento delle opportunità di esercizio dei diritti di opposizione, rettifica e cancellazione. La collaborazione mostrata dalla società statunitense sembra all'apparenza genuina, ma in realtà OpenAI si è inserita nel processo legislativo dell'AI Act con un'intensa attività di influenza[359]. Numerose dichiarazioni rilasciate dai vertici di tale società, così come di altre aziende che sviluppano sistemi simili, hanno amplificato l'attenzione sui rischi potenziali derivanti da un utilizzo improprio dei sistemi di IA, una questione che sembra aver attirato l'interesse dell'opinione pubblica[360].

---

[359] Billy Perrigo, 'Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation', *Time* (20 June 2023) <https://time.com/6288245/openai-eu-lobbying-ai-act/ > accessed 25 June 2023.

[360] Reuters, 'Google AI Pioneer Says He Quit to Speak Freely about Technology's "Dangers"' *Reuters* (9 May 2023) <https://www.reuters.com/technology/google-ai-pioneer-says-he-quit-speak-freely-about-technologys-dangers-2023-05-02/> accessed 17 May 2023; Johana Bhuiyan, 'OpenAI CEO Calls for Laws to Mitigate "Risks of Increasingly Powerful" AI' *The Guardian* (16 May 2023)

L’atto del Garante assume rilevanza nella continua interpretazione dei principi e delle regole del diritto, al fine di prevenire l'obsolescenza normativa. Nonostante il Regolamento UE 2016/679 rappresenti una legislazione recente che mira a regolare i fenomeni tecnologici moderni, è comprensibile che alcuni dei suoi principi, originariamente concepiti per affrontare i danni causati dall'utilizzo dei social network nella società digitale, potrebbero non essere del tutto adeguati a regolare i moderni sistemi di intelligenza artificiale, i quali perseguono obiettivi significativamente diversi rispetto alle tecnologie precedenti. I principi, pur dotati di una certa elasticità e malleabilità, non sembrano completamente adeguati a governare tecnologie così evolute e avanzate, che richiedono probabilmente una riconsiderazione e l'istituzione di un nuovo quadro legislativo che permetta al legislatore europeo di concentrarsi sugli aspetti che veramente possono fare la differenza nella tutela degli individui, in primis la qualità dei dati utilizzati per l’addestramento[361].

L’AI Act, invece, dovrebbe puntare maggiormente sulla sua anima innovativa, facilitando e agevolando lo sviluppo di nuovi sistemi di IA, portando l’UE ad un livello di reale competizione con Cina e USA. Considerando che l'orientamento globale sta progressivamente evolvendo verso una diffusione sempre più ampia di tali sistemi, al fine di aumentare la digitalizzazione, il fenomeno richiede un approccio equilibrato che permetta di bilanciare l'innovazione tecnologica con le necessarie salvaguardie legali e di protezione dei diritti degli individui. L’AI Act potrebbe rappresentare il miglior esempio di regolamentazione attuale se riuscisse a bilanciare la necessità di controllo da parte delle istituzioni sulle zone grigie poste dalle tecnologie più innovative, al fine di tutelare i diritti fondamentali delle persone in rete, e la spinta del processo tecnologico.

<https://www.theguardian.com/technology/2023/may/16/ceo-openai-chatgpt-ai-tech-regulations> accessed 17 May 2023; Cade Metz and Gregory Schmidt, ‘Elon Musk and Others Call for Pause on A.I., Citing “Profound Risks to Society”’ *The New York Times* (29 March 2023) <https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html> accessed 17 May 2023.

[361] Philipp Hacker, ‘A Legal Framework for AI Training Data—from First Principles to the Artificial Intelligence Act’ (2021) 13 Law, Innovation and Technology, 265–268.

Il caso di OpenAI ci induce a esaminare alcune implicazioni su come la decisione dell'Autorità abbia effettivamente considerato, alla luce del GDPR, il funzionamento di ChatGPT e del sistema di IA sottostante, identificando alcune prescrizioni fondate sul rispetto dei principi per la protezione dei dati personali. Così facendo, è stato sollevato il problema della relazione tra i sistemi di IA e la tutela del diritto fondamentale alla protezione dei dati personali.

Secondo il professor Luciano Floridi, gli strumenti come ChatGPT sarebbero paragonabili al giocator di bussolotti [362], nella famosa opera letteraria di Alessandro Manzoni "I Promessi Sposi"[363]. In un famoso episodio del romanzo, Renzo incontra un avvocato e, mentre parla, lo guarda con ingenua ammirazione, come un incauto osservatore in piazza che osserva l'illusionista che, dopo aver riempito la bocca di stoppa, tira fuori un nastro che non finisce mai. I sistemi di IA come ChatGPT sono paragonati a quel truffatore: assimilano dati in quantità massicce per poi produrre un flusso continuo di informazioni, ma di cui non ci si deve fidare ciecamente.

Bisogna chiedersi se il funzionamento di una applicazione fondata sull'IA possa coinvolgere il GDPR, che mira a tutelare il diritto alla protezione dei dati personali, e non si riferisce direttamente alle norme da rispettare per l'operatività nell'UE di applicazioni basate sui sistemi di IA. Si è ampiamente constatato come l'operatività dei sistemi di IA possa coinvolgere il trattamento di dati personali, con conseguente potenziale violazione del GDPR. Dunque, l'applicazione dei principi di correttezza, rettifica e minimizzazione dei dati deve essere ripensata per potersi applicare ai sistemi di IA, mantenendo chiaro l'obiettivo che i sistemi hanno e valutando realmente i rischi che potrebbero derivarne.

Emergono alcune considerazioni conclusive dall'analisi del provvedimento adottato dal Garante italiano. La prima riguarda l'importanza che concerne l'informazione fornita

---

[362] Luciano Floridi, 'AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models' (14 February 2023) 6 <https://papers.ssrn.com/abstract=4358789> accessed 17 May 2023.

[363] Alessandro Manzoni, *I promessi sposi* (Letteratura Italiana Einaudi 1985) 49.

agli utenti del servizio, per renderla più chiara in tutti i sistemi di IA che siano ad uso generale o ad uso speciale. I titolari del trattamento o fornitori sono tenuti a chiarire la natura dei dati trattati e le modalità del trattamento, rispettando il principio del consenso informato, garantendo agli utenti un'idonea informazione non solo sul *quomodo* del trattamento, ma anche sul *quid*, permettendo loro di conoscere sufficienti elementi per opporsi o meno al trattamento stesso. La libertà degli individui deve ispirare la regolamentazione dei sistemi, permettendo agli utenti di poter sempre scegliere e decidere sul trattamento e sull'uso di tali sistemi. La seconda questione riguarda la necessità di concentrarsi sulla qualità dei dati utilizzati per l'addestramento, richiedendo che gli stessi siano corretti, ma nei limiti di quanto permettono le nuove metodologie. La terza questione riguarda la sorveglianza umana, che deve anche prevedere la possibilità di una intelligenza umana avanzata per poter supervisionare i sistemi di IA.

In conclusione, questo contesto normativo, caratterizzato da una sempre più intricata interazione tra diritti individuali e avanzamento tecnologico, evidenzia con estrema chiarezza la necessità di individuare un equilibrio dinamico, che possa rispondere ai pericoli che derivano dall'innovazione di tali sistemi. L'obiettivo finale deve rimanere sempre quello di assicurare un adeguato livello di protezione dei diritti degli utenti, nel pieno rispetto del principio di autodeterminazione informativa, senza ostacolare l'innovazione e il progresso tecnologico.

Conclusione

Nell'ambito di questo studio, abbiamo condotto un'analisi approfondita del quadro normativo per la tutela dei dati personali nell'era dell'IA. Per iniziare, abbiamo sondato l'ampio scenario dell'IA a livello mondiale, evidenziando il ruolo fondamentale del Regolamento UE 2016/679 per la protezione dei dati personali. Ci siamo addentrati nell'evoluzione filosofica e tecnologica dell'IA, analizzando l'importanza della trasformazione del concetto di *privacy* nella società digitale attraverso l'analisi dei casi *Facebook v. Germany e Schrems I e II*. Successivamente, il nostro focus si è spostato sulla legislazione attuale, concentrandoci sul Regolamento Generale sulla Protezione dei Dati (GDPR) e su come si propone di regolare l'IA. Abbiamo esaminato le implicazioni dell'articolo 22 del GDPR per i sistemi di IA e l'importanza della valutazione dell'impatto sulla protezione dei dati personali in tali sistemi, con un approfondimento dedicato all'esempio delle *chatbot*. Inoltre, dopo un'attenta analisi dell'AI Act, abbiamo individuato i punti di convergenza tra il GDPR e l'AI Act, al fine di identificare possibili sinergie e divergenze tra le due normative.

Volgendo la nostra attenzione al concetto di '*General Purpose AI*', recentemente delineato nell'AI Act, abbiamo esaminato le specifiche tecniche dei *foundation models* e il loro ruolo nello sviluppo dei sistemi di IA, sottolineando l'importanza della *privacy* per un corretto sviluppo di queste tecnologie. Abbiamo analizzato inoltre le funzionalità di ChatGPT attraverso il test di Turing, dedicando particolare attenzione al provvedimento adottato dal Garante Italiano della *Privacy* nei confronti di ChatGPT.

Durante tutto il percorso di ricerca, sono emerse diverse criticità nella legislazione, sia attuale che futura, per la protezione dei dati personali nell'ambito dell'IA. Abbiamo evidenziato le sfide principali, tra cui quelle legate alla trasparenza, alla qualità dei dati e alla sorveglianza umana, emergendo la necessità di un approccio più flessibile. Il requisito della trasparenza assume un ruolo di fondamentale importanza, si manifesta attraverso la consapevolezza degli utenti sulla reale identità degli interlocutori con cui interagiscono online e sull'autenticità e veridicità delle relazioni instaurate con essi. In un'era digitale in continua evoluzione, sta diventando sempre più difficile discernere la vera identità di chi si trova dall'altro lato dello schermo, in particolare quando si

interagisce con sistemi di IA come ChatGPT. Questi, infatti, potrebbero raggiungere livelli di imitazione nella conversazione tali da indurre l'utente a credere di interagire con un essere umano, generando così un'ambiguità potenzialmente molto rischiosa. Sembra che il quesito posto da Alan Turing e che abbiamo analizzato nel primo capitolo di questa disamina riaffiori nel contesto della nostra trattazione. *The imitation game*, proposto da Turing, poneva in risalto l'ambiguità che un sistema di IA era in grado di generare nel corso di una conversazione con un individuo, mirando a simulare un dialogo umano tale da illudere un interlocutore. La seconda questione che abbiamo considerato riguardava la qualità dei dati utilizzati per l'addestramento, che abbiamo visto essere molto rigorosa nelle prescrizioni dell'AI Act, richiedendo anche standard difficilmente raggiungibili allo stato dell'arte. Inoltre, abbiamo sottolineato come le tecniche note per modificare i *dataset* e la conoscenza dei moderni GPAI non siano ancora pronte a soddisfare i requisiti stabiliti nella regolamentazione e quelle più idonee sono ancora in fase di sviluppo. L'ultima problematica riguardava le aspettative sulla sorveglianza umana e il fatto che questa non fosse ancora pronta a rispettare le prerogative stabilite dall'AI Act, tenendo conto anche della complessità dei sistemi stessi. Tuttavia, grazie alla combinazione di intelligenza naturale e artificiale, potrebbe essere possibile aiutare le persone a comprendere le dinamiche e la logica che sottendono i processi algoritmici.

Pertanto, il quadro normativo per i sistemi di IA dovrà necessariamente integrare l'AI Act, basandosi su tre principi: la trasparenza, la qualità dei dati e la sorveglianza umana, per permettere agli individui di comprendere meglio la natura e il funzionamento dei sistemi di IA.

Non a caso, la decisione del Garante nei confronti di OpenAI ha obbligato la società a lanciare una campagna informativa non promozionale, veicolata attraverso tutti i canali di comunicazione italiani, per chiarire la natura di ChatGPT come sistema di IA e per divulgare l'opzione di opporsi al trattamento dei dati da parte di OpenAI. L'intervento delle autorità riesce attualmente ad operare solo a livello superficiale su questi sistemi e modelli; ma, attraverso l'AI Act potrebbero essere stabilite delle regole che permettano di minimizzare i rischi per le persone che utilizzano tali sistemi, attraverso misure di trasparenza e il rispetto di alcuni standard di qualità nei dati utilizzati per l'addestramento. È di fondamentale importanza che gli utenti abbiano una

comprensione chiara e trasparente della natura artificiale delle *chatbot* basate su IA con cui interagiscono.

Di fronte all'esperienza passata legata al caso di *Cambridge Analytica*, si è capito come un ambiente digitale poco regolato possa portare a gravi conseguenze, mettendo in evidenza le potenziali ripercussioni negative connesse alla diffusione di informazioni false e alla manipolazione delle persone. D'altra parte, non è possibile interrompere lo sviluppo e l'innovazione dei sistemi di IA, vista la loro potenzialità benefica e il fatto che tutto il mondo sta puntando molto su questa tecnologia.

Di conseguenza, la via più adeguata da seguire risulta essere l'implementazione di regole che obblighino i produttori di sistemi di IA a prevenire l'ambiguità tra l'interazione umana e quella con la macchina, per evitare l'incertezza, o peggio ancora, l'errata convinzione di interagire con un essere umano, quando in realtà ci si sta interfacciando con una chatbot. Affiancando tali obblighi con parametri chiari e ragionevoli da seguire nell'addestramento dei sistemi, che non si limitino a stabilire una lista di requisiti sproporzionati, ma che siano veramente parametri formulati sulla natura dei sistemi di IA. Infine, concentrandosi su una riconsiderazione delle modalità per implementare la sorveglianza umana sui sistemi. In tal modo, si potrebbe tracciare e ritrovare la *retta via* per uno sviluppo all'insegna della sicurezza e dell'innovazione.

Bibliografia

Trattati e convenzioni internazionali

1. Convenzione del Consiglio d'Europa sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale, STCE n. 108, [1981]
2. Consiglio d'Europa, *Convenzione europea dei diritti dell'uomo*, STCE n. 5, 1950
3. Assemblea Generale delle Nazioni Unite, *Dichiarazione Universale dei Diritti Umani*, Risoluzione 219077A, (dicembre 1948).
4. Trattato sull'Unione europea, versione consolidata, [2012]. OJ C 326, 26.10.2012, p. 15-368.
5. Trattato sul funzionamento dell'Unione europea, versione consolidata, [2012]. OJ C 326, 26.10.2012, p. 47-390.
6. Carta dei diritti fondamentali dell'Unione europea, [2012] OJ C 326 del 26.10.2012, p. 391–407.
7. Trattato di Lisbona che modifica il trattato sull'Unione europea e il trattato che istituisce la Comunità europea, firmato a Lisbona il 13 dicembre 2007, GU C 306 del 17.12.2007

Legislazione europea

1. Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati), [2016], OJ L 119, 4.5.2016, p. 1–88

2. Regolamento (UE) 2019/881 Del Parlamento europeo e del Consiglio del 17 aprile 2019 relativo all'ENISA, l'agenzia dell'Unione europea per la Cibersicurezza, e alla certificazione della Cibersicurezza per le tecnologie dell'informazione e della comunicazione, e che abroga il regolamento (UE) n. 526/2013 («regolamento sulla cibersicurezza»), [2019] OJ L 151, p. 15–69
3. Direttiva (UE) 2016/680 del Parlamento Europeo e del Consiglio, del 27 aprile 2016, relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio [2016], OJ L 119/89
4. Direttiva 95/46/CE del Parlamento europeo e del Consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, [1995], OJ L 281, 23.11.1995, p. 31–50.
5. Direttiva (UE) 2019/1152 del Parlamento europeo e del Consiglio, del 20 giugno 2019, relativa a condizioni di lavoro trasparenti e prevedibili nell'Unione europea, O.J. L 186/105

Atti e decisioni della Commissione europea

1. Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle Regioni, *Strategia per il mercato unico digitale in Europa*, COM(2015) 192 final
2. Commissione europea, *Libro Bianco sull'Intelligenza Artificiale – Un approccio europeo all'eccellenza e alla fiducia* COM (2020) 65 final
3. C/2016/4176, Decisione di esecuzione della Commissione 2016/1250 [2016], OJ L 207, 1.8.2016, p. 1-112.
4. Decisione della Commissione 2000/520/CE [2000], OJ L 215, 25.8.2000.
5. High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific disciplines*, [2018].

6. Commissione europea, Directorate General for Communications Networks, Content and Technology, and High Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*. [2019]
7. Commissione europea, A Notification under Article 12 of Regulation (EU) No 1025/20121 (Dec. 2, 2022) (proposing a standardization request to the European Committee for Standardisation (CEN) and the European Committee for Electrotechnical Standardisation (CENELEC), two private standard-setting organizations), [2022]
8. Commissione Europea - Press release, "European Commission and United States Joint Statement on Trans- Atlantic Data Privacy Framework", Brussels, (25 March 2022).
9. Commissione Europea, '"Adequacy Decision on the EU-US Data Privacy Framework", C (2023) 4745 final
10. Commissione Europea, Ursula von Der Leien, "Data Protection: European Commission adopts new adequacy decision for safe and trusted EU-US data flows", 10 July 2023

Legislazione italiana

1. Regio Decreto 16 marzo 1942, n. 262, Approvazione del testo del Codice civile. (042U0262) (GU Serie Generale n.79 del 04-04-1942) [1942].
2. Costituzione della Repubblica Italiana [1948]
3. Decreto Legislativo 27 giugno 2022, n. 104, Attuazione della direttiva (UE) 2019/1152 del Parlamento europeo e del Consiglio del 20 giugno 2019, relativa a condizioni di lavoro trasparenti e prevedibili nell'Unione europea. (22G00113) (GU Serie Generale n.176 del 29-07-2022)
4. Legge 20 maggio 1970, n. 300, Norme sulla tutela della libertà e dignità dei lavoratori, della libertà sindacale e dell'attività sindacale, nei luoghi di lavoro e norme sul collocamento. (GU Serie Generale n.131 del 27-05-1970)
5. Decreto Legislativo 26 maggio 1997, n. 152, Attuazione della direttiva 91/533/CEE concernente l'obbligo del datore di lavoro di informare il lavoratore delle condizioni applicabili al contratto o al rapporto di lavoro. (GU Serie Generale n.135 del 12-06-1997)

Atti e decisioni delle autorità nazionali italiane

1. Garante per la protezione dei dati personali, “Provvedimento del 2 febbraio 2023” [2023], Doc-Web n. 9852214;
2. Garante per la protezione dei dati personali, Comunicato stampa ‘Intelligenza artificiale, dal Garante privacy stop al chatbot “Replika”. Troppi i rischi per i minori e le persone emotivamente fragili’ Doc-Web n. 9852506.
3. Garante per la protezione dei dati personali, Provvedimento dell'11 aprile 2023 doc-web n. 9874702
4. Garante per la protezione dei dati personali, Provvedimento doc-web n. 7496252
5. Garante della protezione dei dati personali, ‘Rapporto finale pubblicato nel 2020’, doc. web n. 9264297
6. Garante della protezione dei dati personali, ‘Deepfake-Vademecum’, doc. web n. 9512226
7. Garante per la protezione dei dati personali, ‘Intelligenza artificiale e ruolo della protezione dei dati personali’. doc-web n. 9855742,
8. Garante per la Protezione dei Dati Personali, ‘Redditometro: le garanzie dell'Autorità a seguito della verifica preliminare sul trattamento di dati personali effettuato dall'Agenzia delle entrate’, [2013] doc-web n. 2765110
9. Garante per la Protezione dei Dati Personali, ‘Provvedimento di autorizzazione al trattamento dei dati personali effettuato attraverso il Sistema di allerta Covid 19- App Immuni a seguito dell’aggiornamento della valutazione di impatto effettuata dal Ministero della salute su cui l’Autorità si era espressa con provvedimento del 1° giugno 2020’ [2021] Doc. web n. 9555987
10. Garante per la Protezione dei Dati Personali, ‘Ordinanza ingiunzione nei confronti di Clearview AI’ [2022], doc-web n. 9751362
11. Garante per la Protezione dei Dati Personali ‘Ordinanza ingiunzione nei confronti di Liceo Artistico Statale di Napoli’ [2020] doc-web n. 9283029
12. Garante della Protezione dei Dati Personali, ‘Provvedimento dell’8 giugno 2023’ [2023], doc. web n. 9896217

13. Garante della Protezione dei Dati Personali, “Relazione Annuale 2022”, 6 luglio 2023, [doc. web n. 9905999]
14. Garante della Protezione dei Dati Personali, ‘Provvedimento del 10 febbraio 2022’, n. 46, [doc. web n. 9751194]
15. Garante per la Protezione dei Dati Personali, ‘Ordinanza ingiunzione nei confronti di Foodinho s.r.l. del 10 giugno 2021’ n. 234 del 2021, [doc. web n. 9675440]
16. Garante per la Protezione dei Dati Personali, ‘Ordinanza ingiunzione nei confronti di Deliveroo Italy s.r.l. del 22 luglio 2021’ n. 285 del 2021, [doc. web n. 9685994]

Casi Corte di Giustizia dell’Unione europea (CGUE)

1. C-252/21: Request for a preliminary ruling from the Oberlandesgericht Düsseldorf (Germany) — *Facebook Inc. and Others v Bundeskartellamt*, [2021] OJ C 320 of 9.8.2021, p. 16–18.
2. C-131/12, *Google Spain SL e Google Inc. c Agencia Española de Protección de Datos (AEPD), Mario Costeja González*, [2014]. ECLI:EU:C:2014:317
3. C-507/17, *Google LLC v. Commission nationale de l'informatique et des libertés (CNIL)* [2019]. ECLI: ECLI:EU:C:2019:772.
4. C-634/21, *OQ contro Land Hessen, con l’intervento di SCHUFA Holding AG,* [2021]
5. C-362/14, *Maximilian Schrems c. Data Protection Commissioner*, [2015], ECLI:EU:C:2015:650
6. C-311/18, *Maximilian Schrems c. Data Protection Commissioner*, [2020], ECLI:EU:C:2020:559
7. C-201/14, *Smaranda Bara e a. / Președintele Casei Naționale de Asigurări de Sănătate e a.*, [2015], ECLI: ECLI:EU:C:2015:638
8. C-252/21, *Meta Platforms Inc. and others c. Bundeskartellamt*, [2023], ECLI:EU:C:2023:537

Casi Corte Europea dei Diritti dell'Uomo

1. *Ciubotaru c Moldova*, n. 27138/04, [2010]
2. *Bernh Larsen Holding AS e a. c. Norvegia*, n. 24117/08, [2013]

Atti delle istituzioni europee e internazionali

1. OECD*, 'Draft Recommendation of the Council on Artificial Intelligence'*, C(2019)34 (2019)
2. Parlamento Europeo, *Risoluzione del Parlamento europeo del 12 settembre 2018 sui sistemi d'arma autonomi (2018/2752(RSP))* (12 settembre 2018)
3. European Data Protection Board, *'Linee guida 3/2019 sul trattamento dei dati personali attraverso dispositivi video'*, [2019]
4. Consiglio dell'Unione Europea, Conclusioni del Consiglio – Plasmare il futuro digitale d'Europa (9 giugno 2020) 8711/20
5. Parlamento Europeo, *Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL))* (20 ottobre 2020)
6. Parlamento Europeo, *Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL))* (20 ottobre 2020);
7. Parlamento Europeo, *Risoluzione del Parlamento europeo del 20 ottobre 2020 sui diritti di proprietà intellettuale per lo sviluppo di tecnologie di intelligenza artificiale (2020/2015(INI))* (20 ottobre 2020);
8. EDPB – GEPD *Parere congiunto 5/2021 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale)*, (18 giugno 2021);
9. Parlamento Europeo, *Risoluzione del Parlamento europeo del 3 maggio 2022 sull'intelligenza artificiale in un'era digitale (2020/2266(INI))* (3 maggio 2022)

10. Parlamento europeo, Commissione per il mercato interno e la protezione dei consumatori e Commissione per le libertà civili, la giustizia e gli affari interni, "*Draft Compromise Amendments on the Draft Report, Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts*", (COM(2021)0206 – C9 0146/2021 – 2021/0106(COD))
11. Opinion of the Committee on Legal Affairs for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), (2023)

Casi corti straniere

1. Caso VI-Kart 1/19 (V), Decisione della Corte regionale superiore di Düsseldorf [2019]
2. Verwaltungsgericht Wiesbaden (Administrative Court, Wiesbaden), 6 K 788/20.WI [2020].
3. Bundesgerichtshof, decisione KVR 69/19, [2020].

Legislazione americana

1. Algorithmic Accountability Act of 2022, S. 3572 (IS)
2. Executive Order n. 14067, 9 March 2022

Atti e decisioni autorità nazionali tedesche

1. Decisione B6-22/16 del Bundeskartellamt [2019]

2. Andreas Mundt, *Press Release del Bundeskartellamt per il caso Facebook Germany*, (7 febbraio 2019).
3. Andreas Mundt, *Press Release del Bundeskartellamt per il caso Facebook Germany*, (4 July 2023)

Atti e decisioni autorità nazionali francesi

1. Commission nationale de l'informatique et des libertés '*The open source PIA software helps to carry out data protection impact assessment*' (30 Giugno 2021).

Atti e decisioni autorità nazionali inglesi

1. Information Commissioner's Office, '*How should we assess security and data minimisation in AI?*' (2021)

Atti e decisioni autorità nazionali australiane

1. Australian Human Rights Commission, Human Rights and Technology Final Report (2021)

Articoli giuridici ed economici

1. Anne SY Cheung and Yongxi Chen, '*From Datafication to Data State: Making Sense of China's Social Credit System and Its Implications*' 47 Law & Social Inquiry (2022).
2. Boris Lubrasky, '*Re-Identification of "Anonymized Data"*', Georgetown Law Technology Review (2016)

3. Carlos Ignacio Gutierrez and others, '*A Proposal for a Definition of General Purpose Artificial Intelligence Systems*' (5 October 2022).
4. Graham Greenleaf, '*The "Brussels Effect" of the EU's "AI Act" on Data Privacy Outside Europe*' (7 June 2021).
5. Glorin Sebastian, 'Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information' (May 2023).
6. Hannah Ruschemeier, '*AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal'* 23 ERA Forum (2023).
7. Jakob Mökander and others, '*The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What Can They Learn from Each Other?*' 32 Minds and Machines (2022).
8. Jennifer Cobbe and Jatinder Singh, '*Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges*' 42 Computer Law & Security Review 105573 (2021).
9. Johann Laux, '*Institutionalised Distrust and Human Oversight of Artificial Intelligence: Toward a Democratic Design of AI Governance under the European Union AI Act*' (3 March 2023)
10. Jorge Constantino, '*Exploring Article 14 of the EU AI Proposal: Human in the Loop Challenges When Overseeing High-Risk AI Systems in Public Service Organisations*' 14 Amsterdam Law Forum (2022).
11. Luciano Floridi and others, '*CapAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act*' (2022).
12. Luciano Floridi, '*AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models*' (14 February 2023).
13. Martin Ebers and others, '*The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)*' 4 J (2021).
14. Matthew S. Erie and Thomas Streinz, '*The Beijing Effect: China's Digital Silk Road as Transnational Data Governance*' 54 New York University Journal of International Law and Politics (2021).

15. Matthew U Scherer, '*Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies'* 29 Harvard Journal of Law & Technology (2015).
16. Mauritz Kop '*EU Artificial Intelligence Act: The European Approach to AI*' Transatlantic Antitrust and IPR Developments (2021).
17. Michael Chui, Eric Hazan, Roger Roberts, Alex Singla, Kate Smaje, Alex Sukharevsky, Lareina Yee, Rodney Zemmel '*The economic potential of generative AI*' McKinsey&Company, (June 2023)
18. Michael Veale and Frederik Zuiderveen Borgesius, '*Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach*' 22 Computer Law Review International (2021).
19. Nathalie A Smuha, '*Beyond the Individual: Governing AI's Societal Harm*' Internet Policy Review 10 (2021)
20. Nathalie A Smuha, '*From a "Race to AI" to a "Race to AI Regulation": Regulatory Competition for Artificial Intelligence*' 13 Law, Innovation and Technology (2021).
21. Noha Lea Halim, Urs Gasser, '*Vectors of AI governance - Juxtaposing the U.S. Algorithmic Accountability Act of 2022 with The EU Artificial Intelligence Act*', (2023)
22. Pauline Kim, '*Data-Driven Discrimination at Work*' 58 William & Mary Law Review (2017).
23. Reuben Binns and Michael Veale, '*Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR*' 11 International Data Privacy Law (2021).
24. Rupprecht Podszun, Philip Marsden, *Restoring balance to digital competition - Sensible rules, effective enforcement, Konrad-Adenauer-Stiftung e. V.* (2020).
25. Samuel D. Warren and Louis D. Brandeis, '*The Right to Privacy'* Harvard Law Review [1890].
26. Sandra Wachter, Brent Mittelstadt and Luciano Floridi, '*Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*' 7 International Data Privacy Law (2017).

27. Teresa Numerico, "*Dobbiamo ripensare l'intelligenza artificiale*" Limes (December 2022)
28. Toby Walsh, "*Will AI end privacy? How do we avoid an Orwellian future*", AI & Society 38 (2023)
29. Yogesh K. Dwivedi et al., "*Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy*", International Journal of Information Management 71 (2023)
30. Philipp Hacker, 'A Legal Framework for AI Training Data—from First Principles to the Artificial Intelligence Act' 13 Law, Innovation and Technology (2021)

Libri e manuali

1. Aristotele, *Politica*, Libro I, IV secolo a.C.
2. Thomas Hobbes, *Leviathan* (prima edizione 1651, Penguin 1985)
3. Alessandro Manzoni, *I promessi sposi* (prima edizione 1827, Letteratura Italiana Einaudi 1985).
4. Isaac Asimov, *'Runaround'* 29 Astounding science fiction (1942).
5. Blaise Pascal, *Machine arithmétique, Oeuvres Complètes de Blaise Pascal*, vol. X, (Louis Lafuma, Editions du Seuil 1972).
6. L. Frank Baum, *The Wonderful Wizard of Oz.* (Michael Patrick Hearn, W. W. Norton & Company 1980)
7. Tim Berners-Lee and Mark Fischetti, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*, (Harper San Francisco 1999).
8. Eric R Kandel, *Principles of Neural Science,* Fifth Edition (McGraw Hill Professional 2013).
9. Stanislas Dehaene, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (Penguin 2014).
10. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015).

11. Yann LeCun, Yoshua Bengio and Geoffrey Hinton, *'Deep Learning'* (Nature 2015).
12. Luigi Laura, *Breve e universale storia degli algoritmi*, (Luiss University Press 2019).
13. Marco Delmastro and Antonio Nicita, *Big data. Come stanno cambiando il nostro mondo* (Il Mulino 2019);
14. Luciano Floridi, *Pensare l'infosfera. La filosofia come design concettuale* (Cortina Raffaello 2020).
15. Anu Bradford, *'The Brussels Effect: How the European Union Rules the World'* [2020] Faculty Books
16. Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (Penguin Publishing Group 2020).
17. Laura DeNardis, *Internet in ogni cosa* (Luiss University Press 2021);
18. Teresa Numerico, *Big data e algoritmi. Prospettive critiche* (Carocci 2021);
19. Giovanni Maria Riccio, Paola Scorza and Ernesto Belisario, *GDPR e normativa privacy: commentario* (2., Wolters Kluwer 2022)
20. Alessandro Pajno, Filippo Donati, and Antonio Perrucci, *Intelligenza artificiale e diritto: una rivoluzione? Diritti fondamentali, dati personali e regolazione*., vol 1 (Il Mulino, 2022)

Articoli scientifici

1. John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, '*A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*', (31 agosto 1955) http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf .
2. Bruce G. Buchanan, '*A (very) brief history of artificial intelligence'* [2005] AI Magazine

3. Daniele Orso, Nicola Federici, Roberto Copetti, Luigi Vetrugno, Tiziana Bove '*Infodemic and the spread of fake news in the COVID-19-era'* [2020] European Journal of Emergency Medicine.
4. Gottfried Wilhelm Leibniz, *A Machine for Doing Multiplication and Division Mathematische Schriften*, Carl Immanuel Gerhardt (Hahn'sche Buchhandlung, 1849).
5. Federico Luigi Menabrea, '*Sketch of the Analytical Engine invented by Charles Babbage, Esq*.' Tradotto da Ada Lovelace, Scientific Memoirs, Vol. III, Richard Taylor and William Francis, (1843).
6. Stephanie Faint, '*The Enigma History and Mathematics'* (MS thesis, University of Waterloo 2016).
7. Alan M. Turing, '*On Computable Numbers, with an Application to the Entscheidungsproblem'* (1937) Proceedings of the London Mathematical Society, vol. 42, no. 1.
8. Alan M. Turing, '*Computing Machinery and Intelligence'*, [1950] Mind 236 https://doi.org/10.1093/mind/LIX.236.433
9. Barry M. Leiner et al. '*A brief history of the Internet'* [2009] ACM SIGCOMM Computer Communication Review.
10. DA Ferrucci, '*Introduction to "This Is Watson*"' (2012) 56 IBM Journal of Research and Development.
11. Nick Seaver, '*Captivating Algorithms: Recommender Systems as Traps'* (2019) 24 Journal of Material Culture.
12. John Jumper and others, '*Highly Accurate Protein Structure Prediction with AlphaFold'* (2021) Nature.
13. Angela Chen, '*Why San Francisco's ban on face recognition is only the start of a long fight*,' in MIT Review, [2019]
14. Rainer Mühlhoff, '*Predictive Privacy: Collective Data Protection in the Context of Artificial Intelligence and Big Data'* (2023) 10 Big Data & Society.
15. Ben Shneiderman, '*Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems'* (2020) 10 ACM Transactions on Interactive Intelligent Systems.
16. Joy Adowaa Buolamwini, '*Gender Shades : Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers'* (Thesis,

Massachusetts Institute of Technology 2017) https://dspace.mit.edu/handle/1721.1/114068 accessed 27 April 2023.

17. Dallas Hill, Christopher D O'Connor and Andrea Slane, '*Police Use of Facial Recognition Technology: The Potential for Engaging the Public through Co-Constructed Policy-Making*' (2022) 24 International Journal of Police Science & Management.
18. Peter Fussey and Daragh Murray, '*Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*' https://repository.essex.ac.uk/24946/ accessed 27 May 2023.
19. Heike Felzmann and others, '*Towards Transparency by Design for Artificial Intelligence*' (2020) 26 Science and Engineering Ethics.
20. Mohammad Hossein Jarrahi, Christoph Lutz and Gemma Newlands, '*Artificial Intelligence, Human Intelligence and Hybrid Intelligence Based on Mutual Augmentation*' (2022) 9 Big Data & Society 20539517221142824.
21. David De Cremer and Garry Kasparov, '*AI Should Augment Human Intelligence, Not Replace It*' [2021] Harvard Business Review https://hbr.org/2021/03/ai-should-augment-human-intelligence-not-replace-it accessed 27 May 2023.
22. Ethan Mollick, '*ChatGPT Is a Tipping Point for AI*' [2022] Harvard Business Review https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai accessed 5 May 2023.
23. Enkelejda Kasneci and others, '*ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*' (2023) 103 Learning and Individual Differences
24. Scott Reed and others, '*A Generalist Agent*' https://arxiv.org/abs/2205.06175 accessed 2 May 2023.
25. Tom B Brown and others, '*Language Models Are Few-Shot Learners*' https://arxiv.org/abs/2005.14165 accessed 5 May 2023.
26. H Holden Thorp, '*ChatGPT Is Fun, but Not an Author*' (2023) 379 Science.
27. Jonathan H Choi and others, '*ChatGPT Goes to Law School*' [2023] SSRN Electronic Journal https://www.ssrn.com/abstract=4335905 accessed 27 April 2023.

28. Fawaz Qasem, '*ChatGPT in Scientific and Academic Research: Future Fears and Reassurances'* (2023) 40 Library Hi Tech News.
29. Luciano Floridi and Massimo Chiriatti, '*GPT-3: Its Nature, Scope, Limits, and Consequences*' (2020) 30 Minds and Machines.
30. Luciano Floridi and others, '*How to Design AI for Social Good: Seven Essential Factors'* (2020) 26 Science and Engineering Ethics.
31. Luciano Floridi, '*Machine Unlearning: Its Nature, Scope, and Importance for a "Delete Culture"*' (arXiv, 24 May 2023) http://arxiv.org/abs/2305.15242 accessed 28 May 2023.
32. Liu, Yi, Gelei Deng, Yuekang Li, Kailong Wang, Tianwei Zhang, Yepang Liu, Haoyu Wang, Yan Zheng, and Yang Liu. "*Prompt Injection attack against LLM-integrated Applications*." arXiv preprint arXiv:2306.05499 (2023);
33. Alexander Wei, Nika Haghtalab, and Jacob Steinhardt. "Jailbroken: *How Does LLM Safety Training Fail?*." arXiv preprint arXiv:2307.02483 (2023)

Blog, siti web e articoli di periodici online
(Ultimo accesso effettuato il 3 settembre 2023)

1. OECD, *Trends and Data overview*, https://oecd.ai/en/trends-and-data
2. Sheehan Matt, '*China's New AI Governance Initiatives Shouldn't Be Ignored*' (Carnegie Endowment for International Peace 4 January 2022) https://carnegieendowment.org/2022/01/04/china-s-new-ai-governance-initiatives-shouldn-t-be-ignored-pub-86127
3. Jorge Liboreiro, '"*The higher the risk, the stricter the rule": Brussels' new draft rules on artificial intelligence*' Euronews (Brussels, 21 aprile 2021), < https://www.euronews.com/my-europe/2021/04/21/the-higher-the-risk-the-stricter-the-rule-brussels-new-draft-rules-on-artificial-intellige>
4. Philip Di Salvo, '*A 30 anni dalla nascita, ecco tutti i rischi che corre il world wide web*' (Wired, 11 marzo 2019), https://www.wired.it/internet/web/2019/03/11/world-wide-web-30-anni/

5. Nicoletta Cottone, '*Italia leader nella robotica: 100 storie di eccellenza*' Il Sole 24 ORE (5 febbraio 2020) https://www.ilsole24ore.com/art/italia-leader-robotica-100-storie-eccellenza-ACSEpOHB
6. Michael Aaron Dennis, '*Defense Advanced Research Projects Agency'*, Encyclopedia Britannica, (23 dicembre 2022), https://www.britannica.com/topic/Defense-Advanced-Research-Projects-Agency
7. World Internet Users Statistics and 2023 World Population Stats (2023) http://internetworldstats.com.
8. Glenn Greenwald, Ewen MacAskill, *NSA Prism program taps in to user data of Apple, Google and others*, The Guardian, (7 giugno 2013), Accessibile al link: https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data
9. Amnesty International, '*Out of Control: Failing EU Laws for Digital Surveillance Export*' (2020) < https://www.amnesty.org/en/documents/eur01/2556/2020/en/ >.
10. Vikram Singh Bisen, '*What Is Human in the Loop Machine Learning: Why & How Used in AI?*' (VSINGHBISEN, 16 August 2022) https://medium.com/vsinghbisen/what-is-human-in-the-loop-machine-learning-why-how-used-in-ai-60c7b44eb2c0.
11. Andreas Häuselmann, '*The ECJ's First Landmark Case on Automated Decision-Making – a Report from the Oral Hearing before the First Chamber*' (European Law Blog, 20 February 2023) https://europeanlawblog.eu/2023/02/20/the-ecjs-first-landmark-case-on-automated-decision-making-a-report-from-the-oral-hearing-before-the-first-chamber/.
12. Dave Lee, '*Tay: Microsoft Issues Apology over Racist Chatbot Fiasco*' BBC News (25 March 2016) https://www.bbc.com/news/technology-35902104;
13. Abby Ohlheiser, '*Trolls Turned Tay, Microsoft's Fun Millennial AI Bot, into a Genocidal Maniac*' Washington Post (25 March 2016) https://www.washingtonpost.com/news/the-intersect/wp/2016/03/24/the-internet-turned-tay-microsofts-fun-millennial-ai-bot-into-a-genocidal-maniac/.
14. Elle Hunt, '*Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter*' The Guardian (24 March 2016)

https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter.

15. Vagelis Papakonstantinou and Paul De Hert, '*EU Lawmaking in the Artificial Intelligent Age: Act-Ification, GDPR Mimesis, and Regulatory Brutality*' [2021] European Law Blog.
16. Treccani Enciclopedia on line, '*fenotipo*' https://www.treccani.it/enciclopedia/fenotipo/.
17. Jeffrey Dastin and Ayanti Bera, '*Amazon Pauses Police Use of Its Facial Recognition Tech for a Year'* Reuters (10 June 2020) https://www.reuters.com/article/us-amazon-com-facial-recognition-idUSKBN23H3EO.
18. Microsoft, *'Response to the European Commission's Consultation on the Artificial Intelligence Act'* (2021) https://blogs.microsoft.com/wp-content/uploads/prod/sites/73/2021/09/microsoft-response-to-the-european-commission-consultation-on-the-artifical-intelligence-act.pdf
19. Access Now, '*Access Now's submission to the European Commission's adoption consultation on the AI Act'* (2021) https://www.accessnow.org/cms/assets/uploads/2021/06/Access-Now-submission-to-the-European-Commissions-adoption-consultation-on-the-AI-Act.pdf
20. Billy Perrigo, '*Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer*' (Time, 18 January 2023) https://time.com/6247678/openai-chatgpt-kenya-workers/.
21. Kif Leswing, '*Bloomberg Plans to Integrate GPT-Style A.I. into Its Terminal*' (CNBC, 13 April 2023) https://www.cnbc.com/2023/04/13/bloomberg-plans-to integrate-gpt-style-ai-into-its-terminal.html
22. Chelsea Jarvie and Karen Renaud, '*Are You over 18? A Snapshot of Current Age Verification Mechanisms*', 2021 Dewald Roode Workshop (2021) https://strathprints.strath.ac.uk/82540/.
23. John Schulman et al., *ChatGPT: Optimizing Language Models for Dialogue*, OPENAI [2022], https://openai.com/blog/chatgpt/

24. Nick Evershed and Josh Taylor, '*AI Can Fool Voice Recognition Used to Verify Identity by Centrelink and Australian Tax Office*' The Guardian (16 March 2023) https://www.theguardian.com/technology/2023/mar/16/voice-system-used-to-verify-identity-by-centrelink-can-be-fooled-by-ai
25. OpenAI, '*March 20 ChatGPT outage: Here's what happened*', <https://openai.com/blog/march-20-chatgpt-outage>
26. EDPB, '*EDPB Resolves Dispute on Transfers by Meta and Creates Task Force on Chat GPT | European Data Protection Board*' https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en.
27. OpenAI, '*Privacy Policy*' https://openai.com/policies/privacy-policy.
28. OpenAI, '*How Your Data Is Used to Improve Model Performance | OpenAI Help Center*' https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance
29. OpenAI, *FAQ* < https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed>
30. Joshua J., '*Data controls FAQ*', < https://help.openai.com/en/articles/7730893-data-controls-faq>
31. Dan H and ThomasW, '*A Bird's Eye View of the ML Field [Pragmatic AI Safety #2]' s AGI and AI Safety* https://www.alignmentforum.org/posts/AtfQFj8umeyBBkkxa/a-bird-s-eye-view-of-the-ml-field-pragmatic-ai-safety-2
32. Reuters, '*Google AI Pioneer Says He Quit to Speak Freely about Technology's "Dangers"*' Reuters (9 May 2023) https://www.reuters.com/technology/google-ai-pioneer-says-he-quit-speak-freely-about-technologys-dangers-2023-05-02/;
33. Johana Bhuiyan, '*OpenAI CEO Calls for Laws to Mitigate "Risks of Increasingly Powerful" AI*' The Guardian (16 May 2023) https://www.theguardian.com/technology/2023/may/16/ceo-openai-chatgpt-ai-tech-regulations;
34. Tom Lamont, '*Meet Watson, the computer set to outsmart the champions of Jeopardy!*', The Guardian (6 February 2011)

35. Cade Metz and Gregory Schmidt, '*Elon Musk and Others Call for Pause on A.I., Citing "Profound Risks to Society"*' The New York Times (29 March 2023) https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html

36. Billy Perrigo, '*Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation*', Time (20 June 2023) https://time.com/6288245/openai-eu-lobbying-ai-act/

37. Max Schrems, "European Commission gives EU-US data transfers third round at CJEU", <https://noyb.eu/en/european-commission-gives-eu-us-data-transfers-third-round-cjeu>, 10 July 2023