



Comprendere per regolare: brevi considerazioni sulla regolazione europea dell'intelligenza artificiale. A proposito del recente volume di N. CRISTIANINI, *Machina sapiens. L'algoritmo che ci ha rubato il segreto della conoscenza*, Bologna, il Mulino, 2024*

di Giulia Taraborrelli**

SOMMARIO: 1. Intelligenza artificiale e diritto. – 2. La costruzione di macchine pensanti. – 3. Intelligenza artificiale, utenti, competizione. – 4. Le macchine e il concetto di intelligenza.

1. Intelligenza artificiale e diritto

Il rapido sviluppo di nuove tecnologie nel corso dell'ultimo decennio ha determinato la necessità di affrontare inedite prospettive interpretative e di soddisfare nuove istanze regolatorie. Elementi e concetti propri dell'ambito scientifico sono stati progressivamente integrati nel dominio giuridico, al fine di valutare adeguatamente le implicazioni delle tecnologie emergenti e formulare una regolazione che ne riconosca efficacemente le opportunità e le sfide associate¹.

Nello Cristianini, professore di Intelligenza Artificiale all'Università di Bath, propone di tracciare alcuni itinerari per definire il concetto di intelligenza artificiale e di modello di linguaggio a partire da una fondamentale domanda, posta da Alan Turing nel 1950, ossia se le macchine siano in grado di pensare.

Per tentare di rispondere al quesito, l'opera è stata ripartita in tre sezioni, utili a tracciare lo sviluppo diacronico di ChatGPT e dei sistemi affini, in modo tale da passare il testimone tra i diversi soggetti che assistono allo sviluppo tecnologico, vale a dire scienziati e società civile, per concludere esplorando lo scenario che si profila davanti alle macchine stesse. Segue un «glossario informale», che fornisce alcuni elementi essenziali per la comprensione dei concetti fondamentali trattati all'interno del volume.

* Contributo sottoposto a *peer review*.

** Dottoranda di ricerca in Diritto pubblico, comparato e internazionale – Sapienza Università di Roma.

¹ D.U. GALETTA, *Human-stupidity-in-the-loop? Riflessioni (di un giurista) sulle potenzialità e i rischi dell'Intelligenza Artificiale*, in *Federalismi.it*, n. 5/2023, IV-XIII, V: «Anche gli studiosi di area umanistica che si avvicinano al tema IA – e i giuristi rientrano ovviamente all'interno di questa categoria - si trovano sovente in una situazione di vero e proprio “deficit cognitivo”: nel senso che discutono di un oggetto per loro largamente sconosciuto».

Lo studio del fenomeno dell'intelligenza artificiale sul piano tecnico è fondamentale per riuscire ad affrontare le innumerevoli sfide sul piano giuridico che tale tecnologia pone, in considerazione della sua ascrizione tra le «disruptive technologies», che non si limitano a adottare paradigmi tradizionali per migliorare prodotti e servizi esistenti, ma che, impiegando processi tecnici innovativi, creano delle innovazioni senza precedenti².

La prima essenziale definizione per orientarsi all'interno dell'opera riguarda proprio il concetto di «intelligenza artificiale», definita dall'Autore come «la scienza e l'arte di costruire macchine che possono agire (agenti), imparare, ragionare, perseguendo obiettivi in modo autonomo, adattandosi a situazioni imprevedute» (p. 152).

L'intelligenza artificiale può essere considerata parte delle *general purpose technologies* (GPTs), ossia «tecnologie che hanno un impatto vasto e vario su molti settori dell'economia, della società, della cultura, del lavoro, capaci pertanto di provocare trasformazioni significative nella produttività e nelle strutture economiche e sociali»³, seguendo ritmi non prevedibili nella diffusione sempre più estesa in prodotti e servizi.

La definizione di *general purpose models*, che corrisponde alla locuzione «modello di IA per finalità generali» contenuta nel Regolamento europeo sull'intelligenza artificiale o AI Act⁴, si riferisce a modelli di intelligenza artificiale che sono in grado di svolgere funzioni disparate sulla base di un addestramento condotto su una vasta quantità di dati⁵. Tale definizione comprende anche i *Large Language Models*, tra cui ChatGPT, che è riconducibile all'area dell'IA generativa, in quanto in grado di produrre nuovi contenuti rielaborando grandi masse di dati⁶.

Il Regolamento, che tra i principali obiettivi intende migliorare il funzionamento del mercato interno attraverso l'istituzione di un quadro giuridico uniforme per quanto riguarda i sistemi di intelligenza artificiale⁷, costituisce una regolazione di tipo orizzontale incentrata sul rischio, poiché l'impianto regolatorio è suddiviso in base al livello di rischio

² Sul concetto di *disruptive innovation*, si vedano J. L. BOWLER, C. M. CHRISTENSEN, *Disruptive Technologies: Catching the Wave*, in *Harvard Business Review*, January-February 1995, 44-55; C. M. CHRISTENSEN, *The Innovator's Dilemma. When New Technologies Cause Great Firms to Fail*, Boston, Harvard Business School Press, 1997; C. M. CHRISTENSEN, R. MCDONALD, E. J. ALTMAN, J. E. PALMER, *Disruptive Innovation: An Intellectual History and Directions for Future Research*, in *Journal of Management Studies*, 2018, 36; C. CAFFARRA, O. LATHAM, *Is Antitrust in Need of Disruption: What Is Disruptive Innovation and What, if Anything, Does Competition Policy Need To Do to React To It?*, in *Rivista Italiana di Antitrust*, n. 1/2018, 9.

³ G. LO SAPIO, *L'Artificial Intelligence Act e la prova di resistenza per la legalità algoritmica*, in *Federalismi.it*, n. 16/2024, 265-290, 266.

⁴ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale).

⁵ Il Considerando n. 97 spiega infatti che «la nozione di modelli di IA per finalità generali dovrebbe essere chiaramente definita e distinta dalla nozione di sistemi di IA per consentire la certezza del diritto. La definizione dovrebbe basarsi sulle principali caratteristiche funzionali di un modello di IA per finalità generali, in particolare la generalità e la capacità di svolgere con competenza un'ampia gamma di compiti distinti. Questi modelli sono solitamente addestrati su grandi quantità di dati con diversi metodi, come l'apprendimento autosupervisionato, non supervisionato o per rinforzo».

⁶ G. LO SAPIO, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, in *Federalismi*, n. 16/2021, 114-127. Per una trattazione in merito alle questioni sottese alla definizione di intelligenza artificiale prescelta all'interno del Regolamento, si veda G. PROIETTI, *Definire l'indefinibile? I sistemi di intelligenza artificiale alla ricerca di un inquadramento sistematico*, in *Contratto e impresa*, n. 3/2024, 882-925.

⁷ Regolamento (UE) 2024/1689, Considerando n. 1.

rappresentato da un determinato sistema di intelligenza artificiale – rischio inaccettabile, ad alto rischio, rischio basso o minimo – e a seconda del soggetto coinvolto a seconda che si tratti di un utilizzatore, uno sviluppatore, un fornitore o un distributore.

I sistemi di intelligenza artificiale vietati dal Regolamento sono caratterizzati da un «rischio inaccettabile», invece i sistemi «ad alto rischio» sono ammessi, seppur subordinati a un controllo *ex ante* e sottoposti a particolari obblighi, alcuni dei quali inerenti alla cybersicurezza e all'uso dei dati⁸. Ai sistemi a rischio limitato sono correlati obblighi di trasparenza, mentre rispetto ai sistemi a basso rischio sussiste un mero invito allo sviluppo di codici di condotta⁹.

Il Capo V del Regolamento è dedicato ai modelli di intelligenza artificiale per finalità generali, stabilendo alcuni obblighi per i fornitori e definendo obblighi ulteriori – analoghi a quelli previsti per i sistemi di AI classificati ad alto rischio – per i fornitori modelli di IA per finalità generali con rischio sistemico, ossia modelli che hanno una «capacità di impatto elevato valutate sulla base di strumenti tecnici e metodologie adeguati»¹⁰.

ChatGPT, essendo un sistema di intelligenza artificiale composto a sua volta da un modello di IA per finalità generali, in quanto tale non è categorizzato come sistema ad alto rischio, ma la Commissione europea, se venisse a conoscenza che un modello di IA per finalità generali presenta un rischio sistemico, potrà decidere di designare il modello come ad alto rischio¹¹.

La dottrina si è a lungo soffermata sulla classificazione in livelli di rischio e sulle conseguenze derivanti dalla diffusione dell'intelligenza artificiale, nel tentativo di valutare se la regolazione europea sia in grado di garantire adeguati livelli di tutela e di minimizzare le esternalità negative connesse al progresso dell'IA, soprattutto in termini di opacità, complessità, dipendenza dai dati e comportamento autonomo della macchina¹².

⁸ L'art. 5 del Regolamento vieta alcune pratiche di IA, tra cui ad esempio un sistema che utilizzi tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli, oppure strumenti che sfruttino le vulnerabilità connesse all'età o alla disabilità fisica o mentale. I sistemi di IA ad alto rischio sono individuati all'art. 6 del Regolamento e consistono in sistemi di IA destinati a essere utilizzati come componenti di sicurezza di prodotti e sistemi di AI che rientrano in uno o più settori critici e casi d'uso se presentano un rischio significativo di danno per la salute umana, la sicurezza o i diritti fondamentali delle persone fisiche. Vi rientrano i sistemi di IA destinati a essere utilizzati nei settori dell'istruzione, della sanità, della selezione del personale, della sicurezza, dell'amministrazione della giustizia e della pubblica amministrazione se incidono negativamente sui diritti fondamentali dei cittadini.

⁹ Il Regolamento riconduce ai sistemi connotati da rischio limitato l'IA impiegata per generare o manipolare contenuti audiovisivi o per fornire suggerimenti personalizzati, includendo quindi *deepfakes* e *chatbots*. Rientrano invece nella categoria del rischio minimo o nullo i sistemi di IA utilizzati per scopi ludici o per scopi puramente estetici, come videogiochi o filtri fotografici.

¹⁰ Regolamento (UE) 2024/1689, art. 51.

¹¹ D. CLEMENTI, *Generare e non creare? Spunti per una comparazione sulla regolazione dell'intelligenza artificiale generativa tra Stati Uniti, Repubblica Popolare Cinese e Unione Europea*, in *Rivista di diritti comparati*, n. 2/2024, 371-410, 404.

¹² Tra i contributi più recenti in materia, si veda G. PESCE, *L'intelligenza artificiale alla prova del diritto europeo: verso il diritto della paura?*, in *Rivista di diritto amministrativo*, n. 1/2024, 14-26; C. NOVELLI, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in *Federalismi.it*, n. 2/2024, 95-113; G. SMORTO, *La distribuzione del rischio e la tutela dei diritti nel regolamento europeo sull'intelligenza artificiale. Una riflessione critica*, in *Il foro italiano*, n. 5/2024, 208-220. Per un esame della proposta di regolamento sull'intelligenza artificiale, si veda F. RODI, *Gli interventi dell'Unione europea in materia di intelligenza artificiale e robotica: problemi e prospettive*, in G. ALPA (a cura di), *Diritto e intelligenza artificiale*, Pisa, Pacini, 2020, 187-210; L. PARONA, *Prospettive europee e internazionali di regolazione dell'intelligenza artificiale tra principi etici, soft law e self-regulation*, in *Rivista di regolazione dei mercati*, n. 1/2020, 70-93.

Le difficoltà sul piano regolatorio sono accentuate in relazione ai modelli di linguaggio, che in ragione della loro flessibilità possono essere impiegati per svolgere i compiti più disparati e che hanno dimostrato, nel giugno 2023, di non poter soddisfare i criteri imposti dal Regolamento sull'intelligenza artificiale se non attraverso l'adattamento delle politiche industriali del settore¹³.

Alla luce costanti incertezze relative alla disciplina dell'intelligenza artificiale, alcuni interessanti spunti di riflessione potrebbero risiedere in una comprensione approfondita delle caratteristiche tecniche proprie delle ultime versioni di ChatGPT e delle fasi che hanno contribuito al suo sviluppo, il quale ha registrato un'accelerazione significativa negli ultimi mesi¹⁴.

2. La costruzione di macchine pensanti

L'Autore introduce il volume affermando che non è noto il funzionamento esatto di ChatGPT e dei sistemi affini, ma che questi si concretizzano come modelli di linguaggio. Pur essendo sistemi piuttosto semplici dal punto di vista tecnico, sono dotati di un'enorme «intelligenza» grazie alla quantità straordinaria di testi su cui il sistema è stato «addestrato», documenti che non erano mai stati messi in relazione prima d'ora. Il rilievo del fenomeno emerge da due constatazioni: il comportamento degli attuali sistemi di intelligenza artificiale non è paragonabile ai risultati ottenuti con le tecnologie precedenti e lo scarto in termini di conoscenza si è manifestato spontaneamente, eludendo le previsioni degli stessi programmatori.

Gli albori dello sviluppo dei sistemi artificiali sono segnati dal tentativo di soddisfare il «gioco dell'imitazione» teorizzato da Alan Turing: una macchina può essere considerata pensante solo se si dimostra in grado di tenere una conversazione in linguaggio naturale¹⁵ e su qualsiasi argomento fingendosi un essere umano e riuscendo a ingannare l'intervistatore riguardo alla propria identità. Per soddisfare il «test di Turing», nel corso della seconda metà del XX secolo la comunità scientifica ha tentato di infondere nei sistemi di intelligenza artificiale la capacità di comprendere e generare il linguaggio umano e, al contempo, di fornire loro quante più conoscenze possibili sul funzionamento del mondo. Tali sistemi erano in grado di rispondere alle domande formulate dall'utente, ma non avevano la

¹³ R. BOMMISANI, K. KLYMAN, D. ZHANG, P. LIANG, *Do Foundation Model Providers Comply with the Draft EU AI Act?*, Human-Centered Artificial Intelligence, Stanford University, 2023.

¹⁴ In merito alla difficoltà definitoria dell'intelligenza artificiale, M. U. SCHERER, *Regulating Artificial Intelligence Systems: risks, challenges competencies, and strategies*, in Harvard Journal of Law & Technology, n. 2/2016, 353-400. Sul tema, si veda C. EQUIZI, *Intelligenza artificiale: profili di opportunità e di criticità nella irrinunciabile tutela dei diritti fondamentali*, in *Diritti fondamentali*, n. 1/2024, 313-331.

¹⁵ I linguaggi naturali sono stati definiti da G. CAVAGGION, M. OROFINO, *Lingua e Costituzione: l'irrompere dei linguaggi algoritmici*, in Rivista AIC, n. 4/2023, 155-184, 171-172 come «i linguaggi, emersi “naturalmente”, concretamente e abitualmente usati nella comunicazione fra individui di un dato gruppo sociale», in opposizione rispetto al linguaggio algoritmico, mera «sequenza ordinata di passi ben definiti per risolvere un problema o eseguire una determinata operazione». Sul tema, si veda anche F. MARTELLI, R. NAVIGLI, *Il senso di interagire con la macchina in linguaggio naturale*, in Gnosis, n. 2/2018, 81-87. In merito al dialogo fra linguaggio algoritmico e linguaggio normativo, si veda S. SALARDI, *Intelligenza artificiale e semantica del cambiamento: una lettura critica*, Torino, Giappichelli, 2023.

capacità di dialogare, come nel caso del sistema Eliza, sviluppato da Joseph Weizenbaum nel 1966, o del sistema CYC, ideato da Douglas Lenat nel 1984.

Un progresso significativo in termini di modelli di linguaggio e dei necessari «modelli del mondo» è stato registrato solo nel 2017, con lo sviluppo dell'algoritmo Transformer, appartenente alla famiglia di metodi di apprendimento chiamati «reti neurali». Sfruttando un nuovo tipo di processore (la GPU, *Graphic Processing Unit*), Transformer ha consentito ai computer di analizzare enormi quantità di testo in tempi molto ristretti e senza la necessità di alcun intervento umano, aprendo la strada alla intelligenza artificiale detta «generativa», poiché è capace di produrre contenuti nuovi in testi linguistici, musica, immagini o video rispetto ai dati di addestramento¹⁶.

L'Autore delinea quindi i tre livelli di cui si discute in relazione alle tecnologie di intelligenza artificiale: l'agente (ad esempio, ChatGPT), il modello interno che l'agente impiega per adottare le proprie decisioni (ad esempio, GPT-3) e l'algoritmo, come nel caso di Transformer, che crea il modello sulla base di una serie di dati. Questi ultimi possono essere qualificati come «etichettati» o «annotati» quando sono stati accuratamente vagliati da un supervisore – di qui l'apprendimento «supervisionato» – a differenza dei dati «grezzi», acquisiti direttamente dalla macchina, disponibili in grandi quantità, ma di costo e di qualità inferiori.

I ricercatori di OpenAI, impegnati nello sviluppo di un modello generale di linguaggio, nel 2018 hanno rivelato di aver rivoluzionato la tecnica di sviluppo delle macchine intelligenti ripartendo l'addestramento in una prima fase, volta a creare un modello di linguaggio sulla base di grandi quantità di testo grezzo – economico e facilmente reperibile – e in una seconda fase di «raffinamento», che aveva insegnato all'agente compiti specifici grazie a dati curati a mano. La prima fase, che ha rivelato una vera novità, è stata condotta da una versione implementata dell'algoritmo Transformer, addestrato sulla base di una quantità enorme di testi, dimostrando l'innovativa abilità del modello di linguaggio di trasferire ad altri compiti, che in genere richiedono l'impiego di dati costosi, le conoscenze linguistiche apprese addestrandosi su una serie di testi.

Considerando che il modello di linguaggio in questione era stato creato preaddestrando l'algoritmo Transformer in modo generativo, OpenAI ha scelto di chiamarlo *Generatively Pertained Transformer*, o GPT.

Mentre misuravano l'effetto di impiegare diverse quantità di dati nelle diverse fasi dell'addestramento, i ricercatori hanno notato due comportamenti anomali da parte di GPT. Il sistema ha mostrato di avere la capacità di apprendimento senza supervisione esplicita, rispondendo alle domande anche prima della fase di raffinamento, sulla base delle conoscenze acquisite nella sola fase di preaddestramento, senza quindi la necessità dei costosi dati «etichettati». GPT ha anche dimostrato un'elevata capacità di apprendimento di nuove abilità senza la necessità di migliaia di esempi, superando notevolmente i tempi di apprendimento algoritmici necessari fino a quel momento per addestrare un algoritmo.

¹⁶ G. LO SAPIO, *L'Artificial Intelligence Act e la prova di resistenza per la legalità algoritmica*, cit., 268.

Nel 2020 OpenAI ha ritenuto che la capacità più sorprendente di ChatGPT, quella di imparare sulla base del contesto, avrebbe potuto migliorare con l'aumento delle dimensioni dei modelli o dei dati a disposizione. Grazie alla dotazione di un supercomputer da parte di Microsoft è stato possibile sviluppare GPT-3, addestrato su un corpus di circa 500 miliardi di parole in tempi estremamente brevi, contro i 355 anni che sarebbero stati necessari impiegando normali GPU.

Al termine dell'addestramento, OpenAI ha annunciato che GPT-3, pur essendo stato costruito come mero modello linguistico, è in grado di eseguire compiti complessi sulla base di dati generici, esaminando solo pochi esempi del compito richiesto. Non è tuttavia noto in che modo sia stato possibile raggiungere tale risultato, poiché le abilità non derivano dall'algoritmo, ma dal modo in cui questo interagisce con l'infinità di dati impiegati per alimentarlo, trovando concordanze e connessioni nuove e non prevedibili nella fase di addestramento.

Tali circostanze sollevano alcune considerazioni in merito all'imprevedibilità dei sistemi di intelligenza artificiale e al loro rapporto con i dati. Se infatti non è possibile ricostruire il processo decisionale del sistema, allora l'intelligenza artificiale presenta una natura fortemente opaca, complicando l'intervento umano su di essa. La dottrina ha individuato nel carattere di «black box» della macchina l'ostacolo alla possibilità di ricostruirne la logica e di comprendere quali dati di addestramento siano stati utilizzati per elaborare una determinata decisione, rendendo impossibile anche la correzione di errori derivanti da tali dati¹⁷.

Un ulteriore elemento che concorre a definire l'opacità dei sistemi di intelligenza artificiale corrisponde alla loro naturale «instabilità», ossia il fatto che un *software* per definizione è in continua evoluzione, soggetto ad aggiornamenti e modifiche per scongiurarne il malfunzionamento, tracciando una «opacità a geometria variabile»¹⁸.

Il *black box effect* è stato mitigato nel Regolamento europeo sull'intelligenza artificiale con misure volte a garantire il funzionamento trasparente della macchina e il controllo umano sull'*output* del sistema¹⁹, risulta difficile riuscire nell'intento quando gli stessi sviluppatori non sono in grado di comprendere come la macchina abbia potuto raggiungere determinati risultati.

¹⁷ B. MARCHETTI, *Intelligenza artificiale, poteri pubblici e rule of law*, in *Rivista italiana di Diritto pubblico comunitario*, n. 1/2024, 49-64, 55-56; A. MONREALE, *Rischi etico-legali dell'Intelligenza Artificiale*, in *Scienza, tecnologia e forme di produzione giuridica*, in *DPCE online*, n. 3/2020, 3391-3398, 3396-3397; L. TORCHIA, *Lo stato digitale. Una introduzione*, Bologna, il Mulino, 132: «da spiegazione delle inferenze operate dall'algoritmo non può infatti esaurire la motivazione della decisione presa e, peraltro, nel caso degli algoritmi che operano secondo il c.d. *machine learning* non supervisionato è impossibile verificare *ex post* quali dati sono stati utilizzati per giungere ad un determinato esito, con la conseguente opacità delle inferenze, che restano quindi prive di spiegazione». Sul concetto di *black box*, si vedano: F. PASQUALE, *The Black Box Society The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015; F. GIANNOTTI, D. PEDERESCHI, «*Explainable AI*». *Aprire le scatole nere per una Intelligenza Artificiale (AI) umana*, in *Gnosis*, n. 2/2019, 37-45; M. CARABANTES, *Black-box artificial intelligence: an epistemological and critical analysis*, in *AI & Society*, n. 35/2020, 309-317.

¹⁸ G. LO SAPIO, *La black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, cit., 121; C. CASONATO, *L'intelligenza artificiale ed il diritto pubblico comparato ed europeo*, in *Diritto pubblico comparato ed europeo*, n. 1/2022, 169-179, 173.

¹⁹ G. LEMME, *La proposta di Regolamento europeo sulla intelligenza artificiale e la gestione dei rischi: una battaglia che può essere vinta?*, in *Rivista trimestrale di diritto dell'economia*, n. 1/2024, 259-275, 232.

Permane inoltre il problema dell'enorme affidamento dei sistemi di intelligenza artificiale sulla quantità e sulla qualità dei dati di addestramento, che ne rappresentano la materia prima indispensabile²⁰. Nonostante l'art. 15 dell'AI Act imponga il rispetto di un adeguato livello di accuratezza, robustezza e cybersicurezza per i sistemi di IA ad alto rischio, non sono presenti garanzie in materia di aggiornamento e attendibilità dei dati impiegati dai sistemi di intelligenza artificiale, poiché questi impiegano informazioni che sono reperibili su Internet, rischiando quindi di riflettere pregiudizi e *bias* sociali²¹.

3. Intelligenza artificiale, utenti e competizione

Nella seconda parte del volume, l'Autore analizza il percorso di ChatGPT dai laboratori di OpenAI fino alla sua apertura al pubblico, avvenuta nel novembre 2022, poi seguita dal lancio di un ulteriore modello di linguaggio da parte di Google, chiamato Bard.

In seguito all'intervento di numerosi «collaudatori», che avevano migliorato la qualità delle risposte di ChatGPT in modo tale da renderle il più pertinenti possibile nel contesto della conversazione, il sistema si è distinto davanti all'opinione pubblica per la sua capacità di ricordare le battute precedenti del dialogo, per aver adeguato il tono e lo stile alla situazione e per la sua insolita abilità di ragionamento.

Nel 2023 aziende e testate giornalistiche hanno dimostrato che una percentuale sorprendentemente alta di utenti di *chatbots* non è in grado di riconoscere di stare conversando con una macchina oppure con un essere umano, rivelando la tendenza a proiettare sull'agente tratti tipicamente umani, come l'empatia o la comprensione, quando l'interlocuzione avviene per via testuale. In relazione a tale fenomeno, in dottrina è stato ritenuto che l'intervento umano sia insufficiente per una supervisione effettiva rispetto ai meccanismi di intelligenza artificiale per via del c.d. *automation bias*, ossia la tendenza a considerare sempre corretto il giudizio della macchina per via dell'elevato grado di fiducia riposto nell'automazione²².

Il legislatore europeo si è mostrato sensibile al problema, come testimoniato dall'art. 14.4, lett. b) dell'AI Act, disponendo che i soggetti preposti alla sorveglianza sui sistemi di intelligenza artificiale debbano restare consapevoli «della possibile tendenza a fare

²⁰ L. CALIFANO, *ChatGPT e Meta EDI: spunti problematici su profili regolatori e ruolo delle autorità di controllo di protezione dati*, in *Federalismi.it*, n. 10/2023, IV-XV, V; F.M. MANCIOPPI, *La regolamentazione dell'intelligenza artificiale come opzione per la salvaguardia dei valori fondamentali dell'UE*, in *Federalismi.it*, n. 7/2024, 122-134, 118: «Negli ultimi anni, abbiamo assistito ad un cambiamento drastico della funzione dei dati: ne sono esempio l'aumento numerico (da cui deriva un ampio ammontare di dati non utilizzati o non pienamente sfruttati a causa di ostacoli tecnici e incertezze giuridiche), di valore (non solo economico) e di utilità nel progresso dell'intelligenza artificiale». Sul tema si veda anche D. IACOVELLI, M. FONTANA, *Nuove sfide della tecnologia e gestione dei rischi nella proposta di regolamento europeo sull'intelligenza artificiale: set di training, algoritmi e profilazione dei dati. Rilievi critici*, in *Il diritto dell'economia*, n. 3/2022, 107-137.

²¹ L. ROMANÒ, *La responsabilità penale al tempo di ChatGPT: prospettive de iure condendo in tema di gestione del rischio da intelligenza artificiale generativa*, in *Sistema penale*, 17 maggio 2023, 8; P. HACKER, A. ENGEL, T. LIST, *Understanding and Regulating ChatGPT, and Other Large Generative AI Models*, *Verfblog*, 2023.

²² K. YEUNG, *Algorithmic Regulation: A Critical Interrogation*, in *King's College London Law School Research Paper*, n. 27/2017, 25: «humans are highly susceptible to 'automation bias', referring to the tendency of individuals to defer to computational judgments, even when they are capable of recognizing that the situation calls for another choice».

automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio, in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche».

La complessità dell'interazione fra uomo e macchina rivela le proporzioni della lesività che può essere ascritta alla fase di «allineamento» della macchina, processo che consiste nel raccordo tra gli obiettivi perseguiti dall'agente e quelli dell'utente e che dipende dalle finalità perseguite dal produttore. L'Autore precisa che sono del tutto inesplorate le soluzioni sul piano etico e giuridico per quanto riguarda la possibilità che una macchina possa essere programmata con l'obiettivo di costruire un rapporto emotivo con l'utente, nonostante sia da escludere che un agente intelligente possa essere programmato esplicitamente con l'obiettivo di ingannare l'interlocutore umano riguardo alla propria identità. Il Regolamento sull'intelligenza artificiale impone infatti ai fornitori di sistemi di IA per finalità generali di garantire «che gli output del sistema di IA siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente», nel tentativo di contrastare le ricadute negative dei «deepfakes»²³.

L'attività di «allineamento», tuttavia, non è in grado di risolvere il problema principale posto dall'intelligenza artificiale, ovvero il fatto che «non sappiamo quali informazioni sono contenute in questi modelli, [...] che cosa sappiano di noi e del mondo, e non abbiamo ancora un metodo perfetto per controllare il loro comportamento» (p. 73). Sono infatti emersi casi di «jailbreaking», in cui il *chatbot*, confuso da alcune richieste degli utenti, ha rivelato informazioni pericolose nonostante fosse stato programmato per non farlo, in quanto l'operazione di allineamento si limita a reprimere le informazioni nocive senza però eliminarle.

Un altro capo del problema riguarda l'accuratezza di ChatGPT che, non essendo un sistema perfetto, risente di numerose «confabulazioni» e altri errori relativi alle informazioni fornite, dovuti al fatto che le risposte generate dalla macchina sono legate all'analisi di numerose fonti e che, a volte, una fonte potrebbe essere il risultato della combinazione di più documenti e non di un testo univoco, rivelando l'evidente problema sociale legato alla fiducia che numerosi utenti ripongono nelle risposte date dal sistema.

In letteratura si è parlato di «inaffidabilità» dei *chatbots*, che può variare da lieve – ossia risposte che contengono informazioni imprecise o incomplete – a grave, cioè risposte completamente sbagliate, seppur plausibili, come nei casi di «allucinazioni»²⁴.

²³ Regolamento (UE) 2024/1689, art. 50.2. Sul fenomeno *deep fake*, si vedano V. AZZALI, N. ELLECOSTA, *La questione "deepfake" in Italia: una panoramica*, in *MediaLaws*, n. 3/2023, 72-89; M. CAZZANIGA, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai "deepfakes"*, in *MediaLaws*, n. 1/2023, 170-187.

²⁴ A. DIURNI, G. RICCIO, *ChatGPT: Challenges and Legal Issues in Advanced Conversational AI*, in *The Italian Law Journal*, vol. 9, n. 2/2024, 473-497, 475. S. SAPIENZA, *ChatGPT, dati e identità personale: rischi, bilanciamenti, regolazione*, in *Osservatorio costituzionale*, n. 5/2023, 47-71, 55: «l'inferenza sulla base del quale un LLM restituisce il proprio *output* potrebbe non rispecchiare il dato personale originale perché generata sulla base di correlazioni statistiche e non causali o fattuali. Questo è il caso delle c.d. *allucinazioni* dei LLM, ossia i casi in cui un modello statistico considera appropriato un *output* basato su una regolarità statistica che associa una combinazione {nome, cognome} ad alcuni termini, ma che si rivela essere totalmente infondato da un punto di vista fattuale. Naturalmente, non tutte le allucinazioni sono uguali. Alcune, ad esempio, potrebbero risultare in un mero errore fattuale (ad es. l'anno di nascita) o originare informazioni mendaci

Nonostante i tentativi degli sviluppatori di gestire nel modo migliore possibile i contenuti prodotti dai modelli di linguaggio, è stato ritenuto che ogni soluzione di tipo tecnico non potrà mai impedire completamente la diffusione di risposte contenenti informazioni manipolate, *fake news* o incitamento all'odio²⁵. Sul piano giuridico, inoltre, i *chatbots* eludono l'ambito di applicazione del Digital Services Act (DSA), che impone obblighi a intermediari e piattaforme *online* in materia di contenuti illegali, trasparenza e disinformazione²⁶.

Il DSA non trova infatti applicazione in riferimento a ChatGPT, poiché questo rientra nei servizi di comunicazione interpersonale, e non di «diffusione al pubblico»²⁷, così come non rientra nemmeno tra i «servizi intermediari»²⁸, in quanto si limita a comunicare con l'utente ma non pubblica contenuti prodotti da altri, nonostante le risposte non siano originali ma rielaborazioni di dati preesistenti²⁹. Non mancano proposte volte ad applicare il DSA a ChatGPT, sia poiché i contenuti prodotti da tale sistema potrebbero essere diffusi su piattaforme online, sia poiché potrebbe essere qualificato come una «diffusione al pubblico» se rispondesse alla stessa domanda posta da un numero infinito di utenti³⁰.

Una tesi minoritaria vorrebbe inoltre ascrivere i *chatbots* alla nozione di motori di ricerca *online* adottata dal DSA, poiché in entrambi i casi il funzionamento del sistema si basa su una richiesta iniziale avanzata dagli utenti ed è in grado di restituire informazioni inerenti al contenuto richiesto in qualsiasi formato³¹.

L'obiettivo dell'assoggettamento di ChatGPT e dei sistemi affini al regime di responsabilità previsto dal Digital Services Act potrebbe coincidere con una migliore repressione dei contenuti illegali, soprattutto se si evita che questi vengano impiegati nelle fasi relative all'addestramento continuo del modello, contribuendo alla costruzione di una intelligenza artificiale generativa affidabile e a dimensione umana.

Il volume spiega inoltre che i ricercatori temono che gli agenti autonomi possano adottare soluzioni contrarie ai valori e agli interessi umani, svolgendo una serie di passaggi intermedi in buona fede, con conseguenze particolarmente lesive nel caso di manipolazione dei mercati finanziari o dell'opinione pubblica. Dovrà quindi essere sviluppato un meccanismo di allineamento in grado di incidere non solo sugli obiettivi finali ma anche sugli obiettivi intermedi dell'agente intelligente.

su persone fisiche corrispondenti a speciali categorie di dati (ad es. la preferenza sessuale) o su reati che l'interessato non ha commesso».

²⁵ P. HACKER, A. ENGEL, T. LIST, *Understanding and Regulating ChatGPT, and Other Large Generative AI Models*, cit.

²⁶ Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali).

²⁷ Regolamento (UE) 2022/2065, Considerando n. 14.

²⁸ Regolamento (UE) 2022/2065, artt. 2-3.

²⁹ G. DE MINICO, *Too many rules or zero rules for the ChatGPT?*, in *BioLaw Journal – Rivista di BioDiritto*, n. 2/2023, 473-483, 481. In merito a tale ultimo punto, l'Autrice, con riferimento alla sentenza della Court of Justice (Grand Chamber), *L'Oréal and Others v. eBay International AG*, Case C-324/09, 12 July 2011, precisa che: «in line with the CJEC, this is true even if its words and thoughts are not authentic, because the Chat merely synthesizes and re-elaborates data already existing on the Internet. In fact, the Judge has been in favor of denying the liability privilege to the platform which has helped with promoting user-generated contents».

³⁰ G. DE MINICO, *Too many rules or zero rules for the ChatGPT?*, cit., 481.

³¹ M. BASSINI, *Intelligenza artificiale generativa: alcune questioni problematiche*, in *MediaLaws*, n. 2/2023, 391-397, 394-395.

L'elemento conclusivo della seconda parte del lavoro riguarda la competizione tra aziende innescata dallo sviluppo di modelli linguistici costantemente più potenti, poiché i *chatbots* sono in competizione con i motori di ricerca tradizionali, primo fra tutti Google, che rischia di perdere il ruolo chiave di mediatore delle informazioni disponibili su Internet.

Negli ultimi tre anni numerose società, tra cui le statunitensi OpenAI, Meta e Microsoft e la cinese Baidu, si sono contese il primato nello sviluppo del modello di linguaggio più potente, basato sulla ricerca di algoritmi più efficienti e sulla maggiore quantità di documenti di addestramento acquisiti, limitati però da risorse computazionali e qualità dei dati. Ogni settore industriale che prevede l'impiego di tecnologie di intelligenza artificiale dovrà dotarsi di GPU, *hardware* la cui disponibilità è limitata a causa della rapida crescita della richiesta. La principale azienda produttrice di GPU, la NVIDIA, si trova quindi a detenere una posizione di mercato particolarmente strategica, al punto che a novembre 2024 è divenuta la prima azienda al mondo per capitalizzazione³².

La rapida affermazione dell'intelligenza artificiale ha innescato una competizione sia tra grandi società del settore tecnologico sia tra Stati, che stanno investendo risorse pubbliche per accaparrarsi risorse umane e materiali e stanno elaborando strategie normative per guidare il progresso tecnologico³³. Oltre a perseguire l'obiettivo di contrastare la frammentazione normativa in materia, l'Unione europea sta tentando di affermare la propria egemonia attraverso la disciplina del settore tecnologico e di rendere il modello europeo un punto di riferimento globale, seguendo un processo denominato «effetto Bruxelles»³⁴, nel tentativo di supplire, almeno parzialmente, al divario tecnologico e infrastrutturale detenuto da alcune economie mondiali, soprattutto Stati Uniti e Cina³⁵.

Se l'Unione non può competere con altri attori globali in termini di produzione tecnologica, il legislatore europeo mira a delineare un modello regolatorio in netta opposizione all'approccio dirigistico cinese e all'approccio orientato al mercato, nonché di generale sfiducia verso la regolazione di stampo statunitense, sebbene si sia verificata una parziale inversione di tendenza in materia di intelligenza artificiale³⁶.

³² N. RANDEWICH, *Nvidia surpasses \$3.6 trillion market value after Trump win*, in Reuters, 7 novembre 2024. Per alcune considerazioni in materia di peso di borsa, di mercato e sociale delle grandi società attive nel settore tecnologico, si veda L. TORCHIA, *Poteri pubblici e poteri privati nel mondo digitale*, in il Mulino, *Rivista trimestrale di cultura e di politica*, n. 1/2024, 14-33.

³³ G. LO SAPIO, *L'Artificial Intelligence Act e la prova di resistenza per la legalità algoritmica*, cit., 273-274.

³⁴ A. BRADFORD, *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, Oxford, 2020.

³⁵ G. BUTTARELLI, *La regolazione delle piattaforme digitali: il ruolo delle istituzioni pubbliche*, in *Giornale di diritto amministrativo*, n. 1/2023, 116-127, 120. Per un'ampia trattazione sul tema, si veda S. HERMES, E. CLEMONS, M. SCHREIECK, S. PFAB, M. MITRE, M. BOHM, H. KRUMAR, *Breeding grounds of digital platforms: exploring the sources of American platform domination, China's platform self-sufficiency, and Europe's platform gap*, in *Proceedings of the European conference on information Systems (ECIS)*, n. 132/2020, 18.

³⁶ G. FINOCCHIARO, *La regolazione dell'intelligenza artificiale*, in *Rivista trimestrale di diritto pubblico*, n. 4/2022, 1085-1099, 1089-1091. In materia di tripartizione in tre modelli regolatori, si veda A. BRADFORD, *Digital Empires. The global battle to regulate technology*, New York, Oxford University Press, 2023. In merito all'inversione di tendenza rispetto alla tradizionale inerzia regolamentare statunitense, il riferimento è a due documenti adottati dalla Casa Bianca nel 2023, *Blueprint for an AI Bill of Rights e AI Risk Management Framework*, a cui fa riferimento G. LO SAPIO, *L'Artificial Intelligence Act e la prova di resistenza per la legalità algoritmica*, cit., nota n. 30. Sulle diversità di approcci alla regolazione dell'intelligenza artificiale tra Stati Uniti e Unione europea, si vedano G. ALPA, *L'intelligenza artificiale. Il contesto giuridico*, Modena, Mucchi, 2021; E. CHITI E B. MARCHETTI, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di intelligenza artificiale*, in *Rivista della regolazione dei mercati*, n. 1/2020, 29-50; S. ACETO DI CAPRIGLIA, *Intelligenza artificiale; una sfida globale tra rischi*,

L'oligopolio economico detenuto dalle grandi imprese, principalmente statunitensi, che si contendono il mercato digitale, compresa anche l'IA generativa, ha potuto infatti consolidarsi grazie all'inerzia regolamentare statunitense, soprattutto attraverso l'*Administration's Telecommunications Act* del 1996, approvato durante la presidenza Clinton, che ha permesso un'esenzione dai vincoli tradizionalmente imposti alle società, come il controllo dei prezzi, le restrizioni all'accesso alle infrastrutture e ai servizi, le limitazioni alla posizione e alla crescita nel mercato di riferimento, consentendo lo sviluppo di *start-up* in grandi società con un'enorme rilevanza sul piano globale³⁷.

Nonostante le numerose difficoltà connesse allo sviluppo di una regolazione efficiente in materia di intelligenza artificiale, all'AI Act deve essere riconosciuto il merito di essere il primo atto normativo a livello mondiale che ambisce a regolare l'intero settore, rappresentando «un buon compromesso sul piano della sostenibilità giuridica, economica e tecnologica delle sue regole»³⁸

Per concludere la seconda parte del volume, l'Autore riflette sulle perplessità avanzate da scienziati e legislatori riguardo alla delega decisionale in favore di macchine il cui funzionamento è ancora in parte ignoto e che si rischia di non riuscire a controllare, con il rischio di condizionare i canali di informazione e di sostituire il genere umano. L'antidoto contro una reazione inutilmente rigida al progresso tecnologico può essere soltanto una maggiore comprensione e studio dello strumento tecnico.

4. Le macchine e il concetto di intelligenza

L'ultima parte dell'opera è dedicata allo studio di quanto sia noto del funzionamento degli agenti di intelligenza artificiale, sottolineando la necessità di sviluppare rapidamente nuove categorie concettuali per comprendere il funzionamento delle macchine intelligenti, poiché il loro controllo è legato a doppio filo alla loro comprensione.

A destare preoccupazioni non è esclusivamente la rapidità dello sviluppo tecnologico, ma anche la direzione che sta prendendo. I filosofi della scienza trattano infatti di «singolarità tecnologica» per riferirsi alla possibilità che la crescita tecnologica possa diventare talmente rapida da risultare incomprensibile e, di conseguenza, incontrollabile per gli esseri umani. All'aumentare delle conoscenze immagazzinate dalla macchina, aumenterebbe esponenzialmente la facilità con cui può acquisirne delle altre, con il rischio che l'intelligenza umana possa essere superata facilmente in un futuro prossimo.

prospettive e responsabilità. Le soluzioni assunte dai governi unionale, statunitense e sinico. Uno studio comparato, in *Federalismi.it*, n. 9/2024, 39.

³⁷ L. TORCHIA, *Lo stato digitale. Una introduzione*, cit., 14; B. CAROTTI, *Digitalizzazione europea e politica industriale: ancora molto rumore*, in *Rivista trimestrale di diritto pubblico*, n. 4/2022, 997-1013; S. DA EMPOLI, *L'economia di ChatGPT. Tra false paure e veri rischi*, Milano, Egea, 2023, 127; G. GARDINI, *Le regole dell'informazione. Pluralismo e libertà nell'era dell'intelligenza artificiale*, Torino, Giappichelli, 2024, XXI.

³⁸ B. MARCHETTI, *Intelligenza artificiale, poteri pubblici e rule of law*, cit., 56. G. FINOCCHIARO, *La regolazione dell'intelligenza artificiale*, cit., 1094.

Per studiare il funzionamento interno di un modello di linguaggio è possibile adottare come riferimento soltanto modelli molto piccoli e semplici da addestrare, come per esempio BERT, con l'obiettivo di ottenere un modello universalmente valido.

Durante la fase di addestramento, gli organi presenti all'interno di ciascun modulo del modello di linguaggio si specializzano spontaneamente, imparando compiti e assumendo ruoli non deducibili dallo studio dell'algoritmo in astratto, poiché il funzionamento del modello dipende dall'interazione tra algoritmo e ambiente. Il ruolo di ogni singolo organo può dunque essere apprezzato solo empiricamente esaminando un modello già addestrato.

Considerando che le abilità misurate in GPT-4 non si limitano alla padronanza del linguaggio, ma comprendono anche la risoluzione di numerosi compiti complessi che spaziano attraverso diverse discipline e che richiedono l'acquisizione di un «senso comune» o «teoria della mente», ChatGPT è stato ascritto a pieno titolo come una versione iniziale di un sistema di intelligenza artificiale generale (AGI), vale a dire sistemi che dimostrano ampie capacità di ragionamento, pianificazione e apprendimento dall'esperienza a un livello pari o superiore rispetto a quello umano. L'insieme di idee e relazioni che il modello di linguaggio manifesta nei dialoghi con l'utente costituisce la sua rappresentazione del mondo, che evidenzia una serie di concordanze e connessioni che gli esseri umani non sono in grado di cogliere. Rimane comunque il fatto che al momento non esistono garanzie che l'agente sviluppi solo le competenze utili all'essere umano: il sistema potrebbe celare numerosi esiti pericolosi.

Le enormi abilità manifestate dai modelli di linguaggio in ogni campo del sapere umano sono emerse spontaneamente mentre il sistema veniva addestrato per svolgere esercizi meramente linguistici. Non è dunque noto di cosa saranno capaci tali sistemi in futuro e gli studi che possono essere condotti su modelli più semplici si rivelano inutili in tal senso, considerando che il comportamento del sistema è strettamente legato alle sue dimensioni. Nei modelli piccoli molte abilità non si manifestano, ma all'aumentare delle dimensioni del modello alcune di queste affiorano gradualmente, mentre altre, denominate «capacità emergenti» appaiono improvvisamente.

Le capacità emergenti sono state studiate in via esclusivamente empirica, ma non è possibile prevedere con certezza gli sviluppi futuri, considerando la corsa al costante aumento delle dimensioni dei modelli. Non è quindi possibile predire il comportamento di questi strumenti all'aumentare delle dimensioni, così come i ricercatori non riescono a dare una spiegazione al comportamento assunto dai modelli di linguaggio negli ultimi mesi.

Il legislatore europeo ha considerato approfonditamente il rischio di derive legate all'apprendimento «non supervisionato» degli agenti intelligenti. Il contributo umano è infatti considerato nel Regolamento europeo sull'intelligenza artificiale fondamentale per assolvere alle funzioni di controllo sulle decisioni automatizzate, poiché a una maggiore complessità è connessa una maggiore opacità e imprevedibilità degli esiti³⁹.

³⁹ AI Act, considerando 6-bis: «La funzione e gli output di molti di questi sistemi di IA si basano su relazioni matematiche astratte che per gli esseri umani risultano difficili da comprendere e monitorare e i cui input specifici sono difficili da rintracciare. Tali caratteristiche complesse e opache (elemento “scatola nera”) incidono sulla rendicontabilità e sulla spiegabilità».

Ai sensi dell'art. 14.1 dell'AI Act, i sistemi di IA ad alto rischio devono essere progettati e sviluppati in modo tale da poter essere sottoposti a un'efficace supervisione da parte di persone fisiche in misura proporzionale ai rischi associati a tali sistemi. È quindi previsto un veto nei confronti dei sistemi che non consentono un intervento diretto dell'essere umano, in quanto perseguono una linea di ragionamento autonoma e non lineare che sfugge alla comprensione umana e non ne consente la supervisione immediata⁴⁰.

Nell'epilogo, l'Autore sostiene che le linee di indagine più urgenti in materia di modelli di linguaggio riguardano le abilità emergenti, la ridefinizione del concetto di intelligenza e le difficoltà insite nel processo di allineamento degli obiettivi della macchina con quelli del produttore. La gestione dei conflitti e la soluzione degli interrogativi principali in materia di agenti intelligenti dovrà essere necessariamente risolta a livello politico.

L'intelligenza artificiale generativa non sembrerebbe essere disciplinata in ambito europeo in modo totalmente coerente con la natura dello strumento e delle sue caratteristiche, come dimostrato dalla vicenda legata alla temporanea sospensione del trattamento dei dati personali degli utenti italiani imposta a OpenAI da parte del Garante per la protezione dei dati personali, misura che ha indotto la società al temporaneo oscuramento di ChatGPT sul territorio italiano, servizio successivamente riavviato a seguito del rafforzamento dei meccanismi a garanzia della trasparenza nell'uso dei dati personali degli interessati⁴¹.

L'insufficienza degli attuali paradigmi regolatori per supplire alle numerose criticità sollevate dall'intelligenza artificiale può essere ascritta a numerosi fattori, principalmente la rapidità dello sviluppo tecnologico, che rende una regolazione affidata alle assemblee parlamentari troppo lenta, ma anche la difficile ricerca di un equilibrio tra la rigidità delle norme e l'approccio precauzionale, nel tentativo di non inibire lo sviluppo tecnologico.⁴²

Problemi simili riguarderebbero anche lo sviluppo di una co-regolazione tra attori pubblici e privati, che si porrebbe come strumento di regolazione «ibrida» o «collaborativa», in una posizione intermedia fra disciplina normativa e regole tecniche, di cui sono

⁴⁰ M. BARBERIO, *L'uso dell'intelligenza artificiale nell'art. 30 del d.lgs. 36/2023 alla prova dell'AI Act dell'Unione europea*, in *Rivista italiana di informatica e diritto*, n. 2/2023, 253-264, 257. Sul tema si veda anche J. PONCE SOLÉ, *Il regolamento dell'unione europea sull'intelligenza artificiale, la discrezionalità amministrativa e la riserva di umanità*, in *Rivista trimestrale di diritto pubblico*, n. 3/2024, 825-855.

⁴¹ L. MEGALE, *Il Garante della privacy contro ChatGPT: quale ruolo per le autorità pubbliche nel bilanciare sostegno all'innovazione e tutela dei diritti?*, in *Giornale di diritto amministrativo*, n. 3/2023, 403-413, 408: «In virtù della natura di ChatGPT, con particolare riferimento alla sua modalità di costruzione e al fatto che il modello è già stato addestrato, è poco realistico che possano valere il consenso al trattamento, l'esecuzione di un contratto, gli obblighi di legge, gli interessi vitali della persona, il legittimo interesse del titolare o terzi, o l'interesse pubblico (elencati dall'art. 6 RGPD come elementi per la base giuridica del trattamento). [...] Per di più, la natura dell'elemento del "consenso" (considerando che questo dovrebbe essere libero, informato, specifico, inequivocabile, verificabile e revocabile) e le altre prescrizioni del secondo provvedimento del Garante non tengono conto del cosiddetto paradosso della privacy, che costituisce una delle tante ricadute dei limiti cognitivi delle persone». Sulla vicenda, si vedano anche A. CORDELLA, F. GUALDI, *Regulating generative AI: The limits of technology-neutral regulatory frameworks. Insights from Italy's intervention on ChatGPT*, in *Government Information Quarterly*, n. 41/2024, 15; G.M. RIVA, *Diritto e intelligenza artificiale generativa: l'istruttoria del Garante per la protezione dei dati italiano su "OpenAI e ChatGPT"*, in *Cyberspazio e diritto*, n. 1/2024, 153-176; S. TROZZI, *Il principio della finalità del trattamento dei dati personali alla prova dei recenti sviluppi in tema di intelligenza artificiale: il caso ChatGPT e la neuroprivacy*, in *Federalismi.it*, n. 1/2024, 197-231.

⁴² B. MARCHETTI, *Intelligenza artificiale, poteri pubblici e rule of law*, cit., 62.

depositarie le società sviluppatrici di intelligenza artificiale⁴³. Una co-produzione delle norme tecniche si rivela però fragile a causa della sua costante necessità di aggiornamento, dei continui avanzamenti della comunità scientifica, della sua natura prettamente specialistica – in opposizione all’astrattezza propria delle norme giuridiche – nonché per via del suo carattere eteronomo⁴⁴.

Lo scenario rimane inevitabilmente aperto, ma è significativo lo sforzo regolatorio intrapreso dall’Unione attraverso l’AI Act, che si propone di promuovere un’intelligenza artificiale antropocentrica e affidabile, garantendo allo stesso tempo un elevato livello di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell’Unione europea, nonché di proteggere i cittadini dagli effetti nocivi dei sistemi di IA nell’Unione e di promuovere l’innovazione⁴⁵.

Il «punto di non ritorno» nell’evoluzione dell’umanità è rappresentato dal carattere di autonomia dei sistemi di intelligenza artificiale, che ne costituisce un elemento fondante a norma dell’art. 3 del Regolamento⁴⁶. Nel caso in cui le capacità dei modelli di linguaggio dovessero continuare a crescere esponenzialmente, alimentate da tutta la conoscenza possibile fino a raggiungere un’interazione diretta con il mondo reale, tali modelli potrebbero superare l’intelligenza umana. Se la macchina dovesse possedere il segreto della conoscenza, potremmo avere di fronte a un’era nuova, successiva a quella dell’Homo sapiens, su cui incombe la *Machina sapiens* (p. 146)⁴⁷.

⁴³ A. SIMONCINI, *La co-regolazione delle piattaforme digitali*, in *Rivista trimestrale di diritto pubblico*, n. 4/2022, 1031-1049, 1032. A. IANNUZZI, *Paradigmi normativi per la disciplina della tecnologia: auto-regolazione, co-regolazione ed etero-regolazione*, in «Bilancio Comunità Persona», n. 2/2023, 91-107, 99.

⁴⁴ A. IANNUZZI, *Paradigmi normativi per la disciplina della tecnologia*, cit., 99-100.

⁴⁵ Regolamento (UE) 2024/1689, Considerando n. 1.

⁴⁶ G. LO SAPIO, *L’Artificial Intelligence Act e la prova di resistenza per la legalità algoritmica*, cit., 276.

⁴⁷ Sulla rivoluzione in termini di modalità di comunicazione e sulla conoscenza all’epoca dell’intelligenza artificiale, C. EQUIZI, *Intelligenza artificiale: profili di opportunità e di criticità nella irrinunciabile tutela dei diritti fondamentali*, cit., 322.