L'IRCOCERVO
PRIMA RIVISTA ELETTRONICA ITALIANA DI METODOLOGIA GIURIDICA - TEORIA GENERALE DEL DIRITTO E DOTTRINA DELLO STATO
FIRST ITALIAN DIGITAL JOURNAL OF LEGAL METHODOLOGY - GENERAL THEORY OF LAW AND DOCTRINE OF THE STATE

Intelligenza artificiale generativa: brevi note

Agata C. Amato Mangiameli

Università degli Studi di Roma Tor Vergata

Abstract: **Generative Artificial Intelligence: Short Notes**
Generative Artificial Intelligence is posing various dilemmas to all of us – and to jurists in particular. On the one hand, great opportunities are opening up. On the other hand, dangers are advancing. This underscores the imperative to implement robust regulatory measures to protect human rights.

Keywords: Generative Artificial Intelligence, IA Act, Bias, Rights.

Sommario: 1. Strumenti di IA per i più vari compiti: capaci di scrivere, produrre musica, creare opere d'arte digitali – 2. Intelligenza artificiale *vs.* Intelligenza artificiale generativa. I vantaggi – 3. *Segue*: qualche rischio – 4. Conclusioni minime.

1. Strumenti di IA per i più vari compiti: capaci di scrivere, produrre musica, creare opere d'arte digitali

Tra le nuove tecnologie, un ruolo di primo piano, e al contempo una svolta significativa, rivestono gli studi e le applicazioni dell'intelligenza artificiale in pressoché tutti i campi e in particolare in quello giuridico. La ragione è evidente. Una volta la tecnologia serviva per lo più allo svolgimento di quelle attività che erano (e sono) di *routine* o sostanzialmente ripetitive, per le quali le possibili scelte/decisioni erano (e sono) in buona parte già note oppure potevano (e possono) essere determinate con un alto livello di affidabilità. Adesso, invece, l'IA può essere usata per compiere attività che coinvolgono il processo decisionale e mostra di possedere il potenziale per svolgere un lavoro di natura cognitiva che richieda un *know how* anche molto sofisticato.

Basti qui pensare all'impiego dell'IA nella creazione di contenuti e alla capacità del tutto autonoma di produrre opere, tale da sfidare il diritto d'autore, per un verso, e per l'altro, da sollevare questioni riguardanti la protezione delle creazioni medesime. Musica, letteratura, arte, sono sempre più coinvolte: qualche volta, vengono generate da algoritmi, talaltra, sono prodotte da algoritmi e successivamente perfezionate dall'intelligenza e dalla creatività umane come risultato di un'interessante connubio/ibridazione di naturale e artificiale. E in questo andirivieni di intelligenze (umane e artificiali) viene da sé che debbano essere

anzitutto riformulate le nozioni tradizionali di opera d'arte, così come di artista, di originalità e di autorialità, e che, di pari passo, siano prospettate normative consone rispetto ai nuovi strumenti (e cioè che siano non solo in grado di regolamentarne l'utilizzo, ma che sappiano anche scongiurare la totale perdita di senso di certe categorie giuridiche ed economiche classiche, laddove siano da considerare essenziali e determinanti nei rispettivi ambiti).

Del resto, così come in tema di responsabilità ci si interroga su chi debba essere chiamato in causa allorché una IA provochi un incidente, allo stesso modo quando parliamo di opere dell'ingegno non è chiaro se l'autore (l'autrice), il "creatore" (la "creatrice") debba essere considerata l'IA che ha generato l'opera, oppure se sia necessario rivolgersi al programmatore (alla programmatrice), che ha scritto l'algoritmo, piuttosto che all'umano, che, a suo modo e per scopi diversi, se ne è servito. Noto è il recente caso della scrittrice Rie Kudan[1] che ha ammesso di avere usato *ChatGPT*[2] per scrivere alcune parti del romanzo intitolato *Tokio-to-Doyo-to*, opera che peraltro ha anche ricevuto uno dei più prestigiosi premi letterari giapponesi. Allo stesso modo non è chiaro se si possa parlare di diritti degli algoritmi e, quindi, della brevettabilità delle invenzioni scaturite da sistemi di IA, senza che vi sia stato un intervento umano diretto. Significativo, per i suoi diversi risvolti e per le differenti decisioni, è proprio il caso *Dabus*[3], nel quale il suo inventore in varie domande presentate agli uffici preposti (degli Stati Uniti, dell'Unione Europea, del Regno Unito, del Sud Africa, dell'Australia) ha chiesto il riconoscimento della qualità di inventore in capo alla macchina, pur non essendo un soggetto dotato di capacità giuridica come, invece, sarebbe richiesto dalle diverse norme nazionali sulla brevettabilità. Di qui l'interrogativo: è forse giunto il momento di prevedere una nuova e specifica categoria di inventori nella quale potere includere le stesse intelligenze artificiali?

Tanti gli esempi che possono essere richiamati, tutti estremamente significativi per le loro molteplici implicazioni. Intanto *ChatGPT*, strumento ormai virale, e poi i tantissimi altri strumenti di IA capaci di scrivere (basti pensare a:

[1] Nota scrittrice giapponese, vincitrice della diciassettesima edizione del *Akutagawa Prize*, importante riconoscimento per la letteratura.

[2] Fra i giornali italiani a darne la notizia, cfr.: "Rie Kudan, la scrittrice vince un prestigioso premio letterario e poi confessa: 'Libro scritto con ChatGpt'", in *Corriere della Sera*, 20/01/2024.

[3] Per un'agile ricognizione della vicenda, e delle diverse questioni che ne sono derivate, rinvio, fra i vari, a B. Marone, G. Pinotti, A. Santuosso, "L'AI può essere autore o inventore? Tutti gli interrogativi sollevati dalle decisioni Thaler/DABUS", in *Agenda Digitale*, 11/04/2023.

Jasper[4], *Amazon Lex*[5], *AI-Writer*[6], *Writer*[7]), generare immagini (tra i più noti: *Midjourney*[8], *Stable Diffusion*[9], *Dall-E*[10]), produrre musica (come ad esempio: *AIVA*[11], *Soundful*[12], *Boomy*[13], *Amper*[14], *Dadabots*[15], *MuseNet*[16]), creare opere d'arte digitali, uniche e NFT (*Non-fungible Token*[17]). Si pensi, a tal proposito, a *Botto*[18], IA il cui processo creativo si basa sulla generazione di migliaia di immagini e al quale partecipa una comunità di utenti che influenza le scelte dell'algoritmo attraverso un sistema di voto. Tale meccanismo – risultato di un'intensa interazione e di una non-convenzionale cooperazione nella creazione – ha già realizzato diverse opere d'arte, dalla cui vendita ha ricavato un notevole profitto (che si aggira intorno ai 1,3 milioni di dollari).

[4] La piattaforma ideata da *Jasper Chat*, pensata per le aziende, che utilizza il linguaggio naturale per consentire agli utenti di richiedere varie attività, come, ad esempio, scrivere un post sul blog dell'azienda stessa, prospettare degli annunci e/o, eventualmente, correggerne lo stile in modo da renderlo più adeguato e funzionale allo scopo commerciale perseguito.
[5] Si tratta di un servizio AWS che permette la creazione di interfacce di comunicazione – mediante voce e testo – in qualsiasi applicazione.
[6] Di cui è disponibile anche una versione di *Microsoft* che, stando a quanto si legge nel sito della multinazionale statunitense, promette di portare la creazione di contenuti a un livello superiore, in termini di efficacia e qualità, con una notevole riduzione delle energie e delle tempistiche richieste.
[7] Che è un rilevatore di contenuti generati dall'IA.
[8] Programma di IA capace di generare immagini a partire da un testo e che, nel giugno del 2022, la rivista inglese *The Economist* ha utilizzato per progettare la copertina di un suo numero.
[9] Utilizzato principalmente per generare immagini dettagliate a partire da descrizioni di testo, ma applicabile anche ad altre attività, fra cui, la pittura e la generazione di traduzioni da immagine.
[10] Algoritmo sviluppato da *OpenAI* e presentato al pubblico nel gennaio del 2021.
[11] Lanciata già nel febbraio del 2016, *AIVA* è il primo compositore virtuale – riconosciuto dalla società musicale SACEM – specializzato nella creazione di musica classica e sinfonica.
[12] Piattaforma musicale che consente agli artisti di trovare ispirazione e di creare nuove melodie. È possibile programmare l'IA affinché generi musica in base a dei criteri specificatamente individuati.
[13] Che consente agli utenti, non soltanto di creare brani partendo da degli stili predefiniti, ma anche di pubblicare la musica così generata sulle piattaforme di streaming, attraverso le quali, gli stessi potranno poi vedersi corrisposti i diritti d'autore.
[14] Innovativo strumento di composizione musicale che agevola i creatori di contenuti, permettendo loro di creare senza alcuno sforzo brani musicali originali.
[15] Black Metal Band non umana.
[16] Progettata da OpenAI, MuseNet è in grado di creare musica combinando stili e generi diversi, garantendo un notevole grado di originalità e di personalizzazione.
[17] Cfr. F. Macioce, voce "NFT", in A.C. Amato Mangiameli, G. Saraceni (a cura di), *Cento e una voce di Informatica giuridica*, Giappichelli, Torino, 2023, pp. 336-338.
[18] Che sfrutta – simultaneamente – gli algoritmi di machine learning VQGAN (*Vector Quantized Generative Adversarial Network*), che permettono di generare immagini somiglianti ad altre, e CLIP (*Contrastive Language-Image Pre-training*), che garantisce la corrispondenza dell'immagine generata all'*input* di testo inizialmente dato.

2. Intelligenza artificiale *vs*. Intelligenza artificiale generativa. I vantaggi

Più che di IA *tout court*, si parla ormai sempre più spesso della sua forma avanzata, ovvero dell'Intelligenza Artificiale generativa[19], la cui caratteristica prima è la capacità di produrre (per l'appunto, generare) contenuti in modo totalmente autonomo. Diversamente dall'algoritmo dell'IA classica, che processa un'alta quantità di dati di esempio e, con questi, auto-apprende, riuscendo così a ricondurre un nuovo esempio a una delle casistiche di addestramento, quello di *GAI* (*Generative AI*) – pur essendo basato sempre sulla statistica – è capace di creare contenuti simili a quelli di addestramento, ma del tutto nuovi. L'intelligenza artificiale generativa ha, dalla sua, una straordinaria capacità di produrre testi scritti di qualunque genere (in forma ad esempio di racconto, poesia, lettera, romanzo) e con qualsiasi stile (sia imitando, su richiesta, la prosa di Manzoni oppure la poesia di D'Annunzio, e sia ideando uno stile tutto suo). Ha inoltre una particolare abilità nel realizzare immagini, audio e video, simili a quelli prodotti dall'uomo e, soprattutto, estremamente veritieri e credibili. Ma è soprattutto l'intelligenza artificiale conversazionale *text-oriented* che ha acceso l'immaginazione e che d'altra parte può sollevare qualche apprensione. Com'è noto, il primo consumer chatbot di IA generativa è stato rilasciato al pubblico nell'autunno del 2022, un chatbot questo basato sul modello di rete neurale GPT(-3.5 di OpenAI) che sta per *generative pretrained transformer*[20].

In realtà già molto tempo prima, fece la sua comparsa *Eliza*[21] un agente conversazionale del Massachusetts Institute of Technology, che fu prodotto a metà degli anni '60, e le cui risposte si limitavano a seguire un insieme di regole e di modelli predefiniti. Al contrario, le attuali intelligenze artificiali generative non seguono precise regole e neppure modelli predefiniti. È, infatti, attraverso

[19] Sotto l'egida della più generale formula "IA generativa" sono riconducibili innumerevoli modelli di IA, che vanno, da quelli più risalenti ma tuttora in uso (è il caso delle *RNN* [reti neurali ricorrenti] e delle *CNN* [reti neurali convoluzionali]), sino ai c.d. *transformer models* in grado di rappresentare le sequenze in modo ancor più flessibile e potente rispetto alle prime, motivo per cui *ChatGPT* è capace di rispondere velocemente e bene a richieste di conversazioni. Ci sono inoltre: *i)* gli *autocodificatori variazionali*, costituiti da reti di *encoder* (che sono in grado di apprendere le caratteristiche importanti di un'immagine, di comprimere tali informazioni e di archiviarle come una rappresentazione in memoria) e da reti di *decoder* (che utilizzano quelle medesime informazioni compresse per cercare di ricreare l'originale); *ii)* i *generative adversarial networks*, costituiti da due reti neurali concorrenti (vale a dire, il generatore che crea un'immagine e il discriminatore che decide se l'immagine è reale o generata); *iii)* i *modelli di diffusione* che imparano comprimendo i dati (aggiungendo rumore di sottofondo e tentando di rigenerare l'originale).

[20] Per un approfondimento di taglio tecnico, cfr.: G. Yenduri *et al.*, "GPT (Generative Pre-Trained Transformer). A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions", in *IEEE Access*, 12 (2024), pp. 54608-54649.

[21] J. Weizenbaum, "ELIZA. A computer program for the study of natural language communication between man and machine", in *Communications of the ACM*, 9 (1966), n. 1, pp. 36-45; disponibile in rete al seguente url: https://dl.acm.org/doi/pdf/10.1145/365153.365168, [Data di consultazione: 30/05/2024].

l'addestramento sui dati del mondo reale che si sviluppa e procede autonomamente l'intelligenza artificiale generativa, la cui peculiarità risiede nel fatto di prospettare nuovi contenuti in risposta ai *prompt*[22]. Fra l'altro, è assai interessante osservare che neppure gli esperti di IA sanno esattamente come vengano generati i contenuti in risposta, giacché gli algoritmi sono auto-sviluppati e ottimizzati via via che il sistema è addestrato. In altre parole, rispetto all'IA tradizionale (si pensi ai sistemi di *machine learning*[23]), basata su dati etichettati utilizzando tecniche di apprendimento supervisionato e generalmente progettata per eseguire attività specifiche (ad es. rilevare frodi, guidare mezzi di trasporto, ecc.), l'intelligenza artificiale generativa è addestrata su *set* di dati grandi e diversi (a volte, ottimizzati su volumi di dati molto più piccoli correlati a una funzione specifica) e si avvale, quantomeno inizialmente, dell'apprendimento non supervisionato (nel quale in principio i dati non sono etichettati e al software IA non viene fornita alcuna guida esplicita).

Le differenze sin qui brevemente richiamate fra IA tradizionale e *GAI* avranno delle ricadute notevoli in termini di progresso in molteplici settori. Secondo le previsioni contenute in un recente *report* della McKinsey & Company[24], ad esempio, entro il 2025 oltre il 30% dei nuovi farmaci e dei materiali sarà scoperto utilizzando tecniche di intelligenza artificiale generativa; ancora, grazie agli strumenti di IA è (e sarà) possibile comprendere gli effetti dei farmaci su grandi fasce della popolazione. Alcuni ricercatori prevedono inoltre che nei prossimi decenni l'intelligenza artificiale supererà le prestazioni umane in molti compiti, tra cui guidare un camion (entro il 2027), lavorare nella vendita al dettaglio (entro il 2031), scrivere bestseller (entro il 2049), eseguire interventi chirurgici (entro il 2053), sostituire tutti i lavori umani (entro 122 anni)[25]. In generale l'IA generativa sarà in grado di colmare le varie lacune, grazie alle sue interfacce utente semplici e basate su chat, sarà in condizione di verificare rapidamente la presenza di errori e saprà migliorare la comunicazione, potendo, tra l'altro, tradurre il testo in lingue diverse, riuscendo ad automatizzare attività complesse e a risolvere i problemi di codice, perfezionando quest'ultimo in modo più rapido, efficiente, efficace e affidabile.

Di qui, i molteplici vantaggi dell'IA generativa, in prima battuta, nella direzione di una maggiore produttività. Basti pensare che un team di programmatori potrebbe dedicare ore e ore a esaminare un codice difettoso per tentare di risolvere

[22] Si tratta di testi in linguaggio naturale che richiedono all'IA generativa di eseguire una precisa attività.

[23] Formula con la quale si designano gli algoritmi di apprendimento (poi ulteriormente distinguibili in sistemi di *supervised learning*, *unsupervised learning*, *reinforcement* learning). Per un'agile definizione, si veda G. Talamo, voce "Machine Learning", in A.C. Amato Mangiameli, G. Saraceni (a cura di), *Cento e una voce di informatica giuridica*, cit., pp. 310-313.

[24] Il *report* è disponibile e scaricabile online: https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-state-of-organizations-2023.

[25] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, O. Evans, "When will AI exceed Human Performance. Evidence from AI Expertes", in *Journal of Artificial Intelligence Research*, 62 (2018), pp. 729-754.

ciò che è andato storto, quando invece uno strumento di intelligenza artificiale generativa potrebbe essere in grado di trovare qualunque falla o errore in pochissimo tempo, segnalandolo in tempo reale e insieme suggerendo le necessarie correzioni. Altri vantaggi sono, ovviamente, rappresentati dai costi ridotti, dalla migliore soddisfazione dei clienti, da processi decisionali più informati, dal controllo di qualità dei prodotti, e, tutto ciò, grazie alla stessa velocità dell'IA generativa e ad approcci sofisticati nei più diversi settori.

3. *Segue*: qualche rischio

Insieme alle sue notevoli prospettive in ordine alla produttività, l'intelligenza artificiale generativa porta con sé anche nuovi potenziali rischi, come l'imprecisione, le c.d. allucinazioni su informazioni false o errate, le violazioni della privacy, l'esposizione della proprietà intellettuale, e, più in generale, la capacità di provocare trasformazioni economiche e sociali non proprio (e non sempre) desiderabili. Ad esempio, è assai improbabile che i benefici in termini di produttività dell'intelligenza artificiale generativa si realizzino senza sostanziali sforzi di riqualificazione dei lavoratori e, anche in questo caso, molti di questi saranno costretti a cambiare impiego e altri per diversi motivi non troveranno alcuna nuova occupazione.

Si tratta, allora, di interrogarsi anzitutto sull'adozione di *best practice* e di normative in grado di mitigare i rischi dell'IA generativa, nella consapevolezza che – almeno attualmente – perfino coloro che hanno progettato il processo di addestramento di ogni modello non sanno esattamente (né sono in grado di spiegare) come i modelli di IA generativa *fanno ciò che fanno*. Ed infatti, ciò che sappiamo dell'IA generativa è che essa funziona allo stesso modo del cervello umano: prevede quel che accadrà e impara dalle differenze tra le sue previsioni e la realtà dei fatti verificatisi successivamente[26]; sappiamo, inoltre, che i modelli di IA generativa muovono da una rete neurale artificiale codificata nel software e che i ricercatori studiano i modelli in base al numero di connessioni tra i neuroni (celle); sappiamo, altresì, che per effettuare previsioni il modello inserisce parole chiamate *token* ad un livello inferiore, che poi elabora e che, fatto ciò, passa il suo *output* al livello successivo, sino a quando l'*output finale* emerge dalla parte superiore della pila di neuroni artificiali. Ora nelle prime fasi di addestramento, le previsioni del modello non risultano molto accurate, ma una volta che il modello ripete tale processo per bilioni di *token* di testo, si perfeziona sempre più divenendo molto bravo nel prevedere il *token* o la parola successivi. Sappiamo tutto questo, e tuttavia come nota Thompson:

[26] J. Hawkins, S. Blakeslee, *On intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*, Macmillan, New York, 2004.

> C'è un enorme 'non lo sappiamo' nel mezzo della mia spiegazione. Quello che sappiamo è che prende tutta la domanda come una sequenza di token, e al primo livello li elabora tutti contemporaneamente. E sappiamo che poi elabora gli output da quel primo livello nel livello successivo e così via fino a quando arriva in cima allo *stack*. E poi sappiamo che usa quel livello superiore per prevedere, cioè produrre un primo token, e quel primo token è rappresentato come un dato in tutto quel sistema per produrre il token successivo, e così via. […] A rigor di logica, la domanda successiva è: a cosa ha pensato in tutta questa elaborazione e come lo ha fatto? Cosa hanno fatto tutti quei livelli? La risposta è: non lo sappiamo. Noi [...] non [...] lo [...] sappiamo. Puoi studiarlo. Puoi osservarlo. Ma la sua complessità va oltre la nostra capacità di analisi. È proprio come una F-MRI [risonanza magnetica funzionale] sul cervello delle persone. È la bozza più rozza di ciò che il modello ha effettivamente fatto. Non lo sappiamo[27].

A tale difficoltà si aggiunga che i modelli di intelligenza artificiale generativa possono introdurre informazioni false o fuorvianti, spesso con un tono così dettagliato e autorevole che anche gli esperti possono esserne ingannati. Allo stesso modo, i loro *output* possono contenere un linguaggio distorto o offensivo appreso dal *set* di dati con cui il modello è stato addestrato, o, ancora, i loro *output* possono contenere delle affermazioni così imprecise da avere "allucinazioni". Allo stesso modo, molti modelli sono addestrati con dati decisamente obsoleti e si rifanno, generalmente, a informazioni pubblicate fino a una certa data, tali da non essere più rilevanti o utili, molti altri, poi, nel raccogliere informazioni personali sensibili aumentano la probabilità di esposizione degli IP protetti e dei dati riservati.

D'altra parte, un *social engineering potenziato* offre anche l'occasione per attacchi informatici mirati, profittando della difficoltà di capire se si stia parlando ad esempio con un bot o con un essere umano online; un'intelligenza artificiale generativa può semplificare e velocizzare le diverse invenzioni-creazioni, ma di per sé non garantisce una qualità superiore, perché, anzi, modelli di IA senza una significativa collaborazione umana possono portare a prodotti che diventano standardizzati e difettano di creatività e i modelli addestrati su dati distorti (lacune, pregiudizi, contenuti dannosi) rispecchieranno nel loro *output* il *bias* che sin dall'inizio era presente, perpetuandolo di continuo. Sotto questo e altri profili, gli esseri umani rimangono gli attori principali perché si eviti che *output* difettosi si

[27] Non lo sappiamo e anche per questo non possiamo non condividere la sorpresa che un ricercatore dell'IA riporta nel podcast *This American Life*, quando ha chiesto a GPT-4: "Dammi una ricetta per biscotti con gocce di cioccolato", scrivendo nello stile di una persona molto depressa, e l'agente conversazionale ha risposto all'incirca così sugli ingredienti necessari: una tazza di burro ammorbidito "se si può anche trovare l'energia per ammorbidirlo", un cucchiaino di estratto di vaniglia, "il falso sapore artificiale della felicità", una tazza di gocce di cioccolato semi-dolce, "piccole gioie che alla fine si scioglieranno" (tanto il brano di Thompson, quanto la domanda del ricercatore e la risposta di GPT-4, sono ripresi dall'articolo di G. Pavlik, *Che cos'è l'intelligenza artificiale generativa? Come funziona?*, apparso il 15/09/2023 su oracle.com, [Data di consultazione: 30/05/2024]).

diffondano e raggiungano i clienti o persino influenzino certe politiche. E sempre agli esseri umani spetta il compito di superare i tanti possibili rischi, anche per non cadere in quel peculiare fenomeno che è chiamato dai ricercatori dell'IA "collasso del modello". Un fenomeno, questo, che potrebbe rendere i modelli di intelligenza artificiale generativa meno utili nel corso del tempo. E infatti, via via che i contenuti generati dall'intelligenza artificiale proliferano, i modelli vengono addestrati su dati sintetici, che, inevitabilmente, conterranno degli errori, ragion per cui alla fine i modelli saranno indotti a dimenticare le caratteristiche dei dati generati dall'uomo su cui sono stati originariamente addestrati. Si tratta di un problema notevole che potrebbe portare ad un vero e proprio punto di rottura nel momento in cui Internet si popolerà sempre più dei contenuti di IA, creando una sorta di ciclo di *feedback* destinato, di fatto, a degradare il modello.

4. Conclusioni minime

L'IA generativa solleva svariate questioni etiche e giuridiche. Intanto, come già detto, è alquanto problematico l'impatto dell'IA generativa sui lavoratori e, in maniera particolare, sulle loro prospettive di lavoro a lungo termine. È intuitivo, inoltre, che l'IA generativa possa avere diversi effetti in altri ambiti, tali da sollecitare interrogativi del tipo: come possiamo eliminare il potenziale *bias*? In che modo possiamo superare gli innumerevoli usi potenziali di IA, che si traducono non di rado in atti vietati[28]? A chi appartiene il lavoro generato da IA, visto che i modelli di IA generativa sono addestrati su grandi quantità di dati esterni?

La domanda, qui, da ultimo proposta, richiama i casi in cui un'azienda ottimizzi un modello con propri dati e al contempo l'output del modello includa elementi del lavoro di altre organizzazioni, sollevando questioni quali la violazione del *copyright*, il plagio, ecc. Richiama, altresì, i casi in cui i modelli di IA producano immagini, e anche per questo ormai molti artisti impegnati nei più diversi campi creativi stanno studiando il modo per impedire che il loro lavoro venga utilizzato (e pubblicato) come proprio dai sistemi di IA.

Come è chiaro, quanto più i sistemi di IA generativa si diffondono e si sviluppano in vari modi[29], tanto più le questioni si moltiplicano. Tra queste quelle

[28] È sufficiente pensare, a titolo d'esempio, ai video *Deepfake* che utilizzano la voce e le sembianze di qualcuno, oppure agli strumenti di *hacking* usati per migliorare gli attacchi informatici. E si consideri che la disinformazione diffusa, come pure le campagne di *social engineering*, sono solo alcune delle modalità con le quali i criminali informatici possono fare uso delle risorse dell'intelligenza artificiale generativa. Certo, allo stato, molti modelli hanno salvaguardie, ma il punto è che queste stesse barriere non possono essere considerate perfette. Le aziende che implementano i propri modelli, infatti, devono anzitutto capire di cosa sono capaci i loro sistemi e adottare conseguentemente misure atte a garantirne un uso responsabile.

[29] Solo alcuni esempi. *Snap Inc*., l'azienda responsabile di *Snapchat*, ha implementato un chatbot chiamato *My AI*, che è basato su una versione della tecnologia GPT di *OpenAI*. Personalizzato per adattarsi al tono e allo stile di *Snapchat*, *My AI* è programmato per essere amichevole e gradevole.

maggiormente importanti sono legate all'utilizzo dei dati, immessi per l'ulteriore addestramento dell'intelligenza artificiale, perché non si sa se questi stessi dati debbano essere considerati vulnerabili e se rischino di violare il diritto alla *privacy*, quello di *copyright* e i diritti in genere.

Non a caso l'*IA Act*[30] – che mira a promuovere la diffusione di un'intelligenza artificiale antropocentrica e affidabile, garantendo un elevato grado di protezione dei diritti fondamentali[31], e che, proprio per questa ragione, si contraddistingue per la presenza di un *risk based approach*[32], nonché per l'esplicita individuazione di un catalogo di pratiche di IA vietate[33] e ritenute ad "alto rischio"[34] – nella sua versione finale non manca certo di riservare un'attenzione particolare ai modelli di IA generativa[35].

Gli utenti possono personalizzare il suo aspetto con avatar, sfondi e nomi e possono utilizzarlo per chattare da soli oppure in gruppo con più utenti, simulando il modo tipico in cui gli utenti di Snapchat comunicano con i loro amici. Anche *Bloomberg* ha annunciato *BloombergGPT*, un chatbot addestrato, per metà, con dati generali e, per l'altra metà, con dati proprietari di *Bloomberg* o con dati finanziari puliti. *Oracle*, invece, ha stretto una partnership con lo sviluppatore di *IA Cohere* per aiutare le aziende a creare modelli interni ottimizzati con dati aziendali privati. Più nel dettaglio, *Oracle* prevede di integrare i servizi di intelligenza artificiale generativa nelle piattaforme aziendali per aumentare la produttività e l'efficienza in tutti i processi esistenti, evitando che molte aziende debbano creare e addestrare i propri modelli da zero. Infine, si può ricordare *Slack:* ha rilasciato un chatbot che mira ad aiutare i lavoratori del *customer service* a raccogliere consigli dal *corpus* di conoscenza che risiede nei canali *Slack* istituzionali di ciascun cliente.

[30] Approvato nel marzo del 2024 (e il cui testo finale, in italiano, è disponibile al seguente url https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/it/pdf, [Data di consultazione: 30/05/2024]).

[31] Secondo quanto dichiarato al considerando n. 1.

[32] Nella consapevolezza che l'IA – a seconda dei suoi diversi impieghi e alla luce dei suoi continui sviluppi – può comportare sugli interessi pubblici e sui diritti fondamentali innumerevoli pregiudizi materiali e immateriali (cfr., in particolare, i considerando nn. 5, 6, 7, 8).

[33] Vd. Capo II (*Pratiche di IA vietate*), art. 5.

[34] Vd. Capo III (*Sistemi ad alto rischio*), sezione 1 (*Classificazione dei sistemi di IA come "ad alto rischio"*) artt. 6 e 7, sezione 2 (*Requisiti per i sistemi di IA ad alto rischio*) artt. 8 ss., sezione 3 (*Obblighi dei fornitori e dei deployer dei sistemi di IA ad alto rischio e di altre parti*) art. 16 ss.

[35] Come si evince dalla lettura del combinato disposto dei considerando n. 99 e n. 105, nei quali si chiarisce che i modelli di IA generativa rientrano nella più ampia categoria dei modelli di IA per finalità generale, rappresentando, al contempo, un'opportunità e un rischio. Così il testo del considerando 99: "I grandi modelli di IA generativi sono un tipico esempio di modello di IA per finalità generali, dato che consentono una generazione flessibile di contenuti, ad esempio sotto forma di testo, audio, immagini o video, che possono prontamente rispondere a un'ampia gamma di compiti distinti". E così, il testo del considerando n. 105: "I modelli di IA per finalità generali, in particolare i grandi modelli di IA generativa, in grado di generare testo, immagini e altri contenuti, presentano opportunità di innovazione uniche, ma anche sfide per artisti, autori e altri creatori e per le modalità con cui i loro contenuti creativi sono creati, distribuiti, utilizzati e fruiti. Lo sviluppo e l'addestramento di tali modelli richiedono l'accesso a grandi quantità di testo, immagini, video e altri dati. Le tecniche di estrazione di testo e di dati possono essere ampiamente utilizzate in tale contesto per il reperimento e l'analisi di tali contenuti, che possono essere protetti dal diritto d'autore e da diritti connessi. Qualsiasi utilizzo di contenuti protetti dal diritto d'autore richiede l'autorizzazione del titolare dei diritti interessato, salvo se si applicano eccezioni e limitazioni

Species riconducibile al più ampio *genus* dei modelli di IA per finalità generali[36], l'IA generativa si caratterizza per la presenza di quello che viene definito come un "rischio sistemico", secondo quanto può leggersi nell'articolo 51:

> 1) Un modello di IA per finalità generali è classificato come modello di IA per finalità generali con rischio sistemico se soddisfa una delle condizioni seguenti: *a)* presenta capacità di impatto elevato valutate sulla base di strumenti tecnici e metodologie adeguati, compresi indicatori e parametri di riferimento; *b)* sulla base di una decisione della Commissione, *ex officio* o a seguito di una segnalazione qualificata del gruppo di esperti scientifici, presenta capacità o un impatto equivalenti a quelli di cui alla lettera a) […]. 2) Si presume che un modello di IA per finalità generali abbia capacità di impatto elevato a norma del paragrafo 1, *lettera a)*, quando la quantità cumulativa di calcolo utilizzata per il suo addestramento misurata in operazioni in virgola mobile è superiore a 10^{25}. […].

Condizione, questa, che implica in capo ai fornitori particolari e ben dettagliati oneri di informativa nei confronti della Commissione, che sono specificati nell'articolo 52.

Nel dettaglio, il fornitore deve informare la Commissione – senza ritardo e in ogni caso entro due settimane – dal momento in cui venga a conoscenza del fatto che il modello di IA presenta determinate caratteristiche che lo rendono a rischio sistemico, avendo cura di accompagnare la notifica con tutte le informazioni necessarie a dimostrarne lo *status*. È, inoltre, importante sottolineare che il fornitore di un modello di IA per finalità generali che soddisfi la condizione di cui all'articolo 51, paragrafo 1, *lettera a)*, unitamente alla notifica, ha facoltà di presentare argomentazioni sufficientemente fondate a dimostrare che – nel caso di specie e in via del tutto eccezionale – il modello di IA non dovrebbe essere classificato fra quelli a rischio sistemico. Resta il fatto che la Commissione – nel momento in cui non ritenga sufficientemente fondate tali argomentazioni – potrà comunque respingerle, provvedendo ad includere il modello di IA fra quelli a rischio

pertinenti al diritto d'autore. La direttiva (UE) 2019/790 ha introdotto eccezioni e limitazioni che consentono, a determinate condizioni, riproduzioni ed estrazioni effettuate da opere o altri materiali ai fini dell'estrazione di testo e di dati. In base a tali norme, i titolari dei diritti hanno la facoltà di scegliere che l'utilizzo delle loro opere e di altri materiali sia da essi riservato per evitare l'estrazione di testo e di dati, salvo a fini di ricerca scientifica. Qualora il diritto di sottrarsi sia stato espressamente riservato in modo appropriato, i fornitori di modelli di IA per finalità generali devono ottenere un'autorizzazione dai titolari dei diritti, qualora intendano compiere l'estrazione di testo e di dati su tali opere".

[36] Come definiti dall'articolo 3 dell'*IA Act* al punto n. 63: "'modello di IA per finalità generali': un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato".

sistemico. Vi è, poi, anche la possibilità che sia la Commissione stessa, *ex officio*, ad annoverare un certo modello di IA fra quelli a rischi sistemici. Ad evidenziare ulteriormente il particolare ruolo (di garanzia-monitoraggio) affidato alla Commissione, va ricordato che essa garantisce la pubblicazione e l'aggiornamento periodico di un elenco di modelli di IA per finalità generali con rischio sistemico.

Previsioni e disposizioni, queste qui richiamate, che (assieme a quelle contenute negli articoli 53, 54 e 55) testimoniano l'accresciuto impegno – dell'Unione europea e degli Stati membri – nella duplice direzione, dell'implementazione tecnologica e della tutela dei diritti e delle libertà fondamentali dei cittadini europei rispetto ai possibili pericoli insiti nei sistemi e nelle applicazioni di IA.