OpenAI e ChatGPT: funzionalità, evoluzione e questioni aperte

OpenAI and ChatGPT: functionalities, evolution and challenges

Rossana Morriello
Politecnico di Torino
rossana.morriello@polito.it

| abstract

L'obiettivo dell'articolo è tracciare un profilo critico di OpenAI e dei suoi principali prodotti, collocandolo all'interno della discussione generale sull'intelligenza artificiale e in una prospettiva diacronica di costante dialettica tra l'uomo e la tecnologia. In particolare, ChatGPT-3 ha catalizzato il dibattito sulle potenzialità e sulle questioni aperte dell'intelligenza artificiale e dai sistemi di IA generativa. Ripercorrere l'evoluzione del modello e dell'organizzazione che lo ha rilasciato è il primo passo per rapportarsi agli strumenti di IA con un approccio costruttivo ed equilibrato, che consenta di comprenderne la rilevanza e la necessità di governarli per un uso democratico e inclusivo. Nello specifico l'articolo vuole essere un avvio introduttivo per successivi approfondimenti in ambito bibliografico e documentario.

The aim of the article is to draw a critical profile of OpenAI and its main products, positioning it within the general discussion on artificial intelligence and in the diachronic perspective of the relations between man and machine. Particularly, ChatGPT-3 has catalyzed the AI debate and increased general awareness of the potential of similar tools and of the open issues to be resolved and regulated. Retracing the evolution of this model and of the organization that released it, is the first step in order to understand relevance and foster the need to govern them for a democratic and inclusive use. The article aims to be a starting point for further studies from a bibliographic and documentary point of view.

DOI 10.36158/97888929573674

Introduzione

In un articolo del 1964, Derek J. De Solla Price ripercorreva le tappe della dialettica tra uomo e macchina, e del rapporto tra il mondo reale e l'immaginazione inventiva che spinge a superare la realtà proiettandola verso il futuro, per dimostrare come sia una costante nella storia dell'umanità la tendenza a creare tecnologie in grado di riprodurre le caratteristiche del mondo naturale, animale e umano e superarne i limiti. Questo percorso comincia da rappresentazioni quali le figure pittoriche egizie che simulavano il movimento, le ali di Dedalo che consentono a Icaro di spiccare il volo, o le invenzioni antropomorfe come la statua semovente raffigurante il defunto che Marco Antonio aveva fatto erigere e mostrare pubblicamente al funerale di Giulio Cesare, e passa per i numerosi

orologi, meridiane, astrolabi e altri meccanismi ideati per simulare il movimento cosmico nell'antichità e nel Medioevo (Price, 1964; cfr. anche Price, 1959). Nei secoli successivi troviamo ulteriori esempi nelle differenti forme di automi e macchine in grado di compiere movimenti di scrittura, di danza, di esecuzione musicale. Alle applicazioni pratiche si accompagnano le riflessioni filosofiche, quali il dualismo mente-corpo cartesiano (Kantosalo, Falk & Jordanous, 2021), lungo una linea di continuità ininterrotta che giunge fino al dibattito contemporaneo sull'intelligenza artificiale. L'IA è dunque la fase più recente di un processo di evoluzione tecnologica continuo, che ha subito un'accelerazione nella seconda metà del XX secolo, e a ogni innovazione ha visto schierarsi apocalittici e integrati.

Tra questi ultimi si colloca il matematico e informatico di origine ungherese John G. Kemeny, il quale nel 1955, quando si ponevano le basi dell'intelligenza artificiale, in un articolo assumeva una posizione lungimirante rispetto alle teorizzazioni sulle potenzialità della tecnologia, poiché non escludeva la possibilità della macchine di compiere azioni umane e perfino di riprodursi in uno scenario futuro. Kemeny lasciava aperta ogni eventualità, nella consapevolezza di non poter conoscere a priori gli sviluppi della tecnologia (Kemeny, 1955). Lo studioso ungherese affermava come proprio per questo motivo occorresse evitare del tutto di chiedersi cosa le macchine avrebbero potuto fare in futuro e, in uno specifico esempio nel suo articolo, si riferiva alla capacità delle macchine un giorno di scrivere sonetti. Un esempio sagace, dal momento che oggi le macchine, con l'intelligenza artificiale, sono in grado di scrivere componimenti poetici, nonché la conferma della ragionevolezza delle sue affermazioni[1]. Lo sviluppo tecnologico procede per piccoli passi, spesso attraverso ricerche distinte su singoli ambiti che vengono svolte senza troppo clamore pubblico, ma che una volta messe in relazione e raggiunto il momento di maturità, producono un avanzamento significativo nella scienza. Si tratta di un processo continuo e non è dunque possibile conoscere a priori le mete raggiungibili.

L'intelligenza artificiale non è differente da questo punto di vista poiché alle mete odierne si è giunti per tappe, prove, esperimenti. Per ricordarne solo qualcuna, il Perceptron (o Percettrone), fu un modello di rete neurale artificiale introdotto nel 1958 dallo psicologo statunitense Frank Rosenblatt che veniva pubblicizzato come la prima macchina capace di creare idee originali (Hong, 2004, p. 56). In particolare, ispirandosi alle teorie messe a punto dal neurofisiologo Warren McCulloch e dal matematico Walter Pitts, Rosenblatt tentò di simulare il funzionamento dei neuroni artificialmente. Il Perceptron era in grado di riconoscere schemi di lettere e parole, attribuendo un valore 1 (immagine riconosciuta) e 0 (immagine non riconosciuta) al dato, e poneva le basi per le odierni reti neurali. Un altro esempio è Cybertron K-100, un prodotto sviluppato nel 1961 dalla Raytheon che sapeva distinguere gli impulsi sonori provenienti da una fonte esterna grazie a un dataset di alcune migliaia di suoni che il suo creatore, Richard Witt, aveva fornito come campione alla macchina. Witt aveva poi allenato la macchina tramite un bottone che consentiva di confermare oppure rifiutare le risposte fornite da Cybertron in forma di segnale luminoso. Si trattava di una prima rudimentale applicazione di tecniche di machine learning. ELIZA, il programma creato da Joseph Weizenbaum negli anni Sessanta, era, invece, una primordiale forma di chatbot che sperimentava elaborazioni del linguaggio naturale per conversazioni tra uomo e macchina[2]. Avvicinandosi ai nostri

1. Cfr. Rich (2022). Nell'articolo sono riportati alcuni poemi che il *New Yorker* ha chiesto di creare a ChatGPT-3 nello stile di Philip Larkin, Mark Twain e perfino Shakespeare. Certo, mancano di sentimento e passione (Hunter, 2023), ma rimane il fatto che non è facile distinguerle per la persona comune e nemmeno per l'esperto di poesia, se non perché quest'ultimo conosce tutte le poesie di un autore e può rendersene conto.
2. Cfr. https://www.masswerk.at/eliza/ e l'articolo del creatore, professore di informatica al MIT (Weizenbaum, 1966). Sulla storia di ELIZA anche (Bassett, 2029).

DOI 10.36158/97888929573674

giorni, nel 2007 Robbie Allen fondò StatSheet, una piattaforma in grado di raccogliere, elaborare e restituire in forma narrativa le statistiche sportive, con la possibilità di esportare le statistiche, i grafici, i report e altre forme di visualizzazione dei dati per integrarle in altri siti web e blog. StatSheet (poi Automated Insights con l'espansione oltre lo sport), si basava su un sistema di Natural Language Generation (NLG), un campo di studi interdisciplinare tra intelligenza artificiale e linguistica computazionale, volto a creare algoritmi capaci di trasformare un set di dati, anche non linguistici, in una forma narrativa.

Si tratta solo di alcune delle tante ricerche sviluppate e sperimentate in passato, e magari poi abbandonate senza riuscire ad andare oltre la fase di test. Fino a tempi recenti mancavano due grandi leve di accelerazione, oggi disponibili, il notevolissimo aumento del potere computazionale dei computer e la disponibilità di grandi quantità di dati, i big data del web, che con l'inizio del millennio hanno aperto nuove frontiere di ricerca e di sviluppo. Diverse delle sperimentazioni abbandonate sono state infatti riprese e completate.

Gli esempi precedenti avevano lo scopo di testimoniare come nel campo delle tecnologie non si possa parlare di fenomeni o di scoperte improvvise ma piuttosto di processi, la cui durata può essere più o meno lunga e la cui interpretazione può risultare a volte parziale, perlomeno fino a quando i processi non acquisiscono un certo grado di maturità. Ciò implica che i processi tecnologici in corso non vanno sottovalutati o trattati come fenomeni transitori a causa della loro fallacia, dovuta alla condizione di non essere ancora sufficientemente maturi. La natura della scienza è mettere a sistema i risultati, anche parziali, della ricerca, attraverso le connessioni tra i ricercatori e i loro lavori, finché a un certo punto grazie a una particolare innovazione, una nuova interfaccia, o semplicemente perché la società in quel momento è pronta per accogliere un prodotto, questo diventa popolare, spostandosi dal chiuso dei laboratori alla disponibilità pubblica. È quanto è accaduto con OpenAI, e in particolare con ChatGPT, che dal suo rilascio a novembre 2022 ha catalizzato l'attenzione dell'opinione pubblica ed è divenuto oggi il "soggetto" mediatico principale nel dibattito sull'intelligenza artificiale, del quale si discute ampiamente in ambiti sia specialistici sia generalisti, soffermandosi di volta in volta su alcune delle innumerevoli tessere di un mosaico complesso.

L'obiettivo di questo articolo è tentare di ricomporre, almeno in parte, una prima parte del mosaico, delineando un profilo di evoluzione della società OpenAI e descrivendone sinteticamente i prodotti principali, così da strutturare e ordinare logicamente la quantità di informazioni, notizie, opinioni, sovente contrastanti e antitetiche, in circolazione. In questo modo, si vuole dar conto della complessità e costruire una base informativa da cui partire per elaborazioni future e per un più consapevole posizionamento all'interno della discussione generale sull'intelligenza artificiale. Lo sguardo è rivolto alle possibili applicazioni in ambito bibliotecario e documentario dove i sistemi generativi di IA avranno un progressivo impatto su tante attività, dal reference alla gestione dei metadati, dall'accesso alle collezioni all'information literacy. Riprendendo le parole di Kemeny, interrogarsi su cosa può fare e non può fare uno strumento come ChatGPT può apparire aleatorio proprio perché non possiamo sapere quali saranno le evoluzioni dell'IA, in un'epoca in cui peraltro l'avanzamento della tecnologia è ben più rapido di settant'anni fa e lo stato dell'arte attuale dell'intelligenza artificiale può risultare superato nell'arco di pochi mesi se non settimane. Tuttavia, conoscere il contesto in cui nascono e si sviluppano le applicazioni di intelligenza artificiale e comprendere come ci si è arrivati appare un prerequisito indispensabile per rapportarvisi con un approccio costruttivo e poterle governare, seguendone l'evoluzione. Conviene specificare che in questa sede

DOI 10.36158/97888929573674

non si intende andare a fondo sulle specifiche applicazioni, sulle potenzialità positive e sui molti aspetti critici, tanto meno in ambito documentario e bibliotecario, ma si vuole cominciare a portarne alla luce alcune tra le più rilevanti, con lo scopo di contribuire a creare un quadro di riferimento chiaro per favorire successivi approfondimenti.

Breve profilo di OpenAI

L'azienda che recentemente ha più contribuito a far emergere nella discussione pubblica il tema dell'intelligenza artificiale è dunque OpenAI, fondata nel 2015 come società non profit con un finanziamento di un miliardo di dollari da un gruppo di investitori della Silicon Valley tra i quali Sam Altman, Peter Thiel, Jessica Livingston, Elon Musk, il co-fondatore di LinkedIn Reid Hoffman, e con il supporto delle aziende Amazon Web Services (AWS), Infosys, and YC Research. Elon Musk, uno dei personaggi più controversi nell'attuale dibattito sui media digitali, è formalmente uscito dal board degli investitori nel 2018[3], ma continua a sovvenzionare l'organizzazione con delle donazioni. L'obiettivo dichiarato di OpenAI è di implementare e diffondere un'intelligenza artificiale amichevole e sicura, a beneficio dell'umanità[4].

Dopo solo un anno dalla fondazione, nell'aprile del 2016, OpenAI ha rilasciato una versione beta pubblica di OpenAI Gym, una piattaforma di ricerca sul reinforcement learning, o apprendimento per rinforzo, di cui ha poi affidato lo sviluppo all'organizzazione non profit Farma Foundation, specializzata in questo settore di ricerca[5]. Nella definizione dei padri del reinforcement learning, Richard Sutton e Andrew Barto, si tratta di un'area di ricerca del machine learning e delle reti neurali nella quale la macchina apprende attraverso l'interazione con l'ambiente circostante (Sutton & Barto, 2014). È uno dei tre principali paradigmi di apprendimento del machine learning ma diversamente dall'apprendimento supervisionato, in cui la macchina apprende da un set di dati etichettati e classificati, e dall'apprendimento non supervisionato in cui la macchina estrae le regole strutturali da dati non etichettati che clusterizza e classifica a posteriori, nell'apprendimento per rinforzo è l'interazione con l'ambiente che consente all'agente di prendere le decisioni migliori rispetto a un obiettivo. La macchina non cerca la struttura sottostante ma attraverso prove ed errori associati a degli input impara a riconoscere le azioni che consentono di ottenere la massima ricompensa sul lungo periodo. La ricerca sull'apprendimento per rinforzo in OpenAI porta alla release nel dicembre 2016 di Universe, una piattaforma per il training dell'algoritmo di reinforcement learning. Su Universe l'agente di intelligenza artificiale può interagire con gli ambienti virtuali dei videogame per apprendere come operare non solo in situazioni previste e programmate ma anche in situazioni nuove e impreviste. Come dichiara la stessa OpenAI, è questa la frontiera a cui tende la ricerca, ovvero l'AGI (Artificial General Intelligence), o intelligenza artificiale forte, che a differenza dei modelli attuali (detti di intelligenza artificiale debole) che lavorano su un problema per risolverlo o per offrire uno specifico servizio, dovrebbe essere in grado di portare a termine qualsiasi compito.

Nel 2019 la società californiana cambia la natura senza scopo di lucro e diventa capped profit, ovvero ibrida tra profit e non profit, con la creazione della spin-off OpenAI LP (Limi-

3. Ufficialmente per conflitto di interessi ma secondo alcune fonti dopo i tentativi respinti di prendere il controllo dell'organizzazione, cfr. Vincent (2023).
4. Le informazioni e i dati su OpenAI e sulle sue applicazioni riportate in questo articolo, dove non diversamente citato, sono prese dal sito https://openai.com/.
5. Gymnasium https://gymnasium.farama.org/.

DOI 10.36158/97888929573674

ted Partnership). L'operazione serve ad attirare e gestire proficuamente gli investimenti esterni. Secondo la nuova formula, gli investitori e i collaboratori avranno un ritorno di investimento ma non superiore a un certo tetto, o "cap", se la compagnia riuscirà a ottenere i risultati attesi. Oltre tale tetto i guadagni verranno convogliati verso l'originaria OpenAI non profit. La trasformazione permette di accogliere un grosso finanziamento di Microsoft, pari a un altro miliardo di dollari, che contestualmente avvia una partnership per l'implementazione della piattaforma di cloud computing Microsoft Azure.

Nel 2021 viene rilasciata una prima versione di Azure OpenAI e il 16 gennaio 2023 è stata annunciata la disponibilità della piattaforma per tutti gli sviluppatori e le aziende tech nella versione attuale che integra i prodotti di OpenAI GPT-3.5, Codex[6], e DALL-E 2, e quindi anche la possibilità di usare ChatGPT-3. Tre mesi dopo, l'azienda di Bill Gates ha annunciato l'introduzione dell'IA generativa in Microsoft 365, la versione cloud della suite Office, con MS 365 Copilot, un assistente virtuale che svolge compiti quali elaborare le presentazioni powerpoint, creare una sintesi degli argomenti trattati in una riunione su Teams, creare una bozza di un documento Word, rivederla, farne un riassunto[7]. Diverse fonti sostengono che al primo investimento si sia affiancato nel 2023 un ulteriore investimento di dieci miliardi di dollari da parte di Microsoft (Metz & Weise, 2023).

I prodotti attualmente più noti e avanzati di OpenAI sono i modelli di reti neurali artificiali GPT-3, cui a marzo 2023 è seguita GPT-4, con la derivata ChatGPT, e il sistema generativo di immagini DALL-E. In precedenza sono stati realizzati vari altri prodotti, come OpenAI Five, un sistema di IA basato su algoritmi di reinforcement learning per il gaming, analogo al noto Deep Blue della IBM che nel 1997 sconfisse il campione di scacchi Garry Kasparov. Il progetto Learning Dexterity, avviato nel 2018 per insegnare ad arti robotici a eseguire compiti umani come afferrare un oggetto. Prima di giungere a perfezionare ChatGPT-3, si è passati per ChatGPT-2, un modello di IA non supervisionata, rilasciato nel 2019.

L'elenco non esaustivo di sperimentazioni cui si è accennato sopra intende dimostrare come OpenAI sia una sorta di piattaforma aperta, un incubatore di idee che possono essere realizzate internamente, oppure in collaborazione con aziende esterne, alle quali a volte viene poi ceduto lo sviluppo, se ritenuto non percorribile da OpenAI. L'insieme delle sperimentazioni sui vari fronti alimenta i prodotti di punta come ChatGPT e DALL-E. Dal 2015 OpenAI ha operato senza troppi clamori, lavorando sulle diverse tecnologie, fino al momento in cui un modello ritenuto sufficientemente maturo come GPT-3 è stato rilasciato con l'interfaccia amichevole della chat in grado di elaborare e rispondere a input di testo in linguaggio naturale, così da consentire un training massivo, ovvero quello che milioni di persone stanno compiendo su ChatGPT-3 in molti paesi del mondo, quantificato in cento milioni di visitatori nei soli primi tre mesi dal rilascio pubblico[8].

ChatGPT e DALL-E: opportunità e rischi

GPT (Generative Pre-trained Transformer) è il modello di rete neurale artificiale, realizzato nel 2020, sul quale sono costruire le chat di OpenAI, e in particolare la tan-

6. Il sistema di OpenAI che traduce il linguaggio naturale in codice https://openai.com/blog/openai-codex.
7. *Introducing Microsoft 365 Copilot – your copilot for work*, March 16, 2023, https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/.
8. Reuters stima che a gennaio 2023 ChatGPT abbia avuto cento milioni di utilizzatori a due mesi dal lancio, e 13 milioni di visitatori unici al giorno https://www.reuters.com/technology/ChatGPTsets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

DOI 10.36158/97888929573674

to discussa ChatGPT-3. GPT è un tipo di rete neurale definito Large Language Model (LLM), ovvero modelli linguistici di grandi dimensioni, in quanto usano enormi quantità di dati per il training, dai quali derivano modelli statistici con tecniche di elaborazione del linguaggio naturale (NLP) e generazione del linguaggio naturale (NLG). Sulla base della modellizzazione statistica della rete neurale soggiacente, strumenti come ChatGPT sono in grado di interagire ai prompt testuali immessi per rispondere a quesiti, tradurre brani, e anche generare testi dalle caratteristiche simili a quelli che potrebbero essere creati da un essere umano. ChatGPT-3 è la versione della chatbot di OpenAI rilasciata per la prima volta pubblicamente il 30 novembre 2022 e diventata in breve virale. L'aspetto innovativo di ChatGPT-3 è l'essere dotata di un'interfaccia semplice e amichevole, che consente interrogazioni tramite il linguaggio naturale a chiunque e non soltanto a chi padroneggia linguaggi di programmazione informatica. Il testo di risposta al prompt viene generato sulla base dei dati di training con i quali il modello è stato implementato, pari a 570 GB di dati corrispondenti a 300 miliardi di parole, e alle elaborazioni algoritmiche della rete neurale GPT che utilizza 175 miliardi di parametri[9]. Prima di essere rilasciato pubblicamente il modello è stato dunque oggetto di training sui dati, anche con intervento umano[10], inclusi processi di *adversarial training*, un addestramento su attacchi simulati da chatbot avversarie che lo rende in grado di riconoscere attacchi malevoli e ingannevoli (Hao, 2020). Il successo di ChatGPT-3 ha senza dubbio amplificato il dibattito sulla tecnologia e sul futuro dell'intelligenza artificiale.

Una parte consistente del dibattito si focalizza sull'obiettivo di stabilire se e quanto ChatGPT sia intelligente, partendo dal presupposto che vi sia consenso su cos'è l'intelligenza umana. In realtà, gli studi su questo fronte ci dicono che non vi è una definizione univoca e considerata accettabile in tutte le epoche di intelligenza, né ancora ai nostri giorni vi è un consenso unanime su quali elementi la caratterizzino. Nelle differenti epoche della storia dell'umanità sono stati individuati elementi diversi alla base della nozione di intelligenza, e diversa ne è anche l'interpretazione tra culture differenti. In generale, vi sono molti approcci, e fattori variabili a seconda dell'approccio, che può essere biologico, sociologico, antropologico, epistemologico, e così via (Cianciolo & Sternberg, 2004; Cornoldi, 2007). L'approccio computazionale è uno di questi. Tuttavia, non mi soffermerò sulla questione dell'intelligenza di ChatGPT-3 né entrerò nel dibattito su quanto sia intelligente l'intelligenza artificiale, rimandando alle numerose sedi che lo hanno già ospitato[11]. Si tratta di "un'intelligenza", se così la si vuole definire, certamente diversa da quella umana, di tipo prevalentemente comunicativo più che cognitivo, in quanto genera risposte rielaborando i dati già esistenti in base a distribuzioni statistiche. Un algoritmo può tradurre un testo in cinese senza conoscere il cinese e correggere un testo in inglese senza comprenderne il significato (Esposito, 2021), così come riconosce un'immagine senza comprendere che si tratta di un'immagine ma soltanto computando i pixel nella loro distribuzione statistica. Il *Manifesto on algorithmic humanitarianism* di Dan McQuillan ribadisce come non vi sia intelligenza nell'intelligenza artificiale né apprendimento nel machine learning ma si tratti solo di un processo di minimizzazione matematica e dunque non abbia senso chiedersi perché un sistema di ML abbia risposto in un modo piuttosto che in un altro (McQuillan, 2018).

9. I dati e le informazioni sui ChatGPT sono tratti dal sito stesso (OpenAI, 2022).
10. Oggetto di contestazione per l'uso di lavoratori sottopagati, cfr. Perrigo (2023).
11. La rivista *Limes* ha intitolato un intero fascicolo 12/2022 all'argomento, con il titolo *L'intelligenza non è artificiale*. Cfr. anche, tra i numerosi scritti (Floridi, 2021; 2023).

DOI 10.36158/97888929573674

La definizione "intelligenza artificiale" coniata da John McCarthy nel 1956 sembra dunque non adattarsi alle dinamiche della società contemporanea, ma non vi è dubbio che ChatGPT-3 abbia concretizzato ciò che era alla base delle prime ricerche sull'IA, e degli obiettivi esplicitati nel noto articolo di Alan Turing su *Mind* (Turing, 1950), ovvero realizzare programmi in grado di imitare il comportamento umano in modo da rendere indistinguibile il risultato prodotto da una macchina da quello prodotto da un umano. Nell'opinione pubblica, al di fuori dei laboratori scientifici, ciò era rimasto finora a un livello di immaginario, confinato ai mondi distopici della fantascienza, ma oggi con ChatGPT-3 appare reale. Come conseguenza, nella realtà si riversano i timori che hanno riempito le pagine della science fiction, in primis lo scenario in cui la macchina prende il sopravvento sull'uomo. Un immaginario entrato nella creazione letteraria nel XIX secolo, in un'altra fase di grandi trasformazioni industriali e tecnologiche, peraltro sfociate in azioni violente contro le macchine durante il luddismo inglese. L'affinarsi delle capacità performanti delle macchine e soprattutto l'avvicinarsi di queste alle caratteristiche umane, come l'intelligenza, genera opposizione, per esempio il cosiddetto effetto "uncanny valley", teorizzato nel 1970 dallo studioso giapponese di robotica Masahiro Mori, ovvero una reazione di avversione e di non accettazione da parte di alcune persone. Vi assistiamo in relazione a ChatGPT-3, per esempio, con le molte voci nettamente apocalittiche e chiaramente caratterizzate da timori irrazionali, che sfociano negli appelli a fermare lo sviluppo dell'IA, come se ciò fosse possibile, senza minimamente considerare le potenzialità positive dell'intelligenza artificiale, soprattutto se si riuscirà a condurla lungo una linea di crescita democratica e inclusiva. Proprio la fase di crescita sarà, difatti, cruciale. Allo stato attuale i sistemi di IA apprendono come se fossero in uno stadio infantile, per imitazione e rielaborazione dei dati in strutture semplici, ma è ragionevole pensare che in tempi relativamente brevi possano raggiungere l'età adulta, perfezionando le loro capacità. Il grande training a cui l'algoritmo di OpenAI è sottoposto quotidianamente dai milioni di persone che interrogano la chat, migliora progressivamente la sua capacità di fornire risposte corrette. I feedback sono preziosi e difatti ogni risposta fornita da ChatGPT-3 è corredata dalla richiesta di esprimere un giudizio, positivo o negativo, sulla qualità della risposta e di spiegare i motivi per i quali la si ritiene adeguata oppure no, permettendo così a OpenAI di raccogliere informazioni utili per perfezionare l'algoritmo[12].

Dall'evoluzione di ChatGPT-3 è nata infatti in pochi mesi ChatGPT-4, con funzionalità migliorate ma con un passo indietro rispetto alla promessa di apertura e condivisione, in quanto è a pagamento. Il comunicato che a marzo 2023 ne annunciava il rilascio spiega che la chiusura è dovuta al timore di un uso non etico dell'IA e alla concorrenza. Vi sono ormai molte aziende al lavoro su prodotti simili, a cominciare da Anthropic, la società fondata nel 2021 da Dario Amodei, in precedenza vice presidente del settore ricerca di OpenAI, che quando ne è fuoruscito ha portato con sé alcuni esperti, tra cui uno dei creatori di ChatGPT-3. Partner nelle imprese di Anthropic è Google che nel 2022 ha investito 300 milioni di dollari nella start-up. A febbraio 2023 Anthropic ha annunciato il rilascio di Claude, una chatbot molto simile a ChatGPT-3, integrata nella app Poe. Il colosso di Mountain View ha rilasciato anche Google Bard, la risposta a ChatGPT integrata in Chrome. Mentre Bing AI è la nuova versione del motore di ricerca di Microsoft perfezionata con le funzionalità di ChatGPT nel browser MS Edge. Anche la fondazione Mozilla ha annunciato a marzo 2023 la creazione della start-up Mozilla.ai allo lo scopo di

12. Ciò viene esplicitamente dichiarato con un avviso nella chat che spiega: «our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve».

DOI 10.36158/97888929573674

costruire un ecosistema di intelligenza artificiale affidabile e open source, controllata da un agency umana, improntata a principi di accountability, trasparenza e apertura, e usata per essere utile alla comunità. Le potenzialità dei modelli di IA generativa non risiedono solo nelle soluzioni autonome come ChatGPT-3, ma soprattutto nella loro integrazione in piattaforme e strumenti già esistenti che dispongono di grandi quantità di dati, come i motori di ricerca, i repositories come GitHub CoPilot[13], oppure nei sistemi di videoconferenza come Webex[14], nelle piattaforme e content management system come WordPress in cui è possibile generare testi per blog e siti web con AI generativa, e nelle numerose altre applicazioni in cui GPT è già stata introdotta o che stanno lavorando per introdurla.

In quest'ottica di integrazione delle funzionalità, ChatGPT-4 compie offre un'implementazione, unendo l'IA generativa di testi e l'IA generativa di immagini, consentendo di generare sia testo sia immagine da un input testuale o da immagini. I sistemi generativi text-to-image sono un altro grande potenziale dell'IA. OpenAI possiede DALL-E, una rete neurale rilasciata a gennaio 2021 e ora alla versione 2. Anche nell'ambito del text-to-image la concorrenza è alta, con sistemi come Midjourney, Stable Diffusion, Google Imagen, DreamBooth, per citare solo i più noti, e con i primi passi dei sistemi generativi text-to-video, come Gen-2[15]. Le capacità di tali reti neurali sono impressionanti, come si evince dalle immagini di altissima qualità generate, indistinguibili da immagini vere, e peraltro create nell'arco di pochi secondi.

Com'è ampiamente già emerso, i sistemi di diffusione di questo tipo portano con sé il rischio di manipolazione delle immagini in contesti comunicativi di entertainment ma anche scientifici, la cui finalità può essere artistica, creativa e ludica, ma anche ingannevole e fraudolenta. Non ci troviamo di fronte a una novità poiché in ambito giornalistico, artistico e anche scientifico la manipolazione delle immagini, così come la manipolazione dei dati in generale, è un problema che risale all'epoca pre-digitale, ed è tra le principali cause di violazione dell'etica e integrità della ricerca (Morriello, 2022; Stokel-Walker, 2023). Con l'IA le tecniche di falsificazione e manipolazione sono più sofisticate e diventa più difficile distinguere le contraffazioni, ma soprattutto gli strumenti sono alla portata di tutti. Al contrario le applicazioni di intelligenza artificiale che aiutano a individuare le contraffazioni non sono ancora così diffuse. Anche la produzione di contenuti testuali falsi e offensivi è un problema che ha già dato luogo a denunce a OpenAI per diffamazione (Belanger, 2023). Sono certamente tutti aspetti preoccupanti che devono essere regolamentati e non sottovalutati, ma la narrazione mediatica odierna tende a oscurare i numerosi lati positivi dell'intelligenza artificiale. Accanto alle innegabili criticità, gli usi e le potenzialità positive della cosiddetta AI for Good sono innumerevoli. Basti pensare alle applicazioni in ambito medico, soprattutto per i sistemi di generazione di immagini (Liu, Lu, Zhang et al., 2021; Dahmen, Kayaalp, Ollivier et al., 2023) oppure alle numerose potenzialità in ambito educativo, che peraltro l'UNESCO ha elencato in un report specificamente dedicato a ChatGPT (UNESCO, 2023). Pur non tralasciando i numerosi rischi, che vanno dall'integrità della ricerca e della didattica ai bias cognitivi e alle discriminazioni di genere, dall'accessibilità alla concentrazione in mani private e commerciali, l'UNESCO descrive il potenziale positivo di ChatGPT in termini di stimolo e supporto alle attività di apprendimento degli studenti, di ausilio in diverse fasi del ciclo di vita della ricerca scientifica, di facilitatore nel lavoro collaborativo. Inoltre, oggi con l'intelligenza

13. GitHub Copilot https://github.com/features/copilot.
14. Webex ChatGPT https://github.com/features/copilot.
15. Gen-2 https://research.runwayml.com/gen2.

DOI 10.36158/97888929573674

artificiale si progettano le architetture per il metaverso, che non sono solo esercizi di stile ma facilitano l'analisi e la risoluzione di problemi nel mondo reale. La rilevanza è testimoniata dall'interesse dei grandi studi architettonici come Zaha Hadid che ha progettato un'intera città nel metaverso[16].

Un'ulteriore frontiera di sviluppo sulla quale OpenAI è già al lavoro è l'intelligenza artificiale emozionale (Emotional AI)[17], una branca dell'IA che si occupa di rilevare le emozioni attraverso dati raccolti dall'espressione facciale, cui si possono poi aggiungere ulteriori elementi di analisi, quali il movimento facciale e l'inflessione della voce. A differenza della sentiment analysis, che estrae dati sulle emozioni da parole e testi, l'IAE usa come fonti le immagini attingendo a grosse banche dati, come per esempio IMDB-WIKI[18], una raccolta di immagini etichettate e categorizzate già pronte per permettere il training dei modelli di IA. Alla base degli studi sull'intelligenza artificiale emozionale, risiede la teoria delle emozioni universali, comuni a tutti i popoli e a tutte le culture, elaborata dallo psicologo americano Paul Ekman, secondo la quale le microespressioni facciali riconducono ad alcune costanti: rabbia, disprezzo, disgusto, piacere, paura, tristezza, sorpresa[19].

Dalla breve analisi della sua espansione in questa prima parte dell'articolo, si evince come l'aspetto caratterizzante di OpenAI sia il suo modello organizzativo di piattaforma aperta, che sperimenta continuamente. ChatGPT è il luogo in cui attualmente le diverse sperimentazioni trovano una sintesi e dal quale probabilmente evolveranno rapidamente. Tim Berners-Lee, l'inventore del web, nell'intravedere il potenziale di ChatGPT, ha dichiarato che un giorno ognuno di noi avrà un assistente personale come ChatGPT (Kharpal, 2023). Come profetizzata Kemeny settant'anni fa, non possiamo avere contezza di cosa potrà fare l'intelligenza artificiale in futuro, ma di certo possiamo affermare che il punto nodale sarà riuscire a governarla.

Le questioni aperte: autenticità, autorialità, copyright, privacy

Tra le innumerevoli questioni che richiedono riflessione e azione, l'autenticità è di certo centrale, tanto che la stessa OpenAI sta lavorando alla creazione di una sorta di filigrana digitale, che consenta di capire se un oggetto digitale è stato prodotto da un'intelligenza artificiale, e ne ha realizzato una versione demo dotata di un'agevole maschera di ricerca nella quale si può inserire il testo da sottoporre a verifica[20]. In seguito al successo di ChatGPT-3, si stanno sviluppando e diffondendo strumenti per rilevare l'uso dell'IA, come GPTZero[21]. Tuttavia, anche laddove risulti possibile individuare l'uso dell'intelligenza artificiale in un testo scritto o in un'immagine in maniera inequivocabile, è necessario stabilire quale percentuale di "assistenza" di un'intelligenza artificiale possa essere considerata plagio, contraffazione o altra forma di violazione dell'integrità del testo o dell'immagine. Le sollecitazioni emergenti sono diverse. Innanzitutto, occorre chiedersi

16. *Zaha Hadid Architects Reveals Its "Cyber-Urban" Metaverse City*, World Architecture https://worldarchitecture.org/article-links/enhmn/zaha-hadid-architects-reveals-its-cyber-urban-metaverse-city.html.
17. Boost GPT-3 conversational skill with Emotion AI, https://community.openai.com/t/boost-gpt-3-conversational-skill-with-emotion-ai/18771.
18. *IMDB-WIKI – 500k+ face images with age and gender labels*, https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/.
19. Successivamente ampliate dallo stesso autore. Per approfondire cfr. https://www.paulekman.com/universal-emotions/.
20. https://openai-openai-detector.hf.space/.
21. GPTZero https://gptzero.me/ è uno strumento basato sull'intelligenza artificiale creato all'inizio del 2023 da Edward Tian, studente della Princeton University, allo scopo di individuare i testi scritti con ChatGPT-3 o simili attraverso le variabili statistiche della "perplexity" e "burstiness", ovvero misurando la complessità e la varietà delle frasi usate nel testo, partendo dal presupposto che la scrittura umana presenta maggiore complessità e varietà.

DOI 10.36158/97888929573674

come rilevare l'intervento di ChatGPT-3 nelle specifiche parti di un prodotto editoriale, e come trattare il suo utilizzo per migliorare un testo precedentemente scritto da una persona. Inoltre, in che misura lo si dovrà considerare come un tipo di intervento non accettabile? E come comportarsi quando l'uso di ChatGPT viene dichiarato o addirittura l'applicazione di IA viene elencata tra gli autori?

In ambito scientifico, si è discusso molto dell'esperimento dell'editore Springer di creazione di un volume con IA (Lana, 2022), ma nel frattempo le istanze poste dall'uso dell'IA generativa sono divenute molteplici, a cominciare dai numerosi articoli scientifici in cui ChatGPT compare come co-autore (Stokel-Walker, 2023)[22]. Gli editori stanno prendendo posizione a riguardo. Nella maggior parte dei casi tendono ad accettare l'uso di strumenti di IA, ma richiedono di dichiararlo nell'articolo. Tuttavia, non ne riconoscono il ruolo di co-autore. Sono su questa posizione, tra gli altri, Elsevier, Springer, Taylor & Francis, IEEE, JAMA Network[23], nonché COPE, l'associazione britannica di editori scientifici che si occupa di questioni di etica e integrità della ricerca (COPE, 2023). La rivista *Science*, invece, più radicalmente non accetta l'uso dei sistemi di IA in qualsiasi parte del testo, se non concordato con il direttore della rivista, considerandolo come una forma di cattiva condotta scientifica (Holden Thorp, 2023). La scelta di non ammettere l'uso di ChatGPT in un lavoro scientifico rischia però di generare un effetto distorsivo amplificato, soprattutto alla luce della difficoltà oggettiva di rilevare l'uso dell'IA negli articoli e negli altri prodotti della ricerca. All'Università di Chicago è stato condotto un esperimento sottoponendo a verifica 50 abstract di articoli medici creati con ChatGPT partendo da un input di una selezione di abstract veri pubblicati nelle riviste *JAMA*, *The New England Journal of Medicine*, *The BMJ*, *The Lancet and Nature* (Else, 2023). La verifica è stata condotta sottoponendo gli abstract generati con ChatGPT a un sistema anti-plagio, a un AI detector e chiedendo a un gruppo di ricercatori medici di riconoscere gli abstract generati artificialmente. Come prevedibile, le risposte dei ricercatori sono state le meno precise nell'individuare l'uso dell'IA. Si può dunque immaginare la difficoltà di una verifica analoga all'interno di un processo di peer review, nel quale diviene progressivamente più complicato rilevare le manipolazioni e i casi di violazione di etica e integrità della ricerca, anche per l'uso frequente di strumenti informatici sofisticati, per esempio rispetto al numero crescente dei paper mills (Morriello, 2022)[24]. Se da un lato, infatti, l'uso di applicazioni di intelligenza artificiale può aiutare i processi di revisione tra pari, come sottolinea l'UNESCO nel report sopra citato (UNESCO, 2023), è pur vero che le questioni già di per sé complesse dell'etica e integrità della ricerca e delle criticità odierne della peer review (Capaccioni, Guerrini, Morriello, 2023) si complicano ulteriormente con l'introduzione di strumenti come ChatGPT nella scrittura degli articoli (Stokel-Walker, 2023). La scelta dei maggiori editori di tentare di normalizzare, e in questo modo regolamentare, quella che diventerà con probabilità una pratica comune appare la più lungimirante. Difatti, anche gli enti che elaborano i principali stili citazionali si stanno adeguando al nuovo contesto e hanno pubblicato delle linee guida su come citare ChatGPT secondo i diversi stili (MLA, 2023; The Chicago Manual of Style, 2023; APA, McAdoo, 2023).

Al contrario, al di fuori delle reti scientifiche, qualsiasi forma di controllo risulta più complessa e la consapevolezza appare minore. Su Amazon sono già in vendita centinaia di libri in formato kindle, e in alcuni casi anche, o esclusivamente, in formato cartaceo a stampa, nei quali ChatGPT-3 compare come autore o co-autore, e che co-

22. Cfr. come esempio Kung, Cheatham, ChatGPT et al., 2022.
23. Elsevier 2023; Taylor & Francis 2023; IEEE 2023; Flanagin, Bibbins-Domingo, Berkwits, Christiansen, 2023.
24. E senza voler aprire in questa sede la questione delle implicazioni per la valutazione della ricerca.

DOI 10.36158/97888929573674

prono diversi generi, dalle storie per bambini[25], alla narrativa per adulti[26], dalla poesia[27] alla manualistica[28], inclusa la guida all'uso di ChatGPT-3 scritta da ChatGPT-3, che vi compare come unico autore e che viene descritta come «ChatGPT's Guide to Utilizing ChatGPT, written and Taught by ChatGPT»[29]. Tra gli esempi citati in nota (tratti da Amazon.it), è emblematico il volume *Lazy Keto Cookery*, una raccolta di ricette scritte da ChatGPT-3 con la revisione dell'autore umano (o almeno presunto tale), nel quale in quarta di copertina compare una breve nota biografica di ChatGPT-3 accanto a quella all'altro autore. Sono pubblicazioni indipendenti ovviamente, per la maggior parte edite in self-publishing, ma dotate di ISBN. Solo in alcuni casi compare il nome di un editore, del quale però non si trova traccia in altre sedi. D'altronde, ChatGPT-3 potrebbe anche attribuirsi un editore nel processo generativo. Sulla piattaforma di vendita online mancano però alcune informazioni bibliografiche fondamentali, come la data di pubblicazione. Dal punto di vista bibliografico la comparsa di libri con ChatGPT-3 come autore comporta la necessità di porre attenzione e ridefinire le modalità e gli strumenti di organizzazione dell'informazione per far fronte a tali nuove istanze e, in ambito bibliotecario, di farvi fronte tramite un ripensamento di alcune attività, quali la catalogazione, peraltro secondo linee di sviluppo già in corso nel solco della metadatazione (Guerrini, 2020; 2022). La questione dell'autenticità si collega peraltro all'autorialità, e alle responsabilità correlate, nonché a come gestire dal punto di vista bibliografico l'agency dell'intelligenza artificiale (Lana, 2022).

Un altro grande ambito di riflessione relativo all'IA riguarda il copyright, nella duplice connotazione di protezione dei dati di input e di definizione dei diritti per gli output. Non sempre i dati immessi nei sistemi di modellazione sono privi di copyright o rilasciati con licenze Creative Commons che ne consentano il riuso. Stable Diffusion è un sistema aperto di deep learning generativo da testo a immagini, rilasciato nel 2022 come risultato di una collaborazione tra Stability AI, CompVis LMU, Runway con EleutherAI e LAION, allenato su un dataset pubblico di LAION derivato da Common Crawl e basato su 890 milioni di parametri e 160 milioni di immagini (Carlini, 2023). Common Crawl è un'organizzazione non profit che scandaglia il web per creare dataset e archivi che poi rende disponibili pubblicamente e gratuitamente. L'archivio di cinque miliardi di oggetti digitali, nell'ordine di petabyte di dati, è ora ospitato dal sistema di archiviazione S3 di Amazon. Qualsiasi azienda si può appoggiare ai dati di Common Crawl. A differenza di OpenAI, il cui training set non è noto (probabilmente anche per evitare problemi legati al copyright), il dataset usato da Stability AI è open access. Naturalmente è tutt'altro che facile da consultare per un comune cittadino ma qualche esperto lo ha fatto e, con l'ausilio delle stesse tecniche di IA, ha individuato la provenienza delle immagini (Baio, 2022), provocando la reazione degli artisti che ne sono gli autori e degli enti che ne detengono i diritti, come Getty Image che ha intentato una causa contro Stability AI per aver copiato ed elaborato milioni di immagini protette da copyright, senza chiedere l'autorizzazione né dichiarare in alcun modo la opere originali usate. Ma l'argomento è più complesso di

25. Per esempio, *Ava the Ballerina Alligator: A Silly and Inspiring Tale of Dreams and Determination,* di Bill Hackett (autore), Assistant – ChatGPT-3 OpenAI (autore), Iram Adnan (illustratore).
26. Per esempio, *Un viaje a través de los hongos mágicos: Conversaciones con Don Antonio, Ser de Luz (Spanish Edition),* di ChatGPT-3 OPENIA (autore), Jesús Morán Morán (autore), Gaby OPENIA (prefatore).
27. Come *Echoes of the Universe* (english edition) Formato Kindle, edizione inglese di Dawson Hunt (autore), ChatGPT AI (autore), Stable Diffusion AI (illustratore).
28. Per esempio "*Lazy Keto Cookery: 60 Beginner-Friendly Recipes by ChatGPT-3, Prompts by Rick Major": 60 Easy Recipes for Beginners on the Lazy Keto Diet,* di Chat GPT (autore), Rick Major (autore), disponibile solo a stampa in brossura a 49,16 euro e rilegato a 54,08 euro.
29. *Unleashing the Power of ChatGPT: 10x Your Business in 2023 with Cutting Edge Technology AI Technology*, Formato Kindle, edizione inglese, autore ChatGPT by OpenAI.

DOI 10.36158/97888929573674

quanto appaia. Innanzitutto perché i sistemi generativi usano perlopiù immagini pubblicamente disponibili sul web e, sebbene si possa trattare di un uso diverso da quello per il quale i dati sono stati resi pubblici e dunque a rischio di violazione dei diritti, nei fatti bisognerebbe capire quale parte di questi dati viene usata dai sistemi di IA e in che modo, in quanto si tratta di milioni di immagini e petabyte di dati, elaborate secondo milioni di parametri. Definire con precisione quali e quanti dati vengono usati rispetto a uno specifico output è nella maggior parte dei casi impossibile, vista l'opacità degli algoritmi, anche nei confronti degli stessi programmatori che li hanno creati.

L'altro aspetto della questione del copyright riguarda il prodotto dell'IA generativa. Quello che i sistemi di AI producono non è semplicemente una ricombinazione degli elementi memorizzati, c'è qualcosa di più, anche se non sempre è particolarmente creativo o innovativo, ma di certo generano qualcosa di nuovo. Una poesia nello stile di Philip Larkin non è una poesia di Philip Larkin così come un quadro nello stile di Cézanne non è un quadro di Cézanne. L'IA generativa crea un nuovo spazio nella dimensione dei diritti che dovrà essere codificato. Per esempio, ha occupato le pagine di alcuni giornali la notizia della creazione del manga *Cyberpunk: Peach John* con l'uso di Midjourney come illustratore (Holland, 2023). L'opera ha sollevato in primo luogo la questione del copyright dei dati immessi nel sistema generativo ma, ipotizzando che non vi fosse violazione del copyright in questo senso, occorre chiedersi chi sia il detentore del copyright del manga. Una domanda alla quale la legislazione vigente non è pronta a rispondere, e non solo in Italia.

Nel mese di febbraio 2023, lo United States Copyright Office (USCO) si è espresso sulla tutela della proprietà intellettuale della graphic novel *Zarya of the Dawn*, creata dall'artista Kristina Kashtanova con l'ausilio di Midjourney per le immagini. Kashtanova chiedeva l'applicazione delle leggi sul copyright riconoscendosi come detentrice dei diritti sull'opera nel suo insieme, ma l'istanza è stata respinta dall'USCO che non le ha riconosciuto i diritti sull'intera opera e ha limitato la tutela al testo redatto dell'autrice e alla sua selezione e combinazione di testo e immagini. L'USCO ha rilevato come l'intervento creativo umano in un sistema di AI generativa non sia sufficiente per dar luogo alla protezione del copyright dell'intera opera. Nonostante Kashtanova abbia spiegato come i risultati ottenuti con Midjourney siano stati da lei modificati e adattati, l'USCO ha ritenuto che questi non fossero elementi sufficienti per poter considerare il suo operato come "autorialità tradizionale" delle immagini. Secondo l'ente statunitense, i risultati ottenuti tramite la piattaforma di AI pur essendo conseguenti a prompt testuali immessi dalla donna vengono generati sulla base di un'elaborazione algoritmica il cui risultato non è prevedibile. Nella sua lettera l'USCO effettua una comparazione con la fotografia e con il lavoro del fotografo, rilevando come rispetto all'opera fotografica il lavoro di Kashtanova non presenti un grado sufficiente di creatività (Wolfson, 2023). L'istituzione statunitense aggiunge, inoltre, che nei casi in cui vi fosse un livello maggiore di creatività, tale da permettere all'autore di governare i contenuti dell'IA, allora si potrebbe considerare la tutela complessiva dell'opera. La sentenza tocca un punto cruciale, ovvero come definire il livello di creatività in un'opera generata totalmente o parzialmente con IA e soprattutto come distinguere l'apporto e l'entità di creatività umana dal lavoro generativo dell'IA. Per poterlo fare occorre, come minimo, un altro sistema di intelligenza artificiale, ma finora i sistemi di IA detection non sono così sofisticati da riuscire ad arrivare a tale livello di dettaglio, riuscendo a stabilire al massimo la quota percentuale di intervento dell'IA e non sempre con precisione. Un altro aspetto rilevante portato alla luce dal caso di *Zarya of the Dawn* è il riferimento delle legislazioni nazionali a concetti come quello espresso

DOI 10.36158/97888929573674

da USCO di "autorialità tradizionale", intesa principalmente come un'attività umana, in quanto le leggi sono nate quando questa era l'unica opzione possibile. Le disposizioni legislative appaiono non più adeguate al contesto attuale, e ciò accadeva anche prima della diffusione dell'IA. Il concetto di autore nel mondo digitale, in particolare se non umano, non viene chiarito dalle norme, con l'unica eccezione di un piccola apertura nella direttiva 2001/29/CE (c.d. direttiva Infosoc o copyright) che parla di «autore dell'opera o qualunque altro titolare di diritti» (Lavagnini, 2022).

Peraltro, in assenza di riferimenti legislativi, è evidente come la situazione sia al momento di grande confusione. Da un lato abbiamo gli editori scientifici che chiedono di dichiarare l'uso di sistemi di IA negli articoli ma non mettono in dubbio la proprietà intellettuale dell'autore umano (mentre il copyright è di solito in mano agli editori stessi), cui viene riconosciuta l'unica autorialità possibile, dall'altro invece vi sono sentenze come quella di USCO che non riconoscono "l'autorialità tradizionale" di un'opera elaborata con IA all'autrice umana. Al contempo, sono già in vendita su Amazon centinaia di libri in cui ChatGPT-3 viene dichiarato autore o co-autore, Midjourney come illustratore, e così via, come negli esempi sopra citati. A chi va attribuito dunque il copyright di queste pubblicazioni?

Il ritardo legislativo nel regolamentare il nuovo contesto digitale, ancor prima dell'intelligenza artificiale, è chiaramente un ritardo di consapevolezza da parte delle istituzioni pubbliche governative in generale, che non favorisce una corretta metabolizzazione dei sistemi di IA nella società. Le sollecitazioni per una revisione delle leggi arrivano da tempo da diverse fonti. L'UNESCO nel 2022 ha emanato le *Raccomandazioni sull'etica dell'intelligenza artificiale*, chiedendo ai governi di implementarle in tempi brevi (UNESCO, 2022). La Comunità europea sta lavorando a un regolamento che procede a rilento anche per la difficoltà dei rappresentanti politici di giungere a un accordo su alcune parti (Martorana & Savella, 2023). Di fronte alla rapida diffusione di strumenti come ChatGPT-3 la regolamentazione è un'esigenza fondamentale e non procrastinabile. Non si tratta semplicemente di riformare le leggi ma anche di condurre una riflessione profonda su quali aspetti debbano prevalere, se il diritto d'autore di un singolo, oppure la possibilità offerta a tutti di accedere ai contenuti generati dai sistemi di IA seguendo una linea di pensiero avviata dal movimento open access, se si possa mantenere il concetto di autore e editore "tradizionale" in un'epoca in cui entrambi questi ruoli sono in trasformazione oppure si debba adeguarli ai nuovi contesti. Oltre al copyright, vi sono poi le questioni legate alla privacy dei dati e delle immagini utilizzate (Carlini et al., 2023). In assenza di una legislazione adeguata e di una politica attenta agli sviluppi tecnologici, è comprensibile che le istituzioni possano avere difficoltà nella gestione di tecnologie fortemente impattanti, con il rischio di giungere a soluzioni estreme come la chiusura di ChatGPT-3 in Italia in seguito ai rilievi fatti dal Garante della Privacy. Nel mese di aprile 2023, il Garante per Protezione dei dati personali ha chiesto a OpenAI chiarimenti sulla conformità al GDPR (General Data Protection Regulation), in particolare in relazione al rispetto della privacy dei dati usati per il modello ChatGPT-3 e alla tutela dei dati dei minori. Il Garante ha contestato l'assenza di un'informativa per gli utenti e, nonostante OpenAI dichiari che il servizio è rivolto a chi ha più di 13 anni, la mancanza di filtri in grado di far rispettare tale limitazione. Il Garante ha chiesto la documentazione e garanzie in tal senso. OpenAI ha chiuso temporaneamente l'accesso a ChatGPT-3 in Italia[30] La narrazione mediatica,

30. I vari comunicati sulla vicenda sono disponibili sul sito del Garante https://garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847.

DOI 10.36158/97888929573674

peraltro, ha posto l'Italia tra i paesi che hanno bannato ChatGPT-3, fraintendendo l'azione del Garante. OpenAI ha risposto ai rilievi del Garante adeguandosi e confermando dunque che regolamentare l'IA e impartire una determinata direzione è possibile.

Una governance pubblica e per il bene comune dell'intelligenza artificiale è realizzabile e auspicabile perché l'alternativa è lasciarla in mano ai privati e all'uso non regolamentato. Permane il rischio anche per l'IA di assistere alla concentrazione dello sviluppo e dei prodotti nelle mani di pochi colossi privati, proprio com'è accaduto con le aziende GAFAM per il web. Un'ipotesi che riprodurrebbe le attuali criticità di un ambiente digitale globale colonizzato da operatori e produttori di risorse in cui predomina un mercato oligopolistico, certi modelli di business, contenuti in lingua inglese e un approccio che tende a schiacciare le diversità. I dati utilizzati dall'IA sono al momento quelli del web, e dunque risentono dei limiti e bias già presenti nel web, senza nessuna forma di verifica rispetto alla produzione di contenuti falsi e fuorvianti. Appare ormai chiaro come la tecnologia sia anche una questione politica e di potere, di sviluppo sostenibile e di giustizia sociale, e questo dovrebbe spingere i governi a occuparsene maggiormente, se non altro in ottica strategica. In Cina ChatGPT-3 è stata vietata, in quanto "strumento di propaganda americana" ma i colossi tech cinesi come Alibaba, la multinazionale alle spalle del principale motore di ricerca Baidu, e Tencent, creatrice dell'app di messaggistica WeChat, stanno lavorando allo sviluppo di sistemi analoghi a quello di OpenAI[31].

Un'intelligenza artificiale indirizzata al bene comune dovrebbe configurarsi non certo come un oligopolio ma come un sistema diffuso e aperto, trasparente e etico, a cominciare dai dati immessi per svolgere i training. Le questioni da risolvere e gli aspetti da regolamentare sono molti, i rischi da non sottovalutare. Solo attribuendo la dovuta priorità a questi temi, con la conseguente ridefinizione dell'agenda pubblica, si potrà tentare di stare al passo con le evoluzioni tecnologiche e riuscire a governarle. Qualcuno ha cominciato a farlo, altri, come l'Italia, tardano. Il Regno Unito ha varato nel 2021 una strategia nazionale per l'intelligenza artificiale (UK Government, 2021), seguita nei mesi successivi al rilascio di ChatGPT-3 di un White Paper (UK Government, 2023). Un analogo documento strategico sull'intelligenza artificiale è stato varato dal governo tedesco nel 2018 (German Federal Government, 2018). I due piani strategici definiscono con precisione la posizione dei governi rispetto ai temi cruciali quali l'uso etico e inclusivo dell'IA, l'uso dei dati nel rispetto dei diritti e della privacy, la proposta di standard, con l'intento di promuovere la leadership del paese su questi aspetti dell'innovazione tecnologica ma sostenendo la necessità di giungere a un giusto equilibrio tra le istanze di sviluppo e il rispetto dei principi dell'IA rivolta al bene comune. A differenza dei due esempi citati, la strategia nazionale per l'intelligenza artificiale lanciata dall'Italia nel 2021 (Governo Italiano, 2021) ha un'impostazione completamente diversa in quanto si tratta solo una raccolta di dati sullo stato della ricerca sull'IA in Italia, e il Centro nazionale per l'intelligenza artificiale fatica a mettersi in moto.

La concentrazione nella mani dei privati dell'IA e l'assenza di una linea di indirizzo pubblica e democratica vedrebbe concretizzarsi i rischi, come la perdita di posti di lavoro, l'aumento delle diseguaglianze, una generalizzata ingiustizia sociale aggravata dall'IA. Per poter portare l'intelligenza artificiale verso un'altra direzione, sfruttando il potenziale positivo e di arricchimento che offre, al contrario, in termini di creazione di nuovi posti di lavoro, di possibilità di colmare in parte le diseguaglianze, di garantire un futuro

31. La Cina vieta ChatGPT: è uno strumento di propaganda americana, https://forbes.it/2023/02/22/cina-ChatGPT-bloccato-intelligenza-artificiale-propaganda-americana/.

DOI 10.36158/97888929573674

sostenibile alle società attraverso il contributo sulle grandi sfide come il cambiamento climatico, sono necessarie politiche mirate e competenze adeguate nei vari settori. Su questo aspetto andrebbero concentrate le maggiori risorse. La mancanza di conoscenza e competenze sull'intelligenza artificiale è uno dei fattori di rallentamento rispetto alla capacità di governarla e di farne un uso positivo nei diversi ambiti di applicazione. Quella viene ormai definita "algorhitmic literacy" dovrebbe essere insegnata nelle scuole e nelle università. Le biblioteche vi potrebbero contribuire poiché sono in grado di raggiungere il pubblico che è al di fuori dei percorsi formativi cui potrebbero offrire corsi di alfabetizzazione algoritmica, a condizione che i bibliotecari siano formati per farlo. Ma questo discorso aprirebbe un nuovo tema che lascio a un approfondimento futuro.

Bibliografia

Bassett, C. (2019). The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present. *AI & Society*, *34*, 803-812. https://doi.org/10.1007/s00146-018-0825-9.

Baio, A. (2022). *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator*. *Waxy.org*. https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/.

Belanger, A. (2023). OpenAI threatened with landmark defamation lawsuit over ChatGPT false claims. *Ars Technica*. https://arstechnica.com/tech-policy/2023/04/openai-may-be-sued-after-ChatGPTfalsely-says-aussie-mayor-is-an-ex-con/.

Capaccioni, A., Guerrini, M., Morriello, R. (a cura di) (2023). *La peer review: un processo in necessaria trasformazione*. Numero speciale di *JLIS.It*, *14*/1. https://www.jlis.it/index.php/jlis/issue/view/37.

Carlini, N. et al. (2023). Extracting Training Data from Diffusion Models. Versione depositata in *arXiv* il 30 gennaio. https://doi.org/10.48550/arXiv.2301.13188.

Cianciolo, A.T., Sternberg, R.J. (2004). *Breve storia dell'intelligenza*. Il Mulino.

COPE (Committee on Publication Ethics) (2023). Authorship and AI Tools. Position Statement, 13 February 2023. https://publicationethics.org/cope-position-statements/ai-author.

Cornoldi, C. (2007). *L'intelligenza*. Il Mulino.

Dal Dosso, S. (2023). Impossible architectures between digital worlds and artificial intelligence. *Domus*. https://www.domusweb.it/en/architecture/gallery/2023/03/15/impossible-architectures-digital-worlds-and-artificial-intelligence.html.

Dahmen, J., Kayaalp, M.E., Ollivier, M. et al. (2023). Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. *Knee Surg Sports Traumatol Arthrosc*, *31*, 1187-1189. https://doi.org/10.1007/s00167-023-07355-6.

Else, H. (2023). Abstracts Written by ChatGPT fool Scientists. *Nature*, *613*, 423.

Elsevier (2023), *The use of AI and AI-assisted writing technologies in scientific writing*. https://www.elsevier.com/about/policies/publishing-ethics/the-use-of-ai-and-ai-assisted-writing-technologies-in-scientific-writing.

Esposito, E. (2021). Dall'intelligenza artificiale alla comunicazione artificiale. *Aut aut*, *392*, 20-34.

Flanagin A., Bibbins-Domingo K., Berkwits M., Christiansen S.L. (2023). Nonhuman "Authors" and Implications for the Integrity of Scientific Publication and Medical Knowledge. *JAMA*, *329*(8), 637-639. https://doi.org/10.1001/jama.2023.1344.

DOI 10.36158/97888929573674

Floridi, L. (2023). AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models (February 14), preprint disponibile su SSRN. https://ssrn.com/abstract=4358789 or http://dx.doi.org/10.2139/ssrn.4358789. Articolo pubblicato su *Philosophy and Technology*, 2023. https://doi.org/10.1007/s13347-023-00621-y.

Floridi, L. (2021). Intelligenza artificiale: il divorzio tra azione e intelligenza. *Aut aut*, *392*. 35-50.

German Federal Government. (2018). National AI Strategy. https://www.ki-strategie-deutschland.de/home.html.

Governo Italiano. (2021). Programma Strategico Intelligenza Artificiale 2022-2024. https://assets.innovazione.gov.it/1637777289-programma-strategico-iaweb.pdf.

Guerrini, M. (2022). *Metadatazione: la catalogazione in era digitale*. Editrice Bibliografica.

Guerrini, M. (2020). *Dalla catalogazione alla metadatazione: tracce di un percorso*. Editrice Bibliografica.

Hao, K. (2020). A new way to train AI systems could keep them safer from hackers. *MIT Technology Review*, *10*. https://www.technologyreview.com/2020/07/10/1005048/ai-deep-learning-safe-from-hackers-adversarial-attacks/.

Holden Thorp, H. (2023). ChatGPT is fun, but not an author. *Science*, 379(663), 313. https://www.science.org/doi/10.1126/science.adg7879.

Holland, O. (2023). This is Japan's first AI-generated manga comic. But is it art?, *CNN*. https://edition.cnn.com/style/article/japan-first-ai-generated-manga-art-intl-hnk/index.html.

Hong, S. (2004). Man and Machine in the 1960s. *Techné*, *7*(3), 50-78.

Hunter, W. (2023). What Poets Know That ChatGPT Doesn't. *The Atlantic*. https://www.theatlantic.com/books/archive/2023/02/ChatGPTai-technology-writing-poetry/673035/.

IEEE (2023). Submission and Peer Review Policies. https://journals.ieeeauthorcenter.ieee.org/become-an-ieee-journal-author/publishing-ethics/guidelines-and-policies/submission-and-peer-review-policies/.

Kantosalo, A., Falk, M. & Jordanous, A. (2021). Embodiment in 18th Century Depictions of Human-Machine Co-Creativity. *Frontiers in Robotics and AI*, 8(662036). 10.3389/frobt.2021.662036.

Kemeny, J.G. (1955). Man Viewed as a Machine. *Scientific American*, *192*(4), 58-67.

Kharpal, A. (2023). The inventor of the web thinks everyone will have their own personal A.I. assistants like ChatGPT. *CNBC*. https://www.cnbc.com/2023/02/17/tim-berners-lee-thinks-we-will-have-our-own-ai-assistants-like-chatgpt.html.

Kung, H.T., Cheatham, M., ChatGPT et al. (2022). Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models, posted in MedrXiv 21 December 2022. https://doi.org/10.1101/2022.12.19.22283643. Now published in *PLOS Digital Health*. https://doi.org/10.1371/journal.pdig.0000198.

Lana, M. (2022). L'agency dei sistemi di intelligenza artificiale. Un punto di vista bibliografico. *DigitCult*, 7(1). https://doi.org/10.36158/97888929552576.

Lavagnini, S. (2022). Un'opera creata dall'IA può essere protetta da diritto d'autore? La Corte di Washington verso la decisione. *Agenda Digitale*. https://www.agendadigitale.eu/cultura-digitale/unopera-creata-dallia-puo-essere-protetta-da-diritto-dautore-la-corte-di-washington-verso-la-decisione/.

Liu, Pr., Lu, L., Zhang, Jy. et al. (2021). Application of Artificial Intelligence in Medicine: An Overview. *CURR MED SCI*, 41, 1105-1115. https://doi.org/10.1007/s11596-021-2474-3.

DOI 10.36158/97888929573674

Martorana, M. & Savella, R. (2023). Intelligenza artificiale: orientamento del Consiglio europeo e ultimi sviluppi nella definizione del Regolamento. *Altalex*. https://www.altalex.com/documents/news/2023/02/15/intelligenza-artificiale-orientamento-consiglio-europeo-ultimi-sviluppi-definizione-regolamento.

APA, McAdoo, T. (2023). How to cite ChatGPT, American Psychological Association. https://apastyle.apa.org/blog/how-to-cite-chatgpt.

McQuillan, D. (2018). Manifesto on algorithmic humanitarianism. *OpenDemocracy*. https://www.opendemocracy.net/en/manifesto-on-algorithmic-humanitarianism/.

Metz, C. & Weise, K. (2023). Microsoft to Invest $10 Billion in OpenAI, the Creator of ChatGPT. *The New York Times*. https://www.nytimes.com/2023/01/23/business/microsoft-ChatGPTartificial-intelligence.html.

Modern Language Association of America (MLA) (2023). How do I cite generative AI in MLA style? https://style.mla.org/citing-generative-ai/.

Morriello, R. (2022). *Dalla pirateria dei libri all'editoria predatoria. Un percorso tra storia della stampa ed etica della comunicazione scientifica*. Ledizioni. Open access su Zenodo https://zenodo.org/record/7614728.

OpenAI. (2022). *Introducing ChatGPT*, **https://openai.com/blog/chatgpt**.

Perrigo, B. (2023). OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic, *Time*. https://time.com/6247678/openai-ChatGPTkenya-workers/.

Price, D.J. de Solla, (1964). Automata and the Origins of Mechanism and Mechanistic Philosophy, *Technology and Culture*, *4*(1), 9-23.

Price, D.J. de Solla (1959). An Ancient Greek Computer. *Scientific American 200*(6), 60-67.

Rich, S. (2022). The New Poem-Making Machinery. *The New Yorker*. https://www.newyorker.com/culture/culture-desk/the-new-poem-making-machinery.

Stokel-Walker, C. (2023). ChatGPT listed as Author on Research Papers. *Nature*, *613*(7945), 620-621. https://doi.org/10.1038/d41586-023-00107-z.

Sutton, R.S. & Barto, A.G. (2014). Reinforcement Learning: An Introduction. The MIT Press. https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf.

Taylor & Francis (2023). Taylor & Francis Clarifies the Responsible use of AI Tools in Academic Content Creation. https://newsroom.taylorandfrancisgroup.com/taylor-francis-clarifies-the-responsible-use-of-ai-tools-in-academic-content-creation/.

Turing, A.M. (1950). Computing machinery and intelligence, *Mind*, *49*(236), 433-460. https://academic.oup.com/mind/article/LIX/236/433/986238.

UK Government. (2021). National AI Strategy. Last updated 18 December 2022. https://www.gov.uk/government/publications/national-ai-strategy.

UK Government. (2023). Policy paper. AI regulation: a pro-innovation approach. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach.

UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137.

UNESCO. (2023). ChatGPT and Artificial Intelligence in Higher Education. https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf.

University of Chicago (2023). The Chicago Manual of Style. Citation, Documentation of Sources. https://www.chicagomanualofstyle.org/qanda/data/faq/topics/Documentation/faq0422.html.

Vincent, J. (2023). Elon Musk reportedly tried and failed to take over OpenAI in 2018. *The Verge*. https://www.theverge.com/2023/3/24/23654701/openai-elon-musk-failed-takeover-report-closed-open-source.

DOI 10.36158/97888929573674

Vincent, J. (2023). ChatGPT can't be credited as an author, says world's largest academic publisher, *The Verge*. https://www.theverge.com/2023/1/26/23570967/ChatGPT-author-scientific-papers-springer-nature-ban.

Weizenbaum, J. (1966). ELIZA. A Computer Program For the Study of Natural Language Communication Between Man and Machine. *Communications of the ACM*, *9*(1), 36-45. https://doi.org/10.1145/365153.365168.

Wolfson, W. (2023). Zarya of the Dawn: US Copyright Office Affirms Limits on Copyright of AI Outputs. *Creative Commons Blog*. https://creativecommons.org/2023/02/27/zarya-of-the-dawn-us-copyright-office-affirms-limits-on-copyright-of-ai-outputs/.

DOI 10.36158/97888929573674