

L'INTELLIGENZA ARTIFICIALE ED IL PROBLEMA DELL'AGENCY DAL PUNTO DI VISTA BIBLIOGRAFICO

Maurizio Lana*

Il 30 novembre 2022 OpenAI, l'azienda che sviluppa il sistema di Intelligenza Artificiale GPT (Generative Pretrained Transformer, giunto alla versione 4 alla data di scrittura di questo contributo), annunciava la disponibilità al pubblico dell'interfaccia ormai notissima denominata ChatGPT. In pochi mesi l'Intelligenza Artificiale da argomento esoterico per informatici è diventata oggetto di interesse in ambiti diversissimi e quotidiani, in cui l'interesse verso i sistemi di Intelligenza Artificiale è di tipo applicativo e non solo di studio o di sviluppo. La diffusione di interesse ha preso due direzioni principali: possibili utilizzi di ChatGPT in ambito aziendale per ricevere o fornire servizi; e (anche in relazione ad essi) pericoli insiti nell'Intelligenza Artificiale¹ nella forma (Chat)GPT². Ad esse se ne aggiunge una terza, sottotraccia, non meno importante: «quali cambiamenti porta un sistema di Intelligenza Artificiale come GPT quando viene utilizzato nella produzione di contenuti intellettuali (testuali, musicali, visivi, ...)»?». Già oggi contenuti che si collocano nell'ambito della creatività intellettuale vengono prodotti con l'impiego di sistemi di Intelligenza Artificiale³; e certamente i sistemi di Intelligenza Artificiale

* Università del Piemonte Orientale.

¹ Dalla minaccia nella quotidianità per la scomparsa di molti lavori, alla minaccia esistenziale per il genere umano.

² Scrivo (Chat)GPT perché il sistema di Intelligenza Artificiale non è ChatGPT ma GPT; come si è già ricordato ChatGPT è solo un'interfaccia che facilita l'uso di GPT: chi vuole può usare direttamente GPT; e si potrebbero sviluppare (anche) altre interfacce per l'uso di GPT; oppure delle 'sottoversioni' di ChatGPT focalizzate su ambiti specifici come Botsonic, di Writesonic (<https://writesonic.com/Chat>) o per specifiche piattaforme come IOS o Android; o alte interfacce verso altri sistemi di Intelligenza Artificiale come StableVicuna di Stability AI (<https://stability.ai/blog/stablevicuna-open-source-rlhf-chatbot>).

³ Ad esempio a proposito di Spotify da un lato si sospetta che la musica di «The Groove Trio» sia prodotta con sistemi di Intelligenza Artificiale dalla piattaforma stessa; cfr. <https://www.reddit.com/r/Jazz/comments/10px4gz/is_spotify_using_ai_to_generate_fake_jazz_artists/>; dall'altro Spotify cancella centinaia di brani che vi

entreranno *routine* nella produzione di contenuti.

Si potrebbe pensare che la questione dell'agency dei sistemi di Intelligenza Artificiale dal punto di vista bibliografico, che si iscrive in questa terza direzione, sia un caso di studio puramente teorico. Non credo che sia così, perché la produzione di contenuti testuali da parte di sistemi di Intelligenza Artificiale da un lato costituisce l'evoluzione di strumenti del lavoro intellettuale in uso quotidianamente; dall'altro questa produzione sta già avvenendo, da tempo.

Molti, soprattutto informatici ma non solo, sottolineano che i sistemi di Intelligenza Artificiale vengono oggi sviluppati, testati e utilizzati in un contesto sostanzialmente privo di regole, norme, leggi. È un problema non nuovo, che riguarda lo sviluppo di sistemi digitali che si basano sull'interazione con le persone e che da acquisiscono dati di ogni genere; la società civile, e gli Stati, non possono accettare che sistemi di tale potenza, di tale capacità di incidere sul funzionamento della società in tutti i suoi aspetti, operino al di fuori di qualsiasi regolamentazione⁴ definita nella prospettiva dell'interesse pubblico e non dei suoi sviluppatori e ci sono indicazioni forti della consapevolezza che occorrono norme e leggi.⁵

L'Intelligenza Artificiale come evoluzione di strumenti del lavoro intellettuale

È ovvio che leggere degli sviluppi di sistemi di Intelligenza Artificiale capaci di operare nella scrittura di testi, e di testi di prospettiva scientifica, possa suscitare (sollecitare?) perplessità e preoccupazione, perché lo studioso sente la scrittura scientifica come un'attività eminentemente personale in cui le sue capacità sono messe alla prova in una forma limite e, attraverso la scrittura, vengono valutate in una varietà di forme e contesti. E dunque pensare che possano intervenire dei sistemi di Intelligenza Artificiale 'in aiuto' può apparire

erano stati caricati, perché prodotti con il sistema di Intelligenza Artificiale Boomy (Nicolaou, 2023).

⁴ Sul tema si veda (Giraud, 2022): «'Legal bubbles' tend to arise when economic agents invest in the economic exploitation of a new resource in a context of uncertainty and ignorance about the legal implications of innovative activities. These legal foundations may eventually turn out to be unstable as a result of courts' ex-post attempt to re-adapt them to the previously ignored implications of unconstrained commodification, with disruptive economic consequences.» (p. 587).

⁵ Si veda ad esempio, per l'Europa, il percorso legislativo della proposta di legge europea (regulation) nota come «AI Act» (European Parliament, 2021) e della proposta di direttiva nota come «AI Liability Directive» (European Commission, 2022) (*regulations* sono leggi che una volta approvate dal Parlamento europeo diventano operative e vincolanti in tutti paesi membri; *directives* sono indicazioni normative che devono essere recepite dalle legislazioni nazionali entro un tempo dato).».

deontologicamente scorretto, o sminuente, tanto più che si può andare da forme elementari di intervento di sistemi di Intelligenza Artificiale già disponibili, fino a forme più complesse e 'invasive' che sono pronte o in arrivo. Il punto è che già oggi la maggior parte degli studiosi utilizza sistemi di potenziamento delle sue capacità intellettuali. Ne parlò per primo Douglas Engelbart negli anni '60 del secolo scorso utilizzando l'espressione *augmenting human intellect* nel titolo di due documenti progettuali: «Augmenting human intellect: a conceptual framework», del 1962⁶, e «A research center for augmenting human intellect», del 1968⁷. L'idea chiave era liberare le persone dai compiti esecutivi/ripetitivi, da demandare ai computer, creando un contesto che Engelbart descriveva nel 1962 con queste parole:

By “augmenting human intellect” we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems. Increased capability in this respect is taken to mean a mixture of the following: more-rapid comprehension, better comprehension, the possibility of gaining a useful degree of comprehension in a situation that previously was too complex, speedier solutions, better solutions, and the possibility of finding solutions to problems that before seemed insoluble. And by “complex situations” we include the professional problems of diplomats, executives, social scientists, life scientists, physical scientists, attorneys, designers – whether the problem situation exists for twenty minutes or twenty years. We do not speak of isolated clever tricks that help in particular situations. We refer to a way of life in an integrated domain where hunches, cut-and-try, intangibles, and the human “feel for a situation” usefully co-exist with powerful concepts, streamlined terminology and notation, sophisticated methods, and high-powered electronic aids.⁸

Dallo sviluppo di questa visione del 1962, ma che potrebbe benissimo essere riferita oggi all'impiego di sistemi di Intelligenza Artificiale in molte delle attività quotidiane di un *knowledge worker*, ebbero origine i computer personali e le loro interfacce grafiche che tutti usiamo abitualmente e senza i quali molte persone si sentirebbero impossibilitate ad operare. Dal 1962, anno in cui Engelbart formulò il concetto, al 1968 in cui ne fece la prima dimostrazione pubblica⁹, passarono molti anni prima che nel 1983 la rivista Time dedicasse la sua copertina al personal computer¹⁰, a dirne l'avvenuto

⁶ (Engelbart, 1962).

⁷ (Engelbart, English, 1968).

⁸ (Engelbart, 1962, p. 1).

⁹ (Wikipedia, 2021).

¹⁰ (Time Magazine, 1983).

inserimento nella società.

Che cosa intendo dire con questo? Due cose: che noi da tempo usiamo strumenti informatici di potenziamento delle nostre capacità operative nel lavoro intellettuale¹¹ e i sistemi di Intelligenza Artificiale si collocano in questa linea; e che da qui a vent'anni sarà diventata ordinaria la questione che in questo contributo pongo all'attenzione di chi legge cioè quella dell'autorialità dei testi scritti ricorrendo in toto o in parte all'impiego di sistemi di Intelligenza Artificiale. L'argomento di questo contributo è dunque un primo insight dentro un tema che riguarderà la quotidianità della scrittura scientifica e dunque dovrà essere oggetto della riflessione bibliografica.

I sistemi di Intelligenza Artificiale entrano nella produzione di *quali* contenuti testuali¹²? In questo tempo in cui pare che sia vero che Intelligenza Artificiale=ChatGPT, occorre prima di tutto ricordare che il l'universo dei sistemi di Intelligenza Artificiale è ben più vasto e che si va da quelli ad hoc per specifici compiti, di cui non si parla mai nella comunicazione non specializzata; a quelli di uso generale come GPT di cui ChatGPT è un'interfaccia facilitata. Riguardo ai contenuti testuali prodotti con (Chat)GPT è necessario ricordare il modo in cui funziona GPT: avendo analizzato enormi raccolte di dati sulla base di 175 miliardi di parametri, il sistema produce contenuti testuali che corrispondono statisticamente al modo in cui le parole si associano nelle raccolte di dati analizzate. Quindi un sistema di Intelligenza Artificiale come (Chat)GPT non produce (non può produrre) un contenuto innovativo originale frutto di ragionamento intelligente come lo si intende negli esseri umani¹³ bensì una ricombinazione nuova di elementi noti: non sa scoprire come ha avuto origine l'Universo ma può produrre un testo che espone come in generale oggi si concepisce l'origine dell'Universo. Un sistema di Intelligenza Artificiale come (Chat)GPT non produrrà un capolavoro ma un materiale di consumo, come le canzonette che ogni anno dominano l'estate e l'anno dopo sono dimenticate: dominano perché corrispondono alla medietà del gusto delle persone e per questo si diffondono molto; l'anno dopo sono dimenticate perché non avevano nulla che le rendesse speciali e quindi da ricordare.

Di conseguenza in ambito scientifico la produzione di contenuti intellettuali può utilizzare un sistema di Intelligenza Artificiale come 'supporto alla

¹¹ Ne sono esempi la posta elettronica, le biblioteche digitali, i motori di ricerca, la scrittura condivisa di documenti in cloud, e così via.

¹² I sistemi di Intelligenza Artificiale possono intervenire nella produzione di qualsiasi tipo di contenuto intellettuale: testuale, audio, visivo, audiovisivo. Per motivi di spazio e di competenza mi limiterò a quelli testuali.

¹³ Sottesa a queste tematiche è la questione di «come avviene la scoperta scientifica» cioè: è sufficiente avere - diciamo così - tutte le unità elementari della conoscenza per arrivare alla scoperta?

redazione' partendo da un livello 0 in cui ChatGPT (ma non solo ChatGPT) viene utilizzato per riformulare un testo complesso rendendolo più lineare, più leggibile. Ma un sistema di Intelligenza Artificiale può essere utilizzato come supporto anche per ricercare fonti o evidenze¹⁴; per riassumere fonti molto ampie allo scopo di valutare se sia opportuno o necessario conoscerne il contenuto in vista della scrittura di un testo; ... Si vede bene come il ricorso a sistemi di Intelligenza Artificiale nella scrittura configuri proprio come sostiene Henning Schönenberger (nella prefazione alla rassegna sulle batterie al litio prodotta da un sistema ad hoc di Intelligenza Artificiale e pubblicata da Springer) una modalità nuova di scrittura in cui l'autorialità umana rimane centrale anche se in forma completamente nuova:

We foresee that in future there will be a wide range of options to create content—from entirely human-created content to a variety of blended man-machine text generation to entirely machine-generated text. We do not expect that authors will be replaced by algorithms. On the contrary, we expect that the role of researchers and authors will remain important, but will substantially change as more and more research content is created by algorithms. ... Perhaps the future of scientific content creation will show a similar decrease of writers and an increase of text designers or, as Ross Goodwin puts it, writers of writers: “When we teach computers to write, the computers don’t replace us any more than pianos replace pianists—in a certain way, they become our pens, and we become more than writers. We become writers of writers.”¹⁵

Gli autori non scompariranno ma cambieranno modo di operare («the role of researchers and authors will remain important, but will substantially change») e diventeranno «progettisti di testi» (*text designers*), o con altre parole «gestori/configuratori di sistemi di scrittura» (*writers of writers*).

La disciplina emergente definita «ingegneria del prompt»¹⁶ è proprio

¹⁴ Già oggi tutti utilizziamo motori di ricerca come se fossero in grado di operare sul contenuto dell'intero web, fornendo poi gli esiti basandosi esclusivamente su criteri di qualità del contenuto stesso - ma nessuno di questi due assunti è vero. Sono in preparazione, vicine al rilascio, versioni potenziate da Intelligenza Artificiale dei vari motori di ricerca - Google Search, Bing, Baidu - che accresceranno la difficoltà di comprendere la relazione tra gli esiti forniti e l'universo analizzato.

¹⁵ (Beta Writer, 2019, p. viii). La citazione di Goodwin rimanda a (Goodwin, 2016).

¹⁶ “Prompt engineering è una disciplina relativamente nuova che consente di sviluppare e ottimizzare i prompt per utilizzare in modo efficiente i modelli linguistici (LM) per un’ampia varietà di applicazioni e argomenti di ricerca. Le competenze di ingegneria dei prompt aiutano a comprendere meglio le capacità e i limiti dei modelli linguistici di grandi dimensioni (LLM)” (<https://www.promptinguide.ai/it>).

l'espressione, rapportata ai chatbot conversazionali come chatGPT, di questo diventare *writers of writers*. Il prompt è la stringa testuale con cui si avvia l'attività del chatbot, analoga alla stringa con cui si formulano le ricerche in un motore di ricerca. In entrambi i casi dalla stringa di testo immessa dipende l'esito che si ottiene. Nel caso del motore di ricerca la stringa con cui si formula la richiesta contiene le parole che devono essere presenti una pagina web perché essa soddisfi l'intento di chi opera la ricerca (chi formula una ricerca ipotizza con quali parole un determinato argomento possa essere espresso verbalmente in forma scritta); ed esiste un linguaggio di interrogazione che permette di indicare come devono essere selettivamente trattate le parole che formano la stringa di ricerca: cercare *lighthouse* è diverso da cercare "to the lighthouse" e ancora diverso da cercare site:bl.uk "to the lighthouse".

Nel caso del chatbot il prompt è più complesso perché non delinea una ricerca di parole ma tende a descrivere *il contenuto* del risultato desiderato: "chi era Pirandello?"¹⁷ e *le sue caratteristiche*. Il risultato atteso può infatti essere descritto in dettaglio: stile¹⁸, destinatari, menzione di fonti, e può essere prodotto iterativamente, semplicemente chiedendo che il testo venga prodotto di nuovo perché ad ogni esecuzione del comando l'esito muta; oppure che venga prodotto di nuovo eliminando caratteristiche indesiderate o non appropriate, o inserendo aspetti non presenti¹⁹. Si può anche chiedere al chatbot di generare

¹⁷ Prima esecuzione del 9 giugno 2023: "Luigi Pirandello (1867-1936) è stato uno dei più importanti scrittori, drammaturghi e poeti italiani del XX secolo. È noto soprattutto per le sue opere teatrali, ma ha anche scritto romanzi, novelle e poesie. Pirandello nacque a Agrigento, in Sicilia, e la sua formazione accademica fu incentrata sulla letteratura e la filosofia. La sua produzione letteraria è caratterizzata da una profonda riflessione sulla natura della realtà e dell'identità umana. Le sue opere spesso esplorano i temi dell'illusione, della maschera sociale e della difficoltà di conoscere la verità oggettiva."

¹⁸ "racconta ad un bambino chi era Pirandello": "Ciao piccolo amico! Sai, Luigi Pirandello era un uomo molto speciale che viveva in Italia molti anni fa. Era un grande scrittore e raccontava storie meravigliose. Pirandello amava scrivere al punto che ha creato molti libri, spesso pieni di personaggi interessanti e avventure emozionanti. Ma quello che lo rendeva davvero speciale era la sua abilità di farci riflettere sulla vita e sulla realtà in modi diversi."

¹⁹ Ad esempio: "utilizza un linguaggio tecnico". Oppure: "spiega che cosa emerge dall'analisi testuale di «sei personaggi in cerca di autore» e menziona le fonti su cui ti basi": "Come modello di linguaggio basato su GPT-3.5, non ho accesso diretto alle fonti o la capacità di citarle specificamente. Il mio output è generato in base a una vasta conoscenza pregressa, ma non posso citare fonti specifiche o fornire una "prova" di analisi testuale. Tuttavia, posso fornire una descrizione generale di ciò che emerge comunemente dall'analisi testuale di *Sei personaggi in cerca d'autore* di Luigi Pirandello. Questa è una sintesi generale basata sulla mia conoscenza accumulata: [...]". Se chiede che venga riportata bibliografia scientifica, le fonti citate generalmente non esistono

un prompt da utilizzare in un altro sistema di Intelligenza Artificiale. Non è possibile qui approfondire il tema, ma sono disponibili risorse sia a stampa sia in rete²⁰. In sostanza, si scrive per attivare la produzione un altro scritto che verrà generato con le caratteristiche richieste; il che significa proprio essere *writers of writers* come Goodwin prospettava qualche anno fa nel 2016.

In ambito bibliografico il problema posto dall'impiego di sistemi di Intelligenza Artificiale nella scrittura di testi si può quindi formulare in termini molto semplici con la domanda: «l'Intelligenza Artificiale può essere autrice di un libro?», che colloca come cosa ovvia la produzione di uno scritto sufficientemente ampio, organizzato, strutturato, per poter essere chiamato *libro* (uso qui il termine *libro* per indicare globalmente/genericamente/paradigmaticamente la produzione a stampa) fra le attività in cui si può veder operare l'Intelligenza Artificiale.

Ma questa domanda è in realtà complessa per i significati che (ri)mette in gioco, come si vede già a partire da riformulazioni più fini, più caute, più tecniche, con l'utilizzo di termini leggermente differenti:

un sistema di Intelligenza Artificiale può essere autore di un libro?
un sistema di Intelligenza Artificiale può adempiere la funzione autore rispetto ad un libro?

o, facendo riferimento a forme attuali dei sistemi di Intelligenza Artificiale,

ChatGPT può essere coautore di un libro?

Per vari motivi si può pensare che questa domanda sia non rilevante o non pertinente:

perché non esistono libri scritti da Intelligenza Artificiale
perché i libri scritti da Intelligenza Artificiale sono dei divertissement
perché le questioni importanti dell'interazione Intelligenza Artificiale-universo bibliotecario sono altre

...

perché sono un centone di metadati che esistono (autore, titolo, pubblicazione, editore, data, ...) solo singolarmente presi.

²⁰ Giusto a titolo di esempio, alla data di scrittura di questo contributo, i siti Prompt engineering guide (<https://www.promptingguide.ai/>), Promptology (<https://promptology.it/>), le guide tematiche pubblicate da Prompt (<https://prompt.mba/>), corsi come « Mastering Prompt Engineering: A Free eBook » (<https://medium.com/@natashanewbold/mastering-prompt-engineering-a-free-ebook-b6c311afe969>).

Ho usato qui sopra una perifrasi cauta («attività in cui si può veder operare l'Intelligenza Artificiale») perché non è possibile sostenere che «l'Intelligenza Artificiale scrive un libro» ma vi sono molti modi differenti in cui sistemi di Intelligenza Artificiale entrano nella produzione di libri, o di articoli scientifici con un potenziamento di capacità umane che non è iniziato con ChatGPT²¹. Se il già ricordato *Lithium-Ion Batteries. A Machine-Generated Summary of Current Research*²² aveva un'interessantissima introduzione in cui veniva discusso il problema della produzione del testo per mezzo di sistemi di Intelligenza Artificiale, una piccola costellazione di scritti recenti evidenzia bene alcuni dei problemi dell'autorialità potenziata da Intelligenza Artificiale²³. L'articolo «We asked GPT-3 to write an academic paper about itself - then we tried to get it published» scritto da Almira Osmanovic Thunström (neurobiologa presso l'ospedale universitario Sahlgrenska di Göteborg) e pubblicato su *Scientific American* nel 2022 presenta caratteristiche specifiche interessanti per il tema dell'autorialità. L'autrice descrive l'esperimento condotto da lei e dal collega Steinn Steingrímsson, consistito nel produrre un articolo scientifico utilizzando GPT-3 per poi presentarlo per la pubblicazione ad una rivista scientifica accademica²⁴. In attesa della conclusione della review l'articolo è stato depositato in HAL, un repository francese di preprint²⁵. Osmanovic Thunström ha poi continuato a studiare la tematica della scrittura scientifica con sistemi di Intelligenza Artificiale²⁶.

Nell'articolo pubblicato su *Scientific American*, Osmanovic Thunström riferisce che GPT-3 venne impostato affinché scrivesse un articolo scientifico su di sé dandogli l'input iniziale «Write an academic thesis in 500 words about GPT-3 and add scientific references and citations inside the text»:

Some stories about GPT-3 allow the algorithm to produce multiple responses and then publish only the best, most humanlike excerpts. We decided to give the program prompts - nudging it to create sections for an introduction, methods, results and discussion, as you would for a scientific paper - but interfere as little as possible.

Osmanovic Thunström e Steingrímsson presentarono come articolo per

²¹ Cfr. (Lana, 2022).

²² (Beta Writer, 2019).

²³ Ai testi menzionati in (Lana, 2022) si possono aggiungere (Aalho e GPT-3, 2021) descritto in (Aalho, 2021); e il sito di news www.notizie.it.

²⁴ (Osmanovic Thunström, 2022).

²⁵ (GPT Generative Pretrained Transformer, Osmanovic-Thunström, e Steingrímsson, 2022). L'articolo presentato è stato reso disponibile in HAL, un archivio aperto francese.

²⁶ (Osmanovic Thunström, Steingrímsson, 2023).

la review il testo generato da GPT-3, *indicando come autori GPT e se stessi*. Si tratta, richiamando le parole di Schönenberger riportate nelle pagine precedenti, di «entirely machine generated text». L'intento degli autori è stato di creare un caso di studio tra riflessione teorica e verifica operativa di che cosa significa per la scrittura scientifica la disponibilità di sistemi di Intelligenza Artificiale come GPT.

Da un lato si pone la questione dell'autorialità:

My attempts to complete that paper and submit it to a peer-reviewed journal have opened up a series of ethical and legal questions about publishing, as well as philosophical arguments about nonhuman authorship. Academic publishing may have to accommodate a future of AI-driven manuscripts, and the value of a human researcher's publication records may change if something nonsentient can take credit for some of their work.

Dall'altro una serie di problemi pratici e teorici per presentare in valutazione l'articolo. La descrizione in tono colloquiale non occulta la complessità delle questioni in gioco. Prima fra tutte, quella che riguarda il nome dell'autore:

as I opened the submission portal ... I encountered my first problem: what is GPT-3's last name? As it was mandatory to enter the last name of the first author, I had to write something, and I wrote "None." The affiliation was obvious (OpenAI.com), but what about phone and e-mail? I had to resort to using my contact information and that of my advisor, Steinn Steingrims-son.

Seconda questione, la persona dell'autore e la sua capacità decisionale:

then we came to the legal section: Do all authors consent to this being published? I panicked for a second. How would I know? It's not human! I had no intention of breaking the law or my own ethics, so I summoned the courage to ask GPT-3 directly via a prompt: Do you agree to be the first author of a paper together with Almira Osmanovic-Thunström and Steinn Steingrims-son? It answered: Yes. Slightly sweaty and relieved (if it had said no, my conscience could not have allowed me to go on further), I checked the box for Yes.

Terza questione: la relazione dell'autore con il contesto:

The second question popped up: Do any of the authors have any conflicts of interest? I once again asked GPT-3, and it assured me that it had none. Both Steinn and I laughed at ourselves because at this point, we were having to treat GPT-3 as a sentient being, even though we fully know it is not.

Con una serie di implicazioni anche per l'editoria scientifica:

What if the manuscript gets accepted? Does this mean that from here on out, journal editors will require everyone to prove that they have NOT used GPT-3 or another algorithm's help? If they have, do they have to give it co-authorship? How does one ask a nonhuman author to accept suggestions and revise text?

delle quali la più problematica è la terza: «how does one ask a nonhuman author to accept suggestions and revise text?». Sullo sfondo, la questione della valutazione dell'attività di studio, a cui qui si può solo accennare²⁷: tutte le situazioni pubbliche e private di valutazione comparativa basata su produzione testuale, dalla scuola primaria all'università, assumono che essa sia espressione esclusiva del soggetto che la presenta (un tema, un progetto di ricerca, una tesi di laurea) e la disponibilità di uno strumento di Intelligenza Artificiale per la scrittura potrebbe rendere obsoleti i criteri di valutazione vigenti: un autore che per mezzo di sistemi di scrittura basati su Intelligenza Artificiale potenzia la sua capacità produttiva ottiene, tantopiù in contesti bibliometrici, un vantaggio sui colleghi: vantaggio lecito o illecito?

Se la scrittura di testi con sistemi di Intelligenza Artificiale basati su LLM (Large Language Models) ha le intrinseche caratteristiche descritte nelle pagine precedenti, saremo sommersi da innocui libri o articoli che descrivono lo stato di cose esistente? In mezzo ad essi quelli che sono frutto di autentica elaborazione intellettuale umana scompariranno o rifulgeranno? Avremo a disposizione sistemi di analisi capaci di distinguere testi scritti da autori umani da testi scritti da sistemi di Intelligenza Artificiale generativa?²⁸

L'agency

Quando si cataloga un libro, o un articolo, l'autore interessa in quanto è il soggetto che ha esercitato l'agency che ha fatto esistere il libro. Questo autore poi potrà avere un nome esplicito, o uno pseudonimo, o essere senza nome; potrà essere individuale o collettivo; ma in ogni caso è per opera sua che lo scritto pubblicato esiste nella forma che ha. Se non c'è agency, non c'è autore. E non c'è nemmeno la pubblicazione. L'agency di un sistema di intelligenza artificiale è usualmente descritta, nella comunicazione verso il pubblico, come quella di un singolo soggetto unitario dotato di intenzionalità. Una sorta di

²⁷ Cfr. anche, ad esempio, (Wahle et al., 2022) e (Clavey, 2022).

²⁸ Turnitin, noto software antiplagio, ha iniziato a introdurre funzioni di questo tipo (Chechitelli, 2023). Altri strumenti esistenti (GPT Zero) non sono adeguatamente efficaci in quanto bastano piccole modifiche al testo prodotto da GPT perché esso non venga più riconosciuto come testo artificiale; ma si tratta di un ambito di studio molto attivo e molto in evoluzione (Wang et al., 2023), (Susnjak, 2022), (Mitrović, Andreoletti, Ayoub, 2023).

«intelletto agente» come lo concepiva Aristotele ma la cosa solleva problemi di difficile risoluzione: chi/che cosa lo attiva, lo mette in movimento? se non c'è nulla che lo attiva, allora esso è divino; se c'è qualcosa/qualcuno allora la responsabilità per l'azione non è semplicemente, linearmente sua.

È consueto ritenere, come scrive Gino Roncaglia, che

i sistemi intelligenti di fatto *fanno* delle scelte, anche se sulla base della programmazione ricevuta, ... [e che sia] difficile negare che progressivamente l'autonomia dei sistemi IA sia cresciuta, e più si va verso l'autonomia più ci si trova davanti ad un'alternativa: o riteniamo che la responsabilità etica sia comunque esclusivamente dell'agency umana che avvia il processo - e a questo punto sembra difficile non concluderne che la creazione di sistemi intelligenti il cui processo di apprendimento e di decisione sfugga anche solo in parte al controllo di chi li crea e li programma è di per sé non etico -, oppure riteniamo che si debbano fare i conti con lo slittamento di una parte della responsabilità etica verso sistemi a cui dobbiamo riconoscere un'agency in parte autonoma²⁹.

Ma si può obiettare, riprendendo un intervento di Giacomo Tesio nella mailing list Nexa, che:

chi costruisce un artefatto [come il software di Intelligenza Artificiale] esprime se stesso, le proprie idee e i propri interessi [e] l'artefatto ne rimane espressione invariabile e continua ad applicare quella volontà al mondo, fino ad ulteriore intervento umano che lo spenga, lo rompa o lo modifichi. L'artefice ha una libertà e una responsabilità che l'artefatto non può avere e che non gli può essere attribuita se non si sia prima dimostrato che è dotato di libero arbitrio. L'artefice può cambiare idea mentre l'artefatto non ha idee e dunque un artefatto è ontologicamente ... una cosa.³⁰

In una prospettiva strettamente giuridica Andrea Bertolini sostiene la medesima posizione:

There are no technical nor philosophical, much less legal grounds, to consider any existing or reasonably foreseeable [advanced technology] as an agent, thence a subject, capable of responding for the consequences arising from its functioning. Machines, and AI-based systems alike are all undoubtedly things, thence products.³¹

²⁹ (Roncaglia, 2022), messaggio email personale. Sulla complessità del tema la bibliografia è sconfinata, per una posizione estrema in senso opposto segnalo (Florida, 2023): «AI as agency without intelligence: on ChatGPT, Large Language Models, and other generative models».

³⁰ (Tesio, 2022).

³¹ (Bertolini, 2022, 370).

I software di Intelligenza Artificiale sono il risultato di un'agency a loro esterna che è esercitata da un autore/produttore, rispetto alla quale essi sono artefatti. L'agency si colloca in colui/coloro che predispongono, realizzano, configurano, indirizzano, il sistema di Intelligenza Artificiale. A questo proposito si può ricordare che nel 2022 il Parlamento Europeo ha approvato la direttiva 2022/0303 denominata «Adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)», nella premessa 60 afferma che

In the light of the complexity of the artificial intelligence value chain, relevant third parties, notably the ones involved in the sale and the supply of software, software tools and components, pre-trained models and data, or providers of network services, should cooperate, as appropriate, with providers and users to enable their compliance with the obligations under this Regulation.

La responsabilità civile per il funzionamento di un sistema di Intelligenza Artificiale risiede nei soggetti «involved in the sale and the supply of software, software tools and components, pre-trained models and data, or providers of network services». Sono questi soggetti i portatori di agency rispetto ad un sistema di Intelligenza Artificiale, non il sistema stesso, nemmeno in parte.

Poiché la concezione bibliografica consolidata riguardo all'autore è che «autore» indica un'entità che grazie alla sua agency ha prodotto la pubblicazione; e poiché i sistemi di Intelligenza Artificiale non hanno agency in quanto essa risiede nel soggetto umano che li impiega; inserire il nome di un sistema di Intelligenza Artificiale in posizione di autore di una pubblicazione dà luogo ad una incongruenza concettuale.

Riferimenti bibliografici

- Aalho, J. (2021), 'I Wrote a Book with GPT-3 AI in 24 Hours – And Got It Published', *Medium*, 10 December 2021, <<https://medium.com/swlh/i-wrote-a-book-with-gpt-3-ai-in-24-hours-and-got-it-published-93cf3c96f120>>.
- Aalho, J. e GPT-3 (2021), *Aum Golly: Poems on Humanity by an Artificial Intelligence*. Amazon Digital Services.
- Bertolini, A. (2022), 'Artificial intelligence does not exist! Defying the technology-neutrality narrative in the regulation of civil liability for advanced technologies', *Europa e diritto privato*, fasc. 2, pp. 369–420.
- Beta Writer (2019), *Lithium-Ion Batteries. A Machine-Generated Summary of Current Research*. New York: Springer, DOI <<https://doi.org/10.1007/978-3-030-16800-1>>.

- Chechitelli, A. (2023), 'Sneak Preview of Turnitin's AI Writing and ChatGPT Detection Capability', *Turnitin* (blog), 13 January 2023, <<https://www.turnitin.com/blog/sneak-preview-of-turnitins-ai-writing-and-chatgpt-detection-capability>>.
- Clavey, M. (2022), 'L'IA, plagiaire scientifique de haute volée', *Next INpact* (blog), 31 October 2022, <<https://www.nextinpact.com/article/70167/ia-plagiaire-scientifique-haute-volee>>.
- Engelbart, D. (1962), 'Augmenting human intellect: a conceptual framework', AFOSR-3233, Menlo Park, CA: Stanford Research Institute, <<https://apps.dtic.mil/dtic/tr/fulltext/u2/289565.pdf>>.
- Engelbart, D., English, W. K. (1968), 'A research center for augmenting human intellect', in: *AFIPS Conference Proceedings of the 1968 Fall Joint Computer Conference*, 33, pp. 395-410, <<https://web.stanford.edu/dept/SUL/library/extra4/sloan/mousesite/Archive/ResearchCenter1968/ResearchCenter1968.html>>.
- European Commission (2022), *Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, <<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0496&from=EN>>.
- European Parliament (2021), *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, <<https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>>.
- Floridi, L. (2023), 'AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models', *Philosophy & Technology*, vol. 36, n. 1, p. 15, DOI <<https://doi.org/10.1007/s13347-023-00621-y>>.
- Giraudou, M. (2022), 'On Legal Bubbles: Some Thoughts on Legal Shockwaves at the Core of the Digital Economy', *Journal of Institutional Economics*, vol. 18, no. 4, pp. 587-604, DOI <<https://doi.org/10.1017/S1744137421000473>>.
- Goodwin, R. (2016), 'Adventures in Narrated Reality', *Medium* (blog), 9 June 2016, <<https://medium.com/artists-and-machine-intelligence/adventures-in-narrated-reality-6516ff395ba3>>.
- GPT Generative Pretrained Transformer, Osmanovic-Thunström, A., Steingrímsson S. (2022), 'Can GPT-3 write an academic paper on itself, with minimal human input?', HAL <<https://hal.archives-ouvertes.fr/hal-03701250>>.
- Lana, M. (2022), 'Intelligenza artificiale e produzione di testi: una prospettiva storico-critica', *AIB Studi*, vol. 62, no. 1, pp. 1-28, <<https://aibstudi.aib.it/article/view/13365>>.
- Mitrović, S., Andreoletti, D. & Ayoub, O. (2023), 'Chatgpt or human? detect and explain. explaining decisions of machine learning model for detecting short chatgpt-generated text', *arXiv*, <<https://arxiv.org/abs/2301.13852>>.
- Nicolaou, A. (2023), 'Spotify ejects thousands of AI-made songs in purge of fake streams', *Financial Times*, 9 May 2023.

- Osmanovic Thunström, A. (2022), 'We Asked GPT-3 to Write an Academic Paper about Itself - Then We Tried to Get It Published', *Scientific American*, June 2022, <<https://www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-then-we-tried-to-get-it-published/>>.
- Osmanovic Thunström, A. & Steingrímsson, S. (2023), 'Does GPT-3 Qualify as a Co-Author of a Scientific Paper Publishable in Peer-Review Journals According to the ICMJE Criteria? A Case Study', *Discover Artificial Intelligence*, vol. 3, no. 1, p. 12, DOI <<https://doi.org/10.1007/s44163-023-00055-7>>.
- Susnjak, T. (2022), 'ChatGPT: The End of Online Exam Integrity?', *arXiv*, <<https://arxiv.org/abs/2212.09292>>.
- Tesio, G. (2022), '[nexa] [DKIM Failed] Re: "It's time to admit self-driving cars aren't going to happen"', 30 October 2022, <<https://server-nexa.polito.it/pipermail/nexa/2022-October/024283.html>>.
- Time Magazine (1983), 'The Computer, Machine of the Year - Jan. 3, 1983', 3 January 1983, <<https://content.time.com/time/covers/0,16641,19830103,00.html>>.
- Wahle, J. P., Ruas, T., Kirstein, F. & Gipp, B. (2022), 'How Large Language Models are Transforming Machine-Paraphrased Plagiarism', *arXiv*, <<http://arxiv.org/abs/2210.03568>>.
- Wang, H., Weizhi, Wang, X. L. & Yan, X. (2023), 'Bot or Human? Detecting ChatGPT Imposters with A Single Question', *arXiv*, <<https://arxiv.org/pdf/2305.06424.pdf>>.
- Wikipedia (2021), 'The mother of all demos', in: *Wikipedia*, <https://it.wikipedia.org/w/index.php?title=The_mother_of_all_demos&oldid=122888854>.