ChatGPT e Intelligenza Artificiale, sì o no?

Autore: Andrea Galluzzi

Gli effetti collaterali dell'avvento di ChatGPT e delle IA di ultima generazione cominciano a farsi sentire, sia per i vantaggi che portano, sia per le questioni etiche e legali che sollevano.

La presentazione al grande pubblico della tecnologia **ChatGPT** [1] (sistema di modellizzazione del linguaggio basato su Intelligenza Artificiale, sviluppato dalla californiana OpenAI) – avvenuta nel novembre scorso ha rappresentato per il mondo tecnologico un evento spartiacque. Come avvenuto anni fa con l'arrivo dell'iPhone – col quale un visionario Steve Jobs presentava un nuovo modo di intendere l'uso dello smartphone – questa tecnologia sta rapidamente imponendosi come un nuovo modo di interagire con le macchine, elevandole dal rango di strumenti a quello di **compagni di lavoro**, aprendo nuovi settori di sviluppo e spingendo molti attori del panorama tecnologico ad aggiornarsi per sfruttarne il successo e i vantaggi. L'ultima versione di ChatGPT, basata sul sistema di modellizzazione linguistica GPT-4 [2], ha affinato ulteriormente la qualità delle sue risposte, aggiunto alcune peculiarità e migliorato il controllo etico dei contenuti. Dall'arrivo di ChatGPT si sono riproposte e moltiplicate le domande sul nostro rapporto con le IA. Il fatto di avere davanti a sé automatismi in grado di interpretare bene il linguaggio e produrre testi o immagini indistinguibili da quelli realizzati da umani pone in evidenza il problema dell'uso malevolo di questi dispositivi. I fronti di discussione più accesi riguardano l'utilizzo dell'IA nel campo della falsificazione di informazioni e l'ingerenza degli automatismi nel sostituire gli umani nei più svariati compiti. La cronaca recente ha messo in luce alcuni fatti importanti che hanno sollevato molte discussioni e una reazione a catena di opinioni diverse. Vediamone alcuni. **Fermare tutto?** Poche settimane fa (più precisamente il 22 marzo scorso) il ***Future of Life Institute*** *(FLI)* [3] – organizzazione senza scopo di lucro che si occupa dei rischi esistenziali legati all'uso delle tecnologie avanzate, con sede negli Stati Uniti) – ha pubblicato, a mo' di mozione, una **lettera aperta** [4] in cui propone a tutti i maggiori laboratori di interrompere per almeno sei mesi lo sviluppo e l'addestramento di IA con livello superiore a GPT-4, in modo da dare tempo alle istituzioni di attuare contromisure legali e politiche per salvaguardare il futuro del genere umano. Le domande poste dalla mozione del *FLI* sono del tutto condivisibili e sembrano fare eco alla lettera che **Sam Altman** (co-fondatore e CEO di OpenAI) aveva scritto già un mese prima [5], in cui sottolineava il carattere etico che deve sottendere lo sviluppo di tutte le IA e la crescente cautela nella divulgazione dei nuovi sistemi, visto il livello raggiunto dall'ultima generazione di queste tecnologie. L'obiettivo è semplice: occorre massimizzare i benefici derivanti dall'IA e allo stesso tempo minimizzare i rischi, facendo in modo che i vantaggi, l'accesso e la ***governance*** che la riguardano siano ampiamente ed equamente condivisi. A scanso di equivoci: non stiamo parlando di sistemi senzienti, in grado cioè di avere auto-consapevolezza di sé e di rivaleggiare con gli esseri umani e competere per il ruolo di specie dominante sul pianeta, ma di **strumenti potenti** che messi in mani sbagliate – o semplicemente ingenue – possono provocare vari danni collaterali, a livello psicologico, sociale, economico, e così via. La mozione in questione conta migliaia di firmatari, fra cui **Elon Musk** (fondatore di Tesla e grande innovatore mediatico), **Steve Wozniak** (co-fondatore di Apple) e parecchi altri co-fondatori della Silicon Valley. Scorrendo la lista dei sottoscrittori si vedono emergere spesso ricercatori, ingegneri, manager e tecnici vari (alcuni appartenenti a colossi informatici come Microsoft, Google, Apple, Meta ecc.). Non è un dato di poco conto. Significa che chi conosce la tecnologia sa cosa questa potrebbe arrivare a fare se utilizzata male. Una presa di posizione individuale, però, non automaticamente corrisponde alla linea di tutta un'azienda. È molto difficile pensare che la mozione possa essere presa seriamente in considerazione dai grandi del web, fondamentalmente per due motivi: 1) sei mesi sono pochi per rendere operativi eventuali accordi legislativi fra Stati su un tema pervasivo come questo; 2) la corsa a modelli di IA sempre più potenti e imprevedibili non si arresterebbe comunque, e sui mercati

lascerebbe indietro gli attuali protagonisti del settore. Inoltre è impossibile arginare un fenomeno ormai pervasivo come questo, viste le implicazioni di troppi interessi di parte. Nel recente report diffuso dall'Università di Stanford sulle tendenze legate all'IA [6] si legge che fino al 2014 i modelli di ***machine learning*** più significativi venivano rilasciati dal mondo accademico. Da allora, l'industria ha preso il sopravvento e nel 2022 si sono contati 32 modelli (fra i più significativi) prodotti da aziende del settore tecnologico rispetto ai soli 3 realizzati in ambito universitario. Costruire sistemi di IA all'avanguardia richiede infatti finanziamenti ingenti, di cui spesso le università non dispongono, al contrario dei mercati, sempre impazienti di riuscire a sfruttare queste nuove risorse. Quello che la mozione del *FLI* intende sottolineare è comunque chiaro: **abbiamo bisogno di tempo per educarci bene allo sviluppo e all'uso corretto di questi strumenti** e imparare a non cadere ingenuamente nelle trappole mediatiche che queste tecnologie possono facilmente contribuire a creare. **Il papa "trapper" e l'ex-presidente arrestato** Nelle ore in cui si discuteva del possibile arresto di **Donald Trump** (lo scorso 20 marzo) il giornalista britannico **Eliot Higgins** (fondatore della piattaforma di giornalismo investigativo Bellingcat) ha diffuso alcune immagini che ritraevano la scena del drammatico arresto dell'ex presidente e del suo conseguente internamento in un carcere di massima sicurezza [7]. Pochi giorni dopo, un improbabile **papa Francesco** in versione "trapper" con addosso un costosissimo puffer bianco è diventato la star di Twitter e di tutto il web, con decine di milioni di visualizzazioni [8]. Oltre al carattere goliardico, i due set di foto hanno in comune due cose: entrambi hanno avuto una grande eco mediatica ed entrambi sono falsi, generati attraverso **Midjourney** [9] (una piattaforma IA specializzata in creazione di immagini a partire da una descrizione testuale). Le foto potevano apparire vere solo ad un occhio ingenuo o poco attento ai dettagli, ma erano molto verosimili, e tanto è bastato per scatenare il web e l'opinione pubblica. La facilità con cui chiunque può utilizzare strumenti come Midjourney e divulgare immagini potenzialmente compromettenti apre domande cruciali sulla necessità di una sensibilizzazione collettiva sul tema della falsificazione delle informazioni. L'IA di Midjourney (e quella delle altre piattaforme simili) mette in mano a chiunque la possibilità di generare immagini secondo la propria fantasia e usufruirne a piacimento. I filtri etici di cui dispongono questi strumenti non riescono ad agire sempre in maniera preventiva, lasciando alle istituzioni il compito – nei casi più gravi – di intervenire a valle di divulgazioni illecite. Ma riuscire ad arginare il fenomeno è praticamente impossibile, visto il proliferare di questi sistemi nei sottoboschi di internet o nelle profondità del *deepweb*. I casi del papa "trapper" e del finto arresto di Trump, anche se di stampo goliardico, appartengono allo stesso fenomeno generato dalle ***fake news*** di tutti i tempi: **costruiscono narrazioni sulle quali si forma la coscienza collettiva del nostro tempo**. Questi fenomeni mediatici contribuiscono cioè a costruire l'orizzonte a cui si tende. In altre parole: essi – come, in generale, tutte le comunicazioni mediatiche – sono responsabili della nostra percezione della realtà e della formazione della coscienza e consapevolezza pubblica. Anche in assenza di una volontà manipolativa, **le narrazioni definiscono ciò che siamo** e "creano" la realtà, cioè le danno senso attraverso il racconto che facciamo di essa. Gli strumenti di oggi ramificano la narrazione in molti modi, rendendo sempre più difficile distinguere il vero dal falso. Da qui il compito di **rafforzare la sensibilità etica individuale e collettiva** per riuscire a dare un senso all'orizzonte che abbiamo davanti. Su questo fronte – con sensibilità diverse da paese a paese – le istituzioni sono andate via via attrezzandosi e hanno maturato nel tempo legislazioni adeguate alle questioni del mondo digitale, aggiornando continuamente le proprie misure per la tutela dei diritti umani. Da questo punto di vista l'Italia ha dimostrato una particolare attenzione ai temi della privacy digitale e fatto più di una volta da apripista, come nella recentissima e discussa **disputa fra il Garante italiano e OpenAI** per la non conformità di ChatGPT con le normative dell'Unione Europea riguardanti la raccolta dei dati personali. **Il blocco di ChatGPT in Italia** Con un provvedimento destinato a fare ancora molto rumore (e probabilmente anche a fare scuola nei libri di diritto digitale) [10] il 30 marzo scorso il **Garante italiano per la privacy** ha contestato a **OpenAI** alcuni illeciti relativi al disallineamento con le normative imposte dal *Regolamento Generale per la Protezione dei Dati Personali* (GDPR). Più precisamente (semplificando) il Garante ha rilevato quattro cose: 1) l'assenza di una informativa

sulla raccolta dei dati operata da OpenAI per addestrare ChatGPT; 2) l'assenza di una idonea base giuridica in relazione al loro trattamento negli algoritmi di addestramento della rete neurale; 3) la mancata corrispondenza che spesso si verifica fra i dati forniti da ChatGPT e quelli reali, determinando un trattamento inesatto dei dati personali; 4) l'assenza di una verifica dell'età degli utenti di ChatGPT che, secondo i termini pubblicati da OpenAI, è riservato a soggetti che abbiano compiuto almeno 13 anni. L'insieme di queste rilevazioni ha fatto sì che l'Autorità Garante intervenisse con il provvedimento urgente di **blocco temporaneo del trattamento** per gli utenti italiani. In altre parole: il Garante per la privacy ha chiesto al produttore di ChatGPT di non utilizzare più i dati dei cittadini del nostro paese (e non di bloccare l'accesso a tutta la piattaforma, cosa che peraltro non sarebbe autorizzata a fare), almeno fino a che non fosse verificata la sua **conformità al GDPR**. Di fronte a questa richiesta, OpenAI si è trovata davanti tre possibilità di reazione [11]: 1) bloccare ChatGPT in Italia; 2) adeguarsi alle normative escludendo i dati dei cittadini italiani; 3) proseguire rischiando le sanzioni indicate dal Garante (che prevedono una multa fino a 20 milioni di euro o fino al 4% del fatturato globale annuo). OpenAI ha scelto la prima, sollevando immediatamente disappunti e polemiche da parte delle migliaia di utenti italiani che usufruivano attivamente del servizio. La scelta di OpenAI è comprensibile: deve **prendere tempo** per capire il da farsi, considerando che molto probabilmente alla richiesta della Autorità Garante italiana – che questa volta è arrivata per prima rispetto al resto del mondo – seguirà la reazione a catena di altri Garanti europei, ripetendo una dinamica simile a quella vista nel caso di **Google Analytics** lo scorso anno [12]. Chi usufruisce di ChatGPT sa bene che il blocco è facilmente aggirabile mascherando il proprio indirizzo IP e che esistono modi molto semplici per farlo. **Ciò che non è aggirabile, invece, è il valore e la tutela dei nostri dati e della salvaguardia dei diritti sul loro utilizzo**. Il concetto di fondo che il Garante vuole (e deve) difendere riguarda il fatto che se una persona rende pubbliche alcune informazioni sul proprio conto, ciò non giustifica l'utilizzo indiscriminato di quei dati da parte di terzi senza il consenso esplicito della persona interessata. La grande "pesca a strascico" effettuata da ChatGPT su *database* di pubblico dominio per addestrare la propria rete neurale sembra sia avvenuta effettivamente nella totale ignoranza di questa regola, con conseguenze potenzialmente enormi dal punto di vista legale. Allo stesso modo, attualmente non è stabilito con precisione se ChatGPT sia in grado di correggere o cancellare le informazioni sbagliate che può generare nei suoi discorsi. Così come un quotidiano è obbligato, per legge, a pubblicare smentite su notizie false – o come Google e gli altri motori di ricerca devono rispettare, per legge, il diritto all'oblio cancellando alcune informazioni dai loro *database* –anche le IA generative devono poter rendere conto dei contenuti che producono. La questione si estende a macchia d'olio su tutti i sistemi che hanno iniziato ad implementare al loro interno gli strumenti di OpenAI e altre tecnologie simili. Da qui una domanda cruciale per **Microsoft**: la versione di ChatGPT integrata nel motore di ricerca *Bing* sarà soggetta alle stesse attenzioni da parte del Garante? La domanda non è di poco conto, anche se l'esperienza legale di Microsoft difficilmente potrebbe essere presa in contropiede. Lo si vede anche dal fatto che – a differenza della versione di ChatGPT gestita direttamente da OpenAI – nel generare le sue risposte testuali *Bing* cita puntualmente le sue fonti. Sarà interessante vedere cosa accadrà a valle dei mille risvolti dei procedimenti in corso. Fra i molti aspetti di queste vicende, in conclusione, ne emerge uno: **chiarificare le basi giuridiche sulle quali costruire il mondo digitale è una attività strettamente necessaria, perché quelle basi svolgono anche una importante funzione educativa**. Oltre a chiarire il valore dei nostri dati (che sono la moneta con la quale paghiamo i servizi "gratuiti" di cui usufruiamo in rete), le basi del diritto ci aiutano a capire il valore delle nostre azioni. Non si tratta dunque di adattare meramente la nostra mentalità allo sviluppo in corso – anche perché la velocità con cui evolvono le tecnologie ci farebbe restare sempre due passi indietro – e neanche frenare ciecamente il progresso. Si tratta invece di continuare a porsi le domande giuste per riuscire a trovare le forme di sviluppo adatte alla sostenibilità della nostra condizione umana. Andrea Galluzzi [1] A. Galluzzi, *Intelligenza artificiale e ChatGPT* (13/02/2023) http://www.cittanuova.it/esperto/2023/2/13/intelligenza-artificiale-chatgpt/ [2] OpenAI, GPT-4

(14/03/2023) https://openai.com/research/gpt-4 [3] Future of Life Institute https://futureoflife.org [4] Future of Life Institute, *Pause Giant AI Experiments: An Open Letter* (22/03/2023) https://futureoflife.org/open-letter/pause-giant-ai-experiments/ [5] S. Altman, *Planning for AGI and beyond* (24/02/2023) https://openai.com/blog/planning-for-agi-and-beyond [6] Stanford University, *Artificial Intelligence Index Report 2023* https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index_Report_2023.pdf [7] A. Belanger, *AI-faked images of Donald Trump's imagined arrest swirl on Twitter* (21/03/2023) https://arstechnica.com/tech-policy/2023/03/fake-ai-generated-images-imagining-donald-trumps-arrest-circulate-on-twitter/ [8] C. Stakel-Walker, *We Spoke To The Guy Who Created The Viral AI Image Of The Pope That Fooled The World* (27/03/2023) https://www.buzzfeednews.com/article/chrisstokelwalker/pope-puffy-jacket-ai-midjourney-image-creator-interview [9] Midjourney https://www.midjourney.com/ [10] Garante per la Protezione dei Dati Personali, Provvedimento del 30 marzo 2023 [9870832] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832 [11] M. Borgobello, *ChatGPT: perché il Garante lo ha bloccato e che succede ora* (31/03/2023) https://www.agendadigitale.eu/sicurezza/privacy/chatgpt-perche-il-garante-lo-ha-bloccato-e-che-succede-ora/ [12] A. Galluzzi, *Google Analytics & Co. e il rispetto della privacy* (13/08/2022) http://www.cittanuova.it/google-analytics-co-rispetto-della-privacy