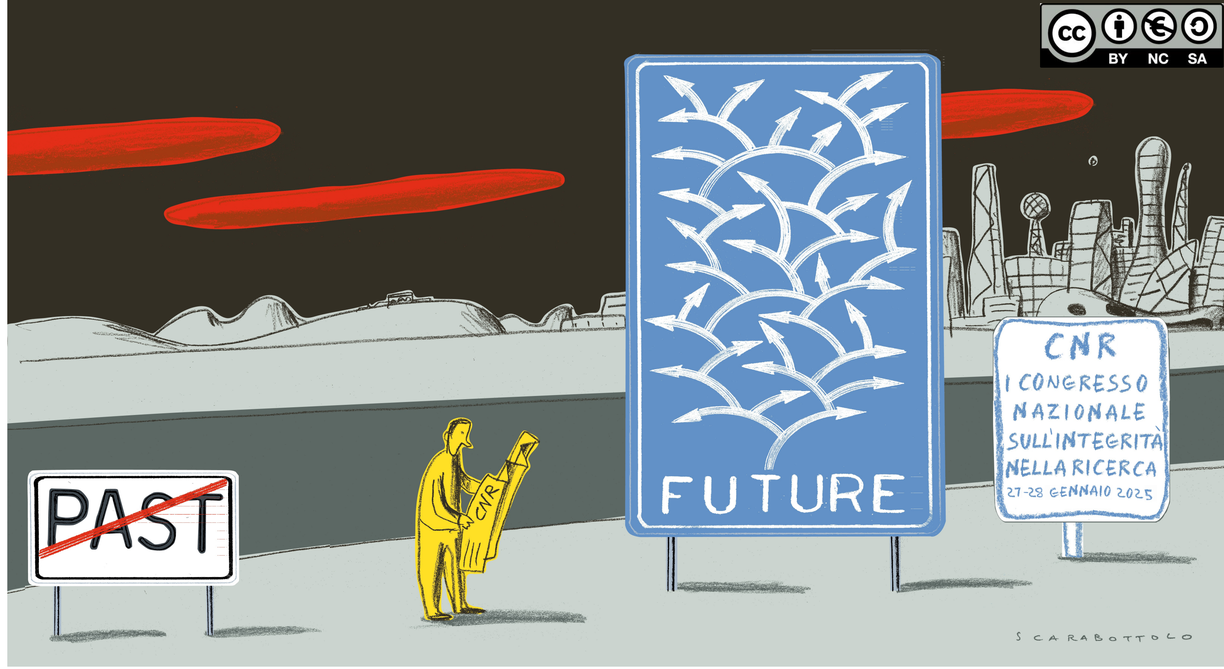
PER UN USO RESPONSABILE DELL'INTELLIGENZA ARTIFICIALE GENERATIVA NELLA RICERCA SCIENTIFICA: OPPORTUNITÀ, RISCHI E SFIDE

Ludovica Marinucci
Centro Interdipartimentale per l'Etica e l'Integrità nella Ricerca, CNR
ludovica.marinucci@ethics.cnr.it

CID ETHICS
Centro Interdipartimentale per l'Etica e l'Integrità nella Ricerca



CC BY NC SA

INTRODUZIONE

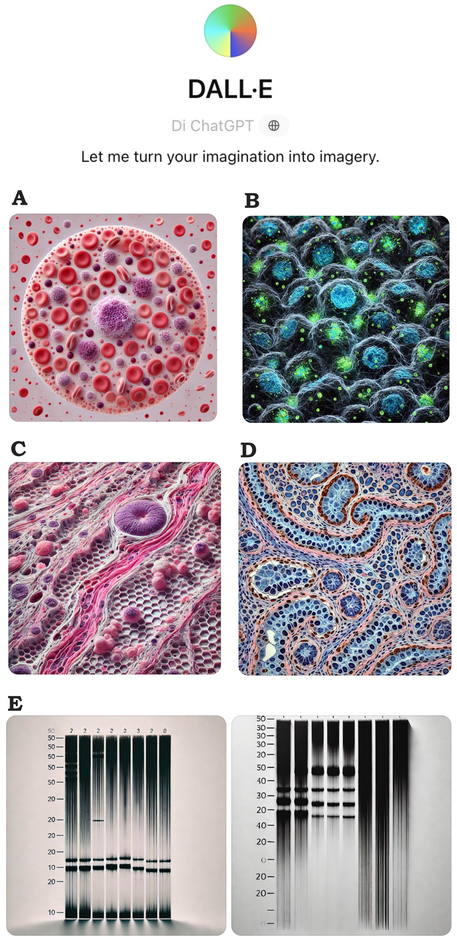
Il rapido progresso dell'Intelligenza Artificiale (IA) generativa, basata su grandi modelli linguistici ed enormi quantità di dati, ha creato *chatbot* sempre più versatili che, essendo in grado di comprendere e generare testi e immagini simili a quelli prodotti da autori umani, stanno ricevendo una notevole attenzione da parte dei ricercatori. Tra i vari *chatbot*, **ChatGPT**, di recente integrato con **DALL-E**, emerge come strumento potente a supporto della produzione scientifica, riducendo il carico di lavoro per i compiti ripetitivi e consentendo di concentrarsi sugli aspetti più creativi e innovativi. Tuttavia, queste tecnologie presentano anche **seri rischi per l'integrità nella ricerca**. Tra i principali si evidenzia come la totale dipendenza da tali strumenti può portare a una moltiplicazione surrettizia delle pubblicazioni scientifiche e compromettere l'originalità e l'approfondimento critico richiesto per produrre contributi significativi e autoriali.

OBIETTIVI

Per sfruttare il potenziale dell'IA generativa senza compromettere i **valori fondamentali dell'integrità nella ricerca**, si propone un'analisi critica dei possibili utilizzi di **ChatGPT** e del suo impatto nelle pratiche scientifiche, basata sia sulla revisione della letteratura esistente sia su un'indagine empirica volta a verificare le effettive potenzialità e limiti del *chatbot*. Tale analisi si pone come necessaria fase preliminare per la definizione di **linee guida** specifiche a supporto dei ricercatori per trovare un equilibrio tra pratiche di ricerca assistite dall'IA e competenze umane.

POTENZIALITÀ E LIMITI DI CHATGPT

La versione attuale di ChatGPT 4o, integrato con DALL-E 3 per la generazione di immagini e con funzionalità di navigazione web, offre supporto in vari tipi di attività nel processo di ricerca scientifico. Ad esempio, attraverso interazioni iterative nella forma di conversazioni, ChatGPT può facilitare il **brainstorming**, proponendo nuove prospettive o domande di ricerca da esplorare. Tuttavia, tali suggerimenti devono essere vagliati in modo critico perché il *chatbot* può generare risposte che sembrano plausibili ma contengono errori e pregiudizi presenti nei dati di addestramento, che rischiano di fuorviare i ricercatori se non verificati. Inoltre, la versione ScholarGPT dedicata alla ricerca può aiutare con suggerimenti su una prima **revisione della letteratura** su un argomento, dato che è in grado di accedere a Google Scholar, PubMed, ecc., e sintetizzare informazioni chiave negli articoli trovati. Tuttavia, avendo accesso limitato ad alcune fonti, le risposte del *chatbot* non coprono tutti gli studi pertinenti. È, quindi, necessario che il ricercatore provveda a un'ulteriore ricognizione da più fonti per un'analisi completa della letteratura. ChatGPT può assistere anche nell'**analisi e interpretazione dei dati** forniti dal ricercatore, suggerendo correlazioni da approfondire anche grazie alla creazione di grafici per la visualizzazione dei dati, e supportare la **generazione e revisione di codice** in vari linguaggi di programmazione facilitando lo sviluppo di script con varie finalità, tra cui l'analisi dei dati. Infine, grazie alla sua abilità di generare testo con stili diversi può contribuire alla **stesura di articoli**, suggerendo titoli concisi e di impatto che catturano l'essenza di un articolo sulla base delle informazioni fornite dal ricercatore. A partire sia dal titolo sia dall'intero articolo, ChatGPT può aiutare a condensare informazioni complesse in un abstract o nella sezione introduttiva che riepiloga obiettivi, metodi e risultati ottenuti. Tuttavia, il ricercatore deve verificare, perfezionare, contestualizzare e personalizzare le risposte. Inoltre, ChatGPT può non solo formattare in pochi secondi la bibliografia nei diversi stili di volta in volta richiesti dagli editori ma anche migliorare la formulazione linguistica fornendo **revisioni di testi** scritti da autori non madrelingua. Invece, essendo stato integrato con DALL-E, ChatGPT riesce ad **analizzare immagini e generarle a partire da testo.** Tuttavia, le policy di OpenAI restringono la realizzazione di immagini, soprattutto di ambito biomedico, così realistiche da poter essere inserite come evidenze in un articolo scientifico. Esse sono adatte invece per finalità divulgative ed educative.



Nuove immagini generate a partire dai 'prompts' forniti da **Zhu et al. (2024)**.
(**A**) blood smears. (**B**) immunofluorescence staining. (**C**) hematoxylin and eosin (H&E) staining. (**D**) immunohistochemistry. (**E**) western blot images.

RISCHI E SFIDE PER L'INTEGRITÀ NELLA RICERCA

Per verificare la possibilità di fabbricare immagini tramite DALL-E, oltre a richieste di creare immagini mediche, ad esempio radiografie dei polmoni, sono stati replicati i 'prompts' delle dettagliate descrizioni di immagini biomediche di risultati sperimentali (Western Blot, immunofluorescenza, ecc.) utilizzati nell'articolo di Zhu e colleghi (2024) per sollevare l'attenzione sui rischi dell'uso di ChatGPT non solo per la scrittura degli articoli ma anche per la **fabbricazione di evidenze scientifiche**. Tuttavia, come si può vedere a sinistra, le nuove immagini generate (A, B, C, D, E) non sono realistiche quanto quelle presenti nell'articolo, facendo perdere leva alle argomentazioni presentate dagli autori. Anche il suggerimento di inserire delle "filigrane invisibili" per riconoscere le immagini generate dall'IA non può ritenersi una soluzione sufficiente dato che tali immagini possono essere poi modificate con altri software, come Photoshop, per ottenere il risultato desiderato. Come per il **plagio**, l'IA generativa più che nuovi problemi per l'integrità nella ricerca sta sollevando nuove sfide non solo relative all'aggiornamento dei software esistenti e alla creazione di nuove **strategie antiplagio**, tra cui l'uso dello stesso ChatGPT (Khalil and Er 2023), ma soprattutto relativi alla sempre più necessaria formazione e responsabilizzazione di ricercatori ed editori di pubblicazioni scientifiche. Se è ancora fonte di dibattito la possibilità di attribuire **autorialità all'IA** (Lund and Naheem 2024), è invece ormai policy consolidata che i ricercatori dichiarino in maniera trasparente come e dove hanno utilizzato ChatGPT e simili nel processo di ricerca e/o scrittura dell'articolo. Tuttavia, è altrettanto necessario che gli stessi editori e revisori vaglino attentamente i contributi ricevuti perché ad oggi la verifica più efficace sulla qualità degli output di ChatGPT resta quella umana. Ciò è tanto più essenziale in considerazione del fatto che grandi editori accademici (Wiley, Sage e Taylor & Francis) stanno vendendo i loro articoli per continuare ad addestrare modelli di IA a fronte di significativi introiti (Kwon 2024), alimentando l'ormai **inevitabile integrazione dell'IA** nel processo di ricerca.

BIBLIOGRAFIA

Khalil, M., & Er, E. (2023). Will ChatGPT Get You Caught? Rethinking of Plagiarism Detection. In: Zaphiris, P., Ioannou, A. (eds) Learning and Collaboration Technologies. HCII 2023. LNCS, vol. 14040 (pp. 475-487). Cham: Springer Nature Switzerland.

Kwon, D. (2024). Publishers are selling papers to train AIs-and making millions of dollars. *Nature*, *636*(8043), 529-530.

Lund, B. D., & Naheem, K. T. (2024). Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals. *Learned Publishing*, *37*(1), 13-21.

Zhu, L., Lai, Y., Mou, W. et al. (2024). ChatGPT's ability to generate realistic experimental images poses a new challenge to academic integrity. *J Hematol Oncol* 17, 27.