Intelligenza artificiale e ricerca scientifica: come salvaguardare la research integrity

30 Settembre 2026

di **Barbara Marchetti**

Università di Trento

In pochi anni, i sistemi di intelligenza artificiale generativa sono entrati stabilmente nelle università e nei centri di ricerca. ChatGPT è stato rapidamente affiancato da strumenti come Gemini, Claude, Perplexity, Elicit, Google Co-Scientist. Questi modelli linguistici di grandi dimensioni possono produrre in pochi secondi testi formalmente plausibili, modulandone stile, registro e complessità. Possono proporre idee, riassumere pubblicazioni, elaborare dati sintetici e assistere il ricercatore in numerose fasi del suo lavoro. I risultati conseguiti nella biologia, nella matematica, nella sismologia dai Large Language Models (LLM) sono sorprendenti. Da AlphaFold all'A-Lab del Lawrende Berkeley National Laboratory, in cui IA e Robot conducono in autonomia esperimenti 24 ore su 24 senza assistenza umana, la ricerca ha abbracciato l'IA nei settori più svariati: i risultati sono apprezzabili ma vengono raggiunti – deve essere sottolineato – a costo di moltissimi fallimenti (P. Sisson, 2026). La qualità linguistica dell'output non va, infatti, confusa con la sua attendibilità scientifica. Gli LLM non comprendono i contenuti di ciò che elaborano, o almeno non li comprendono nello stesso modo in cui li comprende un essere umano, generano risposte sulla base di correlazioni statistiche e selezionano sequenze linguistiche plausibili rispetto al contesto ricevuto: per questa ragione possono produrre informazioni inesatte, inventare riferimenti bibliografici, attribuire un testo ad un autore inesistente o manipolare dati se ciò è necessario per raggiungere il risultato desiderato (le cosiddette allucinazioni). Inoltre, questi modelli non sempre indicano le fonti sulle quali si fondano, possono attingere a contenuti non accademici o non sottoposti a verifica, e possono presentare come certa una risposta soltanto verosimile. È evidente, dunque, che se si intende integrare l'IA generativa nell'attività di ricerca occorre farlo a precise condizioni che variano a seconda degli impieghi che ne fa il ricercatore. Non tutti gli usi dell'IA, infatti, hanno lo stesso peso. Utilizzare un modello per correggere un refuso, migliorare la forma linguistica o tradurre un abstract non equivale a chiedere di formulare un'ipotesi di ricerca, l'interpretazione dei risultati o la scrittura di intere sezioni di un articolo. Nelle prime ipotesi appare evidente il suo impiego ausiliario e preparatorio, nelle seconde si ha invece una sostituzione della macchina al ricercatore. Tuttavia, non sempre il confine è facile da tracciare, e dunque non è chiaro se e quando il suo impiego violi i principi dell'integrità della ricerca.

Una rilevazione pubblicata da *Nature* nel maggio 2025, condotta su circa cinquemila ricercatori, ha mostrato quanto ciò sia controverso: l'uso dell'IA per la traduzione o la revisione di un paper è ampiamente accettato e lecito, mentre le opinioni diventano molto più divise quando il sistema contribuisce direttamente alla scrittura scientifica. La ricerca, inoltre, mette in luce l'assenza di un consenso su quali impieghi debbano essere dichiarati e su come debba essere fatta in concreto tale comunicazione (D. Kwon, 2025).

La distinzione decisiva, dunque, non pare, ancora una volta, tra uso e non uso dell'IA, ma tra un impiego accessorio, controllato e verificabile e un impiego sostitutivo, nel quale al sistema vengono affidate funzioni propriamente scientifiche senza un'effettiva supervisione umana. Quanto più l'IA incide sui contenuti, sul metodo o sui risultati, tanto più rigorosi dovranno essere il controllo del ricercatore e la trasparenza sull'apporto ricevuto.

Lo *European Code of Conduct for Research Integrity* elaborato da ALLEA riconduce l'integrità della ricerca a quattro principi: affidabilità, onestà, rispetto e responsabilità. Nell'edizione del 2023 il Codice ha stabilito che l'uso di servizi esterni, strumenti automatizzati e sistemi di IA deve essere *compatibile* con le regole della disciplina e tale da consentire, quando necessario, *la verifica o la replicazione del lavoro*. Considera, inoltre, una pratica contraria al codice di condotta nascondere l'impiego dell'IA nella creazione di contenuti o nella redazione di pubblicazioni.

Ma quando e per quali ragioni l'uso di IA generativa costituisce una minaccia per l'integrità della ricerca?

Il ricorso all'intelligenza artificiale rileva anzitutto ai fini dell'*authorship*: un sistema di IA non può essere considerato autore in senso proprio, così che la responsabilità non può che rimanere interamente umana. Il

ricercatore deve quindi rispondere delle fonti citate, dell'esattezza dei dati, della correttezza metodologica e delle conclusioni, anche quando una parte del lavoro sia stata realizzata con l'ausilio di strumenti generativi. Da ciò deriva la necessità di una rigorosa *cross-validation*: ogni output rilevante dovrebbe essere verificato dal ricercatore mediante fonti autonome e affidabili. La plausibilità linguistica non può, cioè, diventare un surrogato della prova scientifica. La sorveglianza umana, tuttavia, non è affatto scontata e vanno considerate le difficoltà legate a quella che Walter Quattrociocchi (*L'era dell'epistemia*, Le Scienze, Agosto 2026, 44) chiama *epistemological fault line.* Se negli esseri umani, il giudizio è il risultato di sequenze di operazioni che collegano percezione, esperienza, motivazione, ragionamento e capacità di revisione, nei grandi modelli linguistici "si ha solo una catena di trasformazioni statistiche. Non c'è accesso diretto al mondo, non c'è verifica, non c'è possibilità di sospendere il giudizio". Queste differenze, secondo l'A., attraversano l'intero processo, cosicché ciò che cambia non è tanto il risultato, ma il meccanismo che lo produce. Detto altrimenti, il modo in cui si forma la conoscenza diventa *invisibile e irrilevante*, e ciò che conta è solamente la risposta. La conseguenza è, secondo l'A., che "la capacità di riaprire una risposta, verificarla e ricostruirne i presupposti richiede tempo e competenze elevate", così che l'output costituisce, per la maggioranza, non un punto di partenza – come dovrebbe essere – ma un punto di arrivo.

C'è poi il tema della trasparenza, che si collega al principio dell'onestà della ricerca: chi utilizza IA generativa dovrebbe dichiararlo esplicitamente. Il punto è delicato, dato che in molti settori ancora non esiste una prassi consolidata in tal senso, il che fa presumere che si sia dinnanzi ad un fenomeno di *shadow AI*. Non è chiaro se ciò dipenda dalla preoccupazione che l'impiego di questi modelli possa tradursi in uno stigma per il ricercatore, o se, più semplicemente, non vi sia ancora una adeguata consapevolezza della necessità di dichiararne l'impiego, o se, ancora, si colleghi a talune incertezze circa le modalità con cui assicurare tale trasparenza (ad esempio per qualunque uso, anche minimo dell'IA, o solo per usi sostanziali), ma l'impiego di questi modelli è per ora sottotraccia ed invisibile.

C'è infine il principio del rispetto, nei confronti dei colleghi, di chi partecipa alla ricerca, dell'intero ecosistema scientifico. L'inserimento di materiali in un prompt non è un'operazione neutra: il ricercatore potrebbe comunicare al fornitore testi protetti dal diritto d'autore, dati personali, risultati ancora inediti, informazioni riservate o documenti ricevuti nell'ambito di una revisione tra pari. Occorre pertanto conoscere le condizioni d'uso del servizio del modello, verificare se i contenuti siano utilizzati per l'addestramento e valutare le garanzie offerte in materia di conservazione, sicurezza e localizzazione dei dati.

Sul piano delle regole siamo di fronte ad interventi ancora limitati. Come è noto il regolamento dell'Unione europea (AI Act: reg. Ue 2024/1689) non concerne l'uso dell'IA per scopi di ricerca scientifica salva la previsione di alcuni obblighi in capo ai fornitori di sistemi a finalità generali (General Purpose AI), che però operano per così dire a monte, in fase di sviluppo e addestramento del modello. Qualcosa in più fa la legge italiana n. 132 del 2025, recante disposizioni e deleghe al Governo in materia di intelligenza artificiale, la quale, modificando l'art. 1 della legge sul diritto d'autore, precisa che sono protette le "opere dell'ingegno umano" di carattere creativo, anche quando realizzate con l'ausilio di strumenti di intelligenza artificiale, purché costituiscano il *risultato del lavoro intellettuale dell'autore,* così confermando la centralità dell'apporto umano.

In un contesto poco regolamentato è così l'autoregolazione ad assumere un'importanza decisiva. E in effetti, numerose università italiane e centri di ricerca, oltre che alcune riviste, hanno adottato linee guida, che pur nella varietà delle soluzioni, convergono sui principi della responsabilità personale, del dovere di *cross validation*, della trasparenza e del rispetto delle norme sulla protezione dei dati e del diritto d'autore, cui si somma spesso un richiamo all'uso proporzionato dell'IA generativa, in considerazione del significativo impatto ambientale di questi modelli.

Trasparenza, verifica delle fonti, protezione dei dati e attribuzione chiara dei contributi rappresentano, quindi, non un ostacolo all'innovazione, ma le condizioni necessarie perché l'innovazione non comprometta la fiducia nella scienza. Resta ora da costruire la cultura e la consapevolezza necessarie per comprendere la differenza tra assistenza e sostituzione, fra uso consapevole e automatismo, fra velocità della ricerca e rinuncia al giudizio critico.

È su questo terreno che università, riviste e comunità scientifica sono chiamate a confrontarsi per preservare la loro autorevolezza.

1. Bjelobaba, L. Waddington, M. Perkins, T. Foltýnek, S. Bhattacharyya, D. Weber-Wulff, *Maintaining research integrity in the age of GenAI: an analysis of ethical challenges and recommendations to researchers*, International Journal for Educational Integrity, 2025, 21-18, 1-23;
2. Chauhan, G. Currie*, The Impact of Generative Artificial Intelligence on Research Integrity in Scholarly Publishing, The American Journal of Pathology, Dec. 2024, vol. 194, n. 12, 2234.*
3. Kwon, *Is it ok for AI to write science papers*, Nature, 14 marzo 2025.
4. Ogunniye, A. Hana, *Research Integrity in the Era of Generative AI – A Perspective, https://ukcori.org/wp-content/uploads/2025/05/Research-Integrity-in-the-Era-of-Generative-AI-%E2%80%93-A-Perspective.pdf;*
5. Quattrociocchi*, L'era dell'epistemia*, Le Scienze, (trad Scientific America) Agosto 2026, 44.
6. Sisson, *The lab never sleeps. Can the Science keep up?*, Scientific America, 16 giugno 2026, 2026https://www.scientificamerican.com/article/autonomous-labs-are-running-science-experiments-24-7 ;
7. Holden Thorp, *AI in scientific publishing: slower, worse, and more expansive*, Science, July 2026, 225.
8. Vasconcelos, A. Marusic, *Gen AI and Research Integrity: Where to Now? https://pmc.ncbi.nlm.nih.gov/articles/PMC12018920/*.