La scienza può fidarsi dell'AI? La proposta di GAIDeT

di **Antonella De Robbio**

https://ilbolive.unipd.it/it/news/scienza-ricerca/scienza-puo-fidarsi-dellai-proposta-gaidet

25 GIUGNO 2026

Immagine generata dall'AI

L'avvento dell'intelligenza artificiale generativa ha innescato una trasformazione radicale e senza precedenti nelle pratiche di produzione della conoscenza, scuotendo le fondamenta stesse dell'epistemologia scientifica. Se da un lato i grandi modelli linguistici promettono di democratizzare la ricerca e accelerare la sintesi bibliografica, dall'altro hanno introdotto vulnerabilità sistemiche che minacciano l'integrità dell'intero ecosistema accademico. La manifestazione più evidente di questa transizione tecnologica è rappresentata dalle "**dispercezioni**" (**allucinazioni computazionali**), ovvero la generazione da parte degli algoritmi di dati statistici contraffatti, argomentazioni logiche fallaci e, soprattutto, citazioni bibliografiche interamente inventate ma formalmente inappuntabili.

Il fenomeno, che vede centinaia di migliaia di riferimenti bibliografici fittizi penetrare all'interno della letteratura di ricerca, si è capillarmente infiltrato nei meccanismi della catena della comunicazione scientifica. La contaminazione dei database bibliometrici rischia di innescare un processo di *autofagia informativa*, in cui i futuri modelli algoritmici verranno addestrati su un corpus scientifico già adulterato, con il rischio amplificare esponenzialmente il tasso di errore e di determinare un progressivo collasso dei modelli. Questo scenario delinea una crisi profonda della comunicazione scientifica, definibile come un vero e proprio *meltdown accademico* in cui le disfunzioni dell'intero ecosistema istituzionale della ricerca si riverberano sul logoramento professionale e deontologico dei singoli autori. Da un lato la logica stringente del *publish or perish* costringe da decenni i ricercatori a ritmi produttivi accelerati per accedere a fondi e avanzamenti di carriera, depotenziando l'efficacia della *peer review* e incentivando il ricorso acritico all'AI generativa; dall'altro, tale pressione strutturale si traduce in un diffuso *burnout* professionale: per velocizzare le procedure redazionali più onerose i ricercatori si affidano ad assistenti digitali opachi, esponendosi inconsapevolmente a

sanzioni interdittive capaci di compromettere un'intera carriera per l'"allucinazione" di una macchina.

Alcune istituzioni scientifiche infatti, di fronte a tale inquinamento testuale, hanno avviato processi di modifica dei propri codici di comportamento, chiamando gli autori a rispondere sul piano deontologico e professionale delle anomalie algoritmiche: va letto in questo contesto l'annuncio da parte del **repository open access arXiv** di voler comminare sospensioni fino a un anno per gli autori nei cui manoscritti vengano rintracciate citazioni o riferimenti inesistenti. Come chiarito ufficialmente a metà maggio 2026 dal presidente della sezione di Informatica (*Computer Science chair*) di arXiv, Thomas Dietterich, l'applicazione di questo provvedimento non configura l'introduzione di un nuovo impianto sanzionatorio d'emergenza; al contrario tale misura rappresenta la rigorosa esecuzione di una disposizione già radicata nel codice di condotta della piattaformache sancisce l'obbligo inderogabile in capo agli autori di assumersi la piena e totale responsabilità dei contenuti depositati. Sotto questo profilo l'allucinazione algoritmica non viene trattata come un'anomalia tecnologica esimente, bensì come una formale negligenza nella catena di custodia e validazione scientifica del testo.

Non tutti sono però convinti di un tale approccio. Natalie Khalil, fondatrice di Reviewer3, piattaforma con sede a San Francisco che utilizza l'intelligenza artificiale per aiutare i ricercatori a condurre la revisione paritaria, sostiene che arXiv stia curando il sintomo, non la causa principale: "**Se un ricercatore viene bannato da arXiv, continuerà comunque a fare ricerca altrove**". Una parte consistente della comunità accademica ha comunque accolto con sgomento e indignazione tali provvedimenti, interpretandoli come una reazione sproporzionata che ignora le pressioni strutturali del sistema universitario contemporaneo. Molti studiosi sostengono che l'onere di verificare capillarmente ogni singolo riferimento bibliografico, inclusi i minimi dettagli inseriti nelle appendici o nei materiali supplementari durante le convulse fasi che precedono le scadenze di sottomissione, rappresenti un carico di lavoro insostenibile. Alcuni accademici rivendicano una sorta di immunità o di scudo di responsabilità rispetto alle asserzioni errate generate dal mezzo tecnico, quasi come se l'intelligenza artificiale fosse un'entità autonoma della cui condotta l'autore umano non debba rispondere.

"Le allucinazioni dell'AI non sono semplici errori tecnici: rischiano di contaminare la letteratura scientifica e i futuri modelli algoritmici

Questi atteggiamenti palesano però un certo fraintendimento del concetto di paternità intellettuale. Dall'altro lato, la dirigenza editoriale delle principali piattaforme e riviste scientifiche ribadisce un principio etico fondamentale e non negoziabile: l'assenza di un controllo umano sui risultati generati dai modelli linguistici mina alla radice la fiducia riposta nell'intero manoscritto. Se un autore dimostra negligenza nella verifica degli elementi più facilmente controllabili, come l'esistenza reale di una fonte citata o la veridicità di una nota a piè di pagina, decade la legittimità scientifica dell'intera opera, poiché diventa impossibile stabilire se i dati empirici o le conclusioni teoriche siano il frutto di una rigorosa indagine o di un'ulteriore allucinazione algoritmica.

I filtri di moderazione dei *preprint server* e i processi di *peer review* convenzionali si stanno dimostrando inefficienti nel rilevare le falsificazioni introdotte dall'intelligenza artificiale, poiché i riferimenti inventati vengono spesso inseriti in modo speculativo all'interno di una prosa accademica per il resto impeccabile ed erudita, ingannando anche i revisori più esperti. In questo scenario di vulnerabilità istituzionale e

smarrimento deontologico emerge l'assoluta urgenza di superare la logica della sanzione puramente punitiva ex-post per abbracciare un paradigma di trasparenza proattiva e metodologica.

Diventa così indispensabile fornire ai ricercatori un quadro di riferimento chiaro che permetta loro di utilizzare le potenzialità dell'automazione senza incorrere in violazioni dell'integrità scientifica. È in questo preciso vuoto normativo e procedurale, in un contesto di responsabilità editoriali, che si inserisce la proposta della ***Generative AI Delegation Taxonomy*****, nota con l'acronimo GAIDeT**, sviluppata attraverso un rigoroso processo di costruzione del consenso basato sull'analisi della letteratura scientifica e dei *framework* preesistenti, con l'obiettivo di mappare con precisione millimetrica l'impiego degli algoritmi generativi all'interno del flusso di lavoro della ricerca.

Fino a questo momento, i tentativi di regolamentazione si erano limitati a raccomandazioni generiche o all'adattamento di tassonomie storiche come CRediT (*Contributor Role Taxonomy*), la quale nasce però per catalogare i 14 ruoli contributivi di paternità, escludendo per definizione gli *input* derivanti da sistemi non umani. Altre strutture come quelle proposte dal NIST (*National Institute of Standards and Technology*) hanno introdotto nel 2024 criteri tassonomici che definivano 16 aree di utilizzo dell'AI, indipendenti dalle tecniche e dai domini di applicazione. I compiti venivano così intesi come combinazioni di una o più attività strutturali; tali criteri mancavano tuttavia di una declinazione specifica per le singole e peculiari fasi che compongono l'attività scientifica

GAIDeT supera tali limitazioni offrendo un approccio sì olistico ma strettamente parcellizzato, che scompone l'impresa scientifica in macro-domini e micro-attività, costringendo il ricercatore a **dichiarare non soltanto se ha utilizzato l'intelligenza artificiale, ma anche l'esatta estensione e la natura del compito delegato**. La tassonomia GAIDeT articola l'attività di ricerca scientifica attraverso sette macro-domini fondamentali, che coprono l'intero ciclo di vita del prodotto accademico: la concettualizzazione della ricerca, la revisione della letteratura, la strutturazione della metodologia, l'analisi e l'elaborazione dei dati, la stesura testuale del manoscritto, la supervisione scientifica e, infine, la valutazione dei risvolti etici.

"GAIDeT non chiede di vietare l'AI, ma di documentarne in modo rigoroso il ruolo in ogni fase della ricerca scientifica

All'interno di ciascuna di queste macro-aree, GAIDeT identifica una serie di micro-attività specifiche, associando a ciascuna di esse un determinato livello di supervisione umana richiesto e un protocollo di rendicontazione trasparente. Nel dominio della concettualizzazione, ad esempio, l'uso dell'intelligenza artificiale può spaziare dal mero *brainstorming* per la generazione di idee preliminari fino alla formulazione di ipotesi complesse. Nel campo cruciale della revisione della letteratura, laddove si annida con maggiore frequenza il rischio di citazioni inventate e distorsioni interpretative, la tassonomia impone una distinzione netta tra l'impiego dell'algoritmo come motore di ricerca semantico, come strumento di sintesi di testi preesistenti o come assistente per l'estrazione di metriche. Obbligando l'autore a specificare il ruolo esatto della macchina, GAIDeT sposta il focus dall'ammissibilità astratta dello strumento alla tracciabilità concreta del suo utilizzo, offrendo a revisori ed editori una mappa dettagliata per orientare le proprie verifiche ispettive.

Il nucleo filosofico di GAIDeT risiede nel principio della delega consapevole, secondo cui l'adozione dell'intelligenza artificiale — intesa non come co-autore ma come agente subordinato — non solleva l'essere umano dalla piena responsabilità etica e legale del lavoro. Di fronte alla crisi di *accountability* evidenziata dalle sanzioni dei *repository* di *preprint*, questo modello impone che la supervisione umana si configuri come un processo continuo di validazione epistemica e mai come un mero atto formale. Esigendo la documentazione esplicita e pubblica del grado di coinvolgimento algoritmico in ogni sezione del manoscritto, la tassonomia funge da deterrente procedurale contro l'uso acritico della tecnologia.

L'obbligo di tracciabilità richiama l'autore al dovere di accuratezza, disinnescando alla radice la negligenza che genera il fenomeno delle citazioni fittizie. Per evitare il rischio di un ulteriore aggravio burocratico sui ricercatori, il *framework* teorico è integrato dal *GAIDeT Declaration Generator*, un applicativo digitale ad accesso aperto su GitHub. Attraverso un questionario sistematico, lo strumento traduce l'uso dell'AI in stringhe di metadati e dichiarazioni standardizzate, leggibili sia dall'operatore umano sia dai sistemi di indicizzazione automatica. Questa normalizzazione ottimizza l'intera filiera editoriale: fornisce agli autori un percorso certo per legittimare l'uso della tecnologia al riparo da accuse di frode; consente ai revisori di calibrare lo scetticismo professionale sulle aree a maggiore apporto algoritmico; permette infine a editori e moderatori di implementare filtri di coerenza pre-pubblicazione, riducendo drasticamente i provvedimenti di esclusione o di respingimento dei manoscritti.

L'integrazione di GAIDeT nelle pratiche editoriali offre una risposta strutturale alla polarizzazione tra una **tecnofobia punitiva** – tesa a bandire i modelli linguistici o a sanzionarne gli errori con l'esclusione accademica temporanea – e un'**utopia tecnofila irresponsabile**, che persegue l'automazione rifiutando i doveri della paternità scientifica. La crisi generata dalle sanzioni dei *repository* di *preprint* evidenzia una transizione critica, in cui i tradizionali paradigmi basati sulla fiducia presunta non bastano più a garantire la stabilità della conoscenza di fronte alla produzione testuale sintetica di massa. La soluzione non risiede nel rigetto anacronistico della tecnologia, né nella tolleranza del declino qualitativo per l'urgenza di pubblicare, bensì nella formalizzazione della trasparenza: la dichiarazione dell'uso dell'IA cessa così di essere percepita come un'ammissione di colpa, elevandosi a indicatore di rigore metodologico e onestà intellettuale.

L'adozione diffusa e standardizzata di GAIDeT da parte delle riviste scientifiche, delle società accademiche e delle istituzioni di finanziamento della ricerca permetterebbe di salvaguardare l'integrità della letteratura scientifica globale, sollevando i ricercatori dall'ansia della sanzione cieca e guidandoli verso una collaborazione sicura e trasparente con le tecnologie algoritmiche. La transizione verso una scienza aperta non può prescindere da una chiara codificazione dei rapporti di delega algoritmica, e la **tassonomia GAIDeT traccia la rotta metodologica necessaria per navigare con sicurezza i tormentati mari della rivoluzione dell'intelligenza artificiale**.