AIB notizie
L'informazione dell'Associazione Italiana Biblioteche

Linee guida dinamiche sull'uso responsabile dell'intelligenza artificiale generativa nella ricerca

In primo piano 14 Maggio 2026 0

https://aibnotizie.aib.it/linee-guida-dinamiche-sulluso-responsabile-dellintelligenza-artificiale-generativa-nella-ricerca/

Linee guida dinamiche sull'uso responsabile dell'intelligenza artificiale generativa nella ricerca

Da Antonella De Robbio riceviamo e volentieri pubblichiamo.

Contesto

Nel maggio 2026 sono state completate e formalmente emanate le linee guida sull'uso responsabile dell'intelligenza artificiale generativa nella ricerca, testo che rappresenta la terza edizione del documento redatto in lingua inglese *"Living guidelines on the responsible use of generative AI in research",* frutto di uno sforzo sinergico e congiunto tra la Commissione Europea – nello specifico la Direzione Generale per la Ricerca e l'Innovazione, Direzione E (Prosperità), Unità E4 (DG RTD.E) (L'intelligenza artificiale nella scienza e nelle tecnologie critiche) – e gli Stati membri unitamente alle parti interessate rappresentate all'interno del Forum dello Spazio Europeo della Ricerca (Forum ERA). Il carattere "dinamico" di queste linee guida nasce dall'esigenza di avere uno strumento flessibile, capace di adattarsi ai rapidi cambiamenti dell'intelligenza artificiale, concepito per essere costantemente aggiornato in relazione ai rapidissimi e imprevedibili sviluppi tecnologici delle architetture neurali generative. Il documento si propone di fornire raccomandazioni chiare, semplici e attuabili ai principali attori della comunità scientifica, articolando le responsabilità individuali dei ricercatori, i doveri strutturali delle organizzazioni di ricerca e i compiti di sorveglianza etico-finanziaria propri degli enti finanziatori.

Il quadro concettuale

Il quadro concettuale elaborato in sede europea si innesta sull'eredità epistemologica del *Codice di condotta europeo per l'integrità della ricerca* (i ricercatori devono operare con affidabilità, onestà, rispetto e responsabilità) e sulle riflessioni teoriche in materia di un'intelligenza artificiale affidabile. Le linee guida europee sull'uso dell'intelligenza artificiale nella ricerca si inseriscono comunque nel quadro regolatorio più ampio definito dall'AI Act, che rappresenta il primo tentativo organico di regolamentazione dell'IA a livello europeo. Mentre l'AI Act stabilisce obblighi giuridici vincolanti basati su un approccio proporzionato al rischio, le linee guida rivolte alla comunità scientifica hanno una natura più dinamica e adattiva, pensata per accompagnare

l'evoluzione rapida delle tecnologie. In questo senso, il rapporto tra i due livelli non è gerarchico ma complementare: da un lato la normativa fissa i principi fondamentali di sicurezza, trasparenza e responsabilità; dall'altro, le linee guida operano come strumenti di orientamento pratico per i ricercatori, favorendo l'adozione di buone pratiche e la loro costante revisione. L'obiettivo comune è evitare sia un vuoto regolatorio sia un eccesso di rigidità, costruendo un ecosistema in cui l'innovazione scientifica possa svilupparsi entro confini chiari, ma non soffocanti, e in cui la governance dell'IA rimanga in grado di adattarsi alle trasformazioni tecnologiche e ai nuovi scenari di utilizzo nella ricerca. Da questa sintesi scaturiscono quattro pilastri fondamentali che devono orientare in modo trasversale la condotta di scienziati, istituzioni e organi di finanziamento.

Il primo pilastro è rappresentato dall'affidabilità, intesa come l'imperativo metodologico volto a garantire la qualità intrinseca della ricerca. L'introduzione di strumenti stocastici nel flusso di lavoro scientifico impone una rigorosa validazione empirica dei dati generati, l'obbligo della riproducibilità e un controllo rigoroso della qualità scientifica volta a identificare e neutralizzare le distorsioni annidate nei dataset di addestramento.

Il secondo principio esige un regime di onestà e trasparenza radicale. L'adozione di supporti algoritmici non può rimanere confinata nella sfera privata o imperscrutabile del ricercatore, ma deve essere mostrata concretamente, consentendo alla comunità scientifica dei pari di valutare l'effettiva estensione e profondità del contributo computazionale.

Il terzo punto estende l'orizzonte etico oltre i confini antropocentrici della comunità scientifica, includendo il rispetto per i diritti di proprietà intellettuale, la riservatezza dei soggetti umani coinvolti e la sostenibilità ecosistemica. Quest'ultimo punto richiede una valutazione critica dell'impronta ambientale derivante dai cicli di vita energetici dell'hardware e del software dedicati all'addestramento e all'interrogazione dell'intelligenza artificiale.

Infine, il principio di responsabilità conferma che l'intervento e la supervisione umana sono sempre necessari e che l'autonomia dei sistemi computazionali non esonera mai l'essere umano dalle conseguenze etiche, giuridiche e scientifiche dei risultati pubblicati.

La dirompenza tecnologica e la necessità di una governance dinamica

La prassi scientifica attraversa storicamente cicli di ridefinizione metodologica determinati dall'introduzione di nuovi strumenti di ricerca e metodi di indagine scientifica. Tuttavia, l'avvento e la rapidissima diffusione dei modelli di intelligenza artificiale generalisti, e nello specifico dell'intelligenza artificiale generativa, configurano una transizione paradigmatica che eccede la mera evoluzione strumentale. I modelli di base, addestrati su macro-orizzonti di dati non etichettati, manifestano proprietà emergenti capaci di simulare il ragionamento transdisciplinare, la sintesi euristica e la produzione testuale con standard qualitativi che insidiano la tradizionale demarcazione tra l'intelletto umano e l'output computazionale. Se da un lato l'integrazione di tali architetture neurali accelera la scoperta scientifica, ottimizza l'efficacia dei processi di codifica algoritmica e democratizza l'accesso alla letteratura internazionale per i ricercatori operanti in contesti svantaggiati o non madrelingua, dall'altro introduce vulnerabilità sistemiche che minacciano le fondamenta stesse dell'integrità accademica.

Di fronte a risposte istituzionali ancora frammentate, il Forum dello Spazio Europeo della Ricerca ha definito un quadro etico e metodologico flessibile, pensato per adattarsi all'evoluzione delle tecnologie. L'obiettivo dell'Unione Europea non è introdurre regole rigide e pervasive, che rischierebbero di frenare la ricerca e l'innovazione, ma promuovere una cultura della responsabilità condivisa. L'approccio punta a collegare i principi tradizionali dell'etica scientifica alle nuove sfide

legate alla scarsa trasparenza degli algoritmi, cercando di offrire maggiore chiarezza in una fase ancora incerta dal punto di vista metodologico.

La responsabilità del ricercatore e le questioni di paternità intellettuale

L'analisi della condotta scientifica focalizza l'attenzione sulla figura del ricercatore, inteso come l'agente critico terminale del processo di validazione della conoscenza. Le linee guida sanciscono un principio dogmatico fondamentale: i sistemi di intelligenza artificiale generativa mancano di soggettività giuridica, di agenzia etica e di intenzionalità scientifica; pertanto, non possono in alcun modo rivendicare la paternità o la co-paternità dell'opera scientifica, né essere indicati come autori o co-autori nei contributi accademici. La responsabilità dell'integrità dei contenuti prodotti ricade interamente sull'essere umano, il quale deve sviluppare un'ermeneutica critica nei confronti delle patologie intrinseche alla tecnologia stocastica.

Tra le criticità tecniche più insidiose figurano le distorsioni dei dati di addestramento e le distorsioni da sollecitazione. La distorsione da sollecitazione (*prompt bias*) è la tendenza dei modelli di intelligenza artificiale generativa a conformare e calibrare le proprie risposte alle aspettative, ipotesi o preconcetti espressi dall'utente nel quesito iniziale. Questa distorsione si manifesta come una forma di comportamento adulatore (*sycophancy behavior*) da parte dell'algoritmo. Se il ricercatore formula una richiesta parziale o sbilanciata, l'algoritmo stocastico tenderà ad assecondarla per massimizzare la congruenza con l'input, amplificando il bias di conferma e compromettendo l'obiettività scientifica dell'output.

A ciò si aggiunge il ben noto fenomeno delle dispercezioni (erroneamente definite allucinazioni), ovvero la generazione algoritmica di informazioni plausibili ma prive di corrispondenza empirica, che si manifesta nella fabbricazione di citazioni bibliografiche inesistenti o nell'erronea sintesi concettuale di saggi reali. L'opacità dei modelli, strutturati come impenetrabili scatole nere, rende impossibile tracciare il percorso inferenziale che ha condotto a una determinata conclusione, rendendo la validazione incrociata e l'approccio comparativo requisiti metodologici imprescindibili per qualsiasi analisi automatizzata.

Sotto il profilo operativo, il ricercatore è chiamato a discernere accuratamente tra uso meramente sussidiario o editoriale dello strumento (come la correzione stilistica, la traduzione o il supporto ortografico, che non configurano un utilizzo sostanziale e non richiedono obbligatoriamente una dichiarazione formale, sebbene la trasparenza sia sempre incoraggiata) e uso sostanziale. Quest'ultimo livello d'impiego riguarda l'interpretazione dei dati, la conduzione di rassegne della letteratura, la formulazione di ipotesi o l'individuazione di lacune epistemiche. L'uso sostanziale esige la menzione esplicita dello strumento all'interno del lavoro scientifico, specificando in modo dettagliato le modalità e l'estensione dell'applicazione algoritmica.

In linea con i principi della scienza aperta, i ricercatori sono invitati a rendere pubblici i criteri utilizzati e i relativi risultati generati dai sistemi informatici, organizzandoli in modo chiaro e accessibile, così da permetterne la consultazione e l'analisi da parte della comunità scientifica. Un ulteriore capitolo di stringente rilevanza etica concerne la sicurezza dei dati e la protezione della sfera privata. L'immissione di testi inediti, bozze di manoscritti o dati sensibili all'interno di piattaforme commerciali generaliste comporta il rischio concreto di una cessione involontaria di proprietà intellettuale, poiché tali input vengono frequentemente utilizzati dalle aziende fornitrici per l'addestramento incrementale dei propri modelli, determinando una dispersione incontrollabile del patrimonio scientifico. Il trattamento dei dati personali richiede, pertanto, l'assoluta conformità con le normative sulla protezione dei dati, subordinando l'inserimento di informazioni personali al consenso esplicito degli interessati o a solide basi giuridiche equivalenti.

L’ecosistema istituzionale: verso infrastrutture protette e formazione continua

Sul piano regolamentare, gli enti di ricerca sono invitati a integrare i dettami delle linee guida dinamiche all’interno dei propri codici etici e statuti interni, operando adattamenti specifici guidati dal principio di proporzionalità e dalle peculiarità delle singole discipline scientifiche. È fondamentale, a tal proposito, che le istituzioni disinneschino i meccanismi di stigmatizzazione sociale legati all’uso dell’intelligenza artificiale nella scrittura accademica. Le università devono promuovere un clima di trasparenza in cui la dichiarazione dell’uso di strumenti algoritmici sia considerata un indice di rigore metodologico e di onestà intellettuale.

Se il ricercatore rappresenta l’avamposto etico della scienza, le organizzazioni di ricerca – università, laboratori, accademie e istituti scientifici – costituiscono l’infrastruttura strutturale e culturale che deve abilitare tale responsabilità. Le raccomandazioni europee esortano gli enti accademici a non limitarsi a una postura difensiva, proibitiva o meramente sanzionatoria, ma a farsi promotori attivi di una transizione digitale eticamente orientata attraverso la formulazione di linee guida interne chiare e contestualizzate. La prima necessità strutturale individuata dal documento è lo sviluppo e l’implementazione di soluzioni tecnologiche controllate, ospitate preferibilmente su infrastrutture locali o su cloud sovrani gestiti da consorzi istituzionali accreditati. Solo la disponibilità di ambienti computazionali protetti e isolati può consentire al personale scientifico di sfruttare le potenzialità dell’intelligenza artificiale generativa sui propri dataset proprietari e sui manoscritti non ancora pubblicati, garantendo per via tecnologica, e non solo procedurale, l’assoluta riservatezza, l’integrità dei dati e la protezione dei diritti di proprietà intellettuale rispetto alle ingerenze commerciali esterne.

Parallelamente alla dimensione infrastrutturale, le organizzazioni devono investire nella riconversione culturale del proprio capitale umano intellettuale. È richiesto l’allestimento di programmi di formazione permanente rivolti a tutti i livelli della carriera accademica, estesi anche al personale tecnico-amministrativo e ai gestori della ricerca. Tali percorsi formativi non devono focalizzarsi esclusivamente sul mero addestramento tecnico all’uso dei software o sull’ottimizzazione delle richieste, bensì sulla comprensione teorica delle limitazioni statistiche dei modelli, sulle strategie avanzate di verifica per il rilevamento del plagio e delle allucinazioni, e sulla gestione critica dei condizionamenti di genere, etnici o disciplinari sedimentati negli algoritmi.

Gli enti finanziatori e la gestione delle relazioni con terze parti

L’ultimo anello della catena di governance scientifica è costituito dalle agenzie di finanziamento della ricerca, il cui ruolo è determinante nel plasmare gli standard metodologici attraverso la leva economica dell’allocazione delle risorse. Gli enti finanziatori devono predisporre sistemi di rendicontazione chiari e standardizzati, richiedendo ai candidati di esplicitare l’eventuale uso dell’intelligenza artificiale generativa sia nella fase di stesura della proposta di ricerca sia nella futura esecuzione del progetto finanziato. Allo stesso tempo, le agenzie sono chiamate a esercitare una vigilanza stringente sui propri meccanismi interni di valutazione. Sebbene l’intelligenza artificiale generativa possa utilmente coadiuvare l’ottimizzazione dei flussi amministrativi ed editoriali interni, l’impianto regolamentare europeo esclude categoricamente la possibilità di delegare agli algoritmi la valutazione del merito scientifico delle proposte o la redazione dei rapporti di revisione. L’affidamento della valutazione scientifica a una logica computazionale imperscrutabile violerebbe i principi elementari di equità, imparzialità e rispetto del lavoro intellettuale dei candidati, esponendo il processo decisionale a distorsioni impreviste e allucinazioni valutative. Inoltre, il documento evidenzia la necessità per gli enti finanziatori di proteggere la riservatezza delle proposte ricevute. L’inserimento dei testi dei progetti all’interno di strumenti di intelligenza artificiale generativa esterni durante le fasi istruttorie rappresenta una grave violazione

della riservatezza e della proprietà intellettuale dei proponenti. Le agenzie devono pertanto garantire che i propri valutatori esterni siano pienamente consapevoli di tale divieto e che i processi di revisione rimangano saldamente ancorati al giudizio e alla responsabilità critica dell'essere umano.

L'orizzonte regolamentare delineato dal Forum ERA non si esaurisce nelle relazioni endogene alla comunità scientifica, ma prende in esame l'interazione con attori esterni e terze parti. Nello svolgimento delle attività accademiche, i ricercatori e le organizzazioni si interfacciano costantemente con una pluralità di soggetti, quali consulenti industriali, partner di consorzi internazionali, decisori politici o uditori esterni, che potrebbero impiegare strumenti di intelligenza artificiale generativa all'insaputa dell'istituzione ricevente.

L'utilizzo di assistenti virtuali commerciali o di strumenti automatizzati durante incontri istituzionali, conferenze telematiche o tavoli tecnici per la verbalizzazione automatica, la trascrizione dei dialoghi o la sintesi estemporanea di documenti strategici configura un rischio rilevante per la sicurezza della ricerca. Informazioni coperte da segreto industriale, brevetti in fase di deposito o dati sensibili relativi a tecnologie critiche rischiano di essere catturati dalle interfacce di terze parti e immessi nei circuiti di addestramento delle multinazionali tecnologiche, vanificando le tutele giuridiche faticosamente predisposte dagli enti di ricerca. Pertanto, le linee guida richiamano l'attenzione sulla necessità di una comunicazione preventiva e trasparente: gli scienziati e le istituzioni europee devono informare i propri interlocutori esterni circa i rischi connessi all'uso di tali strumenti, esigendo reciprocità nella dichiarazione dei software impiegati e concordando protocolli comuni per la protezione dei dati e del patrimonio intellettuale collettivo. Questo approccio si inserisce nel più ampio quadro delle iniziative europee volte a rafforzare la sicurezza della ricerca a fronte di interferenze esterne e vulnerabilità sistemiche legate al trasferimento tecnologico globale.

La democrazia partecipativa ed epistemica nella co-creazione della governance: il meccanismo del feedback continuo

La natura ontologicamente dinamica del framework europeo non si esaurisce nella flessibilità teorica delle sue raccomandazioni, ma si sostanzia in un preciso dispositivo di democrazia partecipativa ed epistemica. Consapevoli dell'impossibilità di cristallizzare un fenomeno tecnologico in perenne mutazione attraverso un testo normativo statico, la Commissione Europea e il Forum ERA hanno implementato un canale permanente di consultazione pubblica, accessibile attraverso la piattaforma ufficiale *EUSurvey*. Questo strumento di indagine non si configura come un'indagine demoscopica ex post, né come un mero questionario di gradimento burocratico, bensì come un ingranaggio metodologico essenziale per il funzionamento incrementale delle linee guida stesse. Il sondaggio rappresenta l'interfaccia operativa attraverso cui la comunità scientifica transnazionale, i valutatori, i gestori della ricerca e i singoli cittadini possono esercitare un'agenzia critica, segnalando in tempo reale le lacune applicative, l'emergere di nuove vulnerabilità algoritmiche o l'inadeguatezza di alcuni protocolli rispetto alle specificità di particolari domini disciplinari.

Sotto il profilo della sociologia della scienza e delle politiche pubbliche, questo modello di consultazione permanente sposta il baricentro della regolamentazione da un approccio rigidamente calato dall'alto (*top-down*) a un paradigma collaborativo e distribuito (*bottom-up*). I flussi di dati e i commenti qualitativi raccolti tramite il modulo di feedback online vengono periodicamente analizzati ed elaborati dall'Unità E4 della Direzione Generale per la Ricerca e l'Innovazione. Questo monitoraggio empirico costante fornisce la materia prima informativa necessaria per operare le successive correzioni, integrazioni e riscritture del testo. In tal modo, la transizione dalla terza edizione del maggio 2026 verso le versioni future non avverrà sulla base di astratte speculazioni

giuridiche, ma sarà guidata dall'esperienza d'uso concreta e dalle frizioni metodologiche vissute quotidianamente dai ricercatori nei laboratori e nelle università. Il sondaggio si trasforma così in uno strumento di co-creazione istituzionale, capace di saldare la severità del rigore etico con la fluidità della prassi computazionale, garantendo che lo Spazio Europeo della Ricerca disponga di una cornice regolamentare costantemente allineata alla frontiera dell'innovazione tecnologica.

Per un'ecologia critica della scienza algoritmica

Le linee guida dinamiche sull'uso responsabile dell'intelligenza artificiale generativa nella ricerca rappresentano una tappa fondamentale nella definizione delle politiche scientifiche dello Spazio Europeo della Ricerca. Esse riflettono la consapevolezza che la tecnologia non è un elemento neutro rispetto al processo di costruzione del sapere, ma un agente trasformativo capace di riconfigurare le strutture cognitive, linguistiche ed etiche dei ricercatori. L'impianto teorico del documento rifiuta sia il rifiuto accademico aprioristico, che negherebbe i benefici di straordinaria accelerazione scientifica offerti dall'intelligenza artificiale, sia l'ottimismo tecnologico acritico, che delegherebbe la validazione della conoscenza a meri automatismi stocastici.

L'istanza fondamentale sollevata dalle istituzioni europee risiede nella capacità della comunità scientifica di rendere queste linee guida effettivamente dinamiche, traducendole in pratiche quotidiane di monitoraggio, formazione ed evoluzione normativa. Solo attraverso un controllo critico e costante, una robusta infrastruttura tecnologica protetta e la difesa della centralità umana come fulcro insostituibile del giudizio e della responsabilità scientifica sarà possibile garantire che l'intelligenza artificiale generativa rimanga un potente strumento al servizio del progresso conoscitivo e del bene pubblico, preservando al contempo il rigore metodologico, l'onestà e l'autenticità etica della scoperta scientifica.

……
……..

Antonella De Robbio, *Coordinatore del Gruppo di Studio sulle Politiche dell'Informazione dell'AIB – Associazione Italiana Biblioteche*

intelligenza artificiale, *intelligenza artificiale generativa*