

L'AI addestrata su dataset storici sfida gli umanisti digitali al tavolo dell'*Imitation Game*

Federico Boschetti e Angelo Mario Del Grosso, CNR-ILC

Introduzione e contesto

L'intelligenza artificiale generativa ha superato ormai la fase sperimentale nei laboratori e comincia a entrare nei processi di produzione e fruizione culturale. Gli algoritmi a reti neurali profonde, addestrati su testi e immagini del patrimonio culturale immateriale, non sono più solo strumenti di supporto alla creazione di contenuti: agiscono ormai come co-autori e mediatori cognitivi tra passato e presente.

Questa vicinanza fra umano e artificiale apre questioni decisive. Come preservare l'autenticità quando la simulazione si avvicina asintoticamente all'originale? Come distinguere la verità storica dalla verisimiglianza generativa? E in che misura le macchine, addestrate sui nostri archivi, stanno già reinterpretando – o alterando – quell'Intangible Cultural Heritage che credevamo di custodire?

Il rapporto fra scienze umane e tecnologia non può più essere inteso come applicazione di strumenti digitali a contenuti culturali. Il digitale è ormai parte dei processi di interpretazione. Le scienze umane devono addestrare le macchine a leggere la cultura e, al tempo stesso, imparare da esse nuovi modi di interrogare il passato. In questo contesto, il recupero dell'atteggiamento critico è un'urgenza sociale trans-generazionale: riconoscere il documento autentico da un surrogato sintetico diventa parte della nuova alfabetizzazione culturale.

Siamo passati da una cultura della memoria a una cultura della (ri)generazione automatica, in cui il passato non si conserva ma si ricrea, deformandolo. Gli umanisti digitali hanno un duplice compito: garantire la qualità dei dati storici che alimentano i modelli e mantenere una consapevolezza critica dei processi cognitivi e sociali che tali modelli simulano o amplificano.

Problema di ricerca e domande di discussione

Le AI generative ottimizzate su dataset storici non si limitano a imitare la cultura: la ricreano, filtrandola attraverso bias, pregiudizi, senso comune, stereotipi, censure e omissioni. Da questa prospettiva emergono tre questioni principali.

1. *Dialogo interdisciplinare*. Come rafforzare la collaborazione tra ricercatori in ambito umanistico e informatici, a partire da enti pubblici come il CNR, aprendosi al mondo accademico e all'industria culturale? Quali modelli di valutazione e di cura dei dati possono garantire qualità, interoperabilità e trasparenza nei dataset storici?

Come costruire dataset di dominio annotati con rigore filologico, capaci di rappresentare la complessità culturale del dato umanistico?

2. *Ingegneria della cooperazione*. Come progettare spazi virtuali che favoriscano la collaborazione reale tra discipline, metodologie e scuole di pensiero che si sono sviluppate tra generazioni diverse di studiosi? I Virtual Research Environments, ispirati a modelli come D4Science, possono diventare luoghi dove annotazione, revisione e valutazione siano integrate in un processo continuo? Un helpdesk virtuale con agenti conversazionali di dominio può favorire l'avvicinamento fra comunità di ricerca molto diverse per formazione e obiettivi?

3. *Sostenibilità a lungo termine*. Come rendere durature le risorse digitali nate tramite i finanziamenti post-COVID? È possibile progettare agenti conversazionali di dominio che evolvano in assistenti cognitivi per la ricerca e la didattica? L'AI generativa può contribuire alla manutenzione del software – creando aggiornamenti in autonomia – e il ricorso ad architetture cloud può assicurare la sostenibilità dei progetti digitali di ambito culturale?

Queste domande non richiedono solo soluzioni tecniche, ma una riflessione condivisa su metodo, infrastruttura e sostenibilità in un ecosistema sempre più ibridato fra umano e artificiale.

Approccio proposto: una rivisitazione dell'*Imitation Game* di Turing

Per esplorare il nuovo equilibrio fra intelligenza umana e intelligenza artificiale, proponiamo una rilettura dell'*Imitation Game* di Alan Turing.

Nel test originario, un essere umano deve capire, solo attraverso messaggi scritti, se l'interlocutore è una macchina o un altro essere umano. Il successo dell'AI consiste nell'inganno: convincere l'interlocutore della propria umanità.

Nel campo delle Digital Humanities, la prospettiva si rovescia. Non si tratta più di verificare se la macchina imita tanto bene l'uomo da ingannarlo, ma di comprendere se l'uomo sappia ancora distinguere se stesso dalla macchina.

Immaginiamo un *Imitation Game* triplo, in cui non solo l'essere umano deve discernere la natura dell'interlocutore, ma due AI – una generalista e una specializzata su dataset umanistici – devono a loro volta stabilire se i testi o le risposte provengano da esseri umani o da altri sistemi automatici.

Questo schema richiama la logica delle Generative Adversarial Networks: un modello genera contenuti, un altro li valuta. Ma l'interesse non sta nell'inganno, bensì negli errori di riconoscimento, dove emerge la distanza fra imitazione e comprensione. Ogni errore diventa occasione di conoscenza reciproca: per capire dove si colloca la differenza e come può essere gestita o valorizzata.

Nel laboratorio dell'*Imitation Game*, le macchine apprendono dalle pratiche interpretative degli umanisti; gli umanisti, a loro volta, imparano dalle macchine a riconoscere regolarità e deviazioni.

Si forma così un circolo epistemico virtuoso: l'uomo addestra la macchina, la macchina produce nuove interpretazioni, l'uomo le valuta criticamente e restituisce alla macchina nuovi criteri sia di analisi sia di sintesi. È un processo di apprendimento reciproco più che una competizione.

Discussione e conclusioni

Il nuovo *Imitation Game* tra umanisti e AI generativa solleva una domanda di fondo: cosa significa oggi comprendere un testo, un'immagine, un'opera multimediale?

Se la comprensione si riduce a correlazione statistica, la cultura diventa un insieme di pattern riproducibili; se invece implica riconoscimento, distanza critica e responsabilità interpretativa, la sfida non è tecnologica ma profondamente umana. Da qui derivano tre direzioni operative.

1. *Educazione alla distinzione*. L'umanista digitale dovrà fornire gli strumenti per riconoscere la differenza tra un documento autentico e un contenuto generato non solo in base alla fonte, ma alla coerenza semantica e contestuale. Le competenze metalinguistiche e critiche diventano parte del patrimonio immateriale da trasmettere: non basta sapere che cosa viene detto, ma comprendere come e perché viene detto.

2. *Ingegnerizzazione della cooperazione*. Le infrastrutture digitali devono evolvere in ecosistemi cooperativi, dove produzione, revisione e cura dei dati siano processi integrati. Helpdesk intelligenti e VRE di nuova generazione possono favorire una cooperazione reale, se progettati con consapevolezza dei loro limiti e delle loro implicazioni epistemologiche.

3. *Sostenibilità cognitiva e tecnologica*. La sostenibilità non è solo economica ma cognitiva: riguarda la continuità delle conoscenze incorporate nei sistemi digitali. L'AI può contribuire alla manutenzione e all'aggiornamento delle risorse, ma solo se accompagnata da una curatela umanistica costante, capace di adattare i modelli ai mutamenti culturali.

In conclusione, l'intelligenza artificiale generativa addestrata su dataset storici non è più un semplice strumento, ma un interlocutore della conoscenza. La sfida per gli umanisti non è difendere l'umanità dall'imitazione, ma ridefinirla nella capacità di comprendere e di avere cura: delle parole, delle immagini, dei dati e del passato che continua a riemergere, anche quando parla con voce artificiale ... e la "metasfida" è indovinare in che percentuale l'AI generativa è co-autrice di questo abstract! ;-)