

Nota dell'Associazione Italiana Editori
Audizione VII Commissione su
Tutela diritto d'autore e intelligenza artificiale, Risoluzione 7-00185 Amorese
Camera dei deputati
12 marzo 2023

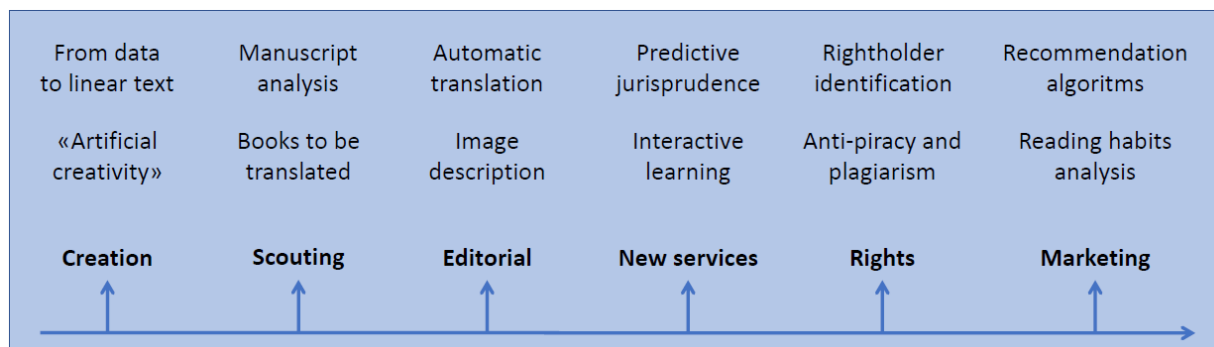
L'Associazione Italiana Editori ringrazia il Presidente e i membri della Commissione per l'opportunità di contribuire con le nostre riflessioni a un tema di grande importanza per lo sviluppo futuro dell'editoria e delle altre industrie culturali e creative e per la loro competitività internazionale.

Quasi in coincidenza con l'audizione in Commissione è giunta all'approvazione finale del Parlamento europeo il Regolamento noto come *Artificial Intelligence Act (AI-Act)*, grazie a cui l'Unione Europea si è dotata di una legislazione quadro sull'Intelligenza Artificiale rispettosa dei diritti dei cittadini e del diritto d'autore e che si pone all'avanguardia nel mondo. Cogliamo l'occasione per ringraziare il Governo italiano per aver sostenuto la richiesta delle industrie creative perché fossero stabilite regole chiare sulla trasparenza delle fonti utilizzate dall'Intelligenza Artificiale per addestrare gli algoritmi. La trasparenza è indispensabile affinché l'innovazione tecnologica possa svilupparsi nel rispetto dei cittadini e in sinergia con le industrie culturali: di fatto, è base della possibilità di analizzare criticamente gli output dell'Intelligenza Artificiale e, per chi detiene i diritti, di sapere quali opere sono utilizzate nello sviluppo di questi strumenti e scegliere così se autorizzarne o meno l'uso.

1. L'intelligenza artificiale nell'editoria libraria e l'esperienza di AIE

L'Intelligenza Artificiale (IA) non è argomento nuovo per gli editori, giacché da tempo lavorano sulle sue applicazioni. AIE se ne occupa in modo fattivo da almeno un lustro, anche guidando gruppi di lavoro internazionali sul tema¹. Già nel 2019 ha organizzato un seminario dal titolo *AI in book communication and publishing* in collaborazione con il prof. Christoph Bläsi (Mainz University); nel marzo 2021, ha rappresentato l'editoria europea nell'incontro *Opportunities and challenges of AI Technologies for the Cultural and Creative Sectors*, organizzato dalla Commissione europea, illustrando la pluralità di applicazioni già allora presenti nel settore, in una tabella che qui riproduciamo:

¹ Nel 2019 AIE ha organizzato un seminario di approfondimento sul tema per la Federazione europea degli editori e nel 2021 ha rappresentato l'industria editoriale europea (libri e giornali) in un workshop sul tema organizzato dalla Commissione europea sulle applicazioni IA alle industrie culturali. Dal 2021 una rappresentante AIE è co-chair del gruppo di lavoro W3C (l'organismo di governance di Internet) che ha scritto i protocolli software per la riserva dei diritti prevista dall'art. 4 della Direttiva sul diritto d'autore nel mercato unico digitale; dal 2022 è membro del gruppo di lavoro dedicato all'IA dell'International Publishers Association



Da tempo, infatti, esistono applicazioni che vanno dalla fase creativa all'analisi di mercato, passando per la selezione dei manoscritti, il supporto alla traduzione, la produzione, l'offerta di nuovi servizi e la gestione dei diritti.

La novità dell'IA in discussione negli ultimi mesi consiste proprio nell'improvviso passaggio dalla dimensione «professionale» a una dimensione «generalista» (*general purpose artificial intelligence*, GPAI, nel gergo internazionale che si va consolidando) con la dirompente comparsa sul mercato di applicazioni facilmente utilizzabili e di largo consumo, basate su dati e modelli linguistici di dimensioni incommensurabilmente superiori alle precedenti.

2. Opportunità e rischi dell'IA nell'editoria libraria

Come tutte le innovazioni tecnologiche, l'IA presenta grandi opportunità che tuttavia potranno essere colte solo se si riuscirà a mitigare i rischi ad essa associati, così da avere un impatto sul comparto editoriale e più in generale sull'economia e sulla vita dei cittadini.

Gli editori e AIE sono consapevoli delle molteplici opportunità offerte dalle nuove generazioni di IA, alcune già esplorate e utilizzate, altre tutte da immaginare. Ogni segmento editoriale potrebbe beneficiare dell'apporto dell'intelligenza artificiale. Ad esempio, nel settore educativo, l'IA è utilizzata nelle soluzioni di *adaptive learning* per la scuola e l'università, in grado di fornire percorsi di apprendimento personalizzato agli studenti, o ancora, per integrare nelle piattaforme editoriali assistenti virtuali alla ricerca e alla didattica e per creare suggerimenti di lettura personalizzati. L'IA è inoltre sperimentata nei processi produttivi per la produzione di metadati, soggettazioni, abstract, parole chiave, tutti strumenti fondamentali per aumentare la visibilità delle pubblicazioni in rete.

In prospettiva, l'IA potrebbe aiutare gli editori a ridurre l'**impatto ambientale** e favorire la transizione verde, migliorando la previsione sia delle tirature, ottimizzando la stampa così evitando la sovrapproduzione, sia della distribuzione delle vendite, al fine di ridurre le rese.

Inoltre, può contribuire alla realizzazione di contenuti più **inclusivi e accessibili** a persone con disabilità. In questo contesto le applicazioni più promettenti riguardano la **velocizzazione di alcune fasi dei processi produttivi** per creare le novità in modo che siano nativamente accessibili o di quelli per la conversione in versione accessibile di pubblicazioni che non lo sono, ad esempio il catalogo. Un altro ambito è quello della **creazione delle descrizioni automatiche delle immagini** che sono un requisito obbligatorio perché una pubblicazione sia accessibile. Allo

stato attuale in entrambi i casi gli output automatici di queste applicazioni non garantiscono una qualità sufficiente ma possono fornire dei “semilavorati” con tempi e costi ridotti rispetto a cui far seguire una revisione da parte di operatori esperti.

Vi sono quindi numerose applicazioni, in parte già presenti sul mercato e in parte in via di sviluppo, che possono avere un importante impatto positivo sia sulle aziende editoriali che sulla società. Tuttavia, in mancanza di regole condivise, ogni settore o segmento editoriale potrebbe essere messo a rischio dalla stessa tecnologia. Per esempio, l’editoria educativa potrebbe trovarsi di fronte alla proliferazione di contenuti automaticamente generati dall’ IA a partire dai propri testi, violando le norme sul diritto d’autore, o originati da fonti non meglio identificate senza controllo di qualità e di veridicità. O ancora, la diffusione sulle grandi piattaforme di titoli generati dagli utenti, interamente grazie all’IA, e che non siano chiaramente segnalati come tali, rischia di confondere i lettori e comprometterne la fiducia. Con un impatto negativo sia sull’intera filiera editoriale sia sulla società.

3. Concorrenza e controllo dei dati nel mercato digitale

Molta dell’efficacia delle applicazioni IA dipende dalla quantità e qualità di dati sui quali gli algoritmi sono addestrati. Le grandi piattaforme (Amazon, nel nostro settore) traggono la loro forza nella capacità di fare previsioni, e i dati sono la benzina di queste previsioni. La quantità di dati da esse detenuti consente di migliorare la qualità dei servizi e di attrarre un maggior numero di utenti. Si instaura così un circolo vizioso che produce posizioni monopolistiche. È ragionevole aspettarsi che lo stesso fenomeno si verifichi con i nuovi servizi di IA generativa, destinati ad acquisire volumi crescenti di dati derivanti dal loro utilizzo.

RISCHIO: che lo spazio dell’IA venga dominato da un numero ristretto di potenti aziende tecnologiche in grado di beneficiare di posizioni monopolistiche nella raccolta e controllo dei dati che alimentano l’intelligenza artificiale.

- **Mitigazione del rischio:** L’Unione europea è intervenuta sul tema con il **Digital Markets Act** (DMA) imponendo alle piattaforme di condividere i dati con i propri partner commerciali. L’attuazione pratica del DMA sarà nei prossimi anni decisiva per evitare che l’IA diventi uno strumento di dominio economico e culturale di poche grandi aziende. Perché sia efficace va accompagnata da incentivi destinati alle imprese culturali per sostenerne gli investimenti e la cooperazione basata sulla condivisione di dati.

RISCHIO: Le applicazioni IA consentono una precisa profilazione degli utenti che le piattaforme sfruttano per far sì che questi ricevano sempre prodotti e servizi di loro interesse ma anche opinioni e notizie il più possibile aderenti al loro punto di vista. Ciò produce una polarizzazione del dibattito culturale e politico e una proliferazione di notizie false e discorsi d’odio.

- **Mitigazione del rischio:** impegno italiano in Europa, in coordinamento con la Commissione Europea, a monitorare l’adozione del Digital Services Act (DSA) e la sua applicazione efficace

da parte delle piattaforme, in particolare riguardo gli obblighi di trasparenza degli algoritmi e di valutazione e mitigazione dei rischi legati ai servizi.

4. Il rispetto dei diritti d'autore da parte degli operatori IA

L'IA ha bisogno di dati e contenuti affidabili per poter produrre risultati affidabili. Vale anche qui il vecchio adagio *"garbage in - garbage out"*, e pertanto i contenuti prodotti dall'industria editoriale sono tra i più appetibili e qualificati candidati a diventare input per i processi di addestramento dei modelli di IA e ancor più cruciali nella fase di miglioramento qualitativo (c.d. *fine tuning process*) e nel loro aggiornamento continuo, tramite le cd *Retrieval-Augmented Generation* (RAG). Ma affinché contenuti affidabili continuino ad alimentare tali modelli è necessario che persone, organizzazioni e imprese siano incentivate a continuare a crearli e produrli e a produrre valore da investire in nuove creazioni.

È quindi indispensabile che **lo sviluppo dell'IA avvenga nel rispetto del diritto d'autore e tramite licenze** che possano facilitare la collaborazione tra aziende tecnologiche e industria dei contenuti. Solo in questo modo è possibile promuovere una IA generativa di qualità, contrastando i fenomeni di disinformazione e i *bias* collegati all'addestramento degli algoritmi.

Per questo l'AIE condivide pienamente lo spirito della **risoluzione dell'On. Amorese** nella misura in cui pone la valorizzazione del diritto d'autore al centro dell'innovazione dell'IA e impegna il Governo a sostenere le iniziative dell'Unione Europea in tal senso.

L'Europa, infatti, ha già posto alcuni principi normativi fondamentali per regolare lo sviluppo dell'IA nel rispetto del diritto d'autore. Si tratta ora di garantirne l'efficace applicazione a livello nazionale.

RISCHIO: Che l'utilizzo di contenuti per l'addestramento delle applicazioni di IA e/o dei modelli che ne sono alla base avvenga in **violazione delle norme sul diritto d'autore** o utilizzando **fonti illegali** quali siti e servizi pirata.

- **Mitigazione del rischio:** il D.Lgs. 177/2021 ha recepito in Italia la Direttiva sul diritto d'autore nel mercato unico digitale che ha posto le basi normative per il **text and data mining (TDM)**, cioè la raccolta di dati per l'addestramento dell'IA. Le eccezioni prevedono che, al di fuori dei casi di uso per ricerca scientifica senza scopo di lucro, i titolari possono riservare i propri diritti esclusivi (opt-out). **L'AI-Act** ha richiamato esplicitamente queste norme, stabilendo l'obbligo per i fornitori di IA di verificare se i diritti per il TDM sono stati riservati e in tal caso, chiedere l'autorizzazione per utilizzare i contenuti per l'addestramento dell'IA. Il regolamento stabilisce inoltre obblighi di trasparenza sulle fonti utilizzate dai sistemi di IA indispensabili per poter appurare se i sistemi di IA hanno rispettato gli opt-out degli aventi diritto e se hanno utilizzato servizi illegali per accedere ai contenuti.

Con l'**opt out**, autori ed editori possono non solo bloccare l'accesso ai contenuti, ma anche offrire **licenze d'uso** per il loro utilizzo nei servizi dell'IA. La riserva dei diritti può essere espressa nei contratti di licenza e, nel caso dei contenuti liberamente accessibili online, tramite

strumenti machine-readable (ossia in un linguaggio comprensibile ai software). Esiste già un protocollo web di facile implementazione per riservare i diritti su tutti i tipi di contenuto online, il **Text and data mining reservation protocol** (TDMRep) sviluppato da un *Community Group* del **W3C**, co-presieduto da Giulia Marangoni di AIE e il centro di ricerca francese EDRLab e promosso dalla Federazione degli editori europei.

Occorre **promuovere la conoscenza degli strumenti di opt-out** e incoraggiare, anche tramite incentivi (es. credito d'imposta sugli investimenti necessari), **l'adozione del protocollo del W3C Community Group** in quanto soluzione aperta, basata sugli standard e le prassi in uso nell'industria dei contenuti, preferendola a soluzioni proprietarie proposte da singole aziende o a soluzioni di opt-out che si applicano solo a specifici servizi di IA.

Occorre inoltre vigilare sulla **corretta applicazione dell'eccezione per il text and data mining**, ad esempio assicurandosi che sistemi di IA sviluppati per fini di ricerca senza scopo di lucro e **successivamente convertiti in prodotti commerciali** rispettino l'opt-out degli aventi diritto.

Infine, è bene ricordare che tale eccezione si applica dal giugno 2021: prima di allora l'addestramento dei sistemi di IA con contenuti protetti sarebbe stato possibile solo tramite autorizzazione degli aventi diritto: pertanto, prima di quella data, l'utilizzo non autorizzato dei contenuti per l'addestramento dell'IA costituisce sempre una violazione del diritto d'autore.

Il Governo italiano ha avuto un ruolo importante nella approvazione dell'**AI Act**. E' ora necessario assicurarsi che le norme siano applicate correttamente. Pur essendo la responsabilità principale in capo alla Commissione, gli Stati Membri potranno svolgere un ruolo importante partecipando alle strutture di governance previste dal regolamento. Per quanto concerne l'obbligo di trasparenza sulle fonti utilizzate per l'addestramento dei sistemi di IA, sarà cruciale la partecipazione **all'Expert Group** degli Stati Membri annunciato dalla Commissione, che dovrà affiancare **l'AI Office** nel predisporre il template che i fornitori di IA dovranno utilizzare per fornire la sintesi dei contenuti utilizzati. Dal design del template dipende infatti l'efficacia della norma, e dovrà consentire agli aventi diritto di ricavare le informazioni realmente necessarie per identificare i propri contenuti e far valere i propri diritti.

Inoltre, occorrerà **vigilare sulla corretta applicazione** di tale obbligo da parte dei fornitori di IA: ad esempio, assicurarsi che la trasparenza delle fonti riguardi *tutti* i dati utilizzati per l'addestramento dei sistemi (e non ad esempio solo quelli usati dopo l'entrata in vigore del regolamento), e che non sia limitata impropriamente tramite il ricorso a **segreti industriali**.

5. Identificazione delle opere generate dall'IA

Nell'identificazione e nella necessità di etichettare delle opere generate dall'IA, è fondamentale tener conto della distinzione tra **uso strumentale** dell'IA, in cui la tecnologia è uno strumento al servizio della creatività umana come nel supporto all'editing (esempio citato nel testo nell'AI

Act), e un **uso sostanziale**, in cui viene meno l'apporto creativo umano e quindi l'opera può dirsi generata interamente dall'IA.

Questo approccio è coerente con la normativa sul diritto d'autore, che individua nelle **creatività umana** il discrimine fondamentale per determinare se un'opera sia soggetta o meno a protezione del DDA. Sia in Europa sia negli Stati Uniti si riconosce che un'opera generata esclusivamente dall'IA non può essere protetta dal diritto d'autore. Al contrario, perché questo accada, è necessario che vi sia un apporto creativo umano.

RISCHIO: La difficoltà nel riconoscere contenuti generati dall'IA (quali notizie false e *deep fake*), con conseguenze, anche gravi, in termini di disinformazione e sicurezza sociale.

- **Mitigazione del rischio:** L'AI-Act stabilisce che i contenuti generati dall'IA siano marcati come tali dagli strumenti utilizzati per crearli, salvo i casi in cui l'IA sia utilizzata come semplice supporto all'editing o non alteri in modo sostanziale i dati forniti dall'autore umano; sono inoltre previsti obblighi di trasparenza per i creatori di *deep fake* e per le notizie generate automaticamente senza revisione umana o controllo editoriale.

RISCHIO: inondazione del mercato con titoli generati automaticamente dall'IA e caricati dagli utenti delle grandi piattaforme (come Amazon), e non immediatamente distinguibili dalle opere frutto della creatività umana.

- **Mitigazione del rischio:** l'AI Act individua più strumenti tramite cui la Commissione può intervenire affinché vi sia trasparenza sui contenuti generati interamente dall'IA, quali l'adozione di atti di esecuzione per l'applicazione e la promozione di codici di condotta. Gli Stati Membri possono fornire un supporto importante monitorando l'operato delle piattaforme e suggerendo soluzioni laddove si riscontrino problemi.

6. Il PNRR e il sostegno all'innovazione

Abbiamo trovato molto interessanti le recenti dichiarazioni del sottosegretario Butti sulla necessità di investire risorse cospicue sulla ricerca e sviluppo, sia delle università sia delle imprese. L'innovazione fondata sull'IA richiede importanti investimenti in ricerca e sviluppo, acquisizione di nuove competenze, riprogettazione dei flussi di lavoro e sviluppo di nuovi prodotti e servizi. In assenza di finanziamenti dedicati, l'innovazione rischia di restare appannaggio di poche grandi imprese, soprattutto in settori come l'editoria, caratterizzati da una forte presenza di PMI. Dedicare una quota degli investimenti destinati alle imprese nel PNRR specificamente a questa attività, con un focus specifico sulle imprese culturali e creative, è a nostro avviso altamente auspicabile. Inoltre, in virtù della natura di tali attività, occorre prevedere forme di sostegno adeguate all'innovazione tipica dell'IA, che richiede investimenti in beni immateriali, che finora non sono ammessi nei programmi di incentivi (Transizione 4.0 e simili) a meno che non fossero legati a investimenti in macchinari.

Inoltre, in riferimento a quanto detto nel paragrafo precedente, contributi diretti a investimenti privati in azioni di sistema nel campo della gestione dati, dell'accessibilità per i cittadini con

disabilità, delle diverse applicazioni dell'IA generativa possono avere un impatto forte sul sistema delle industrie culturali.

Poiché si tratta di investimenti che richiedono la disponibilità di grandi capacità di calcolo, suggeriamo di guardare in questa ottica le opportunità che può aprire la presenza in Italia di Leonardo, uno dei cinque super-computer più potenti al mondo, concepito e gestito dal CINECA. Sono incoraggianti le notizie sulla collaborazione pubblico-privato che il CINECA ha stretto con iGenius, così come il dialogo con la Fondazione FAIR, nata proprio con fondi PNRR, a sua volta fondata sulla collaborazione tra università e imprese. A partire da queste esperienze, riteniamo auspicabile un percorso che ampli queste interlocuzioni alle imprese culturali e a quelle editoriali in particolare.