

ChatGPT e Wikipedia

Susanna Giaccai

La comunità dei volontari che lavora in Wikipedia ha un grande rispetto per il diritto d'autore e di conseguenza vieta di copiare brani coperti dal diritto d'autore nelle voci dell'enciclopedia.

Nelle regole editoriali viene specificato che i testi usati come fonti per la scrittura di voci possono essere riassunti e tale azione viene spiegata così: «il riassunto è un testo in cui si spiega, con parole proprie, il senso complessivo di un altro testo». La spiegazione entra nel dettaglio chiarendo la differenza tra perifrasi e riassunto. La prima prevede la sostituzione di alcune parole con sinonimi, ma in questo caso si avvicina molto al concetto di opera derivata o di plagio e quindi il suo uso è vietato dal copyright; mentre il riassunto è un testo in cui si spiega con parole proprie il senso complessivo di un altro testo; cioè un testo in cui si 'riformula' un altro testo¹.

Entro in questi dettagli perché il tema del riassunto è presente in una discussione avviata il 4 febbraio al Bar di Wikipedia, il luogo in cui i volontari discutono questioni di carattere generale relative alla enciclopedia². Il wikipediano ValterVB chiede: «Mi domandavo se il tanto pubblicizzato [ChatGPT](#) poteva essere usato per Wikipedia, l'idea che mi è venuta nel chiedere la *riformulazione* di un testo.»³

La sua domanda si riferisce a una voce di Wikipedia contenente troppe citazioni e che necessita quindi di essere corretta spostando il sunto delle citazioni nel testo. In questo caso ValterVb aveva inserito le citazioni in ChatGBT chiedendo di 'riformulare' il testo; il

1 Si veda la pagina <https://it.wikipedia.org/wiki/Aiuto:Riformulare_un_testo>

2 Al Bar si accede dal link Bar nella barra di navigazione verticale a sinistra.

3 Si veda la discussione nella pagina <<https://tinyurl.com/56hac68y>> verificato (1/3/23)



risultato era abbastanza buono e quindi egli domanda alla comunità se si tratti di una azione accettabile.

La discussione si snoda tra il tema del diritto di autore e quello delle fonti, elementi fondamentali che devono sempre essere presenti in tutte le voci di Wikipedia.

Ma per seguire lo scambio tra wikipediani sono utili due parole su ChatGPT⁴. Si tratta di un prototipo di *chatbot* basato su intelligenza artificiale e *machine learning* (apprendimento automatico), specializzato nella conversazione con un utente umano. Il software è stato creato da OpenAI, una organizzazione senza fini di lucro per la ricerca sull'intelligenza artificiale con lo scopo di promuovere e sviluppare un'intelligenza artificiale amichevole (*friendly AI*)⁵. Da quando il servizio è nato nel novembre 2022, articoli su ChatGPT iniziano a essere frequenti su quotidiani e blog specialistici; c'è chi sottolinea la preoccupazione dei docenti per la sua diffusione tra gli studenti che consegnano relazioni scritte utilizzando questa AI; altri riferiscono invece di docenti che la ritengono una occasione per addestrare lo spirito critico dei propri allievi a riconoscere il pensiero umano da quello artificiale. Naturalmente sono anche nati subito programmi in grado di identificare un testo scritto da una AI, rispetto a quello scritto da un essere umano⁶.

Le BigTec sono molto interessate in questo sviluppo: Microsoft e Google reagiscono in modo diverso; mentre Microsoft usa ChatGPT per il suo motore di ricerca Bing, Google propone Bard, un sistema diverso costruito pescando notizie dal web e quindi fornendo risultati molto aggiornati. La pecca attribuita a ChatGPT è infatti la sua limitata capacità di offrire notizie aggiornate fondandosi su un sistema di *machine learning* alimentato da fonti più statiche⁷. Infatti la domanda posta a ChatGPT di avere informazioni recenti sulle iniziative di un Museo, ha questa risposta :

«Mi dispiace, ma non sono in grado di fornire informazioni in tempo reale su eventuali recenti iniziative organizzate dal Museo del Tessuto di Prato su Wikipedia. Come modello di linguaggio, la mia conoscenza si ferma al 2021 e non ho la capacità di accedere a notizie o eventi attuali».

4 GPT si scioglie in *Generative Pre-trained Transformer*,

5 Una breve spiegazione del servizio in Wikipedia <<https://it.wikipedia.org/wiki/ChatGPT>> verificato 1/3/23

6 Si veda <<https://tinyurl.com/y89j6y7m>> verificato 1/3/23

7 Cesare Rovelli, Bard, *La nuova intelligenza artificiale di Google e le differenze con ChatGPT*, in Corriere della sera, 7 febbraio 2023 <<https://tinyurl.com/4uy37snk>> verificato 1/3/23

Allarme si sta diffondendo nell'ambito editoriale per il timore di ricevere testi letterari scritti da ChatGPT magari con rielaborazioni di testi di autori di chiara fama⁸ o tra i fumettisti pronti a lanciare un *Manifesto in difesa dei diritti e della creatività dell'essere umano*, in cui si chiede che sia vietato introdurre nel training delle AI dati di persone o opere, per evitare che siano poi riusati dalla intelligenza artificiale per produrre nuove opere⁹. Nell'ambito degli *audiobook* si teme che la voce dei traduttori possa essere copiata per far tradurre altri testi con voce rielaborata da una AI.

Nella discussione che si snoda nel Bar di Wikipedia emergono pareri diversi: Valepert sostiene che non va accettato un testo rielaborato da un AI così come non viene accettata una traduzione automatica di una voce di Wikipedia da un'altra lingua. Negli ultimi anni infatti la comunità wikipediana ha messo a punto un modulo che traduce automaticamente le voci di Wikipedia; questa traduzione automatica però non viene accettata sino a che non siano state fatte considerevoli modifiche per adattarla allo stile della nostra lingua e ai criteri editoriali che la comunità italiana si è data. Ma il riferimento fatto da Valepert al traduttore automatico è improprio, dato che questo traduttore non lede il diritto d'autore poiché i testi di Wikipedia in tutte le lingue sono rilasciati con la licenza CC BY SA e quindi il loro riuso è autorizzato. Contrario anche Frinate che scrive: «pare che la parafrasi di un AI non sia abbastanza originale da costituire una rielaborazione a norma di legge», in poche parole il riassunto fatto da ChaptPGT non è abbastanza 'rielaborazione con parole proprie' ma somiglia più a una opera derivata.

Più aperto l'utente IIPoncio che scrive «Se il materiale in questione non è coperto da copyright, e viene adeguatamente munito di fonti, non vedo perché non possa essere inserito». E qui emergono due questioni da chiarire: il diritto d'autore e le fonti. Chi è l'autore dei testi scritti da ChatGPT? Gianfranco scrive: «il copyright è riconosciuto (soprattutto in diritto latino) all'autore, che è persona, dunque non può essere macchina in quanto priva di capacità giuridica». Qualcuno allora si domanda se il copyright sia della azienda che ha creato ChatGPT. Ma esiste anche il problema di quanto ChatGPT rispetti il diritto d'autore. In effetti lo *scrapper* del Web effettuato da ChaptPGT quanto rispetta il diritto d'autore? quanto del materiale 'ingurgitato' dall'AI è coperto dal diritto d'autore?

⁸ Antonio Prudenzeno, *Quale sarà l'impatto dell'intelligenza artificiale su creatività, comunicazione e ricerche online?*, in Il Libraio, 2 marzo 2023 <<https://tinyurl.com/ak7pa3rn>> verificato 1/3/23

⁹ Si veda il testo del manifesto <<https://tinyurl.com/2p9frcjx>> verificato 1/3/23

Quanto nello scritto prodotto da ChatGPT è frutto di rielaborazione (riassunto), quindi è fuori copyright, e quanto invece è perifrasi, quindi è coperto da copyright?

Le regole di Wikipedia affermano che, nella scrittura delle voci dell'enciclopedia, non è consentito copiare un testo, in quanto sarebbe violazione del copyright, ma è consentito farne un riassunto, una rielaborazione. E tra le competenze di ChatGBT ci sono appunto quelle di:

1. generare riassunti di documenti o articoli lunghi. Ciò può essere utile per ottenere rapidamente una panoramica di un testo senza dover leggere l'intero documento;
2. generare risposte testuali simili a quelle umane, il che lo rende utile per la creazione di chatbot per il servizio clienti, per generare risposte a domande nei forum online o anche per creare contenuti personalizzati per i post sui social media.¹⁰

Quindi, per il punto 1, ChatGPT potrebbe essere accettato in quanto la produzione di un riassunto non infrange il diritto d'autore; mentre per il punto 2 la questione si fa più complessa, oltre che per la diffidenza verso un testo creato da una IA, anche perché non è chiaro chi ne sia l'autore. Rivolgendo la domanda direttamente a ChatGPT, «Di chi è il copyright dei testi creati da ChatGPT?» Si ottiene questa risposta:

«Come assistente virtuale creato da OpenAI, i testi creati da me appartengono a OpenAI e sono protetti dal diritto d'autore. L'uso dei testi creati da ChatGPT è soggetto ai Termini di Servizio di OpenAI, che limitano l'uso dei suoi servizi e dei suoi contenuti a scopi non commerciali. Inoltre, il contenuto prodotto da ChatGPT potrebbe includere informazioni di terzi e citazioni, che sono soggetti a copyright separato dei rispettivi proprietari».

L'ultima frase chiarisce bene che non ci sono garanzie che ChatGPT rispetti il diritto di autore.

Inoltre, per quanto riguarda le capacità di ChatGPT di elaborare nuovi testi, si pone il problema delle fonti; in Wikipedia è infatti essenziale documentare le fonti da cui derivano i concetti descritti nella voce. ValterVB scrive che Sarrow, la AI di Google, sembra sarà in grado di indicare le fonti, ma Frinate dubita, perché riferisce che alcuni test provano che ChatGPT si inventi le fonti¹¹.

¹⁰ Si veda il sito <<https://chatgpt.pro/it/>> verificato 1/3/23

¹¹ Si veda <<https://news.ycombinator.com/item?id=33841672>> verificato 1/3/23

Bramfab prosaicamente scrive: «Ho il sospetto che questa sia una discussione con data di scadenza: IMO (in my opinion) fra un paio d'anni non saremo in grado di distinguere fra un testo umano e uno AI».

Alessandro57 riferisce di una sua chat con ChatGPT di cui segnala gli errori e sottolinea infine che tra le fonti della AI in questo caso c'era anche Wikipedia. Ciò a suo parere toglie ogni dubbio sulla non legittimità di usare ChatGPT per scrivere voci in Wikipedia. Infatti essendo questa una enciclopedia, non è una fonte primaria ma una fonte secondaria in continua evoluzione e modificabile, e anche vandalizzabile da chiunque, e non è corretto usarla come fonte.

Nel frattempo il tema di ChatGPT si è presentato anche nella discussione su una voce dove un utente ha citato ChatGPT come fonte; la reazione è stata immediata nel cancellare il suo commento con la spiegazione: «La pagina di discussione è per interventi umani, non per un programma che non è (ancora) in grado di discutere.»¹².

Termino segnalando, per chi sia interessato ad approfondire, l'articolo del 6 gennaio 2023 di Julio Santillán-Aldana, *ChatGPT's answers to requests about the librarian profession*, del 6 gennaio 2023 disponibile in Zenodo¹³.

Nel frattempo ChatGPT si è evoluta in GPT-4, di cui vedremo le potenzialità nei prossimi mesi o forse solo settimane. Lo stesso Jimmy Wales, fondatore di Wikipedia, ritiene che il tema della scrittura di Wikipedia con ChatGPT si porrà se con la prossima versione GPT-5 sarà in grado di evitare le 'allucinazioni' che l'attuale versione ha ogni tanto¹⁴.

Susanna Giaccai

Collaboratrice in Wikipedia, Wikisource, Wikidata, Commons e bibliotecaria

giaccai@pm.me

12 Si veda in questo confronto di cronologia della pagina di Wikipedia a sinistra il testo di ChatGPT e a destra il commento di chi ha cancellato la frase <<https://tinyurl.com/2p9cbtr9>> verificato 1/3/23

13 Si veda l'articolo qui <<https://zenodo.org/record/7555980#.ZAyra9LMJ8x>> verificato 1/3/23

14 Si veda l'articolo Wikipedia scritta dall'AI? Fondatore valuta possibilità, in ANSA-Hi-tech, 5 aprile 2023 <<https://tinyurl.com/nnzt4wxb>> - verificato 11 aprile 2023