

# La responsabilità è sempre mia

## Riconoscere le allucinazioni

La professoressa di italiano ci ha chiesto di elaborare un confronto scritto tra il libro di Dino Buzzati *La famosa invasione degli orsi in Sicilia* e l'adattamento cinematografico di Lorenzo Mattotti. Quando l'ho consegnato mi pareva che suonasse bene, ma quando la prof me lo ha restituito, mi ha mostrato alcune parti evidenziate in rosso e mi ha chiesto: «Che cosa intendi esattamente con questo passaggio?». Allora mi sono resa conto che non ero in grado di rispondere. E dire che mi ci erano voluti 45 minuti per scrivere quel testo con l'aiuto di un chatbot.

Lo stesso problema lo ha avuto il resto della classe e ovviamente la prof si è accorta che avevamo usato l'intelligenza artificiale. Ma poi ci ha sorpreso dicendo che questo di per sé non era un problema, a patto che la usassimo bene! Tutte le analisi che avevamo fatto erano linguisticamente corrette, ma contenevano alcune parti assurde: oltre a tanti passaggi non pertinenti, c'erano persone e azioni che nel testo e nel film semplicemente... non esistono, come una tale Jolanda, principessa della città dei gatti (!).

Quindi ha fotocopiato il testo di una nostra compagna e l'ha dato a tutta la classe: ognuno ha colorato le allucinazioni prodotte dall'IA in rosso e i passaggi non pertinenti alla domanda in grigio e questo è stato il risultato:

Come è stato possibile per molti di noi non accorgersi di così tanti errori? Ne abbiamo discusso e abbiamo capito che tante persone si erano fidate ciecamente dell'IA, abbagliate dal suo stile elegante. Tutti tendiamo a farlo e a volte dubitiamo prima di noi stessi e solo dopo dell'IA!



### PROVATE ANCHE VOI!

Provate a riprodurre un contesto simile a quello descritto qui sopra.

- Pensate a un libro che avete letto e di cui avete visto l'adattamento cinematografico. Cercate di ricordare quale delle due versioni vi era piaciuta di più e perché. Su quali elementi di confronto si era basato il vostro giudizio?
- Ora chiedete all'IA di fare lo stesso confronto. Ecco una possibile traccia di prompt: *Confronta il libro ... di ... con il film ... di .... Analizza personaggi, luoghi, azioni, temi ecc. in modo contrastivo.* Potete usare **Claude**, **ChatGPT** o **Gemini**.
- Analizzate ora la risposta dell'IA:
  - eliminate le frasi con informazioni accessorie o non pertinenti alla domanda (che chiede di confrontare il libro con il film);
  - evidenziate le affermazioni false; prestate particolare attenzione anche agli aggettivi utilizzati, spesso sono inappropriati;
  - prendetevi il tempo per leggere il testo e comprendere ogni frase.
- Infine correggete o riscrivete tutte le frasi con cui non concordate e che non esprimono le vostre riflessioni e la vostra opinione.

### Ben scritto non è sinonimo di vero

Nell'esercizio appena svolto avete incontrato delle difficoltà a valutare il testo prodotto dall'IA e a individuarne i difetti? Se sì, la spiegazione è nell'*halo effect* ("effetto alone"), un *pregiudizio* che ci porta a generalizzare e ad attribuire automaticamente a un elemento certe qualità in presenza di altre. Facciamo un esempio concreto: i testi stampati sui libri generalmente presentano contenuti attendibili e sono anche caratterizzati da un linguaggio scorrevole e curato. Ecco che allora il nostro cervello tende inconsciamente ad associare a ogni testo ben scritto anche una elevata qualità dei contenuti che non è, invece, necessariamente presente. La stessa cosa accade con i testi generati dall'IA: l'*halo effect* ci induce a **considerare affidabili** le informazioni contenute nei **testi ben scritti** dall'IA, anche quando sono del tutto inventate.

#### CONSIGLIO

Prima di consegnare un testo, immaginate che l'insegnante vi chieda spiegazioni su ogni frase e che voglia sapere esattamente che cosa intendiate con essa. Verificate di poter spiegare sempre con precisione che cosa c'è scritto nel vostro testo.



#### ORA SO...

... che quando scrivo testi utilizzando l'intelligenza artificiale devo:

- eliminare i passaggi irrilevanti e non pertinenti;
- eliminare le allucinazioni, cioè le informazioni che l'IA ha inventato;

- chiedermi se saprei spiegare il significato di ogni passaggio del testo e se condivido tutto ciò che afferma.

Insomma, non mi lascio accecare dallo stile formalmente curato dei chatbot.



## Lasarsi ispirare da allucinazioni e imprecisioni

Utilizzo da molto tempo gli strumenti di scrittura basati sull'intelligenza artificiale a scuola. Ho imparato a filtrare ogni testo prodotto così, chiedendomi che cosa mi sta bene e che cosa invece voglio eliminare perché non è pertinente o è falso. Questo processo è molto utile ed è più prezioso di quanto si pensi: filtrando mi vengono nuove idee, perché anche le cose che decido di eliminare generano nuove associazioni o mi stimolano a fare ulteriori riflessioni.



#### PROVATE ANCHE VOI!

- Usate questo prompt in **ChatGPT**: *Questo fine settimana piove: suggeriscimi 20 possibili attività da fare in casa per non annoiarmi.*
- Selezionate le idee che vi piacciono e scartate le altre. Che cosa resta alla fine? Nel corso di questa selezione vi sono venute in mente nuove idee? Se sì, quali? Parlatene insieme e provate a capire come avete avuto l'ispirazione.



#### ORA SO...

... farmi ispirare anche dagli elementi e dalle idee che scarto.

