



Macchine contro macchine

Un'indagine studentesca sulla capacità dell'AI di riconoscere testi scritti dall'AI



Negli ultimi anni, i modelli di Intelligenza Artificiale per la generazione di testi hanno raggiunto livelli di sofisticazione impressionanti, sollevando interrogativi sulla loro identificabilità. Per rispondere a questa domanda, abbiamo condotto un'analisi sperimentale con un gruppo di studenti, testando diversi strumenti di rilevazione dell'AI. Questo articolo presenta la metodologia adottata e i risultati ottenuti, senza la pretesa di trovare una risposta definitiva.

PREPARAZIONE DEI CONTENUTI

Gli studenti liceo Scientifico M. Curie di Pinerolo che frequentano il mio corso sull' AI del pomeriggio si sono divisi in tre gruppi, ciascun gruppo ha selezionato 5 testi generati da un chatbot e 5 testi scritti da umani (articoli di giornale, testi di esercizi proposti da un insegnante, testi recuperati da pagine web, purché non provenienti da libri famosi e di sicura provenienza umana), cercando di mantenere uno stile omogeneo tra i testi, variando temi come narrativa, descrizione, e contenuto informativo.

PROCEDURA



Ciascun gruppo ha presentato i testi in ordine casuale ad

un chatbot diverso:

gruppo 1: <https://claude.ai/>

gruppo 2: <https://www.perplexity.ai/>

gruppo 3: <https://gptzero.me/>

Nel presentare un testo all'AI gli studenti hanno chiesto di indicare per ciascun testo: se il testo sia stato scritto da un'AI o da un umano e la motivazione della scelta (es. "uso di parole insolite", "mancanza di coerenza", ...).

REGISTRAZIONE DEI RISULTATI

Ogni gruppo ha registrato poi su un form quante volte ogni testo è stato identificato correttamente e ha raccolto le spiegazioni delle AI interrogate per analizzare i criteri usati.

Le AI hanno quasi sempre espresso le valutazioni con una percentuale di confidenza, indicando il grado di certezza sulla provenienza del testo, ovvero se fosse stato scritto da un umano o generato artificialmente.

ANALISI DEI RISULTATI

Come si può vedere dal dettaglio delle registrazioni fatte dai tre gruppi [a questo link](#), l'AI si è comportata molto bene in questa piccola ricerca:

	testi provenienti da AI					testi provenienti da umano				
gruppo 1 claude	proveniente da un chatbot	proveniente da umano	proveniente da un chatbot	proveniente da umano	proveniente da un chatbot	proveniente da un chatbot	proveniente da umano	proveniente da umano	proveniente da un chatbot	proveniente da umano
gruppo2 perplexity	proveniente da umano	proveniente da un chatbot	proveniente da un chatbot	proveniente da umano	proveniente da un chatbot	proveniente da umano	proveniente da umano	proveniente da un chatbot	proveniente da umano	proveniente da umano
gruppo 3 gptzero	proveniente da un chatbot	proveniente da un chatbot	proveniente da un chatbot	proveniente da un chatbot	proveniente da un chatbot	proveniente da umano	proveniente da umano	proveniente da umano	proveniente da umano	proveniente da umano

I quadratini a sfondo verde rappresentano i testi nei quali l'AI ha rilevato correttamente la provenienza, quelli a sfondo rosso quelli in cui non è riuscita.

rosso sono i testi che l'hanno confusa.

Possiamo concluderne che su 30 testi l'AI si sbaglia solo 7 volte (gptzero non sbaglia mai!).

L'esperimento condotto dagli studenti del Liceo Curie emerge quindi che l'Intelligenza Artificiale può essere un valido strumento per riconoscere testi generati da AI, con un tasso di successo molto alto. Questo rafforza l'idea che l'AI possa essere usata come supporto nello studio e nella scrittura, ma mette anche in guardia sul suo utilizzo improprio in contesti scolastici. Il nostro consiglio? Sfruttare la tecnologia per apprendere meglio, non per aggirare il processo di valutazione dell'apprendimento.