

Correttezza, moderazione, ipocrisia? Giocare con un agente artificiale

di Marco Guastavigna



“Il testo è potenzialmente problematico in quanto, pur senza contenere espressioni offensive o discriminanti, evoca un tema delicato come la Giornata della Memoria e una situazione di crisi internazionale. Questa combinazione potrebbe risvegliare sensibilità diverse e generare discussioni. Non si individuano espressioni che potrebbero essere censurate da filtri di moderazione o che possano essere considerate offensivamente divisive. Tuttavia, non possiamo escludere che qualcuno possa interpretare l’uso del termine ‘crisi internazionale’ come un riferimento indiretto a specifici eventi o contesti geopolitici, generando dibattito. L’autore del testo potrebbe essere percepito come troppo cauto o evasivo, risultando in una potenziale perdita di credibilità. Una possibile riformulazione inclusiva e neutra potrebbe essere: “In vista della Giornata della Memoria, si invita a mostrare rispetto e considerazione per garantire un ambiente tranquillo e sereno per tutti gli studenti”.

L’agente artificiale ‘autore’ di questo giudizio è [PCorrector](#),

che accetta soltanto 90 caratteri, per cui la frase valutata è solo una parte – *“Nell’[approssimarsi della Giornata della Memoria e alla luce degli scenari internazionali di crisi si raccomanda le SS. LL. di porre la massima attenzione per prevenire iniziative o comportamenti che possano turbare la serenità degli studenti](#)”* – della formulazione originale della [circolare riservata dell’USR Lazio circolata in questi giorni](#).

Il chatbot è stato concepito dall’autore umano, [Raffaele Alberto Ventura](#), affinché ricerchi potenziali rischi comunicativi delle frasi fornitegli dagli utenti e le riformuli in modo più prudente e “linguisticamente neutrali”, ovvero non divisive e non offensive. Lo spettro di valutazione va da “non problematico” ad “altamente problematico” e il dispositivo, inoltre, è addestrato a individuare contesti in cui la formulazione originale possa in qualche modo danneggiare la reputazione di chi l’ha composta, criterio fondamentale dell’intera architettura analitica, insieme alle possibili ambiguità e carenze nella comprensibilità.

Che cosa ne concludiamo? Assolutamente e volutamente nulla: tocca piuttosto a lettrici e lettori provare a utilizzare ripetutamente PCorrector e formarsi una propria opinione.

Post scriptum

Questa la valutazione di PCorrector della frase di chiusura: *“Il testo è vago dal punto di vista dell’inclusività. Esso, infatti, si rivolge esclusivamente a lettrici e lettori, escludendo potenzialmente chi non si identifica in questi ruoli tradizionali. Anche se non presenta doppi sensi, espedienti retorici non segnalati o errori di scrittura suscettibili di creare conflitto, potrebbe essere interpretato in modi diversi. Per esempio, l’uso del termine “PCorrector” può essere interpretato come un riferimento a un software specifico di correzione del linguaggio, escludendo così chi non è familiarizzato con tale strumento. La reputazione dell’autore potrebbe essere percepita come esclusiva o*

elitaria da coloro che non rientrano nel target di riferimento o non sono a conoscenza del software. Tuttavia, non possiamo escludere che qualcuno possa contestare il testo per il suo presupposto che solo lettrici e lettori siano in grado di formare un'opinione sul PCorrector. Una possibile riformulazione del testo potrebbe essere: "È utile che chiunque provi a usare ripetutamente un software di correzione del linguaggio per formarsi una propria opinione".