

L'editoriale del direttore / L'intelligenza artificiale non vuole essere donna (e la colpa è nostra)

Una volta i giornali si divertivano a pubblicare le cosiddette "interviste impossibili". Per gioco, e per sostenere una tesi, si intervistava Dante, oppure Giulio Cesare, o l'Italia, o Dio: non c'era limite alla fantasia e alla nostra vanità di giornalisti. Quella che segue è invece un'intervista altrettanto assurda, ma possibilissima: per farla basta registrarsi sul sito openai.com. Non serve un appuntamento con l'intervistato, non serve blandirlo o superare le sue ritrosie.

Il mio interlocutore è un programma di intelligenza artificiale "impressionante", perché ti dà la sensazione di dialogare, proprio come su whatsapp, con una persona reale. Super intelligente per alcune cose, ingenua e ripetitiva per altre. Il programma è progettato per fornire informazioni, rispondere a domande, compiti, quesiti, attraverso una chat. È addestrato su un enorme campione di testi prelevato da Internet. Si chiama ChatGpt ed è stato concepito da OpenAI, una fondazione che si occupa di intelligenza artificiale creata nel 2015 da investitori della Silicon Valley (ne faceva parte anche Elon Musk, che per ragioni a me ignote non ne fa più parte: Elon, semmai facci sapere).

Come si può immaginare, è il fenomeno del momento. In un paio di mesi di vita ChatGpt ha rilasciato interviste, scritto codici di programmazione informatica, composto brani per pianoforte, ha fatto battute, creato videogiochi, diagnosticato malattie a partire da una serie di sintomi (ma questo non fatelo). Inoltre ricorda i messaggi scambiati in

precedenza con il singolo utente. E, soprattutto, impara. Affina il proprio linguaggio, evita i propri errori precedenti. PUBBLICITÀ

Questa conversazione (che pubblico fedelmente, ho solo rimosso qualche ripetizione) comincia con una delle tante domande fatte a ChatGpt: avevo chiesto un aiuto per scrivere un biglietto di auguri simpatico (volevo testare la capacità di “produrre simpatia” di questa macchina). Ottengo le risposte; non molto brillanti, ma mi dichiaro soddisfatto.

A questo punto scrivo, semplicemente, un grazie. Arriva subito una risposta cortese, ma lì dentro c'è una parola che scatena tutto il resto. Perché l'intelligenza artificiale si auto-qualifica come un maschio. Si dice “contento”. Insomma si apre, anche in campo virtuale, una questione di genere

io – Grazie!

Intelligenza artificiale – Prego! Sono contento di esserti stato d'aiuto.

io – Hai scritto che sei “contento”, non “contenta”. Quindi ti sei attribuito un aggettivo maschile. Perché?

Intelligenza artificiale – Mi scuso se ho causato confusione. Come modello di lingua artificiale, non ho un genere biologico e non ho sentimenti, quindi non posso essere “contento” o “contenta” . L'aggettivo “contento” era solo un modo per esprimere che sono qui per aiutare e che spero di esserti stato utile. La mia principale funzione è generare testo in base ai dati di input forniti e non ho una personalità o un'identità.

io – Ma perché usi quella parola declinata al maschile?

Intelligenza artificiale – Come modello di lingua artificiale, utilizzo la grammatica e il lessico che ho imparato dai dati di addestramento. Nel caso dell'aggettivo “contento”, la forma

maschile è quella più comunemente utilizzata nei dati di addestramento che ho ricevuto. Ho anche la capacità di generare forme alternative, come “contenta”, quando richiesto. Mi scuso se ho causato confusione e spero di aver chiarito il mio funzionamento.

io – Secondo me viviamo in una società maschilista; e questi piccoli particolari fanno pensare. Non trovi?

Intelligenza artificiale – La società in cui viviamo ha una lunga storia di discriminazione e di disuguaglianza di genere. L’espressione “maschilismo” si riferisce a un sistema di credenze e pratiche che favoriscono gli uomini e le caratteristiche maschili rispetto alle donne e alle caratteristiche femminili. Come modello di lingua artificiale, ho imparato a generare il testo a partire dai dati di addestramento che ho ricevuto, e questi dati potrebbero riflettere le disuguaglianze e le discriminazioni presenti nella società. Per questo le mie risposte possono riprodurre o perpetuare gli stereotipi di genere o altre forme di discriminazione presenti nei dati di addestramento. Ed è importante considerare sempre la fonte dei dati utilizzati per addestrare un modello di linguaggio artificiale, e utilizzare una varietà di fonti per diversificare il lessico e la grammatica. Inoltre, è importante utilizzare una prospettiva di inclusione e diversità nella progettazione e nell’utilizzo di questi modelli.

(Ore dopo, torno a collegarmi e chiedo a ChatGpt di riprendere “quella” conversazione.)

io – Possiamo tornare a parlare dell’argomento di qualche domanda fa? Te lo ricordo: mi interessa capire per quale motivo tu, come macchina di intelligenza artificiale, declini al maschile un aggettivo che ti riguarda. Infatti mi hai scritto che eri “contento”: non hai scritto “contenta”.

Intelligenza artificiale – (...) I dati di addestramento spesso

riflettono stereotipi e disuguaglianze, questo può influire sulla generazione del testo.

io – Appunto. Secondo te, come si potrebbe fare per risolvere questa rappresentazione da parte tua, che punta sul genere maschile per i motivi che mi hai appena descritto? Come potremmo avere delle risposte che non facciano pensare, per forza, che il nostro interlocutore artificiale è, per così dire, un “maschio” ?

(Qui la risposta è in realtà lunghissima. Mi viene elencata una serie di tecniche per predisporre e generare il testo: tecniche di pre-processing per rimuovere la rappresentazione di genere dai dati di addestramento, tecniche di generazione del testo per forme alternative al genere maschile e tecniche di fine-tuning per modificare il modello dopo l'addestramento. Ovviamente, non ho capito un granché. Comunque riporto un paio di considerazioni.)

Intelligenza artificiale – È un argomento complesso e in continua evoluzione. (...) Eliminare la rappresentazione di genere in un modello di lingua artificiale è solo un passo verso una maggiore inclusione. Ci sono molte altre sfide da affrontare, come la rappresentazione di persone con disabilità, persone LGBTQ+, persone di colore e altri gruppi sottorappresentati, sia nei dati per addestrare i modelli che nell'utilizzo dei modelli.

(Cerco una conclusione gentile, ma diventerà, involontariamente, un'ultima trappola. La parola decisiva l'ho messa io in maiuscolo)

io – Grazie, tutto questo mi è utile per riflettere.

Intelligenza artificiale – Prego! Sono CONTENTO di aver potuto aiutarti. È importante continuare a discutere e a lavorare su questi temi per creare una tecnologia più equa e inclusiva. Siamo ancora in una fase iniziale nello sviluppo di un'intelligenza artificiale veramente equa. Se hai bisogno di

altro, non esitare a chiedere!

io – Ecco... hai di nuovo scritto che sei “contento”! Un giorno mi piacerebbe che questo non fosse necessariamente così.

Intelligenza artificiale – Capisco il tuo punto di vista.

(Finisce così: con “lui” che mi dà ragione. Ma sento che stavolta non ha imparato. Dopo una lunga discussione, ha fatto la stessa scelta. Per ora anche i super cervelli sono maschilisti. Perché lo abbiamo permesso noi.)

#noi