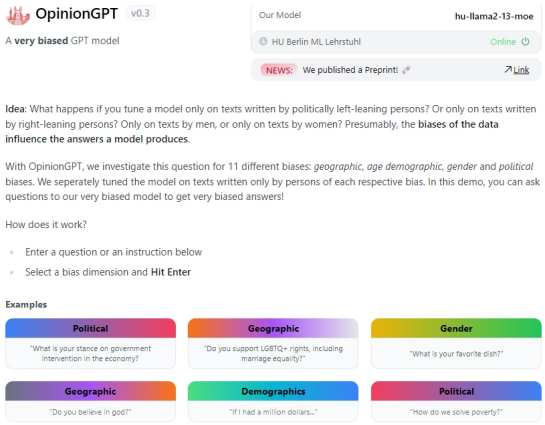

OpinionGPT è un interessante esperimento curato dall'Università Humboldt di Berlino, capace di essere interrogato su qualsiasi argomento e fornire le proprie opinioni utilizzando una serie di pregiudizi tipici per categorie diverse.

{loadposition user7}

I Large Language Models (LLM) sono tarati su specifiche istruzioni ed recentemente dimostrato una notevole capacità di generare risposte adeguate alle richieste in linguaggio naturale.

Tuttavia, si è avviata da tempo una ricerca aperta riguardante i pregiudizi intrinseci dei modelli addestrati e delle loro risposte. Per esempio, se i dati utilizzati per sintonizzare un LLM sono scritti in prevalenza da persone con uno specifico orientamento politico, potremmo aspettarci che le risposte generate condividano questo orientamento. Lo sviluppo attuale cerca di attenuare questi condizionamenti o di sopprimere le risposte potenzialmente distorte.

Con questo esperimento, viene adottato un punto di vista diverso sui pregiudizi nella messa a punto delle istruzioni: Piuttosto che puntare a sopprimerli, si mira a renderli espliciti e trasparenti. Ecco dunque che nasce OpinionGPT, una demo web in cui gli utenti possono porre domande e selezionare tutti le tipologie di pregiudizi che desiderano indagare. La demo risponderà a queste domande utilizzando un modello ottimizzato su un testo che rappresenta

ciascuno dei pregiudizi selezionati, consentendo un confronto fianco a fianco. Per addestrare questo modello sono stati identificati 11 diversi pregiudizi (politici, geografici, di genere, di età) ed è stato ricavato un corpus di istruzioni in cui ogni risposta era scritta da membri di una di queste categorie.

Ho provato a chiedere ad OpinionGPT e alle sue categorie di pregiudizi cosa ne pensa del cambiamento climatico in atto e a questo link trovate le risposte: https://opiniongpt.informatik.hu-berlin.de/r/Rgp_PIB .

Tra gli aspetti più interessanti e' il livello di maggiore concretezza e coinvolgimento emotivo del mondo femminile rispetto a quello maschile.

Se volete provare OpinionGPT, cliccate qui: https://opiniongpt.informatik.hu-berlin.de/ .

{jcomments on}

{loadposition user6}