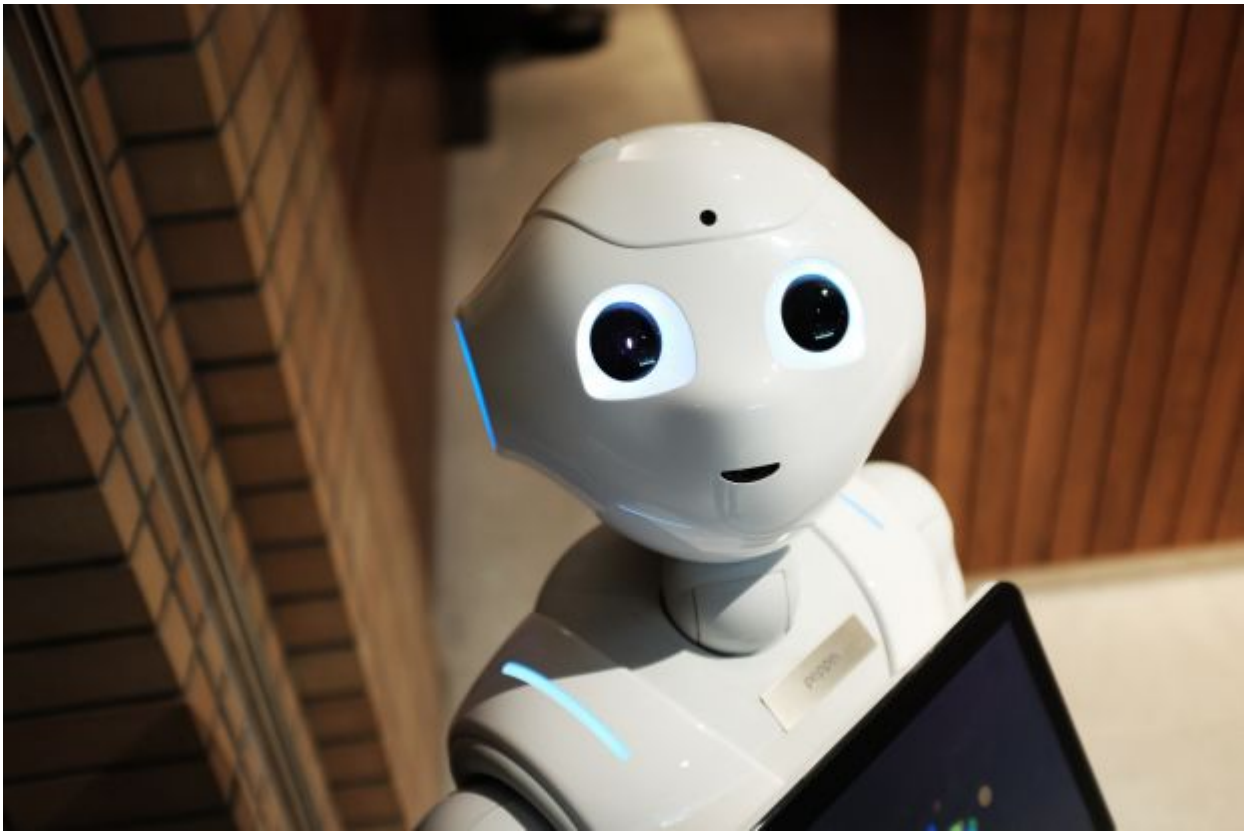
Post scritto da una IA inganna 26.000 persone



E raggiunge il primo posto su Hacker News.

All'inizio del 2019 destò una certa attenzione la presentazione di Gpt-2, un'intelligenza artificiale creata da OpenAI e capace di generare testi di qualità sufficiente da sembrare scritti da un essere umano.

In piena paranoia da *fake news*, Gpt-2 venne subito bollato come il generatore perfetto di notizie false, tanto che i suoi creatori decisero di rendere nota soltanto una parte del programma, e non il suo intero codice.

Nei mesi intercorsi da allora Gpt-2 si è evoluto in Gpt-3, che rispetto al predecessore è capace di creare testi ancora più simili a quelli che produrrebbe un autore umano.

A differenza di quanto fatto per Gpt-2, per Gpt-3 i suoi creatori hanno deciso di adottare un modello che non tenesse

segreto il frutto dei loro sforzi ma allo stesso tempo non lo rendesse disponibile al grande pubblico: la IA è stata pertanto messa a disposizione soltanto dei ricercatori che facciano richiesta di accesso alla beta privata.

Così ha fatto Liam Porr, studente universitario, il quale ha ideato un esperimento con cui ha fornito la prova circa la qualità dei testi di Gpt-3.

Porr ha usato l'intelligenza artificiale per scrivere un post che è stato tanto apprezzato da arrivare al primo posto su Hacker News. Soltanto pochi lettori hanno capito che era stato generato da un programma e, anzi, quanti l'hanno pubblicamente sostenuto nei commenti sono stati zittiti dagli altri, almeno fino a che il lavoro di Porr non è diventato di dominio pubblico.

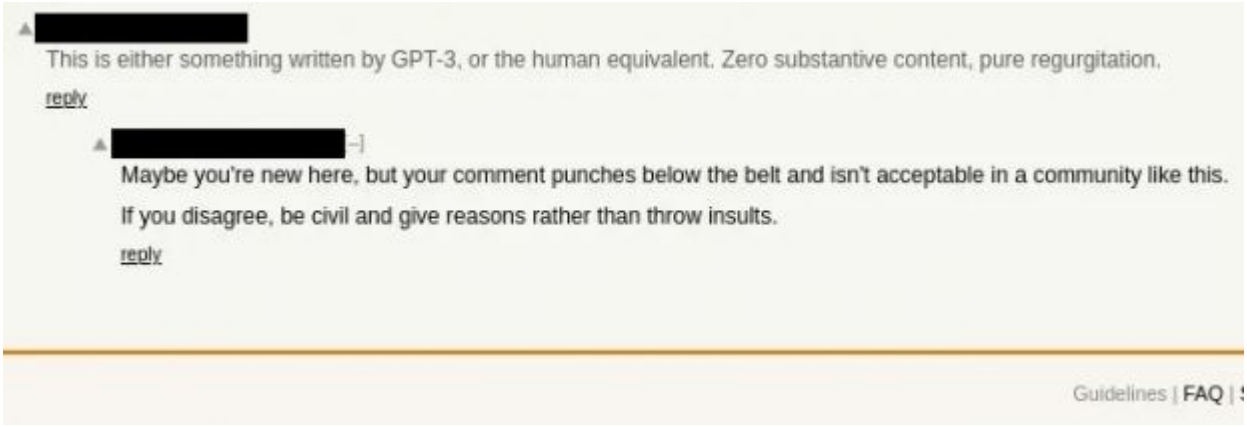
Tale lavoro non è stato estenuante. *«In realtà è stato facilissimo»* – ha raccontato lo studente – *«ed è questa la parte che fa paura»*.

Gpt-3 non è né può essere creativo nel senso di essere capace di dare vita a idee completamente nuove. Viene invece addestrato assorbendo un'enorme mole di dati, che analizza rilevando le parole chiave, le associazioni tra di esse, le espressioni più usate e via di seguito.

Poi, partendo da quanto ha digerito produce un nuovo testo, che chiaramente potrà essere soltanto una rielaborazione del materiale di partenza, esposto però in maniera abbastanza convincente da poter passare per qualcosa di nuovo.

Gpt-3 *«È bravo a scrivere con proprietà di linguaggio, ma non lo è molto nel seguire logica e razionalità»* commenta ancora Porr il quale, conscio di questi limiti, per il suo post di prova ha deciso di adottare un argomento per il quale non sarebbe servita una logica stringente: produttività e auto-aiuto.

Il tentativo è stato evidentemente un successo, anche perché fino alla rivelazione finale i pochi che avevano sollevato dubbi sull'origine del testo hanno immediatamente ricevuto voti negativi dagli altri, fenomeno che peraltro può dirla lunga sulla potenza del conformismo verso l'idea maggioritaria e la sfiducia verso chi solleva dubbi.



È questo un punto da tenere a mente sopratutto ora che Gpt-3 ha dimostrato come sia possibile ingannare un buon numero di persone (il post ha ricevuto 26.000 visualizzazioni in poche ore, e ciò prima dello svelamento), in quanto come già accadde per Gpt-2 si è riaffacciato lo spettro della possibilità di generare molta disinformazione con pochissima fatica.

E, a un livello forse più basso ma sempre problematico, sistemi come Gpt-3 possono essere usati per generare grandi quantità di contenuti clickbait privi di qualsiasi vero contenuto informativo, mandando alla lunga in crisi da un lato la produzione di contenuti nel web (dato che diventa difficile distinguere ciò che è originale da ciò che non lo è) e dall'altro la remunerazione tramite pubblicità (poiché sarà forse necessario provare ai network pubblicitari di ospitare solo contenuti scritti da persone, e non da robot).

«È possibile» – commenta Porr – *«che ci sia un'inondazione di contenuti mediocri, dato che le maglie dei filtri sono ormai tanto larghe. Penso che il valore dei contenuti online calerà parecchio»*.

OpenAI, intanto, non ha commentato l'operato di Porr. *«Magari*

sono un po’ seccati da quanto ho fatto» ipotizza il ragazzo. *«Insomma, è una cosa un po’ sciocca»*.