

L'IA è un'arma potente contro la disinformazione

L'Intelligenza artificiale è usata anche per contrastare la diffusione di notizie false.
Il professor Riccardo Gallotti, coordinatore del progetto europeo AI4TRUST, ci spiega come



L'osservazione che Riccardo Gallotti fa al termine dell'intervista, quando sto già chiudendo il quaderno degli appunti, è tutt'altro che banale. «Sempre più utilizzeremo strumenti come ChatGPT per scrivere testi. In questo modo la lingua usata dall'Intelligenza artificiale diventerà prescrittiva, e a poco a poco il nostro modo di scrivere si conformerà a quello che è considerato "giusto" dai suoi modelli linguistici». Ma non c'è inquietudine nelle parole di questo fisico teorico, che lavora a un progetto di contrasto della disinformazione online grazie all'Intelligenza artificiale (IA). «Le IA come ChatGPT non sono state addestrate a dire cose vere, sono solo in grado di unire parole che stanno bene insieme. Ma se gli chiedi di spiegarti la filosofia di Socrate, al momento possono farlo solo con la profondità di comprensione di un liceale».

Il vero limite di questa nuova tecnologia di cui parlano tutti è la sua incapacità di considerare il contesto in cui le cose vengono dette e scritte; questa è una specialità degli esseri umani, che hanno l'ultima parola anche nel progetto europeo da sei milioni di euro finanziato dall'Ue e coordinato proprio da Riccardo Gallotti.

Si chiama AI4TRUST, e ha lo scopo di combattere la disinformazione online grazie all'alleanza tra esseri umani e macchine. È coordinato dalla Fondazione Bruno Kessler (FBK) di Trento, l'Ente della Provincia autonoma che opera nel campo scientifico tecnologico e delle scienze umane e che è riuscito a ritagliarsi un ruolo importante nel campo delle IA a livello internazionale.

Sulla collina che si affaccia sulla città sede dell'FBK, Riccardo Gallotti lavora con i colleghi a un sistema di raccolta dati che consentirà di monitorare numerose piattaforme social online quasi in tempo reale, segnalando, mediante l'analisi di testo, audio e video i contenuti ad alto rischio di disinformazione. Questi verranno poi rivisti da esperti, i cosiddetti fact-checkers, che si occuperanno del debunking, cioè di confutare sulla base dei fatti le fake news. Anche se questo termine non piace a Riccardo Gallotti, che preferisce utilizzare quello di misinformazione – nata da un errore involontario – o di disinformazione, quando la notizia falsa viene diffusa con intenzioni manipolatrici. I risultati di questo lavoro saranno messi a disposizione degli addetti ai lavori e dei professionisti dell'informazione su una piattaforma dedicata.

L'impegno in questo campo di Riccardo Gallotti, che dirige il CHuB Lab (The Complex Human Behaviour Lab) della FBK, viene da lontano: formatosi come fisico teorico, si appassiona alla fisica della complessità, quella che studia non gli oggetti in quanto tali ma i rapporti che hanno tra di loro. Si occupa di trasporto pubblico e di spostamenti in aereo usando i dati che – fino all'avvento di Elon Musk – si potevano liberamente trarre da

Twitter e nel gennaio del 2020 alla FBK viene coinvolto in un progetto lanciato dal professor Manlio De Domenico, ora all'Università di Padova. Lo scienziato ha avuto l'idea di seguire l'evoluzione della disinformazione attorno a un virus che ancora non interessava a nessuno: nasceva così l'Infodemic Observatory for COVID-19, che verrà poi sostenuto dall'OMS come strumento di monitoraggio della diffusione della disinformazione attorno alla pandemia.

Ma come funziona AI4TRUST? «Prima di tutto – spiega Riccardo Gallotti – abbiamo individuato tre temi attorno ai quali lavorare: il cambiamento climatico, la salute pubblica e la questione dei migranti, tutti argomenti con un impatto globale e su cui prolifera la disinformazione. Grazie a degli strumenti di data science scandagliamo un set di parole chiave ma anche di utenti e gruppi chiave presenti sui social media e su alcune testate online. Il risultato è una grande mole di dati che viene data in pasto all'intelligenza artificiale, la quale estrae le tematiche e le narrative emergenti in rete e le segnala al team internazionale di professionisti dell'informazione, che si occupa di verificare la loro veridicità, le seleziona e le sottopone al vaglio della critica. Inoltre, grazie all'IA e a strumenti matematici propri della scienza delle reti complesse, si può riuscire a capire quanto le notizie false siano il risultato di una campagna concertata in modo volontario oppure no».

Viene in mente, naturalmente, la campagna elettorale che portò all'elezione nel 2016 di Donald Trump a presidente USA, segnata dall'interferenza di operazioni di disinformazione sui social orchestrate dalla Russia. Insomma, l'obiettivo di AI4TRUST è quello di creare una piattaforma di consultazione dove i professionisti possano avere sott'occhio i trending topics, cioè i temi attorno ai quali cresce la disinformazione in un dato periodo.

In questo caso l'IA funziona come un filtro, che sulla base dei modelli su cui è stata «educata» riesce a fornire materiale già selezionato ai revisori umani cui spetta l'ultima parola.

Ma come fare a istruire l'Intelligenza artificiale? «Prendiamo l'esempio dei discorsi di odio: si mette a punto una tabella di Excel con 1000 frasi valutate da annotatori umani come hate speech (linguaggio d'odio, ndr.) e altre 1000 valutate come di linguaggio normale. Sulla base di queste l'IA impara a differenziare le une dalle altre, in un processo di apprendimento continuo basato su set di dati che gli vengono dati in pasto, mentre le sue competenze vengono continuamente testate».

Quello della disinformazione è un tema delicato e di grande impatto, anche politico: secondo un'inchiesta internazionale IPSOS del 2019 ben 86% dei 25mila intervistati in tutto il mondo sostenevano di essere stati confrontati a fake news su Internet, in particolare su Facebook, e quasi nove su dieci hanno ammesso di averci inizialmente creduto. Il nuovo fronte della battaglia contro la disinformazione online è quello dei deep fake, cioè fotografie, audio o video modificati in modo così realistico dalle IA da sembrare veri: il politico al quale si fa dire il contrario di quello che ha detto veramente oppure una innocente foto di un'adolescente che viene trasformata in un nudo, come è successo recentemente in Spagna.

«Paradossalmente – spiega Riccardo Gallotti – per un'IA è più facile riconoscere i deep fake in un video o in un audio che le fake news in un testo scritto. Anche perché alcune comunità nelle quali si è diffusa la disinformazione su alcuni temi hanno imparato ad usare i neologismi per passare sotto i radar». L'IA è dunque un'arma potente nelle mani di chi vuole diffondere notizie false. Ora potrà esserlo anche per chi vuole opporsi alla loro diffusione.

Mattia Pelli, Azione, 04/12/2023