MACCHINE CHE PRENDONO POTERE E IL GIUSTO VALORE DELL'UOMO

PAOLO BENANTI


Avvenire IDEE E COMMENTI 3

Munizioni con uranio impoverito in Ucraina un incubo che ritorna

La decisione britannica di inviare a Kiev proiettili radioattivi pone problemi non solo sanitari. Lo si è già visto nella Guerra del Golfo, nei Balcani e perfino in Sardegna

Ore

Lettera alla pioggia che è vera grazia

Intelligenze artificiali sempre più «agentive» e strabilianti Il mondo delle intelligenze artificiali(AI) è in subbuglio. La settimana scorsa OpenAi leader in questo settore ha rivelato all'umanitàintera l'ultima sua creazione: Gpt-4. Un prodotto dalla sigla incomprensibile ma che in pochissimotempo ha riempito le colonne dei giornali e le pagine dei blog. I risultati che hanno ottenuto aOpenAi sono senz'altro strabilianti. Se leggiamo il documento tecnico che è stato pubblicato almomento del rilascio vediamo capacità inimmaginabili. Gpt-4 può partire da una domanda testualeassociata a un'immagine, mostra di saper riconoscere cosa ci sia nell'immagine e di unire i diversielementi in un quadro di senso: risolvere un problema matematico o rispondere su cosa ci sia di stranoin una foto. Gpt-4 accetta richieste che consistono sia in immagini che in testo, il che –parallelamente all'impostazione di solo testo – consente all'utente di specificare qualsiasi compitodi visione o di linguaggio. In particolare, il modello genera output di testo, dati input costituitida testo e immagini interlacciati in modo arbitrario. Su una serie di domini – tra cui documenti contesto e fotografie, diagrammi o schermate – Gpt-4 mostra capacità simili a quelle degli input di solotesto.

Abbiamo un modello che, figurativamente, unisce "udito", nel senso che sente la nostra richiestascritta, e "vista", mostrando una "percezione artificiale" di un'immagine.

La cosa più strabiliante è che Gpt-4 sa fare cose che i ricercatori non si aspettavano che sapessefare.

Sempre nel documento citato leggiamo: « Nei modelli più potenti emergono spesso nuove capacità.

Alcune di esse sono particolarmente interessanti: la capacità di creare e agire su piani a lungotermine, di accumulare potere e risorse ("ricerca di potere") e di mostrare un comportamento semprepiù "agentivo". In questo contesto, per "agentivo" non si intende l'umanizzazione dei modellilinguistici o il riferimento alla senzienza, ma piuttosto sistemi caratterizzati dalla capacità diraggiungere obiettivi che potrebbero non essere stati specificati concretamente e che non sono apparsinell'addestramento, di concentrarsi sul raggiungimento di obiettivi specifici e quantificabili e difare piani a lungo termine».

Se guardiamo la fonte citata nelle note al testo, si legge, senza mezzi termini: «Usiamo il termineagentività per sottolineare il fatto sempre più evidente che i sistemi di ML [machine learning] nonsono completamente sotto il controllo umano».

Quindi i ricercatori si accorgono che il modo stesso con cui hanno addestrato il sistema – non solol'utilizzo di dati ma il premiare alcune risposte e punirne altre mediante interazione con uomini e

l'uso di modelli di ricompensa basati su regole – trasmette al modello due elementi: la capacità diadottare strategie a lungo termine e la ricerca di potere e risorse nella sua interazione con l'input.
Gli ingegneri che hanno sviluppato Gpt-4 ci dicono che il sistema potrebbe avere un suo «piano internoe potrebbe agire per acquisire le risorse e i modi per ottenerlo ». Tutto questo non perché sia unasorta di genio del male ma probabilmente perché, nell'apprendimento per rinforzo gli addestratori nel"premiare" o "punire" Gpt-4 emettono giudizi, e questi non sono mai solo costruiti su una cosa in séma su una cosa in merito a un fine. Intuitivamente, potremmo pensare che il sistema incorpori questi"microframmenti" di finalità delle valutazioni nel processo di rinforzo facendo emergere una finalitàglobale in una maniera analoga a quella con cui fa emergere informazioni dai dati.
Se questo è quello che sappiamo di Gpt-4, Google con il suo PaLM – un sistema dello stesso tipo –sembra aver raggiunto risultati ancora più potenti e strabilianti, tali da ritardarne il rilasciopubblico per la paura degli effetti.
Ma quanto vale tutto questo? Da un punto di vista commerciale tantissimo. Microsoft, acquisite lelicenze del sistema, sta lanciando una serie di prodotti in cui Gpt-4 aiuterà l'utente a fare i suoicompiti. Chiama questi sistemi Copilot. Dobbiamo abituarci all'idea di una trasformazione di ognicampo del lavoro in cui avremo un copilota che ci guida e ci assiste in ogni ambito della vita? Al dilà dei problemi sociali, la domanda che sorge, allora, è quanto vale l'uomo in questa nuova stagionecaratterizzata dalle AI. A questa domanda ha risposto nei giorni scorsi papa Francesco, che parlandoai partecipanti ai Minerva Dialogues – un tavolo di confronto in cui sono presenti molti attori diprimo piano in questo mondo – ha detto: « Il concetto di dignità umana – questo è il centro – ciimpone di riconoscere e rispettare il fatto che il valore fondamentale di una persona non può esseremisurato da un complesso di dati. [...] Non possiamo permettere che gli algoritmi limitino ocondizionino il rispetto della dignità umana, né che escludano la compassione, la misericordia, ilperdono e, soprattutto, l'apertura alla speranza di un cambiamento della persona». Se la macchina valein senso economico, l'uomo ha un valore che non è misurabile né in valori numerici dei dati né inquelli economici in un bilancio. La dignità umana, al centro dell'algoretica, ci chiede qualirelazioni e quale società vogliamo.