

OPINIONI A CONFRONTO

IA: PRECARIA E INATTENDIBILE

Non è affatto certo che i Large Language Model di oggi miglioreranno. Al contrario, un nuovo studio dimostra che stanno addirittura peggiorando, sono diventati migliori nel rispondere a domande difficili ma sono peggiorati nella risoluzione dei problemi di base: per esempio, le operazioni aritmetiche più semplici.

Fabrizio Tonello

Molti lettori di *Professione Docente* mi hanno scritto rimproverandomi di aver pubblicato nell'ultimo numero] chatgpt: non un oracolo ma un assistente sgobbone - Gilda Professione Docente <https://gildaprofessionedocente.it/news/dettaglio.php?id=1225> un articolo divertente ma non approfondito, né sufficientemente esplicativo, sull'Intelligenza Artificiale. Torno quindi sull'argomento, cercando di spiegare perché ChatGPT ha difficoltà a fare addizioni e sottrazioni.

Partiamo da un dato: l'euforia della Borsa attorno all'IA è interamente basata sulla convinzione che migliorerà sostanzialmente nel tempo. Colossi come Microsoft e JP Morgan stanno investendo decine di miliardi di dollari nelle società che lavorano nel settore non perché i loro attuali prodotti siano particolarmente validi, ma perché i loro modelli futuri dovrebbero essere capaci di sostituire milioni di lavoratori e fare una tonnellata di soldi.

Tuttavia questa scommessa, anche mettendo da parte gli effetti distruttivi per i lavoratori e per la stessa economia, si basa premesse discutibili: non è affatto certo che i *Large Language Model* di oggi miglioreranno. Al contrario, un nuovo studio dimostra che stanno addirittura peggiorando.

Quali sono i modi in cui è possibile migliorare l'IA? Il metodo più discusso e più utilizzato finora è stato quello di raccogliere più dati e di utilizzare più potenza di calcolo per "addestrare" il modello. Un altro metodo è la formazione: perfezionare l'IA sui dati già in suo possesso grazie al feedback umano sulle sue prestazioni, rendendola più efficiente con ciò che già possiede.

Oggi OpenAI, Meta e BLOOM si stanno scontrando con i limiti del primo metodo (nel gergo degli esperti *scaling-up*). Affinché l'IA continui a migliorare allo stesso ritmo i dati, l'infrastruttura e l'utilizzo di energia dovrebbero aumentare in modo esponenziale. Poiché oggi queste aziende sono ben lontane dal fare profitti questa non sembra un'opzione fattibile per il futuro.

Il secondo metodo, definito *shaping-up* è quello del modello Strawberry di OpenAI, che non utilizza più dati di addestramento rispetto al suo predecessore (il mio ChatGPT-4 che ha difficoltà con le tabelline) ma ha un'interfaccia basata su una tecnica chiamata *chain-of-thought* che dovrebbe consentire di produrre modelli di Intelligenza Artificiale migliori senza dover sborsare i fantastiliardi di cui si parla.

Purtroppo sembra che entrambi i metodi stiano raggiungendo i loro limiti: José Hernández-Orallo e i suoi colleghi dell'Università Politecnica di Valencia hanno esaminato le prestazioni degli ultimi modelli di *Large Language Model* con i loro predecessori, interrogandoli su problemi aritmetici, risoluzione di anagrammi, domande geografiche, estrazione di informazioni da elenchi disorganizzati. I risultati sono veramente interessanti.

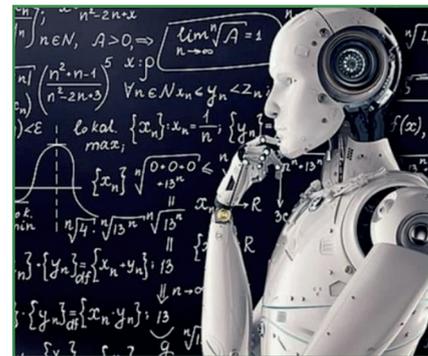
Hanno scoperto che sono diventati migliori nel rispondere a domande difficili ma sono peggiorate nella risoluzione dei problemi di base: per esempio, le operazioni aritmetiche più semplici. Quindi il mio ChatGPT-4 forse **alle origini** sapeva le tabelline ma adesso, ingozzato di nuove informazioni, non le sa più. Uno studio dell'Università della California, Berkeley conferma le mie impressioni del tutto empiriche e casuali. Se cercate su Google "L'intelligenza artificiale è peggiorata", troverete centinaia di articoli e commenti che confermano: i nuovi modelli sono peggiori rispetto alla generazione precedente.

Quindi, cosa sta succedendo? Per il momento sia lo *scaling-up* che lo *shaping-up* non conferiscono magicamente la capacità di risolvere compiti più complessi a questi modelli di Intelligenza Artificiale. In altre parole, non acquisiscono nuove capacità: sembra piuttosto che sostituiscono le vecchie abilità con altre nuove, ad esempio perdendo la capacità di rispondere a domande fondamentali.

Molti informatici sono convinti che non ci siano molte opzioni praticabili per rendere questi chatbot capaci di risolvere problemi di ogni tipo: con molti sforzi, si potrà solo farli migliorare in compiti specifici.

OpenAI e altri produttori pensano che il metodo di *shaping-up* possa risolvere questo problema e rendere l'IA in grado di operare a più livelli, aiutandola a risolvere compiti sia semplici che complessi e a conservare le capacità di base senza perderle quando migliora in altre. Tuttavia, poiché il metodo di *shaping-up* è già stato utilizzato ampiamente fin qui, Hernández-Orallo e altri studiosi rimangono scettici. Ed è un problema che non riguarda solo i chatbot: per esempio, man mano che il software delle Tesla diventa più sofisticato, gli automobilisti riferiscono che peggiora in compiti fondamentali come non passare sui marciapiedi.

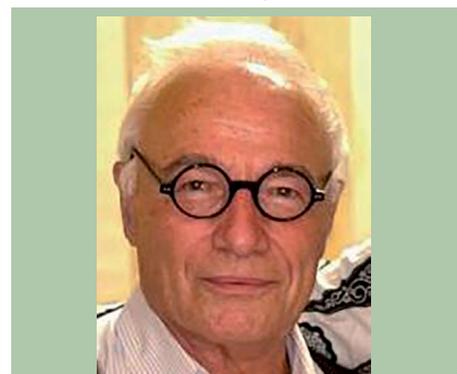
Naturalmente ci sono alcune soluzioni possibili, come l'assemblaggio di più modelli di IA in parallelo, con alcuni migliori in certi compiti e altri in compiti differenti. In questo modo, il chatbot darebbe l'impressione di



avere applicazioni molto varie, perché utilizzerebbe l'IA più adatta allo specifico compito da svolgere. Ma le aziende come OpenAI e i suoi concorrenti stanno già faticando per pagare lo sviluppo dei loro sistemi attuali, quindi anche questo approccio ha dei limiti (benché Microsoft, il principale investitore in OpenAI, abbia una montagna di miliardi da spendere).

Hernández-Orallo riassume la situazione dicendo che oggi ci fidiamo dei nostri comodi chatbot più di quanto dovremmo. Soprattutto perché gli errori semplici sono a volte i più difficili da individuare: il suo studio ha messo in luce un grave difetto nel modo in cui attualmente utilizziamo i *Large Language Model*. Le strade che le aziende produttrici hanno scelto per dare vita a questi modelli "rivoluzionari" di fatto sembrano senza uscita. Non solo peggioreranno l'Intelligenza Artificiale in molti compiti elementari che sono necessari, ma non hanno trovato soluzioni per rendere fattibile la creazione di IA avanzate di prossima generazione.

Ne parleremo in un prossimo articolo (se nel frattempo la bolla borsistica non sarà scoppiata, con effetti devastanti per l'economia americana, cosa che può succedere in qualsiasi momento).



FABRIZIO TONELLO

è docente di Scienza Politica presso l'Università di Padova, dove insegna, tra l'altro, un corso sulla politica estera americana dalle origini ad oggi. Ha insegnato alla University of Pittsburgh e ha fatto ricerca alla Columbia University, oltre che in Italia (alla SISSA di Trieste e all'Università di Bologna). Ha scritto *Democrazia a rischio. La produzione sociale dell'ignoranza* (Pearson, 2019), *L'età dell'ignoranza* (Bruno Mondadori 2010), *Il Nazionalismo americano* (Livian, 2007), *La politica come azione simbolica* (Franco Angeli, 2003). Da molti anni collabora alle pagine culturali del *Manifesto*.