Dopo anni, **OpenAI** rende onore al suo nome pubblicando dei modelli di **intelligenza artificiale** che sono effettivamente “open”, ossia aperti a tutti e gratuiti. Si tratta di due strumenti leggeri e agili che possono essere **gestiti localmente** da computer e server accessibili al grande pubblico, un dettaglio che permetterà agli utilizzatori di preservare la privacy delle interazioni evitando di travasare i dati su servizi cloud che sono abitualmente gestiti da quelle aziende che sono costantemente accusate di sfruttare illecitamente le informazioni raccolte. Nell’attesa del lancio di GPT5, OpenAI ha diffuso in rete **gpt-oss-120b e gpt-oss-20b**, modelli **“open-weight”** che permettono a chi ci lavora sopra di avere piena consapevolezza dei parametri di riferimento adoperati durante l’addestramento del sistema. A livello tecnico, questo dettaglio permette di eseguire i modelli su infrastrutture private, di personalizzarli con dati propri e di integrarli in applicazioni senza appoggiarsi a realtà esterne. A livello accademico, la cosa è interessante perché rende più facile comprendere come un modello finisca a generare certi risultati e certe “allucinazioni”.

L’azienda non aveva più toccato questi livelli di apertura sin dal 14 febbraio **2019**, ovvero dal rilascio iniziale di GPT2. Ai tempi, OpenAI era ancora pienamente una no-profit che si poggiava su ricercatori che credevano nella *mission* accademica dell’organizzazione: costruire un’intelligenza artificiale generale sicura e benefica, condividendo in maniera trasparente con il resto del mondo gli esiti del loro processo di ricerca. Nel marzo del 2019, a neppure un mese dal lancio ufficiale di GPT2, OpenAI ha annunciato la nascita della sua sussidiaria for-profit, **stravolgendo i suoi obiettivi originari** e scatenando una scissione interna che ha poi dato vita alla concorrente Anthropic.

gpt-oss-20b, la più piccola delle nuove varianti, si poggia su 21 miliardi di parametri, i quali vengono ottimizzati da un sistema *mixture-of-experts* (MoE) perché ogni singola unità minima di testo – token – venga elaborata facendo riferimento solamente a 3,6 miliardi di parametri. gpt-oss-120b, il maggiore dei due modelli, scala rispettivamente questi orizzonti a 117 miliardi e a 5,1 miliardi. In termini concreti, vuol dire che gpt-oss-20b può essere **sostenuto da un normale computer** d’alto livello che sia dotato di almeno 16GB di memoria, mentre gpt-oss-120b abbisogna di strumenti che toccano gli 80GB, un requisito decisamente meno comune da soddisfare, ma comunque raggiungibile.

OpenAI non sta però certamente distribuendo gratuitamente nuovi modelli che possano concretamente fare concorrenza ai servizi che vende: per questioni tecniche e commerciali, i due strumenti sono stati progettati per essere leggeri, ma anche **limitati**, inoltre il loro impiego richiede una consapevolezza tecnica di affinamento che va oltre alle capacità del consumatore medio. È inoltre opportuno rimarcare che *open-weight* e *open-source* non sono la stessa cosa: gpt-oss-120b e gpt-oss-20b sono pensati per mostrare i parametri di addestramento, ma non i dati originali di riferimento o il codice di programmazione che è

stato adoperato dall’architettura impiegata. **Non sono “open”** nel senso più assoluto del termine. Si tratta però di limitazioni comprensibili, visto che i giganti del settore – OpenAI compresa – si stanno lanciando in operazioni al limite dello spionaggio per avere la meglio sui propri concorrenti e che i dati di addestramento contengono probabilmente elementi che l’azienda non aveva il diritto di toccare.

A preoccupare c’è anche il fatto che, operando in un contesto lontano dagli occhi e dalle potenzialità di controllo di un gestore, i modelli *open-weight* possano essere affinati per **scopi malevoli**. In tal senso, OpenAI cerca di rassicurare il pubblico sostenendo di aver eseguito test interni utili a verificare che i due strumenti non possano essere impiegati in direzioni rischiose nei contesti biologici e della cybersicurezza. “Il Safety Advisory Group (“SAG”) di OpenAI ha esaminato questi test e ha concluso che […] gpt-oss-120b non ha raggiunto l’High capability nei domini Biological and Chemical Risk o Cyber risk”, conclude l’azienda autoassolvendosi da ogni potenziale malefatta.

Walter Ferri

Giornalista milanese, per *L’Indipendente* si occupa della stesura di articoli di analisi nel campo della tecnologia, dei diritti informatici, della privacy e dei nuovi media, indagando le implicazioni sociali ed etiche delle nuove tecnologie. È coautore e curatore del libro *Sopravvivere nell'era dell'Intelligenza Artificiale*.