IL DIBATTITO SULL'AI

Microsoft ha annunciato che nei prossimi mesi integrerà Copilot, il suo assistente AI, in Windows 11. Sarà quindi possibile, fra le altre cose, chiedere all'assistente di "regolare le impostazioni" o di eseguire altre azioni su un computer (The Verge).
"Stiamo introducendo Windows Copilot, rendendo Windows 11 la prima piattaforma per computer ad annunciare un'assistenza AI centralizzata per aiutare le persone a intervenire facilmente e a realizzare le cose", scrive Microsoft nel suo blog.

In un diverso post, sempre Microsoft spiega il concetto di Copilot (copilota). "Un Copilot è un'applicazione che utilizza la moderna AI e modelli linguistici di grandi dimensioni (LLM) come GPT-4 per assistere le persone in compiti complessi. Microsoft ha introdotto per la prima volta il concetto di Copilot quasi due anni fa con GitHub Copilot, un strumento AI di programmazione che assiste gli sviluppatori nella scrittura del codice, e continuiamo a rilasciare dei Copilot in molte delle attività principali dell'azienda. Riteniamo che il Copilot rappresenti un nuovo paradigma nel software alimentato dall'intelligenza artificiale e un profondo cambiamento nel modo in cui lo stesso software viene sviluppato".

Sul concetto di Copilot torna anche il CTO di Microsoft, Kevin

Scott, colui che sta al cuore di questa trasformazione dell'azienda all'insegna dell'AI. In una intervista a The Verge/Decoder, infatti dice: "Volevamo immaginare come utilizzare questa tecnologia per assistere le persone nel lavoro cognitivo che stanno svolgendo. Il primo che abbiamo costruito è stato GitHub Copilot, uno strumento che aiuta le persone a scrivere codice per svolgere le loro attività di sviluppatori di software. Molto rapidamente ci siamo resi conto che si trattava di un modello per un nuovo tipo di programma, che dunque non ci sarebbe stato solo GitHub Copilot, ma molti [altri] Copilot".

Watermark sui contenuti sintetici

Altro spezzone interessante è quando Scott affronta il tema della generazione di contenuti sintetici e dell'effetto negativo che possono avere, dalla diffusione di disinformazione alla creazione di meccanismi di loop in cui le stesse AI si addestrano su contenuti sintetici, prodotti da altre AI.
"Abbiamo lavorato a un sistema di riconoscimento dei media che consente di inserire un watermark crittografico invisibile nei contenuti audio-visivi", spiega Scott, in modo che, quando si riceve questo contenuto, un software possa decifrare le informazioni che lo riguardano e che indicano da dove arriva.
"È utile per il rilevamento della disinformazione in generale. L'utente può dire: "Voglio consumare solo contenuti di cui capisco la provenienza". O si può dire: "Non voglio consumare contenuti generati dall'intelligenza artificiale", prosegue Scott. Allo stesso modo, il sistema potrebbe essere usato quando si addestra una AI per eliminare contenuti sintetici dai dati di addestramento.

L'intervento di Bengio

Ma se ogni settimana ci sono annunci di prodotto (e non è compito di questa newsletter passarli in rassegna, anche

perché sono tantissimi), ogni settimana arrivano anche nuovi spunti su quello che ho definito il dibattito sull'AI, ovvero la parte di discussione più sociale e politica (che include anche differenti visioni su quelle che sono le capacità tecniche attuali e future di questa tecnologia).

Così, dopo i commenti di Yann LeCun, Geoffrey Hinton, e altri, non poteva mancare un altro dei pionieri della rivoluzione deep learning, Yoshua Bengio. Il vincitore del Premio Turing 2018 ha dichiarato "che la recente corsa di Big Tech al lancio di prodotti di intelligenza artificiale è diventata 'malsana' – scrive il FT – aggiungendo di vedere un "pericolo per i sistemi politici, per la democrazia, per la natura stessa della verità".

A proposito di verità: il problema di queste interviste con esperti è che non sono vere e proprie interviste con domanda e risposta in cui i due interlocutori sono chiaramente separati, ma conversazioni che mescolano virgolettati con un resoconto di quanto detto. Questo rende difficile, specie su temi complessi e tecnici come questi, capire bene le priorità dell'intervistato e le sfumature che attribuisce ad aspetti diversi. Ad ogni modo, quella che traspare dall'articolo è una posizione più pragmatica rispetto ad altri ricercatori del settore. E ha il merito di mettere il dito nella piaga. Scrive il FT: "Bengio ha affermato che la cosa più urgente da fare per le autorità di regolamentazione è rendere i sistemi di AI più trasparenti, anche facendo degli audit sui dati utilizzati per addestrarli e i loro risultati. Inoltre, insieme ad altri colleghi, ha proposto una coalizione internazionale per finanziare la ricerca sull'AI in settori importanti per il pubblico, come il clima e la sanità. "Come gli investimenti nel CERN in Europa o nei programmi spaziali: questa è la scala che dovrebbe essere usata oggi per gli investimenti pubblici nell'AI per portare davvero i benefici dell'intelligenza artificiale a tutti, e non solo per fare un sacco di soldi", ha detto".

Tuttavia Bengio questa settimana affronta anche il tema dei "rischi esistenziali" di una AI autonoma che che possa agire nel mondo in modo catastrofico in un post scritto di suo pugno.
"Il tipo di AI più sicuro che riesco a immaginare è quello privo di qualsiasi capacità di agire in modo autonomo (agency), ma dotata solo di una comprensione scientifica del mondo (il che potrebbe già essere immensamente utile). Credo che dovremmo stare alla larga dai sistemi di AI che assomigliano e si comportano come esseri umani, perché potrebbero diventare delle AI fuori controllo (rogue) e perché potrebbero ingannarci e influenzarci (per promuovere i loro interessi o gli interessi di qualcun altro, non i nostri)"
E ancora: "La sicurezza delle AI richiede ancora molta ricerca, sia a livello tecnico che a livello politico. Ad esempio, vietare i sistemi di AI potenti (ad esempio, quelli superiori alle capacità di GPT-4) a cui viene data autonomia e capacità di agire (agency) sarebbe un buon inizio. Ciò comporterebbe sia una regolamentazione nazionale che accordi internazionali. La motivazione principale che spinge i Paesi in conflitto (come gli Stati Uniti, la Cina e la Russia) a concordare su un trattato di questo tipo è che un'AI fuori controllo (rogue) può essere pericolosa per l'intera umanità, indipendentemente dalla sua nazionalità. Qualcosa di simile alla paura dell'Armageddon nucleare che probabilmente ha motivato l'URSS e gli Stati Uniti a negoziare trattati internazionali sugli armamenti nucleari fin dagli anni Cinquanta".

Il paragone col nucleare

Interessante notare che questi paragoni col nucleare fatti da parte di alcuni ricercatori, imprenditori e politici stanno aumentando (come documentato da settimane in questa newsletter). A rincarare la dose negli ultimi giorni è stato il parlamentare democratico statunitense Seth Moulton, membro della commissione Usa che si occupa di forze armate e difesa.

“Ciò che ci distingue dall’era nucleare è che non appena abbiamo sviluppato le armi nucleari, c’è stato uno sforzo massiccio per limitarne l’uso”, ha detto in una intervista a una newsletter di Politico. “Non ho visto nulla di paragonabile a questo con l’AI. È molto più pericoloso. La Cina sta investendo enormi risorse nell’AI. Putin ha detto che chi vincerà la gara dell’AI controllerà il mondo. Tutti i nostri principali avversari sono in una vera e propria gara con noi sull’AI, e quindi stiamo perdendo la capacità di impostare la definizione di questi standard internazionali”.

Non parliamo abbastanza delle responsabilità attuali

E qui torniamo necessariamente dal piano della ricerca a quello della politica.
Nella scorsa newsletter avevo raccontato l’audizione sull’AI al Senato americano, con le testimonianze, tra gli altri, di Sam Altman, il Ceo di OpenAI (la società che ha rilasciato ChatGPT, DALL-E ecc). E avevo sottolineato due aspetti: l’atmosfera amichevole e la strana propensione a chiedere regole da parte dell’industria. Ci torno questa settimana perché in effetti James Vincent su The Verge si sofferma proprio sulle due questioni. E scrive: “La cosa più insolita dell’audizione del Senato di questa settimana sull’AI è stata l’affabilità. I rappresentanti dell’industria – in primis l’amministratore delegato di OpenAI Sam Altman – si sono trovati d’accordo sulla necessità di regolamentare le nuove tecnologie di AI, mentre i politici sembravano felici di lasciare la responsabilità di redigere le regole alle aziende stesse (…)
Ma la stessa introduzione di un sistema di licenze, come proposto da Altman e altri all’audizione, potrebbe in realtà non avere un effetto immediato, prosegue Vincent. Mentre, durante l’audizione, i rappresentanti dell’industria hanno spesso richiamato l’attenzione su ipotetici danni futuri, prestando scarsa attenzione ai problemi noti che l’AI già

determina.

Gebru e il problema dei discorsi sui rischi esistenziali

Su questo tema in settimana ritorna anche la ricercatrice di AI Timnit Gebru (più volte citata in questa newsletter) intervistata dal Guardian. A proposito del problema di concentrarsi sui rischi della superintelligenza, di una fantomatica AGI (Artificial General Intelligence) e dei cosiddetti "rischi esistenziali" per l'umanità (l'AI è piena di espressioni e termini ambigui il cui utilizzo tradisce però precise visioni filosofiche-politiche) dice:"Questo tipo di conversazione attribuisce capacità d'azione autonoma, [e relativa responsabilità] (agency) a uno strumento invece che agli esseri umani che lo costruiscono". Così si può evitare di assumersi responsabilità, prosegue Gebru. "Si dice: 'Non sono io il problema. È lo strumento. È superpotente. Non sappiamo cosa farà'. No, il problema sei tu. State costruendo qualcosa con caratteristiche precise e lo fate per il vostro profitto. [Tutta questa impostazione] distrae e distoglie l'attenzione dai danni reali e dalle cose che dobbiamo fare. Subito".

Una via africana all'AI?

A questo proposito segnalo una discussione molto interessante su un magazine africano, The Continent, che proprio nella sua ultima edizione si sofferma su una visione differente dell'AI. Cita l'esempio di una startup, Lelapa AI, "fatta da africani, per africani", come sottolinea il suo sito.
"Lelapa AI non sta cercando di creare un programma che ci surclassi tutti. Al contrario, sta creando programmi mirati che utilizzano l'apprendimento automatico e altri strumenti per rispondere a esigenze specifiche", scrive The Continent. "Il suo primo grande progetto, Vulavula, è stato concepito per fornire servizi di traduzione e trascrizione per lingue sottorappresentate in Sudafrica. Invece di raccogliere sul web

i dati di altri, Lelapa AI collabora con linguisti e comunità locali per raccogliere informazioni, permettendo agli stessi di partecipare ai profitti futuri.(…)". Lelapa AI si differenzia dunque da "quei programmi costruiti dall'Occidente su dati provenienti dall'Occidente che rappresentano i loro valori e principi", ha commentato una delle sue fondatrici, Jade Abbott. Che nota come le prospettive e la storia africana siano in gran parte già escluse dai dati utilizzati da OpenAI e dai modelli linguistici di grandi dimensioni di Google. "Questo perché non possono essere facilmente "raccolti dal web(scraped)". Gran parte della storia africana è registrata oralmente o è stata distrutta dai colonizzatori; e le lingue africane non sono supportate (parlate con ChatGPT in Setswana o in isiZulu e le sue risposte saranno in gran parte prive di senso). Per Lelapa, tutto questo rappresenta un'opportunità", scrive ancora The Continent.

Chatbot e religioni

La localizzazione di questi strumenti in una specifica nazione o cultura può assumere però contorni anche molto diversi. Rest of the World racconta di come in India siano nati vari chatbot di AI a sfondo religioso. Ad esempio, GitaGPT. "il chatbot, alimentato dalla tecnologia GPT-3, che fornisce risposte basate sulla Bhagavad Gita", il testo sacro più diffuso fra milioni di indiani. Secondo Rest of World "alcune delle risposte generate dai bot di Gita mancano di filtri per le discriminazioni di casta, la misoginia e persino la [violazione della] legge"