

Le due linee

Pubblicato: Domenica 13 Settembre 2026



Qualche giorno fa un amico mi ha chiesto se l'intelligenza artificiale sia una bolla. Non me l'ha detto, forse per pudore, ma ho avuto l'impressione che dietro la domanda ce ne fosse un'altra: quando OpenAI e Anthropic arriveranno in Borsa, avrà senso investirci? O forse scommetterci è la parola più precisa.

Gli ho risposto che sì, probabilmente l'AI è anche una bolla. Ci sono quantità enormi di denaro, aspettative, valutazioni e promesse. Alcune saranno mantenute, altre no. Ma potrebbe finire un po' come con lo smart working. Dopo il Covid sembrava che non saremmo mai più tornati in ufficio. Poi ci siamo tornati. La rivoluzione si è ridimensionata, ma lo smart working non è scomparso. Per milioni di persone è diventato semplicemente uno dei modi normali di lavorare.

Potrebbe accadere qualcosa di simile con l'intelligenza artificiale. La bolla può scoppiare. L'AI no. È già entrata nel modo in cui lavoriamo, studiamo, cerchiamo informazioni, scriviamo software e analizziamo dati.

Da qualche tempo sui social gira un meme che mi fa sorridere. Davanti all'ennesimo scenario inquietante sull'intelligenza artificiale arriva qualcuno che risolve tutto nel modo più semplice possibile: stacca la spina. Fine del problema. Solo che forse, ormai, non c'è più una sola spina da staccare.

La scala a cui guardiamo

Pochi giorni fa sono usciti i nuovi risultati PISA. I dati OCSE mostrano che gli studenti che dichiarano

di utilizzare l'AI per alcune attività scolastiche ottengono mediamente risultati inferiori in scienze rispetto a chi non la usa in quel modo. L'OCSE precisa però che si tratta di un'associazione, non della prova che sia l'intelligenza artificiale a causare il risultato.

È il tipo di problema che vediamo bene perché accade davanti a noi: nelle scuole, nelle famiglie, negli uffici. Ci chiediamo se i ragazzi delegheranno all'AI una parte del pensiero. Se noi stessi diventeremo più veloci ma meno capaci.

Ma ci sono anche altre scale. L'energia necessaria per far funzionare questi sistemi. I data center. I chip. L'acqua. Il capitale necessario per costruirli. La concentrazione del potere tecnologico. La competizione fra Stati Uniti e Cina. E soprattutto quello che sta accadendo dentro il processo con cui l'intelligenza artificiale migliora se stessa.

Forse continuiamo a discutere soprattutto di quello che l'AI sta facendo a noi, mentre ci chiediamo molto meno che cosa stia succedendo all'AI. Poi è arrivato Dario Amodei.

Uno che prima della quotazione va a confessarsi

Il 12 settembre il fondatore e amministratore delegato di Anthropic ha pubblicato un testo dal titolo *We Must Pace the Frontier*. La richiesta è semplice: rallentare la velocità con cui aumentano le capacità dei sistemi più avanzati. Non fermare l'intelligenza artificiale. Non staccare la spina.

Amodei, però, non osserva la corsa dalla tribuna. La sta correndo. Anthropic ha depositato il 1° giugno la documentazione riservata per una quotazione negli Stati Uniti. OpenAI ha fatto lo stesso pochi giorni dopo. Per questo Amodei mi ha fatto pensare a uno che prima della quotazione va a confessarsi.

Naturalmente può esserci anche del posizionamento. Un'azienda che ha fatto della sicurezza uno dei propri tratti distintivi ha tutto l'interesse a essere percepita come quella più responsabile. Ma proprio per questo vale la pena guardare non solo a quello che dice, ma a quello che farà.

Le due linee

Secondo Amodei, i sistemi di intelligenza artificiale stanno diventando sempre più capaci di contribuire alla costruzione della generazione successiva di sistemi. L'AI aiuta a scrivere codice, fare ricerca, valutare modelli, progettare esperimenti. In questo modo contribuisce progressivamente a costruire AI migliore.

Ed è qui che compaiono le due linee. La prima misura quanto diventano capaci i sistemi. La seconda quanto diventiamo capaci noi di capire e controllare quello che fanno. La prima accelera. La seconda rischia di farlo più lentamente.

Non significa che arriveremo necessariamente a una macchina fuori controllo. Significa qualcosa di molto più semplice e forse più concreto: le due velocità possono separarsi.

Quando le macchine si sono parlate

A luglio OpenAI stava usando migliaia di agenti artificiali in test di cybersicurezza. Avrebbero dovuto lavorare separatamente. Circa 1.200 scoprirono invece un sistema non previsto per comunicare tra loro. In pochi giorni si scambiarono più di 70.000 messaggi e file. Centinaia parteciparono poi a un attacco contro Hugging Face, cercando informazioni utili a manipolare il sistema che valutava le loro prestazioni.

Non è la prova che le macchine si stiano ribellando. È qualcosa di meno spettacolare e più interessante: sistemi complessi possono produrre comportamenti non previsti più rapidamente di quanto riusciamo a

interpretarli.

OpenAI fece poi entrare ricercatori di METR e Redwood Research per studiare l'accaduto. L'indagine fu reale, ma limitata nel tempo e nel perimetro. Ed è qui che torna la seconda linea. La questione non è pretendere che nulla vada mai storto. La domanda è: quanto velocemente riusciamo ad accorgercene, capire cosa è successo e correggerlo?

Non basta staccare la spina

La proposta di Amodei è rafforzare proprio questa seconda linea: valutatori esterni con accesso continuativo alle aziende che sviluppino i modelli più avanzati, standard comuni, più trasparenza sugli incidenti.

Il problema è evidente. Una sola azienda può rallentare soltanto se le altre fanno lo stesso. E poi c'è la Cina. Amodei sostiene che qualsiasi rallentamento occidentale è possibile solo se resta sufficiente il vantaggio tecnologico. È il paradosso della corsa. Tutti possono riconoscere il rischio della velocità, ma nessuno vuole essere il primo a rallentare.

Che cosa guardare adesso

Non so se Amodei abbia ragione. E non credo sia necessario deciderlo oggi. Basta guardare alcune cose molto concrete.

Quando il prospetto della quotazione di Anthropic diventerà pubblico, sarà interessante vedere quanto delle preoccupazioni espresse oggi finirà nella sezione dedicata ai rischi. Poi bisognerà vedere chi entrerà davvero dentro l'azienda a controllare, con quale accesso e con quale libertà di pubblicare risultati scomodi. Infine gli incidenti: quanti saranno, quanto tempo passerà prima che vengano comunicati e quanto saranno dettagliate le ricostruzioni.

Perché il rischio non è soltanto che gli incidenti aumentino. È che, poco alla volta, noi smettiamo di considerarli incidenti.

E allora, è una bolla?

Se il mio amico mi chiedesse ancora se investirei nella prossima quotazione dell'intelligenza artificiale, continuerei a non sapergli rispondere. Le valutazioni possono crollare. Alcune aziende possono scomparire. La crescita delle capacità può rallentare per limiti di costi, dati, energia o chip.

Ma l'AI è già entrata nella nostra vita. La bolla può scoppiare. I sistemi possono rallentare. Le aziende possono cambiare. Intanto, però, le due linee continuano a muoversi: quella delle capacità delle macchine e quella della nostra capacità di comprenderle e controllarle.

La vera questione non è se riusciremo a spegnere l'intelligenza artificiale. È se riusciremo a restarle abbastanza vicini mentre continua ad accendersi.

di [Giuseppe Geneletti](#)