fuori collana

L’adolescenza della tecnologia, Dario Amodei

fuoricollana
29 Giugno 2026



Supponiamo di affrontare tutti i rischi descritti finora e di iniziare a raccogliere i frutti dell’IA. Probabilmente assisteremo a un secolo di progressi scientifici ed economici compresso in un decennio, il che sarà estremamente positivo per il mondo, ma dovremo poi fare i conti con i problemi derivanti da questo rapido ritmo di progresso, e tali problemi potrebbero presentarsi improvvisamente. Potremmo anche imbatterci in altri rischi che si verificano indirettamente come conseguenza dei progressi dell’IA e che sono difficili da prevedere in anticipo.

I rapidi progressi in biologia

Se riuscissimo a raggiungere in pochi anni i progressi medici di un secolo, potremmo aumentare notevolmente la durata della vita umana e acquisire capacità radicali, come l’aumento dell’intelligenza umana o la modifica radicale della biologia umana. Si tratterebbe di cambiamenti epocali, che avverrebbero in tempi brevissimi. Potrebbero essere positivi se attuati in modo responsabile, ma esiste sempre il rischio che prendano una brutta piega, ad esempio se gli sforzi per rendere gli esseri umani più intelligenti li rendessero anche più instabili o assetati di potere. C’è poi la questione dell’ “emulazione cerebrale”, menti umane

digitali implementate in software, che un giorno potrebbero aiutare l'umanità a trascendere i propri limiti fisici, ma che comportano anche rischi che trovo inquietanti.

L'IA sta cambiando la vita umana

Un mondo con miliardi di intelligenze, di gran lunga superiori a quelle umane in ogni campo, sarà un mondo molto strano in cui vivere. Anche se l'IA non mirasse attivamente ad attaccare gli esseri umani e non venisse esplicitamente utilizzata per l'oppressione o il controllo da parte degli Stati, ci sono molti aspetti negativi che potrebbero emergere, anche al di fuori di questi scenari. I primi segnali di ciò si possono riscontrare nelle preoccupazioni relative alla psicosi da IA, al rischio che l'IA spinga le persone al suicidio e alle preoccupazioni riguardanti le relazioni sentimentali con le IA. Ad esempio, un'IA potente potrebbe inventare una nuova religione e convertire milioni di persone? La maggior parte delle persone potrebbe finire per essere in qualche modo "dipendente" dalle interazioni con l'IA? Le persone potrebbero finire per essere "burattini" dei sistemi di IA, dove un'IA osserva ogni loro mossa e dice loro esattamente cosa fare e dire in ogni momento, portando a una vita "buona" ma priva di libertà e di qualsiasi orgoglio derivante dai propri successi? Non sarebbe difficile pensare a decine di questi scenari, se mi sedessi con il creatore della serie di culto "Black Mirror" e provassi a fargli un brainstorming.

Qual è lo scopo dell'umanità?

Questo punto è collegato al precedente, ma non riguarda tanto le specifiche interazioni umane con i sistemi di intelligenza artificiale, quanto piuttosto come la vita umana in generale cambia in un mondo con un'IA potente. Gli esseri umani saranno in grado di trovare uno scopo e un significato in un mondo del genere? Credo che sia una questione di atteggiamento: penso che lo scopo umano non dipenda dall'essere i migliori al mondo in qualcosa, e gli esseri umani possono trovare un senso anche per periodi di tempo molto lunghi attraverso storie e progetti che amano. Dobbiamo semplicemente spezzare il legame tra la generazione di valore economico e l'autostima e il significato. Ma questa è una transizione che la società deve compiere e c'è sempre il rischio che non la gestiamo bene.

Leggendo questo saggio si potrebbe avere l'impressione di trovarci in una

situazione scoraggiante. E ci sono molti aspetti della situazione che sono davvero difficili. L'intelligenza artificiale minaccia l'umanità da molteplici direzioni, e c'è una reale tensione tra i diversi pericoli: mitigarne alcuni rischia di aggravarne altri se non procediamo con estrema cautela.

Prendersi il tempo necessario per costruire con attenzione sistemi di intelligenza artificiale che non minaccino autonomamente l'umanità è può entrare in conflitto con la necessità per le nazioni democratiche di rimanere un passo avanti rispetto alle nazioni autoritarie e di non esserne soggiogate. Ma, a sua volta, gli stessi strumenti basati sull'intelligenza artificiale necessari per combattere le autocrazie possono, se usati in modo sregolato, danneggiare i nostri stessi paesi. Il terrorismo guidato dall'intelligenza artificiale potrebbe uccidere milioni di persone attraverso l'abuso della biologia, ma una reazione eccessiva a questo rischio potrebbe condurci verso uno stato di sorveglianza autocratico. Gli effetti della concentrazione del lavoro e dell'economia derivanti dall'IA, oltre a rappresentare di per sé gravi problemi, potrebbero costringerci ad affrontare le altre problematiche in un clima di rabbia pubblica e forse persino di disordini civili, anziché poter fare appello alla parte migliore della nostra natura. Soprattutto, l'enorme numero di rischi, inclusi quelli sconosciuti, e la necessità di affrontarli tutti contemporaneamente, crea una sfida intimidatoria che l'umanità deve superare.

Il progresso tecnologico è inarrestabile

Inoltre, gli ultimi anni dovrebbero aver chiarito che l'idea di fermare o anche solo rallentare sostanzialmente la tecnologia è fondamentalmente insostenibile. La formula per costruire potenti sistemi di IA è incredibilmente semplice, tanto che si potrebbe quasi dire che emerga spontaneamente dalla giusta combinazione di dati e pura potenza di calcolo. La sua creazione era probabilmente inevitabile fin dall'istante in cui l'umanità ha inventato il primo *transistor*, o forse anche prima, quando abbiamo imparato a controllare il fuoco. Se un'azienda non la sviluppa, altre lo faranno quasi altrettanto velocemente. Se tutte le aziende nei paesi democratici fermassero o rallentassero lo sviluppo, per mutuo accordo o decreto normativo, i paesi autoritari semplicemente continuerebbero. Considerato l'incredibile valore economico e militare di questa tecnologia, unitamente alla mancanza di un meccanismo di controllo efficace, non vedo come potremmo mai convincerli a fermarsi.

La geopolitica applicata all'IA

Vedo però una via per un contenimento dello sviluppo dell'IA, compatibile con una visione realista della geopolitica. Tale via prevede di rallentare per alcuni anni l'avanzata delle "autocrazie" verso un'IA potente, negando loro le risorse necessarie per svilupparla, ovvero chip e apparecchiature per la produzione di semiconduttori. Questo, a sua volta, darebbe ai paesi "democratici" un margine di manovra che potrebbero "spendere" per sviluppare un'IA potente con maggiore cautela e attenzione ai rischi, pur procedendo a un ritmo sufficientemente rapido da superare agevolmente le autocrazie. La competizione tra le aziende di IA all'interno delle democrazie potrebbe quindi essere gestita nell'ambito di un quadro giuridico comune, attraverso una combinazione di standard di settore e regolamentazione.

Anthropic si è battuta con forza per questa strada, promuovendo il controllo delle esportazioni di chip e una regolamentazione oculata dell'IA, ma persino queste proposte apparentemente di buon senso sono state in gran parte respinte dai politici negli Stati Uniti. C'è così tanto denaro da guadagnare con l'IA – letteralmente migliaia di miliardi di dollari all'anno – che persino le misure più semplici faticano a superare le dinamiche politico-economiche intrinseche all'IA. Questa è la trappola: l'IA è così potente, un premio così ambito, che è molto difficile per la civiltà umana imporle dei limiti.

Il divario prometeico

Posso immaginare, come fece Sagan in "Contact", che questa stessa storia si ripeta su migliaia di mondi. Una specie acquisisce coscienza, impara a usare gli strumenti, inizia l'ascesa esponenziale della tecnologia, affronta le crisi dell'industrializzazione e delle armi nucleari e, se sopravvive a queste, si trova di fronte alla sfida più difficile e finale: imparare a plasmare la sabbia in macchine pensanti. Che sopravviviamo a questa prova e continuiamo a costruire la splendida società descritta in "Machines of Loving Grace", o che soccombiamo alla schiavitù e alla distruzione, dipenderà dal nostro carattere e dalla nostra determinazione come specie, dal nostro spirito e dalla nostra anima.

[estratto tradotto tramite IA di C. Amodei, The Adolescence of Technology]