

Il lato oscuro del progresso scientifico: l'ipotesi del mondo vulnerabile

23 Maggio 2023

Lo sviluppo tecnologico ci ha resi e ci renderà sempre più potenti. Siamo abbastanza saggi da utilizzare questo potere per rendere il mondo un posto migliore? O stiamo accelerando verso la nostra rovina?

La vita degli esseri umani è stata radicalmente trasformata, per il meglio, dallo sviluppo scientifico e tecnologico degli ultimi secoli. La mortalità infantile [è crollata](#), come anche la [povertà estrema](#); l'aspettativa di vita media globale è [più che raddoppiata](#) dal 1800; [servizi igienico-sanitari](#), [energia elettrica](#), e [internet](#) sono accessibili a sempre più persone.

Naturalmente, il progresso scientifico ha portato anche nuovi problemi, come il riscaldamento globale e le armi di distruzione di massa, ma tutto sommato, il suo effetto è stato estremamente positivo. Tuttavia, potremmo chiederci: **continuerà a essere positivo anche in futuro?**

Immaginiamo lo sviluppo tecnologico come un'estrazione da un'urna piena di palline. Ogni pallina corrisponde a un'idea o a un'invenzione, e il suo colore indica l'impatto di tale scoperta sul mondo. Le palline bianche sono benefiche, come ad esempio la [soluzione reidratante orale](#) che ha salvato milioni di vite dalle malattie diarroiche; mentre quelle grigie sono moderatamente dannose o armi a doppio taglio, come la fissione nucleare, che ci ha dato [energia pulita](#) ma anche le bombe di Hiroshima e Nagasaki.

C'è infine un terzo tipo di pallina. Non siamo certi della sua esistenza, poiché non è mai stato estratto nell'intera storia dell'umanità. Ma se esistesse, sarebbe **il peggior rischio che corriamo giocando alla lotteria del progresso**: le palline nere. Esse rappresentano tecnologie che, se inventate, distruggerebbero inevitabilmente la nostra civiltà.

Questa metafora è stata proposta da [Nick Bostrom](#) – professore di filosofia a Oxford, direttore del [Future of Humanity Institute](#), e autore del best seller [Superintelligence](#) – nel suo articolo [The Vulnerable World Hypothesis](#): l'ipotesi del mondo vulnerabile. Secondo questa ipotesi, esiste un **livello tecnologico** che – se raggiunto in assenza di straordinarie capacità di vigilanza preventiva e *governance* globale – **renderebbe la devastazione della nostra civiltà**

praticamente certa. In altre parole, un mondo vulnerabile è un mondo in cui l'urna delle possibili invenzioni contiene almeno una pallina nera.

Bostrom distingue quattro tipi di “vulnerabilità della civiltà”, tutti causati da delle palline nere. Il primo prevede una **tecnologia che permetta a piccoli gruppi o individui di causare gravi danni con facilità**. Ad esempio, se assemblare una bomba atomica fosse molto più semplice ed economico di quanto (fortunatamente) davvero è, la nostra civiltà sarebbe terribilmente instabile. La percentuale di persone che deciderebbero, per qualsivoglia motivo, di costruire e usare tale arma sarebbe verosimilmente minima; ma, essendoci otto miliardi di persone sulla Terra, essa basterebbe comunque a causare una catastrofe.

Potrebbero esserci tecnologie di questo tipo nel nostro futuro? La **biologia sintetica** offre un candidato plausibile: il costo di [sequenziare](#) e [sintetizzare](#) il DNA è sempre più contenuto, il che è estremamente positivo per la ricerca, ma potrebbe rendere lo [sviluppo di virus pericolosi](#) facilmente accessibile ai malintenzionati.

Il secondo e il terzo tipo di vulnerabilità prevedono invece delle **tecnologie che incentivano degli attori, anche se benintenzionati, ad agire nocivamente**. In un caso, si tratta di pochi attori così potenti da causare una catastrofe unilateralmente, come gli stati con armi nucleari: se qualche tecnologia avesse permesso agli USA o all'URSS di lanciare il primo colpo nucleare in modo sicuro, compromettendo così la fragile stabilità fornita dalla [distruzione mutua assicurata](#), la Guerra Fredda probabilmente non sarebbe stata così fredda.

Nell'altro caso, si tratta invece di molti individui le cui azioni hanno conseguenze solo marginalmente negative, ma catastrofiche quando combinate tra loro. Il riscaldamento globale è l'esempio più lampante: l'effetto di un singolo viaggio in macchina è trascurabile, ma la somma di tutte le emissioni di gas serra ha e avrà un terribile impatto. Per quanto grave, però, [è improbabile](#) che il riscaldamento globale possa direttamente porre fine alla nostra civiltà, ma è difficile escludere che qualche tecnologia futura possa creare una situazione analoga e peggiore.

Infine, il quarto tipo di vulnerabilità prevede una **tecnologia che pone un rischio ignoto o non ben compreso**, e porta quindi alla catastrofe anche in assenza di malintenzionati o di pessimi incentivi. Una nuova tecnologia può essere quindi rischiosa se crea delle condizioni senza precedenti. Ad esempio, l'obiettivo di varie aziende – tra cui [OpenAI](#), creatrice del celebre ChatGPT – è quello di sviluppare un'**intelligenza artificiale generale** (AGI): un'IA in grado di fare tutto ciò che può fare un essere umano, e magari farlo meglio. Tale tecnologia potrebbe, per la prima volta nella storia dell'umanità, **compromettere la nostra posizione come gli esseri più intelligenti sul pianeta**, con [conseguenze potenzialmente disastrose](#) per tutti noi. Questo esempio è però imperfetto: dati gli incentivi economici a sviluppare un'AGI prima dei concorrenti e gli incentivi strategici a svilupparla prima di altre potenze mondiali, aziende e stati potrebbero decidere di [allentare le misure di sicurezza](#) e rilasciare delle IA non sicure, riportandoci al secondo tipo di vulnerabilità.

Bostrom analizza poi vari modi in cui possiamo mitigare i rischi e sopravvivere a una pallina nera. Bloccare lo sviluppo tecnologico – smettere di estrarre palline dall'urna – non è realistico, né desiderabile. Potremmo però cercare di **rallentare lo sviluppo di tecnologie pericolose e accelerare quello di tecnologie protettive**: l'ordine in cui le palline vengono estratte

potrebbe fare la differenza, se l'urna contiene anche delle palline in grado di proteggerci da quelle nere.

Ma questo e altri metodi non sembrano sufficienti, e i metodi potenzialmente più efficaci sono poco allettanti: per stabilizzare quante più vulnerabilità possibili, può essere infatti necessaria una

combinazione di vigilanza preventiva e governance globale. La prima, plausibilmente ottenuta attraverso una sorveglianza di massa senza precedenti, preverrebbe atti criminali disastrosi. La seconda, forse nella forma di un singolo governo mondiale, risolverebbe facilmente i problemi di coordinazione tra gli stati. Bostrom non si sbilancia riguardo la desiderabilità di tali misure: la sopravvivenza alle palline nere va bilanciata con tutte le problematiche, anche orwelliane, che un simile regime comporterebbe. Inoltre, potrebbero esserci altri modi, ancora inesplorati, per salvaguardare il nostro futuro.

Non sappiamo con certezza se l'urna della creatività umana contiene delle palline nere: la verità dell'ipotesi del mondo vulnerabile è una questione aperta. Per Bostrom, però, **l'evidenza disponibile rende irragionevole essere sicuri che tale ipotesi sia falsa.** E data la posta in gioco, dovremmo forse iniziare a prenderla sul serio.

Se finora non abbiamo mai estratto una pallina nera, non è perché siamo stati particolarmente prudenti o saggi: siamo stati fortunati. Finché la ricerca scientifica e tecnologica continuerà, continueremo a estrarre palline dall'urna, e **affidarsi solamente alla dea bendata potrebbe essere fatale.** Urge una conversazione sul futuro dell'umanità, sui rischi esistenziali posti dalle tecnologie emergenti, e su come possiamo mitigare tali rischi, perché la strategia attuale – sperare che non ci siano palline nere, o ignorare totalmente la possibilità che ce ne siano – è da irresponsabili.

LINK

<https://magazine.fbkeu/it/news/il-lato-oscuro-del-progresso-scientifico-lipotesi-del-mondo-vulnerabile/>

TAG

- #vulcano

AUTORI

- Stocco Luca