

MalpensaNews

L'intelligenza artificiale tra crisi di dati e bolla infrastrutturale: opportunità e rischi per il futuro

Tommaso Guidotti · Sunday, September 8th, 2024

Negli ultimi anni, l'intelligenza artificiale (IA) ha vissuto una crescita esponenziale, divenendo uno dei settori tecnologici più promettenti. Tuttavia, dietro questa corsa all'innovazione, si nascondono due grandi problemi: una crescente crisi di dati e una possibile bolla infrastrutturale. Queste minacce potrebbero rallentare lo sviluppo dell'IA, ma al tempo stesso offrono nuove opportunità per coloro che sapranno adattarsi e innovare.

La crisi dei dati: l'IA rischia di “perdere intelligenza”

Secondo un'inchiesta di The Guardian curata da Jack Apollo George, una delle problematiche più rilevanti nel futuro dell'intelligenza artificiale riguarda una crisi imminente legata alla disponibilità di dati. Entro un periodo compreso tra il 2026 e il 2032, l'intelligenza artificiale rischia di raggiungere un punto critico, dove i modelli avranno esaurito tutto il materiale umano disponibile per l'addestramento. Per comprendere appieno questa crisi, bisogna considerare il funzionamento degli attuali sistemi di IA. Questi modelli si basano su enormi quantità di dati, estratti da testi, immagini, video e altre forme di contenuti prodotti dagli esseri umani nel corso della storia.

Tuttavia, esiste un limite fisico alla quantità di informazioni disponibili: articoli, libri, post sui social media, immagini, video e altre risorse digitali. **Secondo le previsioni, entro il 2032, questi modelli avranno “letto” e assimilato praticamente tutto ciò che è stato prodotto dall'uomo.** A quel punto, non ci saranno più nuovi dati umani da cui trarre insegnamento. Questo significa che i modelli, invece di continuare a crescere e migliorare, potrebbero iniziare a ristagnare o addirittura regredire, essendo costretti a riutilizzare continuamente dati già noti. In altre parole, l'intelligenza artificiale rischia di esaurire il “combustibile” che le permette di migliorare, con gravi conseguenze per la sua evoluzione. A quel punto rimarrebbero solo i dati sintetici, generati artificialmente dalla stessa IA, per continuare l'apprendimento. Ma questo amplificherebbe il problema. Spiega Apollo George:

“Ogni volta che un modello viene addestrato su dati sintetici, perde consapevolezza della “coda lunga” dei “dati minoritari” su cui era stato originariamente addestrato (come parole rare, fatti insoliti, ecc.). La vastità della conoscenza viene erosa e sostituita solo dai dati più probabili: i modelli linguistici di grandi dimensioni sono, in sostanza, macchine sofisticate di predizione del testo. Quindi, quando i dati originali digitali sono già sbilanciati – fortemente orientati verso la lingua inglese, prevalentemente centrati sugli Stati Uniti e pieni di post inaffidabili dai forum – questo pregiudizio verrà semplicemente ripetuto”. Fonte:

<https://www.theguardian.com/technology/article/2024/sep/07/if-journalism-is-going-up-in-smoke-i-might-as-well-get-high-off-the-fumes-confessions-of-a-chatbot-helper>

In questo scenario, diventa cruciale sviluppare nuovi approcci per mantenere viva la crescita dei modelli di intelligenza artificiale. Alcuni suggeriscono di incentivare la creazione di contenuti umani, mentre altri guardano a nuovi ambiti di studio, come la capacità dei modelli di imparare autonomamente attraverso tecniche avanzate di autoapprendimento.

La bolla infrastrutturale: il paradosso degli investimenti nell'IA

Se da un lato i dati scarseggiano, dall'altro le infrastrutture per l'intelligenza artificiale sono cresciute a dismisura, creando una potenziale bolla economica. Secondo un'analisi di David Cahn per Sequoia Capital, la differenza tra gli investimenti nell'infrastruttura IA e i ritorni economici effettivi si sta ampliando. Questa discrepanza è diventata così rilevante da trasformarsi in quello che Cahn definisce "la domanda da 600 miliardi di dollari". Fonte:

<https://www.sequoiacap.com/article/ais-600b-question/>

Per comprendere il problema, è importante sapere che le aziende tech, come Microsoft, Google e Nvidia, hanno investito enormi quantità di capitale per costruire giganteschi data center dotati di GPU (unità di elaborazione grafica) di ultima generazione. Questi data center sono fondamentali per far funzionare le piattaforme IA. Tuttavia, mentre gli investimenti continuano a crescere, i ritorni economici provenienti dalle applicazioni non stanno aumentando allo stesso ritmo. Al di fuori di

giganti come OpenAI (3,4 miliardi di dollari di entrate nel 2024), le nuove aziende stanno faticando a raggiungere ricavi significativi.

Questo squilibrio tra investimento e ritorno è simile a quello che abbiamo visto in passato con la bolla delle dot-com o con la costruzione di ferrovie nel XIX secolo. La storia insegna che le bolle speculative spesso portano a enormi perdite economiche, con molte aziende incapaci di realizzare profitti reali nonostante gli investimenti iniziali. Inoltre, l'infrastruttura IA è soggetta a una rapida svalutazione. Nvidia, per esempio, sta già preparando il lancio del suo chip B100, che sostituirà l'attuale H100 con prestazioni 2,5 volte superiori e un costo "solo" del 25% più alto. Questo significa che i costosi investimenti fatti oggi potrebbero perdere valore molto velocemente.

Innovare per superare le crisi

Nonostante queste sfide, l'intelligenza artificiale resta un settore promettente. La crisi dei dati e la bolla infrastrutturale possono trasformarsi in opportunità per chi saprà affrontarle con creatività e strategia. Per affrontare la carenza di dati reali, è necessario sviluppare nuovi approcci. Collaborazioni tra aziende tecnologiche, istituzioni accademiche e governi potrebbero favorire la condivisione di dati e garantire che i modelli continuino a essere addestrati su dati di alta qualità.

Inoltre, si stanno studiando algoritmi che possano funzionare con meno dati o che siano in grado di imparare da set di dati più piccoli e diversificati, riducendo la dipendenza da enormi quantità di informazioni.

Per evitare che la bolla infrastrutturale scoppi, le aziende dovrebbero focalizzarsi su come rendere più efficienti i sistemi esistenti. Investire in modelli che ottimizzino l'uso delle risorse e riducano i costi può contribuire a mantenere stabile il mercato. Inoltre, lo sviluppo di applicazioni specializzate in settori come la sanità, la finanza o l'ambiente può generare ritorni economici tangibili e sostenibili. Ma la vera soluzione è che l'innovazione si concentri sulla creazione di valore reale per gli utenti. E questo può essere facilitato se le tecnologie diventano open-source e accessibili a giovani imprenditori capaci di nuove intuizioni e in contatto con le diverse esigenze e gradi di maturità di ogni parte del mondo.

“Se stai giocando a poker da trenta minuti e non sai chi sia il gonzo, tu sei il gonzo”, Warren Buffett.

This entry was posted on Sunday, September 8th, 2024 at 9:34 am and is filed under [Scienza e Tecnologia](#)

You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. You can leave a response, or [trackback](#) from your own site.