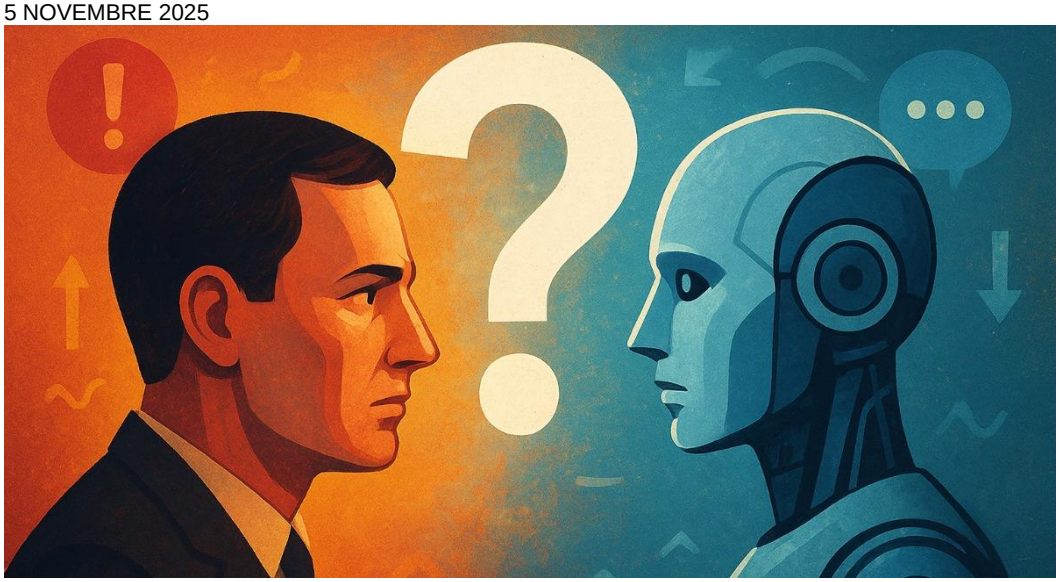
Predire l’imprevedibile: perché l’IA ha sempre smentito i suoi profeti

di **Antonella De Robbio**
https://ilbolive.unipd.it/it/news/scienza-ricerca/predire-limprevedibile-perche-lia-ha-sempre

5 NOVEMBRE 2025



La maggior parte delle previsioni sulle Intelligenze Artificiali (IA), fatte negli ultimi decenni, sono state disattese, spesso in entrambe le direzioni: **alcune troppo ottimistiche, altre troppo catastrofiste**. Le ragioni sono molteplici e si collocano a cavallo tra tecnologia, psicologia, economia e cultura. La costante di queste previsioni è l’ottimismo tecnologico e la sottovalutazione della complessità del mondo reale, spesso combinata con pressioni sociali o economiche che alimentano narrazioni futuristiche. Le previsioni sull’IA falliscono perché l’IA è un fenomeno non lineare, emergente e sociotecnico, dove scienza, economia, cultura e psicologia interagiscono in modo caotico. Ciò che impariamo è che ogni previsione precisa sull’IA è quasi certamente sbagliata, ma l’analisi dei suoi effetti reali – sociali, etici, giuridici – è invece più utile e concreta.

Le profezie mancate non si contano. Già nel 1957 Herbert Simon, premio Nobel per l’economia e uno dei padri fondatori dell’IA, dichiarava che entro vent’anni una macchina sarebbe stata in grado di svolgere qualsiasi lavoro umano, sopravvalutando la possibilità di tradurre in formule logiche la percezione, il linguaggio e la creatività. Pochi anni dopo, nel 1967, Marvin Minsky si mostrò altrettanto ottimista, affermando che entro un decennio avremmo avuto macchine dotate di intelligenza paragonabile a quella umana. Negli anni Settanta, però, l’IA era ancora confinata a sistemi esperti e programmi capaci di giocare a scacchi o a go, mentre ancora oggi ’intelligenza artificiale “generale” rimane un orizzonte lontano. Il matematico John McCarthy, che aveva coniato nel 1956 il termine *Artificial Intelligence*, fu anch’egli vittima di eccessivo entusiasmo, sostenendo che i problemi fondamentali dell’IA sarebbero stati risolti entro una generazione. McCarthy e i suoi contemporanei immaginavano che bastasse formalizzare i processi logici per emulare la mente umana, ma sottovalutarono la complessità del mondo reale, dominato da ambiguità, contesto e percezioni non strutturate. Negli anni Novanta e Duemila Ray Kurzweil, profeta del transumanesimo, annunciò che entro il 2029 avremmo avuto sistemi in grado di superare il test di Turing e di dialogare in modo indistinguibile da un essere

umano. I progressi dei modelli linguistici di grandi dimensioni, come ChatGPT, hanno certamente avvicinato quel traguardo, ma si tratta ancora di algoritmi statistici, privi di coscienza e intenzionalità. Kurzweil aveva confuso la potenza di calcolo con la comprensione: le macchine possono imitare l'intelligenza, ma non *essere* intelligenti nel senso umano del termine.

Anche le previsioni più "pratiche" si sono rivelate fallaci. Negli anni Ottanta molti immaginavano che nei Duemila ogni casa avrebbe avuto un robot personale in grado di cucinare, pulire e accudire i bambini o gli anziani. Oggi i robot domestici più diffusi sono aspirapolvere automatici: nessuno prepara la cena o accompagna i figli a scuola o accudisce gli anziani. I pionieri avevano ignorato la complessità dei movimenti fisici, la percezione ambientale e il costo elevato delle tecnologie necessarie. Infine, negli anni Dieci del Duemila, diverse analisi economiche preannunciarono che l'IA avrebbe sostituito la maggior parte dei lavori entro il 2020. Non è accaduto: l'automazione ha cambiato il modo di lavorare in molti settori, ma non ha cancellato il ruolo umano. Gli esperti avevano sottovalutato la resilienza delle professioni, la lentezza con cui le innovazioni si diffondono e la capacità delle persone di adattarsi e di creare nuove funzioni complementari alle macchine. Perché è così difficile cogliere nel segno quando si parla di ciò che verrà? Certo, il futuro – in qualunque ambito umano – resta per sua natura imprevedibile, ma nel caso dell'intelligenza artificiale c'è qualcosa di più.

In primis dobbiamo considerare l'influenza dei modelli economici, in quanto le previsioni sono spesso state strumentali o ideologiche: gonfiate da start-up e venture capital per attirare investimenti, oppure minimizzate da chi temeva la sostituzione del lavoro umano. La "verità tecnica" si è persa tra narrazioni economiche e storytelling politico. Uno dei nodi è **la difficoltà intrinseca nel prevedere le emergenze dei sistemi complessi.** Un modello linguistico di medie dimensioni si comporta in un modo, ma un LLM di grandi dimensioni mostra comportamenti emergenti che nessuno aveva anticipato, qualitativamente diversi.

Questo rende impossibile predire in anticipo cosa succederà, anche per gli scienziati che lo progettano. In parte la difficoltà nel progettare e di conseguenza poter prevedere effetti a lungo termine è nella **natura non lineare del progresso tecnologico**, non continuo ma costellato di "salti" e "inverni". Negli anni Sessanta e Ottanta si pensava che bastasse aumentare la potenza di calcolo per raggiungere un'IA generale, poi arrivarono gli AI winters per mancanza di risultati, periodi di forte rallentamento o disillusione nello sviluppo e nel finanziamento della ricerca sull'intelligenza artificiale. Il **primo inverno** arrivò negli anni Settanta, quando i modelli simbolici non riuscirono a produrre risultati concreti e i fondi pubblici furono tagliati. Il **secondo inverno** si verificò alla fine degli anni Ottanta dopo il fallimento dei sistemi esperti, troppo rigidi e costosi da mantenere. Seguì una fase di **stagnazione,** fino alla rinascita del deep learning negli anni Dieci del nostro secolo, grazie ai big data e alla potenza delle GPU (*Graphics Processing Units*, *unità di elaborazione grafica*), processori progettati originariamente per gestire la grafica dei videogiochi. In quegli anni la possibilità di sfruttare le GPU per il calcolo scientifico ha permesso di elaborare enormi quantità di dati e di accelerare in modo decisivo lo sviluppo dei modelli di apprendimento automatico, rendendo **possibile una** primavera dell'IA successiva ai due inverni. **Nessuno però aveva previsto in questa forma** il salto improvviso nei cinque anni tra il 2017 e il 2022 (con *transformer*, *LLM*, ChatGPT ecc.).

A incidere profondamente sulle previsioni è stata anche la sovrastima dell'hardware e la sottostima dei dati. Per decenni si è creduto che il limite fosse la potenza di calcolo: in realtà il fattore decisivo è stato l'accesso a dati massivi e puliti (che in realtà puliti non sono), insieme a nuove architetture di rete neurale. Chi faceva previsioni vent'anni fa non poteva immaginare il web come base d'addestramento globale. Oggi però il web è sempre più **inquinato da dati sporchi**: pagine obsolete, siti abbandonati, contenuti duplicati o generati automaticamente, che persistono come **fossili digitali.** Questa massa di informazioni deteriorate riduce la qualità delle fonti e rende più difficile per i sistemi di intelligenza artificiale distinguere ciò che è affidabile da ciò che non lo è. I dati puliti o di qualità – accurati, completi, coerenti, aggiornati e privi di duplicazioni o valori errati – sono fondamentali per ottenere modelli di intelligenza artificiale affidabili.

Dati sporchi o disomogenei portano invece a modelli distorti, poco generalizzabili e inclini a riprodurre bias o errori. La qualità dei dati è quindi più importante della loro quantità: un dataset vastissimo ma incoerente produce risultati peggiori di uno più piccolo ma ben curato. Per questo gran parte del lavoro degli esperti di machine learning consiste nella pulizia, validazione e normalizzazione dei dati, un processo che garantisce che il modello impari regole reali invece di rumore casuale. In medicina, per esempio, la qualità dei dati è cruciale per l'affidabilità delle IA che analizzano immagini radiologiche o istologiche: se le immagini sono sfocate, mal etichettate o provenienti da fonti eterogenee, il modello può diagnosticare in modo errato. Al contrario, dataset accuratamente selezionati e standardizzati permettono di addestrare algoritmi capaci di riconoscere precocemente tumori, lesioni o alterazioni cellulari con precisione paragonabile — e talvolta superiore — a quella umana. In questi casi, la "pulizia" del dato è letteralmente una questione di vita o di errore diagnostico.

Ci sono infine almeno tre fattori che hanno contribuito all'indeterminatezza delle previsioni sull'intelligenza artificiale: **la confusione tra IA "forte" e IA "debole", il bias psicologico e culturale e infine l'effetto "wow" e il conseguente oblio**. Molti esperti e, soprattutto, i media hanno infatti sovrapposto due concetti profondamente diversi. L'IA "forte" indica un'intelligenza paragonabile o superiore a quella umana, autonoma e cosciente, mentre l'IA "debole" si riferisce a sistemi specializzati e statistici, utili ma privi di reale consapevolezza. Oggi i progressi più significativi riguardano quasi esclusivamente quest'ultima — dal machine learning ai modelli linguistici fino alla visione artificiale — ma il linguaggio mediatico ha spesso alimentato aspettative di tipo fantascientifico, generando inevitabili delusioni o allarmismi.

A questo si aggiunge un bias psicologico e culturale: ogni generazione tende a proiettare sulle nuove tecnologie le proprie paure e speranze, come già accaduto con i farmaci o le terapie rivoluzionarie del passato. Negli anni Cinquanta prevaleva la fiducia nel poter creare macchine "razionali", capaci di pensare come l'uomo; negli anni Ottanta si diffuse invece il timore di un imminente dominio delle macchine; oggi le preoccupazioni si concentrano su temi come la disinformazione, l'etica e la disoccupazione cognitiva. Queste oscillazioni emotive influenzano anche gli esperti, che non sono immuni dai bias cognitivi o dalle pressioni mediatiche. Ogni volta che una tecnologia IA raggiunge una capacità nuova (nel gioco degli scacchi o nel go, riconoscimento vocale, traduzione, scrittura), cessa di essere percepita come "vera IA". **È il cosiddetto *AI effect*: una volta risolto un problema, lo si declassa a "banale automazione"**. Così, le previsioni sembrano sbagliate perché spostiamo continuamente l'asticella di ciò che consideriamo "intelligenza". È detto appunto oblio rapido dopo effetto "wow".

Gli errori ricorrenti, da Simon a Kurzweil, mostrano comunque una costante: **l’uomo tende a proiettare sull’IA le proprie paure e speranze, oscillando tra utopia e distopia**. La storia delle previsioni sull’intelligenza artificiale non è dunque una sequenza di fallimenti, ma un percorso di apprendimento collettivo, in cui ogni illusione infranta ha aiutato a comprendere un po’ meglio la natura complessa, ibrida e profondamente umana di questa tecnologia — che parla di noi almeno quanto parla del futuro che immaginiamo.