

World model

RIVOLUZIONE ONLINE – 10 LUGLIO 2026

Parole parole parole: la nuova scommessa dell'intelligenza artificiale punta a superare il limite del linguaggio, per insegnare alle macchine la fisica della realtà.

World model

Tutti ne parlano, nessuno lo fa

‘World model’ è diventata, secondo Yann LeCun (scienziato premio Turing e pioniere delle reti neurali), una buzzword che assomiglia al sesso tra adolescenti: tutti ne parlano, **tutti sostengono di praticarlo, e quasi nessuno sa davvero cosa sia o lo fa sul serio**. È una battuta che ripete nelle sue [interviste](#), e funziona come una bussola: ogni volta che un’azienda dice di avere “un World model”, la prima domanda dovrebbe essere se lo intende sul serio, o se sta solo riciclando un’etichetta di moda.

Ma cos’è un World model?

È un sistema capace di rispondere a una domanda sola: **dato lo stato del mondo adesso, e data un’azione che sto per compiere, cosa succederà un istante dopo?** Serve avere, da qualche parte dentro il sistema, **un’idea di come funziona il mondo**, abbastanza **buona da prevedere le conseguenze di un gesto prima di farlo**, più che generare immagini o testo che suonino plausibili.

Anche un gatto, dice LeCun, ha qualcosa del genere: gli basta per calcolare un salto senza

risolvere equazioni.

La differenza con i Large Language Model

È qui che si misura la distanza dai **LLM**, i modelli linguistici su cui si basano ChatGPT, Claude o Gemini. Per compiti che si risolvono con simboli (matematica, codice) funzionano benissimo, spesso meglio di noi. Il problema, secondo LeCun, riguarda il modo in cui elaborano un pensiero passo per passo: l'unica cosa che si trasmette da un momento di calcolo al successivo è **una parola alla volta, un token. È un canale strettissimo, attraverso cui deve passare tutto il ragionamento. Quando un essere umano pensa, non lo fa mettendo in fila parole una dopo l'altra: si forma un'immagine, un'idea, prima ancora di dirla.** Il vero ragionamento, sostiene LeCun, assomiglia più a questo che a un testo che si scrive da solo.

Nonostante questi limiti strutturali, gli investimenti continuano a premiare i modelli linguistici. L'illusione di umanità è soprattutto una precisa leva commerciale: **una macchina che parla come una persona è più facile da vendere**, da integrare nei flussi di lavoro e da giustificare nei piani di efficienza aziendale. Questa spinta economica **alimenta lo sviluppo di modelli sempre più "umanizzati"**, anche se questo non cambia il limite di fondo descritto sopra: restano modelli fatti per il linguaggio, non per prevedere il mondo fisico.

JEPA

La proposta concreta di LeCun per costruire un World model di questo tipo si chiama [JEPA](#), pubblicata nel 2022, e parte da un'idea controintuitiva: **prevedere ogni singolo pixel di un video, come fanno i modelli generativi, è la strada sbagliata.**

Da capo scienziato di Meta, LeCun aveva provato fino al 2021 esattamente questo, senza risultati soddisfacenti. Il problema non era prevedere cosa sarebbe successo in generale, ma prevederlo fin nei minimi dettagli: l'esatta texture di un muro dietro l'angolo, il preciso disegno

delle foglie di un albero mosse dal vento, sono cose che nessun modello può indovinare pixel per pixel, semplicemente perché quel livello di dettaglio non è deducibile da nulla che si conosca in anticipo. Un sistema addestrato a provarci comunque finisce per inventarseli, non per capire cosa conta davvero in quella scena. La via d'uscita di JEPA, allora, è smettere di provarci. **JEPA impara invece a trattenere solo l'informazione prevedibile, scartando il resto come rumore, e a restituire così la struttura del mondo più che i suoi dettagli minimi**

LeCun lo spiega con un esempio che viene da un suo hobby, la vela: **chi regola una vela si affida a un'idea approssimativa e intuitiva di cosa succederà, più che a un calcolo esplicito delle equazioni di fluidodinamica.** È lo stesso principio dietro l'ambizione di applicare i World model alla robotica e alla guida autonoma.

Costruire un sistema così pone però un problema tecnico chiamato 'collasso': se il sistema può azzerare il proprio errore ignorando i dati e ripetendo sempre la stessa risposta, tende a farlo, perché è la scorciatoia più comoda. Evitarlo ha richiesto anni di lavoro, fino all'uscita di V-JEPA 2, pubblicato dal suo laboratorio in Meta nel giugno 2025, pochi mesi prima delle sue dimissioni: addestrato su oltre un milione di ore di video presi da internet, e poi affinato con appena 62 ore di filmati di un braccio robotico, **è riuscito a far manipolare a un robot oggetti mai visti prima, in ambienti mai visti prima,** con un tasso di successo tra il 65 e l'80%. È la prima prova concreta che la strada di JEPA, applicata alla fisica del mondo reale, funziona davvero, non più solo un'ambizione dichiarata.

A marzo 2026, già fuori da Meta, LeCun ha presentato "[LeWM](#)", un World model con appena 15 milioni di parametri, addestrabile su una singola GPU, che pianifica 48 volte più rapidamente del modello di riferimento della categoria, lavorando su uno spazio latente molto più compatto. Con 15 milioni di parametri si intende un modello minuscolo per gli standard attuali: gli LLM più diffusi ne hanno centinaia di miliardi.

Per LeCun, capire bene la struttura di un problema può contare più che aumentare la

potenza di calcolo dedicata a risolverlo.

Resta un limite più sottile ancora aperto: prevedere cosa succede un istante dopo è una cosa, capire a quale livello vada osservato il problema, i dettagli di un singolo oggetto o il comportamento di un intero sistema, è un'altra. Su questo secondo punto, i modelli che conosciamo non risultano ancora capaci di cavarsela.

J-space

Qualche giorno fa, però, Anthropic ha pubblicato uno studio ([A global workspace in language models](#)) secondo cui **Claude terrebbe a mente alcuni concetti senza mai scriverli**. Alla domanda “quante zampe ha l'animale che tesse le ragnatele”, il modello risponde direttamente “8” ma, a metà elaborazione, al suo interno, si accende il concetto di “ragno”.

Quando i [ricercatori](#) hanno provato a sostituire artificialmente quel concetto con quello di “formica”, la risposta cambiava in “6”.

Si tratterebbe di **un pensiero silenzioso, non ridotto a una parola scritta volta per volta**: proprio il genere di ragionamento che LeCun riserva alla mente umana e nega agli LLM.

Resta un dubbio: quando la tecnica usata da Anthropic accende ‘ragno’, lo fa perché nella rete esiste qualcosa che somiglia a un **concetto autonomo di ragno, o semplicemente perché quel nodo ha la probabilità più alta di comparire in quel contesto?**

Qualunque sia la risposta a quella domanda, il modello non cambia mentre risponde: usa sempre la stessa mappa, costruita una volta per tutte durante l'addestramento, e si limita a percorrerla. Un sistema che imparasse davvero sul momento, aggiornando quella mappa mentre agisce (in termine tecnico, aggiustando i propri pesi in tempo reale), sarebbe tutta un'altra cosa: più vicino a un cervello che si modifica di continuo che a una mappa, per quanto dettagliata, che resta sempre la stessa. Per ora né gli LLM né i primi World model sanno farlo.

Due domande diverse

JEPA e J-space, in fondo, rispondono a domande diverse.

JEPA si chiede se quello che il sistema tiene in mente descriva bene il mondo reale, abbastanza da **prevedere cosa succederà dopo**.

J-space si chiede invece se quel **contenuto**, qualunque cosa sia, **resti a disposizione del resto del sistema, richiamabile e utilizzabile altrove**, invece di restare chiuso in un angolo, indipendentemente da quanto sia accurato rispetto al mondo.

Una domanda riguarda quanto **un'idea** sia **fedele alla realtà**, l'altra quanto sia **a portata di mano**.

Dentro l'argomento di LeCun contro gli LLM convivono due affermazioni diverse: un limite sulla disponibilità (un LLM non può tenere in mente qualcosa senza metterlo subito in parole, per via del canale a un token) e un limite sulla fedeltà al mondo (quel poco che rappresenta resta comunque fatto di parole, non nato da un'esperienza reale). Lo studio di Anthropic tocca solo il primo: mostra un concetto disponibile senza essere scritto parola per parola, ma resta comunque fatto di parole. Per questo la tesi di LeCun ne esce complicata, non smentita: sul terreno che gli interessa di più, prevedere cosa succede in un video o quando un braccio robotico tocca un oggetto, la distanza da JEPA resta la stessa.

Un progetto anche politico

Fin qui la scommessa tecnica. LeCun l'ha però giocata dentro una scommessa più grande, anche politica. Nel 2025 ha lasciato Meta, per fondare [AMI Labs](#), con sede a Parigi. Il lancio ufficiale, il 10 marzo 2026, ha portato con sé un [finanziamento superiore al miliardo di dollari](#), arrivato da investitori di tre continenti diversi. La sede parigina è una scelta di posizionamento: LeCun la rivendica come quella di **un fornitore di tecnologia che non sia né americano né cinese**.

In questa cornice si inserisce [Project Tapestry](#), di cui LeCun è consulente scientifico. Il progetto nasce dentro la AI Alliance, una coalizione che conta oltre duecento organizzazioni; al primo incontro operativo, a Parigi, hanno preso parte una trentina di Istituzioni arrivate da diversi paesi. L'idea è quella di **costruire insieme un modello condiviso tra università, governi e aziende di paesi diversi, senza che nessuno debba condividere i propri dati**. Tra i partecipanti circoleranno solo i risultati dell'addestramento, non i dati stessi: numeri che si combinano via via in un modello comune.

I dati condivisi

Detta così sembra una garanzia solida, ma qui si nascondono due rischi diversi.

Il primo riguarda il momento in cui quei numeri viaggiano da un partecipante all'altro: un aggiornamento condiviso può, in teoria, essere analizzato per risalire a frammenti dei dati che lo hanno prodotto, anche se quei dati non vengono mai spediti direttamente. Project Tapestry affida questo problema a un gruppo di lavoro apposito, con tre strumenti: **cifrare gli scambi** (affinché chi coordina il progetto non possa curiosare nel contributo di un singolo partecipante), aggiungere un po' di **disturbo statistico** ai numeri condivisi (che copre soprattutto i dettagli che compaiono una volta sola), far girare i **calcoli più delicati in ambienti protetti**. Per ora è solo un impegno preso, non una soluzione già pronta: il progetto è ancora ai primi passi.

Il secondo rischio riguarda il modello già finito, non più lo scambio che lo ha costruito. Facciamo un esempio: il fatto che Ciccio tradisca il suo partner resta un segreto, mai scritto da nessuna parte in modo esplicito. Se però alla domanda "secondo te Ciccio lo tradisce?" il sistema risponde "sono sicuro al cento per cento", il segreto è comunque trapelato: non perché qualcuno l'abbia letto, ma perché il modello ha imparato a collegare tra loro segnali che di solito vanno insieme.

Nessuna delle tre misure sopra lo previene: è il prezzo di un modello capace di generalizzare

bene, non un difetto nel modo in cui i pesi vengono scambiati tra i nodi.

Quanto questa architettura riuscirà davvero a proteggere chi vi partecipa? Chi controllerà alla fine il modello condiviso? Ci troveremo davanti a un nuovo Palantir? (vedi l'articolo [Privatizzare il Sovrano](#))

Dal linguaggio al mondo

Gli **LLM** hanno già fatto una **rivoluzione, nel campo del linguaggio**: hanno reso una macchina capace di leggere, scrivere, tradurre, programmare, quasi come farebbe una persona (spesso meglio). È una rivoluzione che riguarda un mondo fatto di **simboli e parole**.

I **World model** promettono un'altra **rivoluzione, nel campo della fisica**: una macchina capace di **anticipare il comportamento del mondo**, più che di descriverlo a parole: come reagirà un oggetto, un corpo, una superficie, prima ancora che accada. In pratica, questo vuol dire prevedere cosa succede quando un braccio robotico solleva un oggetto fragile, quando un'auto frena su una strada bagnata, quando un impianto industriale segnala una vibrazione anomala.

È una scommessa che riguarda **l'agire nel mondo fisico, più che lo scrivere**, e si misurerà nei prossimi anni sul campo, in un laboratorio di robotica o dentro un'auto a guida autonoma, più che nei titoli dei giornali.

Se dovesse funzionare davvero, l'applicazione non si fermerebbe al laboratorio.

Gli **LLM** hanno già spostato la **competizione sul lavoro cognitivo**: chi scrive, traduce, programma, la sente ogni giorno.

Un **World model** maturo sposterebbe la stessa pressione su un **lavoro fisico, non cognitivo**: dalle professioni (magazzinieri, autisti, operai di linea, manutentori) a tanti settori diversi (sanità, difesa, automazione industriale, manifattura, aerospazio, automotive).

È esattamente la soglia che Cai Fang, l'economista cinese di cui abbiamo scritto qui, chiama la **seconda fase dell'automazione**: quella in cui non conta più quanto una persona sappia fare, perché **il lavoro rimasto agli umani si restringe a prescindere dalle competenze di ciascuno.**

Oggi siamo ancora nella prima fase, quella in cui l'AI toglie e crea posti in modo asimmetrico. I World model sono la tecnologia che renderebbe concreta la seconda.

– di Paola Furlan e Andrea Mazzucchi