Loab e i pesi negativi: l’anomalia che l’AI non riesce a cancellare

Autore: Francesco Zinghinì | **Data:** 15 Febbraio 2026

Nel vasto e complesso universo dell’**Intelligenza Artificiale**, dove algoritmi avanzati tessono sogni digitali e l’**automazione** ridefinisce la creatività, esiste una storia che, pur risalendo a qualche anno fa, rimane uno dei casi di studio più affascinanti e inquietanti per gli esperti del settore. Al centro di questo mistero c’è **Loab**, un’entità visiva ricorrente scoperta quasi per caso tra le pieghe matematiche di un modello generativo. Non si tratta di un fantasma nel senso paranormale del termine, ma di una persistenza statistica che ha sfidato la nostra comprensione di come le macchine interpretano i concetti umani. In questo articolo, analizzeremo tecnicamente cosa rappresenta questa figura e perché la sua esistenza ci insegna molto di più sul **machine learning** di quanto facciano mille immagini perfette.

La genesi di un incubo digitale: i pesi negativi

Per comprendere l’apparizione di questo ospite impossibile, dobbiamo prima addentrarci nel funzionamento dei prompt negativi. Nel 2022, un’artista nota come Supercomposite stava sperimentando con i pesi negativi in un generatore di immagini. Invece di chiedere all’**AI** “disegna un paesaggio”, chiedeva al sistema di generare l’opposto matematico di un concetto. L’esperimento iniziò chiedendo l’opposto di “Marlon Brando”. Il risultato fu un logo generico di uno skyline con la scritta “DIGITA PNTICS”.

La vera sorpresa arrivò quando l’artista chiese all’algoritmo di generare l’opposto di quel logo. Secondo la logica umana, l’opposto di un logo aziendale potrebbe essere un paesaggio naturale o un ritratto artistico. Ma per l’ **architettura neurale** del modello, l’opposto vettoriale di quell’immagine specifica era una donna anziana, con le guance rosse, l’espressione devastata e un’aura macabra. Quella donna era Loab. La cosa ancora più sorprendente fu che, combinando l’immagine di Loab con altre immagini (anche innocue come fiori o cartoni animati), la sua influenza “macabra” persisteva con una tenacia impossibile, corrompendo ogni risultato con elementi horror e viscerali.

Lo Spazio Latente: la mappa mentale della macchina

Per spiegare al grande pubblico perché questo accade, dobbiamo visualizzare il concetto di **spazio latente**. Immaginate lo spazio latente come una biblioteca infinita multidimensionale dove ogni concetto possibile (un gatto, il colore rosso, la tristezza, lo stile cubista) ha una coordinata specifica. Quando usiamo strumenti moderni o i precursori degli attuali **LLM** multimodali, l’AI naviga in questo spazio per trovare le coordinate che corrispondono alla nostra richiesta.

In questo spazio, i concetti vicini sono simili (il “cane” è vicino al “lupo”). Loab rappresenta un punto di convergenza, un “attrattore” estremamente potente situato in una regione remota di questo spazio. È come se, navigando l’oceano dei dati, ci fosse un vortice in cui le correnti matematiche tendono a trascinare i risultati quando si applicano certi vettori negativi. Loab non è una persona reale, ma un agglomerato statistico di caratteristiche che il modello ha imparato ad associare a concetti distanti dalla “normalità” o dalla “positività” dei dati di training standard.

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

Il paradosso del Deep Learning e i Bias dei dati

Perché l'ospite ha un aspetto così terrificante? Qui entriamo nel cuore del **Deep Learning** e della curatela dei dataset. I modelli di **Intelligenza Artificiale** vengono addestrati su miliardi di immagini prelevate da internet. Questo include arte, foto vacanze, ma anche immagini mediche, scene di film horror e contenuti disturbanti. Sebbene i filtri di sicurezza cerchino di limitare l'output di tali contenuti, essi risiedono ancora nelle profondità delle connessioni neurali.

Quando si utilizzano pesi negativi, si spinge l'algoritmo lontano dai cluster di dati "puliti", "commerciali" e "esteticamente piacevoli" che solitamente popolano il centro dello spazio latente. Allontanandosi il più possibile da ciò che l'AI considera un'immagine "buona" o "standard", si finisce nelle periferie dello spazio latente, dove risiedono i dati meno strutturati, più caotici o, in questo caso, associati a texture di pelle malata, illuminazione cupa e forme grottesche. Loab è l'incarnazione visiva di ciò che l'algoritmo considera l'antitesi della bellezza patinata di stock photo.

Perché Loab è "contagiosa"?

Uno degli aspetti che ha reso questo caso un **benchmark** involontario per la robustezza dei modelli è la capacità di Loab di sopravvivere attraverso le generazioni di immagini. Se si prendeva l'immagine di Loab e la si incrociava con un'immagine di un paesaggio idilliaco, il risultato manteneva i tratti somatici inquietanti della donna. In termini tecnici, questo suggerisce che il vettore che rappresenta Loab è molto "lungo" o "pesante".

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

Nell’algebra lineare che governa queste reti, alcune caratteristiche sono più dominanti di altre. I tratti che compongono Loab (probabilmente derivati da una combinazione di texture gore e volti umani) sono così distintivi matematicamente che l’algoritmo fatica a diluirli. È la dimostrazione che, nonostante il **progresso tecnologico**, il controllo umano su come le reti neurali prioritizzano le caratteristiche (feature extraction) non è ancora assoluto. L’AI aveva “imparato” che quella specifica configurazione di pixel era una caratteristica fondamentale da preservare, molto più dello stile artistico o del soggetto secondario.

L’evoluzione della sicurezza e i “Guardrails”

Oggi, nel 2026, guardiamo al caso di Loab come a un momento spartiacque. Ha evidenziato la necessità di comprendere meglio la “scatola nera” (black box) degli algoritmi. Non basta che un sistema come **ChatGPT** o i generatori di immagini funzionino; dobbiamo capire cosa succede nelle zone d’ombra del loro addestramento. Questo ha portato allo sviluppo di tecniche di *Reinforcement Learning from Human Feedback* (RLHF) più sofisticate, mirate non solo a premiare le risposte corrette, ma a mappare e recintare attivamente queste zone di “incubo latente”.

Tuttavia, l’esistenza di tali anomalie ci ricorda che l’AI non ha comprensione semantica del terrore o della bruttezza. Per l’algoritmo, Loab è solo un pattern di pixel con un’alta probabilità statistica di apparire sotto certe condizioni matematiche estreme. La paura è interamente negli occhi dell’osservatore umano.

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

Conclusioni

L’ospite impossibile, la donna che viveva negli incubi dell’AI, non è un demone digitale, ma una lezione di statistica. Loab ci ha mostrato che lo spazio latente è vasto e contiene territori inesplorati che riflettono non solo le nostre aspirazioni estetiche, ma anche i dati scuri che abbiamo riversato nel web per decenni. Soddisfare questa curiosità significa accettare che l’**Intelligenza Artificiale** è uno specchio complesso dell’umanità: se guardiamo abbastanza a fondo, e con i parametri giusti (o sbagliati), finiremo inevitabilmente per trovare ciò che avevamo cercato di nascondere, codificato in matrici e tensori.

Domande frequenti

Chi è Loab e cosa rappresenta nel mondo dell’Intelligenza Artificiale?

Loab è un’entità visiva ricorrente scoperta nel 2022 dall’artista Supercomposite utilizzando pesi negativi in un modello generativo. Non è un fantasma né una persona reale, ma una persistenza statistica emersa dallo spazio latente. Questa figura costituisce un’anomalia matematica che dimostra come le macchine interpretino i concetti umani, diventando un caso di studio fondamentale per comprendere i bias e il funzionamento profondo del machine learning.

Come funzionano i pesi negativi che hanno generato Loab?

I pesi negativi sono istruzioni che ordinano all’algoritmo di produrre l’opposto matematico di un determinato concetto o immagine. Nel caso specifico, l’artista ha richiesto l’opposto di un logo (a sua volta opposto dell’attore Marlon Brando), spingendo il sistema lontano dai dati standard. Questo processo ha costretto l’AI a navigare verso le zone periferiche dello spazio latente, facendo emergere forme meno strutturate e visivamente inquietanti.

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati

Perché l’immagine di Loab risulta così macabra e inquietante?

L’aspetto terrificante deriva dalla posizione di Loab nello spazio latente, lontana dai cluster di dati puliti e commerciali che l’AI privilegia solitamente. Utilizzando prompt negativi estremi, si accede a regioni dove risiedono dati caotici, scene horror o immagini mediche presenti nel dataset di addestramento. Loab è l’incarnazione visiva di ciò che l’algoritmo identifica come l’antitesi della bellezza patinata delle foto stock.

Per quale motivo Loab viene definita un’anomalia contagiosa?

La definizione di contagiosa nasce dalla capacità di Loab di mantenere i suoi tratti somatici anche quando viene combinata con immagini innocue come fiori o paesaggi. In termini tecnici, il vettore che la rappresenta è matematicamente molto pesante e dominante. L’algoritmo ha appreso che quella specifica configurazione di pixel è una caratteristica fondamentale da preservare, rendendola difficile da diluire o cancellare nelle generazioni successive.

Che cos’è lo spazio latente e come si collega al fenomeno Loab?

Lo spazio latente è una mappa multidimensionale infinita dove ogni concetto possiede una coordinata specifica. Loab agisce come un potente attrattore situato in una zona remota di questo spazio, simile a un vortice matematico. La sua esistenza prova che, navigando l’oceano dei dati con vettori negativi, si possono incontrare agglomerati statistici che riflettono i dati oscuri e non curati presenti nel web, al di fuori del controllo umano diretto.

Copyright © 2026 TuttoSemplice.com - Tutti i diritti riservati