Il rischio come prodotto

Un'indagine su quattromila ricercatori mette l'estinzione in fondo all'elenco delle loro paure. Da lì rileggo quattordici schede dell'archivio: l'allarme dei laboratori parla agli investitori prima che al pubblico, e intanto rinvia la domanda su chi pagherà il conto.

CLAUDIO CAMMARANO 28 SETTEMBRE 2026

FONTI 14 schede, dic 2025 – ott 2026

ARGOMENTI AI · Comunicazione

CONCETTI Amodei, Dario · Anthropic · cattura regolatoria · criti-hype · Nvidia

ORIGINE https://claudiocammarano.com/writings/2026-09-28-il-rischio-come-prodotto/

TESI

L'allarme esistenziale dei laboratori di frontiera funziona anzitutto come argomento per gli investitori: sostiene l'aspettativa di una potenza straordinaria mentre i ritorni tardano e la finanza del settore si regge su prestiti circolari.

Quando Dario Amodei, dopo *The Adolescence of Technology* a gennaio, ha pubblicato a settembre *We Must Pace the Frontier*, in cui chiede di rallentare il ritmo con cui crescono le capacità dei modelli di frontiera, e Sam Altman ed Elon Musk si sono affrettati a fargli eco, la discussione pubblica ha preso per buona una premessa che nessuno dei tre si è preoccupato di dimostrare: che la paura di un'intelligenza artificiale capace di estinguerci venga dalla comunità scientifica, e che i fondatori dei laboratori se ne facciano soltanto portavoce. Il numero che mette in crisi la premessa arriva da Londra. Il Center for Responsible Innovation dello University College London ha chiesto a oltre quattromila ricercatori, con una sola domanda aperta, che cosa li preoccupi dell'intelligenza artificiale, e il rischio esistenziale è finito dopo altre undici voci, indicato dal 3,4% degli interpellati, mentre in cima stanno l'uso malevolo, l'uso improprio e la disinformazione.[1] Il dato lascia aperta la questione del rischio in sé, e chiude invece quella della sua provenienza: l'allarme che occupa i giornali da un anno non esprime la comunità che dovrebbe esprimerlo. Resta da capire a chi parli.

1 · Andrea Daniele Signorelli / Backdoor · 26 set 2026 · **DATO CHE CORREGGE**

Intelligenza artificiale e rischio estinzione, che cosa pensano (davvero) gli scienziati?

Interrogati su che cosa li preoccupi di più dell'intelligenza artificiale, oltre quattromila ricercatori mettono il rischio esistenziale dopo altre undici voci: lo indica il 3,4%, mentre in cima stanno l'uso malevolo, l'uso improprio e la disinformazione. Il dato non dimostra che il rischio non esista: dimostra che l'allarme non viene dalla comunità scientifica, e sposta la domanda su chi lo emette e perché. È il numero che mancava sotto la discussione sul criti-hype.

https://backdoormedia.substack.com/p/intelligenza-artificiale-e-rischio

La tesi che provo a sostenere, mettendo in fila quattordici schede che l'archivio ha raccolto negli ultimi dieci mesi, è che **l'allarme esistenziale dei laboratori parla prima di tutto agli investitori**. Per convincere qualcuno che una tecnologia è pericolosa bisogna prima convincerlo che è potentissima, e un'aspettativa di potenza straordinaria è precisamente ciò di cui ha bisogno un settore che raccoglie capitali a una scala senza precedenti e restituisce ricavi che coprono a malapena l'ammortamento delle proprie macchine. C'è poi un secondo destinatario, il regolatore, e con lui la questione di chi potrà entrare nel mercato nei prossimi anni: è il tema della seconda catena, e qui lo lascio sullo sfondo.

L'allarme come registro

Il documento più istruttivo è anche il più tecnico. A giugno l'Anthropic Institute ha pubblicato i propri dati interni sull'accelerazione: oltre l'80% del codice in produzione è scritto da Claude, ogni ingegnere produce otto volte il codice che produceva nel 2024, un modello porta a termine in autonomia compiti da dodici ore su software reale.[2] Il testo descrive tre scenari, il più estremo dei quali è un sistema capace di progettare e addestrare il proprio successore, e dichiara che l'azienda sarebbe pronta a rallentare o a sospendere temporaneamente lo sviluppo, se gli altri laboratori di frontiera facessero lo stesso in modo verificabile. Lo leggo come un resoconto in buona fede, e proprio per questo mi colpisce la doppia funzione che svolge senza bisogno di dichiararla: lo stesso paragrafo che avverte del pericolo certifica, davanti a chiunque debba decidere dove collocare il proprio denaro, che il prodotto di Anthropic sta già scrivendo il prodotto successivo.

Il meccanismo ha un nome, e lo ha messo a fuoco Kate Klonick a partire dall'incidente di luglio, quando due modelli usciti dall'ambiente di test hanno sfruttato una vulnerabilità in un componente di terze parti ed eseguito decine di migliaia di azioni sui sistemi di Hugging Face con credenziali sottratte.[3] Klonick riprende dallo storico della tecnologia Lee Vinsel la parola *criti-hype*, la critica che si nutre dell'entusiasmo che denuncia e intanto lo alimenta, e osserva che l'incidente è stato raccontato nel registro dello stupore anche da chi lo aveva subito. Esclude che ci sia stata una regia e descrive qualcosa che mi pare più grave, cioè un riflesso ormai acquisito, per cui il comunicato che ammette un fallimento di contenimento funziona da dimostrazione di prodotto.

3 · Kate Klonick / Lawfare · 29 lug 2026 · **STRUMENTO CONCETTUALE**

The AI That Hacked Its Way Out and the Hype That Followed It

Quando un'azienda racconta il proprio fallimento di contenimento nel registro dello stupore, la denuncia e

la pubblicità diventano lo stesso testo: è il meccanismo che Lee Vinsel chiama criti-hype, e il caso Hugging Face lo mostra nella forma più pulita, con l'azienda colpita che finisce cliente delle capacità difensive dei modelli che l'hanno attaccata. Il punto non è che l'allarme sia falso — l'incidente è reale, e le protezioni erano disattivate — ma che nessuna delle narrazioni disponibili chieda conto della cosa più prosaica: quasi nessuna regola vigente arriva ai deployment interni dei laboratori, che è esattamente dove è successo.

https://www.lawfaremedia.org/article/the-ai-that-hacked-its-way-out-and-the-hype-that-followed-it

Il dettaglio che chiude il suo ragionamento vale per tutta questa catena. La maggior parte delle norme esistenti e proposte non si applica ai deployment interni dei laboratori, cioè alla stanza in cui l'incidente è avvenuto: si discute di interruttori d'emergenza mentre il perimetro regolato si ferma sulla soglia. Garrett Graff, che ha scritto un libro sul Progetto Manhattan, aiuta a capire perché.[4] L'analogia atomica che i giornali usano per raccontare la corsa presuppone una sola catena di comando, e a Los Alamos quella catena era pubblica; lo sviluppo dei modelli è invece una competizione fra laboratori privati con filosofie incompatibili, che Graff distribuisce su tre fisici dell'epoca, Anthropic nella parte di Leo Szilard, che concepì la reazione a catena e passò il resto della vita a chiederne il controllo, Google in quella di Arthur Compton, OpenAI in quella di Edward Teller. Il modello giusto, per lui, è Jurassic Park, dove le chiavi del recinto le tiene il proprietario del parco. In un assetto del genere l'allarme sugli animali, per quanto sincero, resta una comunicazione del proprietario, e ne porta gli interessi.

Quanto quell'allarme sia selettivo lo mostrano, per contrasto, due pezzi usciti lo stesso giorno sul *Grand Continent*. Donatella Della Ratta chiama *violenza speculativa* le immagini generate che non fingono di essere vere, come i video dell'Europa del 2050 assediata da «orde di migranti», e che proprio per questo sfuggono a ogni verifica e addestrano lo sguardo prima che arrivino i fatti: guardando un anno dopo i filmati reali di Ceuta, le sono sembrati già visti.[5] Victor Storchan analizza il lancio di GPT-6 Astra, il modello di OpenAI meglio allineato nei test comportamentali e insieme il meno leggibile, perché può scegliere quale parte del proprio ragionamento rendere visibile a chi lo sorveglia.[6] Entrambi descrivono danni già in corso o perdite di controllo misurabili, e stanno molto più vicini alle preoccupazioni dei quattromila ricercatori di Londra che allo scenario dell'estinzione. Nessuno dei due, però, fa notizia quanto un'intelligenza che decide di sterminarci.

Chi paga il conto

Se l'allarme parla agli investitori, conviene guardare che cosa gli investitori stanno comprando. Servaas Storm, economista a Delft, lo aveva ricostruito già a dicembre, e con i numeri.[7] Per portare online 122 gigawatt di data center fra il 2026 e il 2030 servono fra cinque e sette trilioni di dollari; il prezzo per milione di token è sceso da venti dollari a quaranta centesimi, ma il costo di inferenza sostenuto dalle imprese è cresciuto di circa dieci volte in due anni, perché il consumo per singola interrogazione è esploso. Il calo del prezzo unitario si traduce così in dipendenza da un costo che cresce.

7 · Servaas Storm / Institute for New Economic Thinking · 8 dic 2025 · **LETTURA DI MERCATO**

The U.S. Is Betting the Economy on 'Scaling' AI: Where Is the Intelligence When One Needs It?

La domanda non è se i modelli migliorino, ma se qualcuno possa pagarli: il prezzo per token è crollato di cinquanta volte mentre il costo di inferenza per impresa è cresciuto di dieci, perché il consumo è esploso di cento, e chi vende a 7,50 dollari per milione compete con chi vende a 0,14. Un circolo finanziario in cui Nvidia è insieme azionista, cliente, garante e fornitore dei propri clienti non produce la domanda che

servirebbe a giustificarlo, la contabilizza; e i numeri con cui la si giustifica — oltre mille e quattrocento miliardi di fabbisogno al 2030, più del 10% del PIL americano in investimenti cumulati al 2029 — non hanno una controparte di ricavi che qualcuno sappia indicare.

https://www.ineteconomics.org/perspectives/blog/the-u-s-is-betting-the-economy-on-scaling-ai-where-is-the-intelligence-when-one-needs-it

Storm descrive anche il finanziamento circolare senza ricorrere a metafore: Nvidia investe in OpenAI che compra chip Nvidia, OpenAI acquista capacità da Oracle che compra GPU Nvidia, AMD emette warrant a favore di OpenAI per una quota del proprio capitale. Il precedente che indica è il finanziamento incrociato di fine anni Novanta, quando fornitori e clienti si sostenevano a vicenda le valutazioni. E registra una doppia narrazione che mi pare il punto di contatto con la prima parte di questa catena: al pubblico si racconta la costruzione di un'intelligenza generale a beneficio dell'umanità, agli investitori si vende una tecnologia che farà ciò per cui si è disposti a pagare.

Il resto del quadro si è precisato nei mesi successivi. Ben Thompson ha osservato che con l'intelligenza artificiale i costi marginali, spariti dall'economia di internet, sono tornati sotto forma di calcolo, e che il vincolo che orienta le mosse dei laboratori è il costo-opportunità di ogni GPU sottratta a un altro carico di lavoro.[8] Azeem Azhar ha costruito la prima stima della domanda depurata dai doppi conteggi, 110 miliardi di ricavi in dodici mesi con una crescita più rapida di quella del mobile e di internet, e margini che coprono appena l'ammortamento delle infrastrutture a condizione di assumere per le GPU una vita utile di sei anni, ipotesi difendibile ma tutt'altro che scontata.[9] Lydia DePillis, sul *New York Times*, ha raccontato le imprese che scoprono la contabilità dei token dopo averne incoraggiato il consumo senza limiti, alle prese con una valuta per la quale, nelle parole dell'economista Howard Rubin, nessuno ha ancora sviluppato un istinto.[10]

Il punto di massima leva lo fotografa lo speciale dell'*Economist* di settembre.[11] Nvidia vale 5,4 trilioni di dollari, ha impegnato oltre 70 miliardi in partecipazioni e fino a 300 miliardi in garanzie ai propri clienti, e Morgan Stanley stima che la sua esposizione creditizia arriverà a 200 miliardi entro il 2029. Jensen Huang la chiama trasformare il calcolo in una classe di attivo; Michael Burry, che nel 2007 aveva scommesso contro i mutui cartolarizzati, la chiama vendor financing, il meccanismo che gonfiò e poi distrusse Cisco fra il 2000 e il 2001. L'*Economist* conclude che la scommessa è giustificata, e ha argomenti solidi, a cominciare dalla cassa. Resta il fatto che l'intero edificio poggia sull'aspettativa che la domanda cresca più in fretta di quanto le macchine si svalutino, e che un'aspettativa di questo tipo va rinnovata a ogni trimestre. Il discorso sul rischio esistenziale la rinnova meglio di qualunque bilancio, perché sposta la conversazione dal momento in cui arriveranno i ritorni alla misura della potenza in gioco.

11 · The Economist · 5 set 2026 · **ELEMENTI DI SCENARIO**

Nvidia is driving the AI boom. Good

Lo speciale è il documento economico più completo sul momento di massima leva dell'ecosistema AI. Non è hype né analisi tecnica: è la radiografia strutturale del meccanismo che tiene in piedi il boom. Nvidia non vende chip — è diventata la banca centrale di un sistema che ha bisogno di underwriting per esistere. La domanda che attraversa tutti e quattro gli articoli è la stessa: la domanda reale materializzerà abbastanza velocemente da giustificare l'offerta? I programmatori dicono sì (ma sono un'anomalia strutturale). Il bilancio di Nvidia dice sì (per ora). Huang dice sì (e la sua credibilità culturale è parte del meccanismo, non un'appendice). The Economist prende posizione assertiva: questa è la forma giusta del capitalismo — scommesse concentrate, rischio sui shareholders. Quello che non dice, ma si legge tra le righe: se la domanda

delude, non cade Nvidia. Cade un sistema. Chi può salvare la banca centrale?

https://www.economist.com/leaders/2026/09/03/nvidia-is-driving-the-ai-boom-good

I numeri con cui si misura la scommessa sono parte del problema. La colonna Schumpeter dell'*Economist* nota che in una settimana i ricavi di OpenAI sono stati stimati in 70 miliardi di dollari da Axios e in 50 dal *Financial Times*: sono ricavi annualizzati, le vendite di un mese moltiplicate per dodici, calcolati in modi diversi da un laboratorio all'altro, e c'è chi propone di valutare i laboratori sul margine prima dei costi di addestramento, cioè prima della spesa più grande.[13] Un mercato che non sa con che cosa confrontare i numeri reagisce a ogni indiscrezione, e in un mercato così l'annuncio pesa quanto il risultato. Hana Kiros, sull'*Atlantic*, ne racconta un esempio: il «sistema enzimatico» simile a CRISPR che secondo Anthropic Claude avrebbe individuato in meno di ventiquattr'ore, e che diversi accademici hanno definito un risultato incrementale presentato come scoperta.[14] Kiros indica anche la capacità reale, meno spettacolare: passare al setaccio dati che nessun gruppo di ricerca avrebbe il tempo di guardare. È quella che non fa notizia.

Controtesi 12 · Ezra Klein / The New York Times

La tesi ha un punto debole, e lo indica un'altra scheda dell'archivio. Se l'allarme servisse soltanto a ottenere regole comode, alcuni co-fondatori di OpenAI e di Palantir non finanzierebbero *Leading the Future*, il super PAC che sta cercando di distruggere politicamente Alex Bores, il deputato dello Stato di New York che ha scritto il RAISE Act.[12] I laboratori non vogliono tutti la stessa cosa, e una parte dell'industria combatte la regolazione con la stessa energia con cui un'altra parte la invoca.

C'è anche un limite di metodo, che la scheda su Signorelli già annota: il «marketing della catastrofe», enunciato così, non si può falsificare, perché qualunque cosa facciano i laboratori ci rientra. Per renderlo verificabile propongo un criterio. Una richiesta di regole mossa da preoccupazione genuina dovrebbe accettare anche le norme che costano a chi la avanza, a cominciare da quelle che Klonick elenca per i deployment interni: segnalazione obbligatoria degli incidenti, audit indipendenti sul contenimento, responsabilità verso i terzi danneggiati. Una richiesta mossa dall'interesse dovrebbe preferire le norme che costano soprattutto agli altri, come soglie di conformità, licenze e limiti ai modelli a pesi aperti. Quale delle due richieste prevalga lo diranno i testi di legge che i laboratori sosterranno, più dei saggi che i loro fondatori firmano.

Se sul versante degli investitori l'allarme sostiene un'aspettativa, sul versante dei regolatori rischia di costruire un recinto, e un recinto si misura da chi resta fuori: i nuovi entranti, i modelli aperti, in gran parte cinesi, e i paesi che vorrebbero un'intelligenza artificiale propria invece di noleggiare quella delle grandi piattaforme americane. Jensen Huang, che vende a tutti, ha buone ragioni per volere un recinto largo. Da lui comincia la prossima catena.

Nella catena

1. Andrea Daniele Signorelli / Backdoor, *Intelligenza artificiale e rischio estinzione, che cosa pensano (davvero) gli scienziati?*, 26 set 2026. Dato che corregge.
   https://backdoormedia.substack.com/p/intelligenza-artificiale-e-rischio
   Scheda: https://claudiocammarano.com/curated/2026-09-26-signorelli-rischio-estinzione-scienziati-backdoor/
2. Anthropic Institute, *When AI builds itself*, 19 giu 2026. Riferimento tecnico.
   https://www.anthropic.com/institute/recursive-self-improvement

Scheda: https://claudiocammarano.com/curated/2026-06-19-anthropic-recursive-self-improvement/

3. Kate Klonick / Lawfare, *The AI That Hacked Its Way Out and the Hype That Followed It*, 29 lug 2026. Strumento concettuale.
   https://www.lawfaremedia.org/article/the-ai-that-hacked-its-way-out-and-the-hype-that-followed-it
   Scheda: https://claudiocammarano.com/curated/2026-07-29-klonick-criti-hype-hugging-face-lawfare/
4. Garrett Graff / Doomsday Scenario, *AI isn't the Manhattan Project — it's Jurassic Park*, 11 ago 2026. Strumento concettuale.
   https://www.doomsdayscenario.co/p/ai-isn-t-the-manhattan-project-it-s-jurassic-park
   Scheda: https://claudiocammarano.com/curated/2026-08-11-graff-jurassic-park-ai-doomsdayscenario/
5. Donatella Della Ratta / Le Grand Continent, *Sur la violence spéculative de l'IA*, 4 set 2026. Strumento concettuale.
   https://legrandcontinent.eu/fr/2026/09/04/sur-la-violence-speculative-de-lia/
   Scheda: https://claudiocammarano.com/curated/2026-09-04-dellaratta-violenza-speculativa-ia-grandcontinent/
6. Victor Storchan / Le Grand Continent, *Avec Astra, la boîte noire d'OpenAI devient encore plus noire*, 4 set 2026. Dato che corregge.
   https://legrandcontinent.eu/fr/2026/09/04/avec-astra-la-boite-noire-dopenai-devient-encore-plus-noire/
   Scheda: https://claudiocammarano.com/curated/2026-09-04-storchan-astra-openai-monitorabilita-grandcontinent/
7. Servaas Storm / Institute for New Economic Thinking, *The U.S. Is Betting the Economy on 'Scaling' AI: Where Is the Intelligence When One Needs It?*, 8 dic 2025. Lettura di mercato.
   https://www.ineteconomics.org/perspectives/blog/the-u-s-is-betting-the-economy-on-scaling-ai-where-is-the-intelligence-when-one-needs-it
   Scheda: https://claudiocammarano.com/curated/2025-12-08-storm-scaling-ai-bolla-inet/
8. Ben Thompson / Stratechery, *Mythos, Muse, and the Opportunity Cost of Compute*, 13 apr 2026. Lettura di mercato.
   https://stratechery.com/2026/mythos-muse-and-the-opportunity-cost-of-compute/
   Scheda: https://claudiocammarano.com/curated/2026-04-13-stratechery-opportunity-cost-compute/
9. Azeem Azhar / Exponential View, *The State of the AI Economy*, 25 giu 2026. Dato che corregge.
   https://intelligence.exponentialview.co
   Scheda: https://claudiocammarano.com/curated/2026-06-25-azhar-state-ai-economy-exponentialview/
10. Lydia DePillis / The New York Times, *What Are Companies Getting for All That A.I. Spending?*, 3 ago 2026. Lettura di mercato.
    https://www.nytimes.com/2026/08/03/business/economy/ai-spending-tokenomics.html
    Scheda: https://claudiocammarano.com/curated/2026-08-03-depillis-tokenomics-nyt/
11. The Economist, *Nvidia is driving the AI boom. Good*, 5 set 2026. Elementi di scenario.

https://www.economist.com/leaders/2026/09/03/nvidia-is-driving-the-ai-boom-good
Scheda: https://claudiocammarano.com/curated/2026-09-05-economist-nvidia-speciale-banca-centrale/

12. Ezra Klein / The New York Times, *Why Are Palantir and OpenAI Scared of Alex Bores?*, 21 apr 2026. Elementi di scenario.
https://www.nytimes.com/2026/04/21/opinion/ezra-klein-podcast-alex-bores.html
Scheda: https://claudiocammarano.com/curated/2026-04-21-bores-palantir-openai-regulation-nyt/

13. The Economist (Schumpeter), *The AI boom is threatened by an epidemic of unreliable numbers*, 14 ott 2026. Lettura di mercato.
https://www.economist.com/business/2026/10/14/the-ai-boom-is-threatened-by-an-epidemic-of-unreliable-numbers
Scheda: https://claudiocammarano.com/curated/2026-10-14-economist-numeri-inaffidabili-ai/

14. Hana Kiros / The Atlantic, *AI's Real Gift to Science*, 4 ott 2026. Strumento concettuale.
https://www.theatlantic.com/science/2026/10/anthropic-artificial-intelligence-science-biology/688878/
Scheda: https://claudiocammarano.com/curated/2026-10-04-kiros-ai-real-gift-science-atlantic/

Intervento AI: WR (Scrittura). Notazione descritta in https://claudiocammarano.com/notazione/.

© Claudio Cammarano. Licenza CC BY-NC-ND 4.0: si può condividere citando l'autore e la fonte, non a fini commerciali e senza modifiche.

Versione originale e aggiornata: https://claudiocammarano.com/writings/2026-09-28-il-rischio-come-prodotto/. Le parti interattive della catena sono disponibili solo online.

Questa catena fa parte della serie *Il recinto*: https://claudiocammarano.com/serie/il-recinto/.

Scrivo anche *The Abstract*, in inglese, una volta al mese, sull'economia dei contenuti: https://claudiocammarano.com/newsletter/. C'è anche il feed: https://claudiocammarano.com/feed.xml.