

42 Policy Brief n. 01

L'AI Act ha già i freni. Basta iniziare a usarli.

Il pacing invocato da Dario Amodei coincide, sul terreno del diritto, con il regime dei sistemi ad alto rischio dell'AI Act. Regime che il Digital Omnibus ha rinviato al 2027-2028.

Giuseppe Vaciago, Partner, 42 Law Firm **Matteo Flora**, Founder & CEO, The Fool

Key Points

1. Le contromisure che il dibattito industriale invoca per governare l'accelerazione dei modelli di frontiera (verifica esterna, sorveglianza umana, gestione strutturata del rischio, notifica degli incidenti) sono già codificate nell'AI Act per i sistemi ad alto rischio.
2. Il regime esiste dal 1° agosto 2024. Il Regolamento (UE) 2026/1744 (Digital Omnibus on AI) ne ha rinviato l'entrata in vigore al **2 dicembre 2027** per i sistemi dell'Allegato III e al **2 agosto 2028** per quelli dell'Allegato I.
3. Nel frattempo altre parti dell'AI Act si applicano già, con sanzioni attive. Il fronte politico corretto è la messa in opera del regime esistente, non la sua riscrittura.

Executive Summary

Il 9 settembre 2026 Jacob Coxon, ricercatore di Anthropic, si è dimesso denunciando pubblicamente il ritmo di sviluppo dei modelli di frontiera¹. Tre giorni dopo, il 12 settembre, Dario Amodei ha pubblicato *We Must Pace the Frontier*, saggio in cui identifica quattro famiglie di contromisure per governare quel ritmo². I due gesti non si citano; la convergenza è cronologica, il vocabolario è comune.

Questo brief argomenta che quel vocabolario è già scritto, per la parte che ricade sul diritto del prodotto, nel regime che l'AI Act³ impone ai sistemi ad alto rischio. Delle quattro famiglie che Amodei articola, due valutatori esterni con presenza indipendente, coordinamento tra laboratori su standard e ritmi di rilascio trovano nel testo del regolamento una corrispondenza quasi integrale. Le altre due (coordinamento internazionale sui rischi acuti, misure di sicurezza nazionale) ricadono su terreni che nessun regolamento europeo di prodotto avrebbe potuto occupare per natura, e restano affidate al coordinamento intergovernativo e alla politica di sicurezza degli Stati.

Il regime dei sistemi ad alto rischio è tuttavia rinviato. Il Regolamento (UE) 2026/1744 (Digital Omnibus on AI), pubblicato il 24 luglio 2026, modifica l'art. 113 dell'AI Act e sposta al 2 dicembre 2027 e al 2 agosto 2028 l'applicazione degli obblighi per i sistemi degli Allegati III e I⁴. Ne discende una domanda operativa che questo brief prova a mettere a fuoco: se una richiesta di pacing invoca contromisure che l'AI Act già codifica, il rinvio di quel regime opera, sul piano dell'attuazione europea, in senso contrario a quella richiesta.

I. Le contromisure invocate da Amodei

Il saggio di Amodei articola quattro famiglie di misure². La prima riguarda i valutatori esterni: figure indipendenti con presenza on-site nel laboratorio, accesso ai sistemi in sviluppo, diritto di pubblicare senza controllo editoriale del soggetto valutato. La seconda riguarda il coordinamento tra i laboratori di frontiera, con la definizione di standard comuni di sicurezza e con limiti concordati alla velocità dei rilasci. Queste due prime famiglie appartengono, per natura degli strumenti che invocano, al diritto del prodotto e alla regolazione d'impresa: sono misure che si esprimono in obblighi verificabili su chi sviluppa e su chi distribuisce sistemi.

Le altre due si spostano su piani diversi. La terza famiglia chiede coordinamento globale sui rischi acuti (armi biologiche, chimica ostile, autonomia militare), con valutazioni pre-rilascio e con freni sulla velocità dell'auto-miglioramento ricorsivo. La quarta è propriamente materia di sicurezza nazionale: controllo delle esportazioni sui chip, tutela dei pesi dei modelli da esfiltrazione, contromisure contro la distillazione non autorizzata. Sono terreni di coordinamento intergovernativo e di politica di sicurezza degli Stati sui quali un regolamento europeo di prodotto non è, per natura, lo strumento competente. Sul primo piano l'Unione ha già legiferato; sul secondo no, e non poteva.

II. La corrispondenza con il regime high-risk dell'AI Act

La corrispondenza fra ciò che Amodei chiede a chi sviluppa e distribuisce sistemi di intelligenza artificiale e ciò che l'AI Act impone a chi opera sistemi ad alto rischio non è analogia interpretativa: è mappatura letterale, articolo per articolo. Ogni misura invocata come parte del pacing trova nel regolamento una sede normativa puntuale, un contenuto di obbligo, un onere di conformità documentabile e un impianto sanzionatorio proporzionato. La tabella che segue restituisce questa corrispondenza in forma sinottica.

Misura invocata da Amodei o dalla letteratura di sicurezza	Sede nell'AI Act (Reg. UE 2024/1689)
Gestione iterativa del rischio, documentata e aggiornata	Art. 9, sistema di gestione del rischio
Qualità dei dati, mitigazione dei bias, tracciabilità del dataset	Art. 10, governance dei dati
Documentazione tecnica auditabile	Art. 11 e Allegato IV
Logging automatico e tracciabilità operativa	Art. 12, conservazione delle registrazioni
Trasparenza operativa verso il deployer professionale	Art. 13
Sorveglianza umana effettiva	Art. 14
Accuratezza, robustezza, cibersicurezza	Art. 15
Valutazione di conformità da parte terza (notified body)	Art. 43
Valutazione d'impatto sui diritti fondamentali (deployer pubblici e assimilati)	Art. 27
Sorveglianza post-mercato	Art. 72
Notifica degli incidenti gravi	Art. 73
Contromisure per modelli GPAI con rischio sistemico (test avversari, notifica incidenti, cibersicurezza a livello modello)	Artt. 51-55

Executive Takeaway. *Le contromisure del regime high-risk sono già codificate. Ne resta pendente l'entrata in vigore piena.*

La lettura della tabella restituisce l'ambito del discorso in modo netto. Le famiglie 3 e 4 delle misure di Amodei non compaiono per una ragione strutturale e non per lacuna del legislatore europeo: un regolamento europeo di prodotto disciplina i sistemi in commercio, non i rapporti tra Stati né la sicurezza dei pesi come materia di difesa. Il perimetro dell'AI Act è definito dal tipo di strumento, non da un giudizio politico sulle contromisure necessarie. Sulle prime due famiglie, invece, la copertura è quasi integrale e ciò impone di leggere in modo diverso la richiesta di pacing quando la si porta al livello europeo: quella richiesta, se è

coerente con sé stessa, deve tradursi in accelerazione dell'attuazione del regime esistente e non in domanda di un nuovo intervento normativo.

III. Il calendario dopo il Digital Omnibus

Il Regolamento (UE) 2026/1744 (Digital Omnibus on AI) modifica con l'art. 1, punto 41 le date di applicazione dell'AI Act⁴. Rilevano due date: il **2 dicembre 2027** per l'applicazione integrale delle Sezioni 1, 2 e 3 del Capo III ai sistemi classificati ad alto rischio ai sensi dell'art. 6(2) (Allegato III: impiego, credit scoring, biometria, giustizia, servizi essenziali), e il **2 agosto 2028** per i sistemi classificati ad alto rischio ai sensi dell'art. 6(1) (Allegato I: sistemi incorporati in prodotti già disciplinati da normativa europea di sicurezza).

Il rinvio non svuota il resto dell'AI Act. Sono già vincolanti, e in molti casi già sanzionabili, le pratiche vietate dell'art. 5 (dal 2 febbraio 2025), l'obbligo di alfabetizzazione dei fornitori e degli utilizzatori dell'art. 4 (stessa data), gli obblighi di trasparenza per i sistemi che generano o manipolano contenuti dell'art. 50 (dal 2 agosto 2026), il regime dei fornitori di modelli GPAI con le sanzioni dell'art. 101 attivabili dalla Commissione (dalla stessa data) e il quadro sanzionatorio generale dell'art. 99 par. 4. Le soglie sanzionatorie, per l'art. 99 par. 4 e per l'art. 101, sono fissate in **15.000.000 di euro o al 3 per cento del fatturato mondiale annuo dell'esercizio precedente, il maggiore dei due**³.

Executive Takeaway. *Il regime high-risk è rinviato. Il resto dell'AI Act, no.*

Il Digital Omnibus non deregolamenta: differisce. Un differimento privo dello sblocco degli standard armonizzati e dell'operatività piena dei notified bodies trascina con sé anche gli strumenti che avrebbero potuto renderlo meno costoso⁵.

IV. Implicazioni operative

4.1 Per il policy maker europeo

La priorità è la messa in opera del regime esistente, non l'apertura di un cantiere di riscrittura. Tre leve inutilizzate, in ordine di impatto, meritano attenzione politica prima di ogni ipotesi di nuovo intervento normativo:

1. Pubblicazione tempestiva degli standard armonizzati per l'AI Act.
2. Dotazione operativa piena delle autorità nazionali competenti.
3. Sostegno strutturato al *Code of Practice* per i modelli GPAI.

Senza queste tre leve, il 2 dicembre 2027 arriva con un regime formalmente in vigore ma non ancora attuabile.

4.2 Per il provider e il deployer di sistemi ad alto rischio

L'anticipazione dell'adeguamento è la scelta razionale: chi arriva alla scadenza con i presidi minimi già a regime non subisce l'entrata in vigore, la assorbe. I presidi minimi da attivare prima della scadenza applicabile sono sei, tutti già codificati nel testo del regolamento.

1. Un sistema di gestione del rischio ex art. 9, con ciclo iterativo di identificazione, valutazione e mitigazione e con un registro dei rischi aggiornato periodicamente.
2. Una funzione di governance dei dati ex art. 10, con documentazione delle fonti, dei criteri di selezione, delle misure di mitigazione dei bias e delle limitazioni note.

3. Un dossier di documentazione tecnica ex art. 11 e Allegato IV, in forma auditabile, su architettura, dati di addestramento, prestazioni e limiti d'uso.
4. Misure organizzative e tecniche per una sorveglianza umana effettiva ex art. 14 sui punti di decisione.
5. Una valutazione d'impatto sui diritti fondamentali ex art. 27, obbligatoria per i deployer pubblici e assimilati e consigliabile a titolo prudenziale per gli operatori privati esposti.
6. Un registro degli incidenti gravi con canale di notifica strutturato ex art. 73, con procedura di raccolta, classificazione, escalation e punto di contatto verso l'autorità competente.

La roadmap di anticipo che raccomandiamo, calcolata sulla scadenza applicabile al singolo sistema, si articola su quattro finestre:

1. **Dodici mesi prima della scadenza:** mappatura interna dei sistemi (alto rischio Allegato I, alto rischio Allegato III, GPAI, rischio limitato) e assegnazione dei ruoli (provider, deployer, distributore, importer).
2. **Nove mesi prima:** attivazione dei sei presidi minimi in forma non ancora certificata.
3. **Sei mesi prima:** gap assessment con soggetto terzo e preparazione formale del dossier tecnico ex art. 11.
4. **Tre mesi prima:** dry run con notified body per i sistemi Allegato I e test dei canali di notifica ex art. 73.

4.3 Per Anthropic e per i firmatari del pacing

L'autoregolazione, di cui la Responsible Scaling Policy di Anthropic (arrivata alla versione 3.4 nel luglio 2026) è il caso di scuola⁶, è utile ma priva di forza esecutiva esterna. Il regime high-risk dell'AI Act, pur rinviato, resta l'unico strumento vincolante che una richiesta di regolazione del ritmo possa realisticamente invocare in Europa. Chiedere pacing e sostenere in parallelo il differimento di quel regime è, sul piano operativo, contraddittorio.

V. Considerazioni conclusive

Il regime dei sistemi ad alto rischio dell'AI Act contiene già la parte del vocabolario di pacing che spetta al diritto del prodotto. Il Digital Omnibus ne ha rinviato l'entrata in vigore al 2027 e al 2028 senza sciogliere i nodi che avrebbero reso il rinvio meno costoso: standard armonizzati, capacità operativa dei notified bodies, code of practice GPAI. Nel frattempo, parti significative dell'AI Act sono già vincolanti e producono sanzioni. Il policy maker europeo, il provider, il deployer e i firmatari industriali del pacing hanno di fronte, ciascuno, una mossa di attuazione o di anticipo che non richiede una nuova legge.

****Il pacing dell'intelligenza artificiale non ha bisogno di una nuova legge. Ha bisogno che quella scritta cominci a funzionare.****

Note

- ¹ Charlotte Trueman, *"Gambling with our lives": Anthropic researcher quits, warns against self-improving AI"*, TechCrunch, 9 settembre 2026, <https://techcrunch.com/2026/09/09/gambling-with-our-lives-anthropic-researcher-quits-warns-against-self-improving-ai/>; TIME, intervista a Jacob Coxon, 9 settembre 2026, <https://time.com/article/2026/09/09/ai-anthropic-openai-jacob-coxon/>. Nelle dichiarazioni riportate Coxon parla di *"racing straight to self-improving superintelligence and gambling with our lives"* e osserva che *"things are speeding up, and they're not under control"*. Sul filo interno di Anthropic cfr. anche il commento pubblico di Evan Hubinger (Alignment Science): *"we do not yet have a plan to solve alignment for superintelligence and are not clearly on track to."*
- ² Dario Amodei, *"We Must Pace the Frontier"*, darioamodei.com, 12 settembre 2026, <https://darioamodei.com/post/we-must-pace-the-frontier>. Il saggio articola quattro famiglie di contromisure: embedded evaluators con accesso operativo e diritto di pubblicazione indipendente; coordinamento fra laboratori di frontiera su standard e ritmi di rilascio; coordinamento globale su usi pericolosi con test pre-rilascio e limiti all'auto-miglioramento ricorsivo; misure di sicurezza nazionale (controllo esportazioni, protezione dei pesi, anti-distillazione). Rinvia esplicitamente a Dario Amodei, *"The Urgency of Interpretability"* (aprile 2025) e *"Machines of Loving Grace"* (ottobre 2024) come tappe precedenti dello stesso ragionamento.
- ³ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act), in Gazzetta Ufficiale dell'Unione Europea, entrato in vigore il 1° agosto 2024, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>. Le corrispondenze articolo-per-articolo indicate nella tabella della sezione II sono rinvii al testo consolidato del Regolamento. Sulle sanzioni cfr. art. 99, par. 4 (violazioni degli operatori diverse dall'art. 5, incluse quelle dell'art. 50) e art. 101 (fornitori di modelli GPAI): in entrambi i casi ammende fino a 15.000.000 euro o al 3 per cento del fatturato mondiale annuo dell'esercizio precedente, il maggiore dei due.
- ⁴ Regolamento (UE) 2026/1744 del Parlamento europeo e del Consiglio (Digital Omnibus on AI), in Gazzetta Ufficiale dell'Unione Europea, pubblicato il 24 luglio 2026, entrato in vigore il 27 luglio 2026, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ%3AL_202601744. L'art. 1, punto 41 modifica l'art. 113 dell'AI Act fissando l'applicazione delle Sezioni 1, 2 e 3 del Capo III al 2 dicembre 2027 per i sistemi ad alto rischio ex art. 6(2) e Allegato III, e al 2 agosto 2028 per quelli ex art. 6(1) e Allegato I. Sull'analisi del regime post-Digital Omnibus vedi Oreste Pollicino, Giovanni De Gregorio, Rebecca Pupella, *"The AI Act deadlines after the Digital Omnibus on AI"*, Pollicino AI Advisory Policy Brief, 2 agosto 2026, pp. 1-8, <https://pollicinoaidvisory.eu/en/the-ai-act-deadlines-after-the-digital-omnibus-on-ai-applicable-obligations-and-operational-steps-for-providers-and-deployers-di-o-pollicino-g-de-gregorio-e-r-pupella-02-08-2026-1-8/>; Cloud Security Alliance Labs, *"EU AI Act's High-Risk Deadline: Deferred, Not Cancelled"*, 2026, <https://labs.cloudsecurityalliance.org/research/csa-research-note-eu-ai-act-high-risk-deadline-omnibus-20260/>.
- ⁵ Sull'enforcement gap dell'AI Act e sulle implicazioni operative del rinvio vedi Oreste Pollicino, Rebecca Pupella, *"Gli obblighi per i fornitori di sistemi di IA ad alto rischio ai sensi del Regolamento (UE) 2024/1689 (AI Act)"*, Pollicino AI Advisory Policy Brief, 24 ottobre 2025, pp. 1-6, https://pollicinoaidvisory.eu/wp-content/uploads/2025/10/PB_Pollicino_pupella.pdf; Cloud Security Alliance Labs, *"EU AI Act High-Risk Deadline: Enterprise Readiness Gap"*, 2026, <https://labs.cloudsecurityalliance.org/research/csa-research-note-eu-ai-act-high-risk-compliance-deadline-20/>; Bruegel, *"The European Union AI Act: premature or precocious regulation?"*, Bruegel Analysis, <https://www.bruegel.org/analysis/european-union-ai-act-premature-or-precocious-regulation>; Bruegel, *"The right balance: how to fix European Union artificial intelligence regulation"*, Bruegel Policy Brief, <https://www.bruegel.org/policy-brief/right-balance-how-fix-european-union-artificial-intelligence-regulation>. Sul quadro istituzionale di enforcement cfr. Commissione Europea, *"The enforcement framework of the AI Act"*, DG CNECT, Shaping Europe's Digital Future, <https://digital-strategy.ec.europa.eu/en/policies/enforcement-ai-act>. Per il contesto scientifico indipendente sullo stato dell'arte della sicurezza dei sistemi IA di frontiera vedi Yoshua Bengio et al., *International AI Safety Report 2025*, 29 gennaio 2025, <https://internationalaisafetyreport.org/publication/international-ai-safety-report-2025>.
- ⁶ Anthropic, *Responsible Scaling Policy*, versione 3.4, efficace dall'8 luglio 2026, <https://www.anthropic.com/responsible-scaling-policy>. La versione corrente aggiorna la soglia per l'automated R&D "to better track the threat model of concern" e introduce affinamenti procedurali su risk reporting e revisione esterna. La cronologia delle versioni pubblicate nel 2026 (v3.0 del 24 febbraio, v3.1 del 2 aprile, v3.2 del 29 aprile, v3.3 del 26 maggio, v3.4 dell'8 luglio) documenta un processo iterativo di aggiustamento delle soglie e non un impianto stabilizzato.