

20 OTTOBRE 2026

Rivista.AI

INTELLIGENZA ARTIFICIALE. INNOVAZIONE. IMPATTO REALE.

SERIE
SOVRANITÀ
TECNOLOGICA

VOL. 07

Open Source, il potere di scegliere

Dai modelli aperti al potere industriale: la partita della sovranità tecnologica.

Open Source
e Open Weight

Deep Learning
e Community AI

Geopolitica:
Usa, Cina, Europa

L'Italia:
da follower
a snodo

RESEARCH PAPER



rivista.ai

Serie Sovranità Tecnologica

Documento di analisi strategica,
ottobre 2026

In collaborazione con



Con il contributo di

Andrea David Passarelli

Computer Scientist, Babelscape

Antonio Dina

Executive Editor, Rivista.AI

Fabio Ricceri

Executive Editor, Rivista.AI

Rivista.AI

mail: redazione@rivista.ai

newsletter: rivistaai.substack.com

web: www.rivista.ai

Open Source, il potere di scegliere

Come i modelli aperti stanno cambiando
gli equilibri tecnologici tra Stati Uniti, Cina
ed Europa

a cura di **Franco Bernabè**



La sovranità tecnologica non si misura dal numero di modelli sviluppati o dalle infrastrutture possedute, ma dalla capacità di cambiare tecnologia senza perdere dati, competenze, applicazioni e continuità operativa.

L'open source può rendere possibile questa libertà, ma è la reversibilità delle scelte a renderla concreta.



OPEN SOURCE AI

Accesso alle componenti necessarie e libertà di utilizzo, studio, modifica e condivisione del sistema.

Obiettivo: comprensione e riutilizzo indipendente.



OPEN-WEIGHT AI

Accesso ai parametri appresi dal modello, con possibilità e restrizioni definite dalla licenza.

Obiettivo: accesso ed esecuzione del modello.

Sommario.

PREMESSA	Pag. 4
EXECUTIVE SUMMARY	Pag. 6
PARTE I. LE RADICI: L'OPEN SOURCE NEL SOFTWARE (1984–2015)	
1.1 Origini culturali e filosofiche	Pag. 8
1.2 Il ciclo virtuoso dell'open source in IT	9
1.3 Analisi comparativa: dall'IT all'AI	9
PARTE II. DEEP LEARNING E COMMUNITY AI (2012–2020)	
2.1 I primi segnali	Pag. 10
2.2 La nascita di Hugging Face	10
2.3 Le altre community	10
2.4 La questione delle licenze nell'AI	11
PARTE III. LA DIVISIONE APERTO/CHIUSO: IL CASO OPENAI	
3.1 Nascita e mutazione	Pag. 12
3.2 Gli argomenti contro l'apertura	12
3.3 Critiche e risposte	13
3.4 Contesto regolamentare americano	13
PARTE IV. GEOPOLITICA: USA E CINA	
4.1 Il modello americano	Pag. 14
4.2 Il modello cinese: apertura come strategia	14
4.3 Confronto e implicazioni	14
PARTE V. NVIDIA E NEMOTRON: L'OPEN SOURCE COME LEVA HARDWARE	
5.1 Posizione di mercato	Pag. 16
5.2 La strategia Nemotron	16
5.3 Lezione generale	17
PARTE VI. L'EUROPA: COSA FARE	
6.1 Dal discorso alle istituzioni	Pag. 18
6.2 Una grammatica della sovranità: costruire, servire, far agire	18
6.3 Che cosa non è sovranità	19
6.4 L'Italia: da follower a snodo	19
6.5 Le vulnerabilità	20
6.6 Raccomandazioni strategiche	21
CONCLUSIONI	Pag. 22

Premessa. La sovranità oltre il modello

La sovranità tecnologica non coincide con la capacità di produrre autonomamente ogni componente della tecnologia che utilizziamo. Per un Paese come l'Italia, e in larga misura per l'Europa, sarebbe un obiettivo difficilmente raggiungibile, soprattutto nell'intelligenza artificiale, dove la concentrazione di capitali, infrastrutture di calcolo e competenze industriali ha assunto dimensioni senza precedenti.

La questione è piuttosto comprendere quali dipendenze siamo disposti ad accettare, quali capacità dobbiamo preservare e quali condizioni ci consentano di mantenere un margine effettivo di scelta.

È attorno a questa domanda che si è sviluppato il percorso di ricerca della serie Sovranità Tecnologica di Rivista.AI. Nei lavori precedenti abbiamo progressivamente spostato l'attenzione dalla semplice adozione dell'intelligenza artificiale alla capacità di governarne gli effetti economici, industriali e geopolitici. Abbiamo osservato come il valore strategico di un Paese non dipenda soltanto dalle tecnologie che riesce ad acquisire, ma anche dal patrimonio di competenze, dati, processi produttivi e conoscenze che è in grado di conservare e valorizzare.

In questa prospettiva, *“L'effetto mosaico”* ha contribuito a mettere a fuoco un elemento essenziale: la sovranità non nasce dal controllo isolato di una singola tecnologia, ma dalla capacità di combinare risorse e competenze differenti, costruendo un sistema industriale

che non perda il governo dei propri asset strategici.

Per l'Italia, questa riflessione assume un significato particolare, perché il suo principale vantaggio competitivo non risiede nella disponibilità di grandi modelli linguistici o di infrastrutture paragonabili a quelle degli hyperscaler americani, ma in un patrimonio di conoscenze industriali, capacità ingegneristiche e specializzazioni produttive che l'intelligenza artificiale può contribuire a valorizzare, oppure a rendere più dipendente da tecnologie sviluppate altrove.

Questo settimo paper affronta una nuova dimensione di questo problema: l'apertura dei modelli di intelligenza artificiale.

Per decenni, l'open source ha rappresentato uno dei principali strumenti attraverso i quali il software è diventato un'infrastruttura condivisa dell'economia digitale. La possibilità di accedere al codice, modificarlo, distribuirlo e sviluppare soluzioni indipendenti ha favorito la concorrenza, ridotto alcune barriere all'ingresso e consentito la nascita di ecosistemi tecnologici che nessuna singola impresa avrebbe potuto costruire da sola.

Con l'intelligenza artificiale, tuttavia, il significato dell'apertura è cambiato. La diffusione dei modelli open-weight, i cui pesi possono essere resi disponibili anche quando dati e processi di addestramento rimangono proprietari, ha ampliato le possibilità di

accesso a tecnologie avanzate. Ma ha anche introdotto una distinzione fondamentale: poter utilizzare o modificare un modello non significa necessariamente controllare le condizioni tecnologiche ed economiche che ne rendono possibile il funzionamento.

Un modello aperto può continuare a dipendere da infrastrutture di calcolo, semiconduttori, piattaforme cloud, strumenti di inferenza e servizi controllati da un numero ristretto di operatori.

L'apertura di un componente, quindi, non elimina automaticamente le dipendenze. Può modificarle, trasferirle e, in alcuni casi, rafforzare il potere di chi controlla gli altri livelli della filiera.

È anche per questo che la competizione tra modelli aperti e proprietari non può essere interpretata soltanto come una contrapposizione tra differenti filosofie dello sviluppo tecnologico. Le strategie adottate negli Stati Uniti, in Cina e in Europa mostrano come la disponibilità dei modelli sia diventata una leva economica e geopolitica, attraverso la quale imprese e sistemi industriali cercano di orientare standard, ecosistemi e mercati.

Per l'Europa, e in particolare per l'Italia, la questione non è dunque scegliere ideologicamente tra intelligenza artificiale aperta e chiusa. È comprendere in quali condizioni l'apertura possa tradursi in maggiore autonomia, capacità di innovazione e controllo industriale.

Questo richiede di guardare oltre il modello.

Significa interrogarsi sulla disponibilità dei dati, sulle competenze necessarie per adattare le

tecnologie ai processi produttivi, sui costi dell'inferenza, sulla possibilità di utilizzare infrastrutture differenti e sulla capacità di sostituire un fornitore senza compromettere applicazioni, conoscenze e continuità operativa.

In altre parole, significa introdurre un criterio che attraversa l'intero lavoro: la reversibilità delle scelte tecnologiche.

La sovranità non consiste nel possedere tutto, ma nel conservare la possibilità di cambiare. E questa possibilità non può rimanere una dichiarazione di principio: deve poter essere verificata nelle architetture tecnologiche, nei contratti, nei costi e nelle competenze di cui un Paese dispone.

Con questo paper intendiamo quindi contribuire a una riflessione che riguarda direttamente la competitività dell'industria italiana e la capacità delle istituzioni di orientare le trasformazioni in corso.

Perché il valore dell'open source, nell'era dell'intelligenza artificiale, non si misura soltanto da ciò che rende accessibile, ma soprattutto dalla libertà tecnologica e industriale che consente di costruire.

Franco Bernabè
Presidente TechVisory

Executive Summary

L'intelligenza artificiale sta attraversando una trasformazione che riguarda non soltanto le capacità dei modelli, ma anche le condizioni economiche e tecnologiche del loro utilizzo. La crescente diffusione di modelli open-weight, accessibili e adattabili al di fuori delle piattaforme proprietarie che li hanno sviluppati, sta modificando gli equilibri competitivi tra imprese, infrastrutture e sistemi industriali.

La disponibilità dei pesi rappresenta un'opportunità concreta per ridurre le barriere tecnologiche, diversificare i fornitori e sviluppare capacità autonome di utilizzo e adattamento dei modelli. Questi benefici, tuttavia, non eliminano necessariamente le dipendenze legate alle infrastrutture, agli strumenti e alle condizioni economiche dell'intera filiera.

Il settimo paper della serie Sovranità Tecnologica di Rivista.AI analizza questa trasformazione partendo dall'evoluzione dell'open source nel software, dove la disponibilità del codice e la possibilità di modificarlo hanno favorito la nascita di ecosistemi tecnologici condivisi.

Nell'intelligenza artificiale il quadro è più complesso: un modello non è costituito soltanto dal codice, ma anche dai pesi, dai dati di addestramento, dalle procedure di ottimizzazione e dalle infrastrutture necessarie al suo funzionamento. La disponibilità dei pesi non garantisce quindi la piena riproducibilità del modello, né il controllo delle condizioni nelle quali viene eseguito.

Questa distinzione è diventata centrale nella

competizione internazionale. Negli Stati Uniti convivono strategie proprietarie, basate sul controllo dei modelli e delle piattaforme di accesso, e strategie di apertura finalizzate alla diffusione di ecosistemi tecnologici.

La Cina, attraverso iniziative come DeepSeek e Qwen, ha utilizzato la distribuzione di modelli aperti per accelerare l'adozione internazionale delle proprie tecnologie e rafforzare la presenza dei suoi operatori nella competizione globale. L'apertura dei modelli si configura così non soltanto come una scelta tecnica o culturale, ma come uno strumento di politica industriale e di influenza geopolitica.

Il caso di Nvidia rende particolarmente evidente la dimensione economica del fenomeno. La diffusione di modelli aperti può ampliare il mercato delle infrastrutture necessarie a eseguirli, favorendo chi controlla semiconduttori, software di ottimizzazione e capacità di calcolo. Il valore non scompare quando un modello diventa accessibile: può trasferirsi verso l'hardware, il cloud, i servizi di inferenza e gli strumenti che consentono di integrare l'intelligenza artificiale nei processi produttivi.

Occorre tuttavia tenere presente che l'apertura del modello non equivale all'apertura dell'intero sistema tecnologico. Per valutare la sovranità occorre distinguere almeno tre capacità: costruire e adattare i modelli, eseguirli attraverso infrastrutture controllabili e governare gli ambienti nei quali l'AI agisce, utilizza strumenti, accede ai dati e partecipa

ai processi decisionali. La dipendenza può manifestarsi in ciascuno di questi livelli, anche quando il modello utilizzato è formalmente aperto.

Per l'Europa, questa prospettiva impone di superare una concezione della sovranità limitata alla localizzazione dei dati o allo sviluppo di modelli nazionali. La disponibilità di iniziative europee, infrastrutture di calcolo, comunità open source e competenze industriali costituisce una base importante, ma non sufficiente se le condizioni economiche dell'inferenza, gli strumenti applicativi e le capacità di aggiornamento rimangono concentrate presso fornitori esterni.

L'Italia dispone di competenze scientifiche, iniziative tecnologiche e un patrimonio di conoscenze industriali che possono contribuire a costruire una posizione più autonoma nell'ecosistema europeo dell'intelligenza artificiale.

Il valore strategico di queste risorse dipenderà però dalla capacità di collegarle, evitando la frammentazione dei progetti e favorendo architetture interoperabili, sviluppo di competenze, controllo dei dati e utilizzo industriale delle tecnologie.

In questa prospettiva, il Paper individua alcune priorità per le politiche pubbliche e per le strategie delle imprese: definire criteri verificabili di apertura dei modelli, introdurre requisiti di portabilità e sostituibilità negli acquisti tecnologici, sostenere infrastrutture e competenze per l'inferenza, rafforzare la partecipazione italiana ai progetti open source strategici e utilizzare la domanda pubblica come leva per lo sviluppo di un ecosistema nazionale ed europeo.

Il criterio che unifica queste indicazioni è la reversibilità delle scelte tecnologiche: la possibilità di sostituire modelli, fornitori e componenti dell'infrastruttura senza perdere dati, competenze, applicazioni e continuità operativa, né sostenere costi tali da rendere il cambiamento impraticabile.

La sovranità tecnologica, in definitiva, non si misura soltanto attraverso le tecnologie possedute o utilizzate, ma attraverso la capacità effettiva di governarle.

L'open source può rappresentare uno degli strumenti più importanti per costruire questa capacità, a condizione che l'apertura si traduca in autonomia industriale e non semplicemente in nuove forme di dipendenza.



Parte I. Le radici dell'open source nel software (1984-2015)

1.1 Origini culturali e filosofiche

Per comprendere che cosa significhi oggi parlare di "apertura" nell'intelligenza artificiale occorre risalire alla matrice culturale da cui il concetto è nato: quella del software. Nei laboratori universitari americani degli anni Settanta la circolazione del codice sorgente non era una scelta ideologica ma la norma operativa.

Il software viaggiava insieme all'hardware, i ricercatori si scambiavano programmi come si scambiano articoli, e la cultura hacker che si formava in quegli ambienti poggiava su un presupposto semplice: la conoscenza tecnica cresce quando può essere letta, discussa e migliorata da chiunque abbia la competenza per farlo.

La rottura arriva all'inizio degli anni Ottanta, quando l'industria scopre che il software può

essere un prodotto a sé stante e non un accessorio delle macchine. I sorgenti si chiudono, le licenze diventano restrittive, il codice si trasforma in segreto industriale. È in reazione a questa chiusura che Richard Stallman avvia nel 1983 il progetto GNU e fonda nel 1985 la Free Software Foundation, formalizzando le celebri "quattro libertà": eseguire il programma per qualsiasi scopo, studiarne il funzionamento, redistribuirne copie, modificarlo e diffondere le modifiche. Per Stallman il software libero è una questione etica prima ancora che tecnica: chiudere il codice significa sottrarre agli utenti il controllo sugli strumenti da cui dipendono.

Nel 1998 Bruce Perens ed Eric Raymond fondano la Open Source Initiative e coniano il termine che oggi domina il dibattito. La mossa è tattica e sostanziale: "open source" accantona la carica morale del free software per

parlare la lingua dell'industria, presentando l'apertura come metodo di produzione superiore piuttosto che come imperativo etico. La distinzione tra le due anime, quella ideale e quella pragmatica, non è mai scomparsa e riaffiorerà puntualmente nel dibattito sull'intelligenza artificiale. A dimostrare che il modello poteva scalare e vincere era intanto arrivato Linux, il kernel avviato da Linus Torvalds nel 1991: un progetto nato in una cameretta universitaria finlandese che in due decenni sarebbe diventato la base della quasi totalità dei server, dei supercomputer e dei dispositivi mobili del pianeta.

1.2 Il ciclo virtuoso dell'open source nell'IT

Il successo dell'open source nel software non è un accidente storico ma il risultato di un meccanismo economico preciso, un ciclo virtuoso che si autoalimenta. Il primo ingranaggio è la diffusione dei contributi: quando il codice è aperto, chiunque incontri un problema può risolverlo e restituire la soluzione alla comunità, moltiplicando le energie disponibili ben oltre l'organico di qualsiasi singola azienda. Il secondo è l'ispezione del codice: la verificabilità pubblica alza la qualità e la sicurezza, perché i difetti vengono scoperti e corretti da una platea di revisori che nessun reparto qualità proprietario potrebbe eguagliare. Il terzo è la riduzione del lock-in: chi adotta una tecnologia aperta sa di poterla modificare, migrare, far mantenere da altri, e questa garanzia abbassa il costo percepito dell'adozione, accelerando la diffusione dello standard.

Le licenze sono l'infrastruttura giuridica di questo ciclo, e la loro varietà riflette strategie diverse. La GPL incarna l'approccio copyleft: chi ridistribuisce opere derivate deve mantenerle aperte, vincolando l'ecosistema alla reciprocità. Le licenze permissive - Apache, MIT, BSD - rinunciano a questo vincolo e puntano sulla massima adozione possibile, accettando

che il codice possa essere incorporato anche in prodotti proprietari. La scelta tra i due modelli è una scelta di politica industriale in miniatura: proteggere i beni comuni o massimizzare la diffusione. Il risultato storico è che l'open source è diventato l'infrastruttura invisibile dell'intera economia digitale. Internet funziona su protocolli e implementazioni aperte; il web si regge su server come Apache e Nginx; i database relazionali dominanti - MySQL, PostgreSQL - sono aperti, come lo sono i linguaggi con cui si scrive la gran parte del software contemporaneo. E le grandi aziende, dopo una fase di ostilità, hanno imparato a farne il perno delle proprie strategie con IBM che ha investito su Linux già alla fine degli anni '90, trasformandolo nella spina dorsale della propria offerta enterprise.

1.3 Analisi comparativa: dall'IT all'AI

Le analogie tra il software degli anni '90 e l'intelligenza artificiale degli anni '20 spiegano perché il concetto di apertura sia stato trasferito da un dominio all'altro. In entrambi i casi sono in gioco gli standard di un'intera economia, i costi di ingresso determinano chi può competere e l'innovazione distribuita, alimentata da migliaia di team su una base comune, può avanzare più rapidamente dei laboratori chiusi. Le differenze, tuttavia, sono sostanziali. Addestrare un modello di frontiera richiede investimenti enormemente superiori a quelli necessari per sviluppare un sistema operativo: migliaia di acceleratori, data center dedicati e consumi energetici elevati. Un modello non è soltanto codice, ma il risultato di dati, procedure di ottimizzazione, infrastrutture e competenze specialistiche. La disponibilità di pesi e di altri componenti del modello amplia le possibilità di utilizzo, adattamento e valutazione indipendente, ma non garantisce piena trasparenza e riproducibilità del processo di addestramento.

Parte II. Deep learning e community AI (2012-2020)

2.1 I primi segnali

La stagione contemporanea dell'intelligenza artificiale si apre nel 2012, quando AlexNet dimostra che le reti neurali profonde, allenare su GPU, possono schiacciare la concorrenza nel riconoscimento delle immagini. È una rinascita e avviene in piena tradizione accademica: articoli pubblici, dataset condivisi, codice allegato alle pubblicazioni. La ricerca sul deep learning cresce nei suoi primi anni come cresce la scienza, per accumulazione aperta di risultati verificabili, e questa abitudine segnerà a lungo la cultura del settore anche dopo l'ingresso massiccio del capitale privato.

L'apertura si consolida negli strumenti. Nel 2015 Google rilascia TensorFlow, Meta (allora Facebook) PyTorch nel 2016: i due framework che diventeranno l'infrastruttura di fatto della ricerca mondiale sono open source, regalati dai due giganti che avrebbero potuto tenerli chiusi.

La ragione è la stessa che aveva spinto IBM su Linux: standard aperti attraggono talenti, accelerano l'ecosistema e spostano la competizione sul terreno dove chi li rilascia è più forte. E soprattutto, le scoperte decisive restano pubbliche: il meccanismo di attenzione e l'architettura transformer, presentati nel 2017 nel paper "Attention Is All You Need", sono il fondamento tecnico di tutta la generazione successiva di modelli e sono nati come letteratura aperta, a disposizione di chiunque.

2.2 La nascita di Hugging Face

Se i framework sono l'infrastruttura, la piazza viene costruita quasi per caso. Hugging Face, fondata nel 2016 come startup di chatbot per

teenager, compie attorno al 2020 un pivot che la trasforma nell'hub centrale dell'ecosistema. L'intuizione è replicare per i modelli ciò che GitHub aveva fatto per il codice: uno spazio dove caricare, scoprire, scaricare e discutere modelli, dataset e dimostrazioni interattive, tenuto insieme dalla libreria Transformers, che diventa lo standard con cui la comunità impacchetta e distribuisce le architetture.

L'effetto è una democratizzazione reale, nel senso tecnico del termine: un laboratorio con migliaia di GPU e un dottorando con una scheda consumer operano sulla stessa piattaforma, scaricano gli stessi artefatti, parlano la stessa lingua. La distanza tra frontiera e periferia, che i costi di addestramento tendevano ad allargare senza rimedio, viene compressa dalla circolazione gratuita dei risultati. È il ciclo virtuoso della Parte I che si riaccende in un dominio nuovo.

2.3 Le altre community

Attorno a Hugging Face fiorisce un arcipelago di comunità con vocazioni diverse, che vale la pena distinguere perché anticipa le fratture successive. EleutherAI nasce nel 2020 come collettivo di ricercatori volontari con un obiettivo esplicito: replicare GPT-3, il modello che OpenAI aveva deciso di non rilasciare.

Ne risultano GPT-Neo, GPT-J e soprattutto Pythia, una famiglia di modelli pubblicata con dati e procedure documentate, pensata per rendere la ricerca sui grandi modelli di linguaggio una scienza riproducibile anziché una dimostrazione di potenza. LAION, dal 2021, affronta l'altro collo di bottiglia, i dati,

costruendo LAION-5B, un dataset aperto di miliardi di coppie immagine-testo, senza il quale la generazione di immagini aperta non sarebbe esistita. Su quel dataset nasce nel 2022 Stable Diffusion, frutto della collaborazione tra la startup Stability AI e il gruppo di ricerca CompVis: il modello viene rilasciato apertamente, in contrapposizione frontale a Midjourney e DALL-E che restano chiusi, e nel giro di settimane genera un ecosistema di applicazioni che nessuna roadmap proprietaria avrebbe potuto pianificare.

A completare il panorama stanno le infrastrutture di circolazione e di esecuzione: GitHub come archivio del codice, Kaggle come arena delle competizioni, Papers with Code come indice che lega articoli e implementazioni, e più tardi Ollama e vLLM, che rendono possibile eseguire e servire modelli di grandi dimensioni su hardware accessibile. Due modelli di governance si fronteggiano in questo arcipelago: i collettivi bottom-up, animati da volontari e da una motivazione vicina all'etica hacker delle origini, e le startup-fondazione, in cui l'apertura è anche strategia commerciale. La tensione tra le due anime del 1998 - ideale e pragmatismo - è tornata, con gli stessi contorni.

2.4 La questione delle licenze nell'AI

È qui che l'analogia con il software si incrina. Nell'IT la definizione di open source era stabilizzata: la conformità alle quattro libertà, certificata dall'Open Source Initiative, tracciava un confine binario e verificabile. Nell'AI l'artefatto è composito - codice, pesi, dati, procedure - e l'apertura può riguardarne una parte sola.

Si è così affermata la categoria degli open weight: i pesi del modello vengono resi disponibili, anche quando i dati di addestramento e il codice di training rimangono proprietari.

Chi vi accede può utilizzare il modello, adattarlo alle proprie esigenze, eseguirlo su infrastrutture differenti e valutarne autonomamente prestazioni e comportamento.

Il caso scuola è Llama. Meta lo presenta come aperto e i suoi pesi circolano effettivamente, ma l'apertura effettiva deve essere valutata sulla base della licenza della specifica versione del modello. Le condizioni possono differire tra versioni e famiglie, con conseguenze su utilizzo commerciale, redistribuzione e obblighi imposti agli utilizzatori.

Per questo motivo, la dicitura open-weight non deve essere interpretata come sinonimo di assenza di restrizioni contrattuali, né automaticamente equiparata all'open source secondo i criteri tradizionali. La distinzione è rilevante sul piano delle licenze e delle libertà riconosciute agli utilizzatori, ma non riduce il valore tecnologico e industriale che la disponibilità dei pesi può generare.

Se l'utilizzo, la modifica o la distribuzione di un modello sono soggetti a restrizioni imposte dal produttore, non si può parlare di piena libertà secondo i principi dell'open source. Per chiarire quali requisiti debba rispettare un sistema di intelligenza artificiale per essere considerato realmente aperto, nel 2024 l'Open Source Initiative (OSI) ha pubblicato la Open Source AI Definition.

La definizione ha però suscitato un dibattito, soprattutto perché ammette che i dati di addestramento non siano integralmente accessibili, una scelta contestata da chi considera la trasparenza dei dati indispensabile per una vera apertura.

La questione non è filologica: su che cosa possa legittimamente chiamarsi "aperto" si giocano appalti, politiche industriali e la fiducia di chi costruisce sistemi destinati a durare.



Parte III. La divisione aperto/chiuso: il caso OpenAI

3.1 Nascita e mutazione

Nessuna vicenda incarna la divisione tra aperto e chiuso quanto quella di OpenAI, perché la società è nata dalla parte dell'apertura. Fondata nel 2015 come organizzazione no-profit, con una missione dichiarata, garantire che l'intelligenza artificiale generale fosse "a beneficio di tutta l'umanità" e un modello operativo esplicitamente aperto: pubblicare la ricerca, condividere i risultati, evitare che la concentrazione del potere tecnologico producesse un monopolio privato dell'intelligenza.

La mutazione avviene per gradi e per ragioni che vale la pena riconoscere nella loro forza. Nel 2019 la struttura diventa capped-profit, un ibrido che consente di raccogliere capitale privato mantenendo formalmente il controllo alla no-profit, e si sigla la partnership con Microsoft, che porta miliardi di dollari e, soprattutto, accesso al calcolo.

Il punto di svolta pubblico arriva con GPT-4

nel 2023: il modello non viene rilasciato, e la motivazione ufficiale è la sicurezza, "sarebbe irresponsabile distribuirlo", dati i rischi. Dietro "la motivazione dichiarata se ne affacciano altre, non necessariamente false né incompatibili: i costi di addestramento sono ormai tali che regalare il risultato significa finanziare i concorrenti, la competizione si è fatta frontale e il controllo sull'uso del modello è un argomento che, preso sul serio, ha un suo peso. Ma il fatto resta: l'organizzazione nata per impedire la concentrazione dell'intelligenza è diventata il suo campione più chiuso.

3.2 Gli argomenti contro l'apertura

Gli argomenti a favore della chiusura meritano di essere esposti nella loro versione migliore. Il primo è la sicurezza sostanziale: alcuni modelli di frontiera possono sviluppare capacità, in ambito biologico, chimico, cibernetico che nelle mani di attori malintenzionati rischiano di abbassare la soglia di accesso a conoscenze pericolose. La distribuzione pub-

blica dei pesi rende inoltre difficile impedirne successivamente la circolazione, riducendo le possibilità di controllo diretto da parte dello sviluppatore. Il secondo è economico: la disponibilità dei pesi può facilitare il trasferimento di capacità verso modelli più piccoli, attraverso tecniche di distillazione e adattamento, riducendo i costi di sviluppo per altri operatori. Dal punto di vista delle imprese che hanno sostenuto gli investimenti iniziali, questa possibilità può tradursi in una riduzione del vantaggio competitivo e in una più rapida diffusione delle capacità sviluppate. Il terzo è operativo: un modello distribuito attraverso API proprietarie consente allo sviluppatore di mantenere un controllo centralizzato sull'accesso, sugli aggiornamenti e sulle modalità di utilizzo, mentre la distribuzione dei pesi trasferisce una parte di queste responsabilità ai soggetti che adottano e gestiscono il modello. In questa ricostruzione, la chiusura viene quindi presentata come uno strumento per garantire un controllo più diretto sulla sicurezza e sull'utilizzo della tecnologia.

3.3 Critiche e risposte

Le critiche a questa architettura argomentativa sono altrettanto serie. Yann LeCun, chief AI scientist di Meta, ha formulato la posizione opposta nella sua forma più tagliente: l'AI non può essere sicura se non è aperta, perché la sicurezza di un sistema che nessuno può ispezionare è una promessa, non una proprietà verificabile e il rischio è che la segretezza sia un vantaggio competitivo travestito da precauzione.

L'argomento riprende, consapevolmente o no, la lezione della sicurezza informatica: la sorveglianza collettiva di migliaia di occhi indipendenti ha storicamente prodotto sistemi più robusti dei controlli privati di un singolo fornitore.

Esiste però un contro argomento che merita attenzione: le capacità davvero pericolose

richiedono infrastrutture costose per essere esercitate e l'apertura totale trasferirebbe il vantaggio dai laboratori accreditati, sottoposti a scrutinio pubblico e regolatorio, verso attori che quello scrutinio lo eludono per definizione. Il dibattito, insomma, non si risolve con uno slogan. Ciò che però l'esperienza recente ha chiarito è il prezzo sistemico del modello chiuso: centralizzazione del potere, dipendenza degli ecosistemi da un pugno di fornitori, lock-in che si trasmette dalle API alle applicazioni. Questo prezzo è documentato. La sicurezza promessa dalla chiusura, al contrario, resta un'asserzione che i suoi proponenti non hanno mai reso verificabile.

3.4 Il contesto regolamentare americano

Il dibattito non è rimasto confinato ai laboratori. Negli stessi mesi in cui si consumava la svolta della chiusura, Washington discuteva se e come regolare i modelli di frontiera: audizioni e lettere al Congresso, proposte di licensing - l'idea che per addestrare modelli oltre una certa soglia di potenza servisse un'autorizzazione federale - e la reazione allarmata delle startup e delle comunità open, che in quel disegno vedevano la regolamentazione trasformarsi in barriera all'ingresso a beneficio degli incumbent.

A rendere il quadro più opaco, la tensione con la Cina come motivazione implicita e trasversale: l'argomento che l'America non può permettersi di aprire ciò che Pechino copierebbe, oppure, nella versione speculare, che l'America non può permettersi di chiudere ciò che Pechino aprirebbe. Come si vedrà nella parte successiva, entrambe le versioni contenevano una parte di verità.

Parte IV. Geopolitica: Usa e Cina

4.1 Il modello americano

Negli Stati Uniti la divisione aperto/chiuso non è un destino ma una strategia duplice, perseguita da attori diversi con obiettivi convergenti. Sul fronte chiuso stanno OpenAI, Anthropic e Google, che difendono la rendita sui modelli migliori servendoli come prodotto. Sul fronte aperto sta Meta, che con la famiglia Llama ha scelto di regalare i pesi: non per filantropia, ma perché i pesi aperti sono la piazza d'armi perfetta contro chi il modello lo vende. Se lo standard diventa gratuito, il prezzo dei modelli altrui crolla e se lo standard gratuito lo definisce Meta, il lock-in si sposta lì dove Meta incassa: piattaforme, pubblicità, servizi, engagement.

È la lezione di Android applicata all'intelligenza: aprire il componente che il rivale monetizza per commoditizzarlo, e concentrare il valore sullo strato che si controlla.

La strategia, però, convive con un dibattito di sicurezza nazionale che la taglia trasversalmente: se i migliori pesi americani sono pubblici, non li sta usando anche Pechino?

La domanda è seria e ha una risposta scomoda per entrambi gli schieramenti: sì, li usa, ma li usava comunque, distillando i modelli chiusi e nel frattempo la circolazione aperta mantiene l'ecosistema globale ancorato agli standard americani anziché spingerlo verso alternative.

4.2 Il modello cinese: apertura come strategia

La Cina ha letto la stessa lezione e l'ha applicata con maggiore radicalità. Il caso DeepSeek ha segnato lo spartiacque: un

modello con performance competitive ai vertici, addestrato a costi dichiarati di un ordine di grandezza inferiori, rilasciato apertamente: uno shock che si è propagato dai laboratori ai mercati finanziari nel giro di ore, perché smentiva simultaneamente due assunti: che la frontiera richiedesse necessariamente risorse colossali, e che la chiusura proteggesse il vantaggio americano.

Accanto a DeepSeek, Alibaba ha costruito con la famiglia Qwen la presenza più sistematica: famiglie complete di modelli in ogni taglia e modalità, dominanti nelle classifiche dei modelli aperti, adottati come base di partenza da sviluppatori di mezzo mondo.

Le ragioni strategiche di questa apertura sono trasparenti e per nulla sentimentali. Rilasciare pesi competitivi colma il divario reputazionale con gli Stati Uniti più in fretta di qualsiasi pro-gramma di ricerca nazionale; costruisce soft power tecnologico, ancorando gli sviluppatori globali a stack cinesi; riduce la dipendenza dall'hardware americano, perché un ecosistema che prospera su modelli efficienti ha meno bisogno dei chip che Washington non vende più e ha un rapporto costo-beneficio schiacciante: il rilascio costa quasi nulla rispetto all'addestramento, mentre i ritorni di ecosistema, standard e influenza si accumulano gratuitamente. L'apertura, per Pechino, non è la rinuncia alla rendita: è il mezzo più economico per conquistare la posizione.

4.3 Confronto e implicazioni

Il risultato è un paradosso che definisce la fase presente. Gli Stati Uniti conservano una posizione di primo piano nello sviluppo dei modelli

proprietari più avanzati, mentre la Cina ha fatto della distribuzione di modelli open-weight competitivi uno degli strumenti principali della propria espansione tecnologica. La frontiera senza standard è una posizione più fragile di quanto appaia: lo si è imparato con i sistemi operativi, con i browser, con i protocolli di rete. Chi definisce lo standard gratuito definisce il mercato e chi vende il prodotto migliore fuori dallo standard rischia la nicchia di lusso.

La reazione americana si sta organizzando su due binari: i controlli su esportazioni, chip e una discussione sempre più esplicita sui modelli aperti come commodity strategica, un bene la cui diffusione ha conseguenze geopolitiche comparabili a quelle dell'hardware. È in questo spazio, tra una frontiera chiusa e uno standard aperto conteso tra due potenze, che si colloca la questione europea, che affronteremo nel capitolo successivo.

Matrice di Valutazione della Sovranità Operativa dell'AI

APPLICAZIONE PRATICA E RISCONTRO COMPARATIVO SUI SISTEMI ANALIZZATI

1. Valutazione Comparativa per Tipologia di Architettura AI

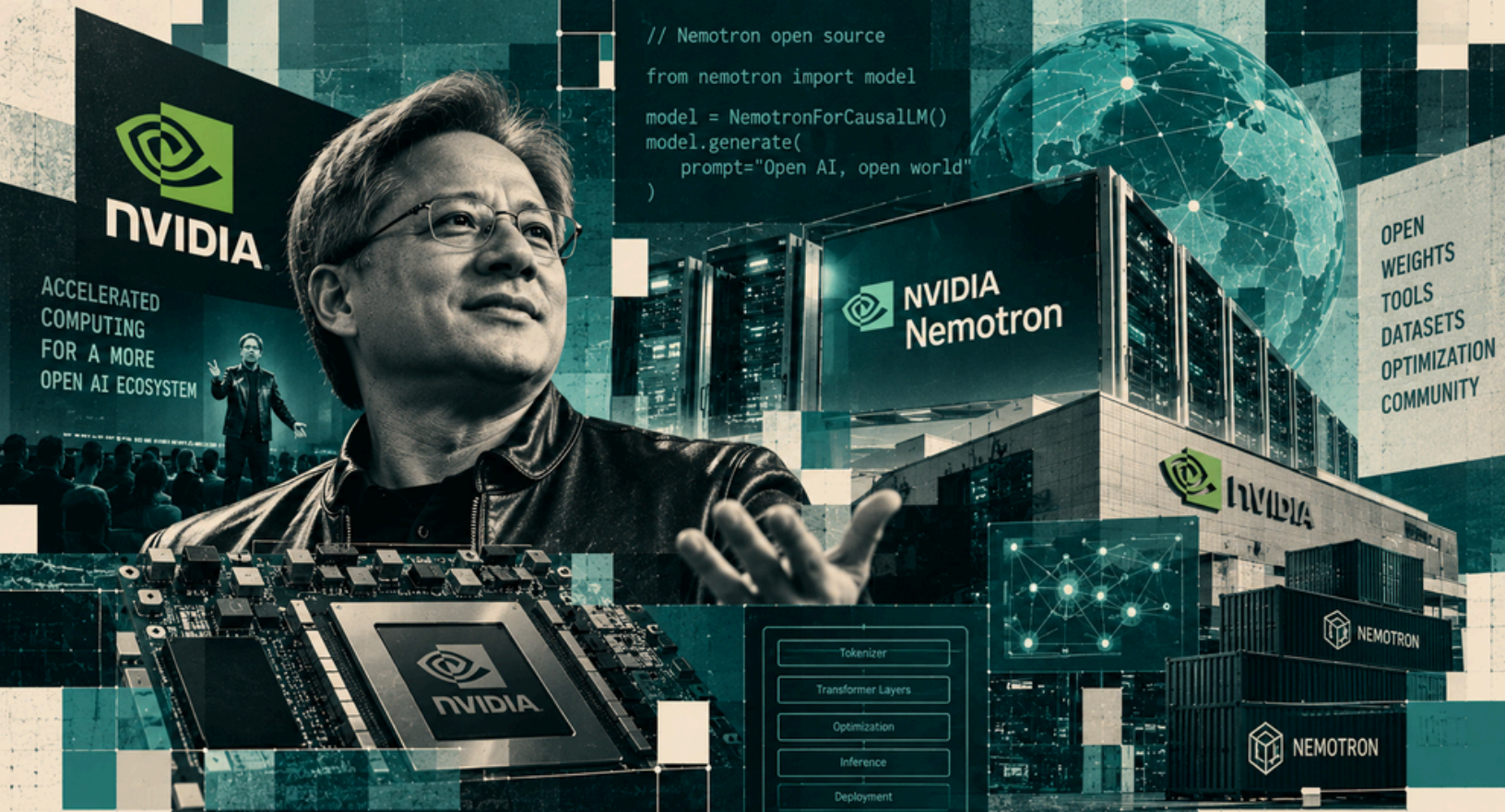
La tabella seguente popola i criteri di sovranità analizzando tre modelli tipologici di adozione: **A) Open-Weights Self-Hosted** (es. Llama 3 on-premise), **B) Managed API Proprietary** (es. OpenAI / Anthropic cloud API), e **C) Modello ibrido / Managed Sovereign Cloud**.

Dimensione	A) Open-Weights Self-Hosted	B) API Proprietary Cloud	C) Cloud Sovrano / Managed
Modello	Alta: Pesì scaricabili, licenza permissiva (es. Llama Community), modificabilità totale.	Bassa: Pesì non accessibili, licenze restrittive, dipendenza totale dal vendor.	Media: Accesso controllato ai pesì o modelli custom accordati contrattualmente.
Dati	Alta: Governance totale in-house, zero data leakage verso terzi, portabilità nativa.	Bassa: Tracciamento dipendente dai termini di servizio del cloud provider estero o locale.	Media-Alta: Storage localizzato con policy di non-training garantite contrattualmente.
Inferenza	Alta: Esecuzione locale via vLLM/Ollama, costi fissi hardware, zero lock-in di API.	Bassa: Pay-per-token vincolato, rischio deprecazione endpoint e modifiche breaking.	Media: Endpoint dedicati con garanzie di SLA e continuità operativa a medio termine.
Infrastruttura	Media: Richiede investimenti in GPU (NVIDIA/AMD) e competenze DevOps/MLOps interne.	Alta: Nessun onere infrastrutturale diretto; gestione totalmente delegata al fornitore.	Alta: Infrastruttura certificata in cloud locali o data center nazionali dedicati.
Continuità Operativa	Alta: Sostituibile istantaneamente con altro modello open a parità di hardware (bassa frizione).	Bassa: Rischio blocco servizio immediato in caso di controversie o interruzioni del vendor.	Media-Alta: Mitigazione del rischio grazie a contratti di escrow e clausole di transizione.

2. Indice Sintetico di Sovranità Operativa

Sintesi quantitativa dei profili di rischio e controllo per ciascuna opzione strategica esaminata:

Architettura	Livello di Sovranità	Costo di Migrazione	Principale Vulnerabilità
A) Open-Weights Self-Hosted	ALTA (Livello 1)	Basso / Medio (solo infrastruttura)	Carenza di competenze interne e obsolescenza hardware.
B) API Proprietary Cloud	BASSA (Livello 3)	Altissimo (riscrittura codice/prompt)	Vendor lock-in, variazioni tariffarie e giurisdizione estera.
C) Cloud Sovrano / Managed	MEDIA (Livello 2)	Medio / Alto (vincoli contrattuali)	Subordinazione tecnologica a intermediari locali.



Parte V. NVIDIA e Nemotron: l'open source come leva hardware

5.1 Posizione di mercato

C'è un attore la cui strategia si comprende solo uscendo dalla dicotomia aperto/chiuso sui modelli, perché il suo prodotto non è il modello ma ciò che lo fa girare. NVIDIA occupa una posizione dominante nell'infrastruttura hardware per l'intelligenza artificiale. Il suo vantaggio non dipende esclusivamente dalle GPU, ma dall'integrazione tra acceleratori, CUDA, librerie software e strumenti di sviluppo e distribuzione.

Questa integrazione crea costi di sostituzione che possono persistere anche quando un'organizzazione decide di cambiare modello AI.

La sua posizione dominante, però, non è priva di minacce. I grandi laboratori e i cloud

provider - i suoi clienti migliori - hanno ogni incentivo a costruire infrastrutture e chip proprietari che riducano la dipendenza dal fornitore; AMD presidia la fascia alternativa del mercato; e i custom chip dei cloud (TPU di Google, Trainium di Amazon e analoghi) sottraggono carichi di lavoro all'ecosistema CUDA. Per NVIDIA il rischio di lungo periodo non è un concorrente che venda GPU migliori: è un mondo in cui i modelli migliori girino su hardware non suo, ottimizzati altrove, serviti da stack che non controlla.

5.2 La strategia Nemotron

È in questa chiave che va letta la famiglia Nemotron, i modelli aperti rilasciati da NVIDIA nel periodo 2024-25 con licenze permissive.

Le ragioni economiche sono immediate: un modello ottimizzato in fabbrica per l'hardware NVIDIA è, di fatto, un dimostratore delle prestazioni delle sue GPU: chi lo adotta scopre che gira meglio su quel silicio, e l'adozione del modello diventa domanda di hardware. I modelli aperti funzionano così da standard "Nvidia-friendly": fissano convenzioni, formati e percorsi di ottimizzazione attorno ai quali si organizza l'ecosistema, e svolgono un marketing tecnico dei chip più efficace di qualsiasi brochure, perché la prova è eseguibile.

Le ragioni strategiche sono altrettanto nitide.

Rilasciare modelli propri colma il divario nella produzione di foundation model senza entrare in guerra di prodotto con i clienti; posiziona NVIDIA come fornitore neutro dell'intero ecosistema, l'unico attore che ha interesse alla prosperità di tutti i costruttori di modelli aperti, perché ognuno di essi è un moltiplicatore della domanda di calcolo e indebolisce strutturalmente i modelli chiusi, il cui successo è la premessa degli stack verticali che minacciano il suo dominio. La logica, ridotta a

una riga, è impeccabile: il modello è gratis, il motore che lo fa girare lo vendiamo noi.

5.3 Lezione generale

Nemotron chiarisce il principio che attraversa tutta la storia raccontata fin qui: l'apertura è una leva, e come ogni leva sposta il valore invece di distruggerlo. Meta la usa per spostarlo dai modelli alle piattaforme, Hugging Face per spostarlo verso l'hub, i cloud provider per spostarlo verso l'infrastruttura venduta a consumo, NVIDIA per spostarlo verso l'hardware. In ogni caso, l'open source ridistribuisce il valore da chi possiede il modello a chi possiede l'infrastruttura che lo rende utilizzabile e questa legge di gravità economica, che la retorica dell'apertura tende a nascondere, è il punto esatto da cui deve partire qualsiasi discussione seria sulla sovranità.

Perché se il valore migra verso l'infrastruttura, la domanda decisiva non è più soltanto "chi possiede il modello", ma "chi possiede tutto il resto". È la domanda a cui l'Europa, dopo anni di attesa, ha cominciato a rispondere.



Parte VI. L'Europa: cosa fare

6.1 Dal discorso alle istituzioni

Il quadro europeo di ottobre 2026 è diverso da quello di dodici mesi prima. Il pacchetto sulla sovranità tecnologica presentato dalla Commissione a giugno comprende semiconduttori, AI, cloud e open source, e include una strategia dedicata all'open source con una mobilitazione di circa due miliardi di euro su sette anni, una rete di uffici per i programmi open source (OSPO) e un'infrastruttura di beni comuni digitali.

Accanto a questo si prefigura un quadro europeo per valutare la sovranità di cloud e AI. Il segnale politico è netto: l'open source non è più trattato come una preferenza culturale della comunità degli sviluppatori, ma come un pilastro dell'autonomia strategica.

Dopo quanto visto nelle parti precedenti, la svolta non sorprende: se l'apertura è il terreno su cui Stati Uniti e Cina stanno giocando la partita degli standard, restarne fuori equivarrebbe a subirne l'esito.

A questa cornice si sono affiancati fatti concreti. Il Frontier AI Grand Challenge, lanciato a febbraio e assegnato a giugno al consorzio EUROPA guidato dall'italiana Domyne, punta a un modello frontier aperto di oltre 400 miliardi di parametri, multilingue nativo sulle 24 lingue dell'Unione, addestrato su infrastruttura europea EuroHPC.

Sul versante dei modelli, i rilasci di operatori europei come Mistral e Aleph Alpha mostrano che la capacità industriale esiste e che la conformità all'AI Act può essere progettata fin dall'origine anziché inseguita a posteriori.

Sul versante associativo, la nascita di una

Sovereign AI Foundation sotto l'ombrello dell'Eclipse Foundation, con una platea di organizzazioni industriali e di ricerca europee, indica che si sta costruendo anche il tessuto di governance neutrale che mancava, lo stesso ruolo che le fondazioni avevano svolto, trent'anni prima, per il software.

Si tratta di un passaggio significativo: dal discorso alle istituzioni, dalle istituzioni ai prodotti. Va però sottolineato che molte delle informazioni più recenti, in particolare date di rilascio, parametri dichiarati e composizione delle coalizioni, provengono da comunicati e fonti giornalistiche e andrebbero verificate sulle fonti primarie prima di essere assunte come base di decisioni di policy.

6.2 Una grammatica della sovranità: costruire, servire, far agire

Per trasformare la cornice politica in strategia serve un vocabolario analitico. Una lettura utile, emersa nel dibattito attorno all'Open Intelligence Summit di Digital Ocean, articola l'intelligenza in tre livelli: costruire (dati, pesi aperti, fine-tuning, post-training), servire (motori di inferenza, routing, caching, uso delle GPU, latenza e throughput) e far agire (agenti, memoria, strumenti, protocolli come MCP). La proposta ha origine commerciale e va letta come tale, ma la sua tassonomia ha un valore analitico che prescinde dall'origine.

La sua forza sta nel mostrare che la sovranità non si perde soltanto scegliendo un fornitore extra-europeo. Si perde quando i dati devono uscire da un perimetro controllabile per alimentare un servizio remoto;

quando il modello è fruibile solo tramite API proprietaria; quando il costo dell'inferenza diventa una variabile opaca decisa dal fornitore; quando memoria, agenti e orchestrazione sono legati a una piattaforma non sostituibile senza riscrivere l'intero sistema. Ciascuna di queste perdite si colloca in un punto diverso della catena, e soltanto la prima riguarda il modello in senso stretto. Ne discendono due osservazioni di carattere strategico.

Il livello "servire" è il più sottovalutato e il più decisivo. La Parte V ha mostrato come l'apertura sposti il valore dal modello all'infrastruttura che lo esegue; visto dalla prospettiva di un Paese, quel principio diventa una questione di politica economica. Possedere i pesi di un modello non significa possedere l'economia dell'AI: servono GPU, memoria, storage, networking, energia, software di orchestrazione e competenze, ed è lì che il collo di bottiglia, e con esso la rendita, si concentra. È una dinamica già vista nella storia dell'industria tecnologica: si libera il componente che tutti guardano mentre il margine migra in quello che pochi considerano abbastanza interessante. Sapere quanto costa far girare un modello, dove sono elaborati i dati, su quali GPU, con quale runtime e con quale facilità un carico di lavoro può migrare, significa trattare la sovranità come una questione di unit economics prima che di orgoglio nazionale.

Il livello "far agire" è il più delicato sul piano della sicurezza e della responsabilità. Con gli agenti, il modello non si limita a rispondere: legge documenti, interroga basi dati, usa API, modifica sistemi, conserva memoria, compie scelte operative. Controllare il solo modello è allora inadeguato; occorre controllare l'harness, gli strumenti, le autorizzazioni, la memoria e i protocolli. Per la pubblica amministrazione e per le infrastrutture critiche, questo è il terreno su cui la sovranità si tra-

duce in responsabilità giuridica e continuità del servizio.

6.3 Che cosa non è sovranità

Da questa impostazione discende una definizione in negativo, utile per evitare errori di politica industriale.

La sovranità non è localizzazione dei dati. Un dato conservato sul territorio nazionale ma elaborato da un sistema che non può essere ispezionato, modificato, migrato o sostituito offre una sovranità decorativa.

Non è nemmeno il possesso di un modello con un nome nazionale. La tentazione di avere un LLM italiano per ogni applicazione possibile è comprensibile ma rischia di essere dispersiva e persino provinciale: un Paese non diventa autonomo perché dispone di un modello che porta il suo nome, ma perché imprese e amministrazioni possono governare il ciclo economico e tecnologico dell'intelligenza.

E non è autarchia. La sovranità non significa possedere tutto, significa non essere prigionieri di niente. La vera unità di misura non è il numero di modelli prodotti né i miliardi investiti in un cluster, ma la capacità di sostituire un componente dello stack senza perdere dati, competenze, memoria, applicazioni e continuità operativa. È un criterio di reversibilità, e come tale misurabile, auditabile e inseribile nei capitolati di gara.

6.4 L'Italia: da follower a snodo

In questo scenario la posizione italiana è più interessante di quanto la narrazione corrente suggerisca. Il Paese guida, attraverso Domyn, il consorzio vincitore della sfida europea sul modello frontier aperto. Esprime con Engineering una proposta di architettura end-to-end, denominata IS-IA, che intende gover-

nare l'intero ciclo dell'AI dall'infrastruttura ai dati, dal foundation model all'orchestrazione nei processi, con al centro il modello EngGPT 2. Annovera progetti di ricerca come Minerva, famiglia di modelli preaddestrati da zero per l'italiano con il contributo di Sapienza NLP e dell'infrastruttura Leonardo del CINECA; iniziative della pubblica amministrazione, come il progetto AI Welfare dell'INPS, finalizzato allo sviluppo di un modello linguistico proprietario per il welfare pubblico italiano; un ecosistema di imprese più piccole, orientate a soluzioni locali, privacy-first e on-premise. Sul piano infrastrutturale si registra la presenza di una AI factory a Bologna e il dibattito su una gigafactory europea.

Il valore di questo insieme non sta nella somma dei singoli progetti, che hanno livelli di apertura, maturità e ambizione molto diversi, ma nella possibilità di una composizione: un modello frontier europeo a cui l'Italia contribuisce in modo determinante, architetture applicative nazionali capaci di adattarlo ai processi di amministrazioni e settori regolati, dati pubblici di qualità, una domanda pubblica in grado di fare da cliente-àncora.

L'Italia può quindi aspirare al ruolo di snodo, non di semplice utilizzatore, a patto di evitare due errori simmetrici: la frammentazione in molte iniziative non interoperabili e la dipendenza di fatto da infrastrutture esterne per il livello che crea più valore.

6.5 Le vulnerabilità

Un'analisi della sovranità tecnologica deve considerare anche le vulnerabilità. Il compute e l'hardware. L'Europa dispone di capacità di calcolo e produzione di semiconduttori molto inferiori a Stati Uniti e Cina. Se l'inferenza dipende da infrastrutture controllate all'estero, l'apertura dei pesi non garantisce da sola autonomia, perché il valore eco-

nomico continua a concentrarsi nei servizi e nell'hardware.

Il contributo upstream. Utilizzare tecnologie aperte non significa partecipare al loro sviluppo. Senza un ruolo nella governance dei progetti e nella definizione di runtime, protocolli e standard, si rimane esposti a decisioni assunte altrove.

I diversi livelli di apertura. Gli open-weight consentono utilizzo, adattamento e valutazione indipendente dei modelli; l'accesso anche a dati e procedure di addestramento amplia trasparenza e riproducibilità. Distinguere questi livelli, come propone la Open Source AI Definition del 2024, permette di valutare l'apertura sulla sostanza anziché sulle dichiarazioni.

Il rischio di etichetta. L'Europa non deve trasformare l'open intelligence in un marchio industriale, finanziando modelli propri mentre infrastrutture, inferenza e strumenti agentici restano controllati da realtà estere.

La dipendenza dall'offerta estera. L'accessibilità dei modelli può cambiare per decisioni commerciali o geopolitiche. Nel 2026 restrizioni statunitensi hanno temporaneamente limitato l'accesso ad alcuni modelli Anthropic; nel 2024 Meta ha imposto limitazioni europee ai modelli multimodali Llama 3.2. Il 16 ottobre 2026 anche la Cina ha proposto controlli sull'esportazione di tecnologie AI avanzate, inclusi i pesi di determinati modelli, pur escludendo quelli già pubblicati in open source e l'accesso tramite API. Situazioni differenti che mostrano come la disponibilità delle tecnologie e delle loro future evoluzioni possa dipendere da decisioni esterne.

La competizione esterna. I modelli open-weight cinesi e americani offrono opportunità concrete, ma la loro diffusione risponde anche a strategie industriali e geopolitiche.

L'interesse dei fornitori a espandere i propri ecosistemi non coincide necessariamente con quello europeo a preservare autonomia e reversibilità delle scelte.

6.6 Raccomandazioni strategiche

Tradurre la sovranità tecnologica in scelte operative non significa introdurre nuovi vincoli normativi o amministrativi, ma creare le condizioni affinché imprese e istituzioni possano conservare alternative tecnologiche effettive. La reversibilità delle scelte costituisce il principio guida di sette direttrici strategiche per l'Italia e per l'Europa.

1. Favorire la trasparenza delle condizioni tecnologiche. Promuovere criteri semplici e condivisi per distinguere i diversi livelli di apertura dei modelli, dai pesi al codice, dai dati agli strumenti di esecuzione, senza introdurre nuovi obblighi generalizzati. L'obiettivo è consentire a imprese e amministrazioni di valutare consapevolmente le soluzioni disponibili.
2. Assumere la reversibilità come criterio di scelta tecnologica. Negli acquisti pubblici e nelle strategie di investimento, valorizzare soluzioni che consentano di cambiare modelli, fornitori e infrastrutture preservando dati, applicazioni, competenze e continuità operativa. La verifica della portabilità dovrebbe essere proporzionata alla criticità dei sistemi, evitando requisiti e procedure che aumentino inutilmente i costi di adozione.
3. Rafforzare le capacità di inferenza. Sostenere lo sviluppo di competenze, infrastrutture e strumenti che consentano a imprese e amministrazioni di scegliere dove e come eseguire i modelli, valutandone prestazioni e condizioni di utilizzo. L'obiettivo è ampliare le alternative disponibili, non imporre specifiche soluzioni tecnologiche o localizzazioni geografiche.

4. Favorire l'adozione responsabile degli agenti nella pubblica amministrazione. Promuovere architetture interoperabili, strumenti di controllo degli accessi e soluzioni che consentano di verificare le operazioni compiute dai sistemi AI, valorizzando gli standard esistenti, evitando la duplicazione degli obblighi normativi e amministrativi.

5. Rafforzare la partecipazione italiana agli ecosistemi open source. Favorire il contributo di imprese, università e centri di ricerca ai progetti internazionali che sviluppino modelli, strumenti e standard aperti, utilizzando prioritariamente programmi e risorse già disponibili. La partecipazione alle comunità tecnologiche deve diventare uno strumento di influenza e acquisizione di competenze, non un nuovo adempimento.

6. Favorire l'interoperabilità tra i progetti. Collegare le iniziative europee, le architetture nazionali e i progetti della pubblica amministrazione, valorizzando il riuso delle tecnologie e delle competenze, evitando duplicazioni e favorendo la collaborazione tra soggetti pubblici e privati.

7. Sviluppare capacità tecnologiche e infrastrutturali strategiche. Concentrare gli investimenti europei nelle infrastrutture di calcolo, nella ricerca e nelle competenze industriali che possono ridurre le dipendenze più critiche, favorendo l'accesso condiviso alle risorse e la capacità di valutare autonomamente i modelli avanzati. L'obiettivo non è replicare ogni componente della filiera globale, ma preservare alternative tecnologiche credibili.

Conclusioni. L'apertura non basta

La storia dell'open source dimostra che l'apertura di una tecnologia può modificare profondamente gli equilibri di un mercato. L'accesso al codice, la possibilità di studiarlo, modificarlo e distribuirlo hanno favorito ecosistemi industriali che hanno trasformato l'informatica, riducendo le barriere all'ingresso e sottraendo componenti tecnologiche al controllo esclusivo di singoli fornitori. Ma la stessa storia insegna che l'apertura non elimina il potere economico: ne modifica la distribuzione, creando nuove opportunità e dipendenze.

Nell'intelligenza artificiale questa dinamica assume una rilevanza ancora maggiore. La diffusione dei modelli open-weight consente a imprese, università e amministrazioni di utilizzare e adattare tecnologie avanzate un tempo accessibili prevalentemente attraverso servizi proprietari. È un cambiamento significativo, che amplia le possibilità di scelta delle imprese, favorisce la concorrenza e accelera l'innovazione, contribuendo a una distribuzione più ampia delle capacità tecnologiche. La trasformazione del mercato dei modelli, tuttavia, non comporta automaticamente un'analoga redistribuzione del controllo sulle infrastrutture e sugli altri componenti della filiera.

La disponibilità dei pesi non garantisce infatti l'accesso ai dati di addestramento, la riproducibilità del modello o il controllo delle infrastrutture necessarie alla sua esecuzione.

Quando l'apertura si estende ai modelli, agli strumenti di inferenza e agli ambienti in cui operano gli agenti, può diventare una vera strategia industriale. Per sostenerla, tuttavia, servono laboratori, capacità di calcolo, semiconduttori e competenze. Per l'Italia e per l'Europa la posta non è semplicemente pro-

durre più modelli, ma decidere dove risiede l'intelligenza economica del Paese e chi può controllarla, modificarla o trasferirla.

La sovranità tecnologica dipende quindi dalla capacità di imprese e istituzioni di governare le tecnologie utilizzate, comprenderne le dipendenze e conservare alternative praticabili quando cambiano le condizioni economiche, contrattuali o geopolitiche.

Per l'Italia questa prospettiva assume un significato particolare. Come evidenziato nei precedenti lavori della serie Sovranità Tecnologica, il principale patrimonio competitivo nazionale risiede nelle competenze industriali, nei dati produttivi, nelle conoscenze ingegneristiche e nelle specializzazioni manifatturiere. L'intelligenza artificiale può amplificarne il valore, ma anche trasferire il controllo su processi e conoscenze verso piattaforme sviluppate e gestite altrove.

Il rischio non consiste nell'utilizzare tecnologie straniere, condizione inevitabile in una economia interdipendente, ma nel costruire applicazioni e processi dai quali diventi impossibile separarsi senza perdere dati, competenze, investimenti e continuità operativa. Questa distinzione permette di superare tanto l'illusione dell'autosufficienza nazionale quanto l'idea che l'accesso alle tecnologie globali renda superflua una politica industriale.

L'open source può offrire una risposta concreta, purché venga considerato non come un'etichetta, ma come parte di una strategia più ampia. Modelli modificabili, strumenti interoperabili, protocolli aperti e infrastrut-

ture controllabili possono ampliare la libertà di scelta delle imprese e delle amministrazioni, ma richiedono investimenti, competenze e condizioni contrattuali adeguate. Non basta poter scaricare un modello. Occorre saperlo adattare, integrare nei processi, valutarne le prestazioni, governarne i costi e sostituirlo quando necessario. Bisogna conoscere dove vengono elaborati i dati, chi controlla le infrastrutture, come operano gli agenti e quali conseguenze avrebbe l'interruzione di un servizio. Senza queste capacità, l'apertura rischia di tradursi in una disponibilità soltanto apparente, mentre le decisioni strategiche rimangono altrove.

È qui che la riflessione deve tradursi in politica industriale. Sostenere l'open source nell'AI non significa soltanto finanziare nuovi modelli o incentivarne l'adozione, ma creare le condizioni affinché imprese, centri di ricerca e amministrazioni sviluppino competenze proprie, contribuiscano agli ecosistemi tecnologici, controllino i dati strategici e dispongano di alternative credibili ai fornitori dominanti.

La domanda pubblica può svolgere un ruolo decisivo. Attraverso gli acquisti tecnologici, le istituzioni contribuiscono a determinare quali competenze industriali si sviluppino, quali standard si consolidano e quali dipendenze diventano strutturali. Analogamente, gli incentivi alle imprese dovrebbero essere valutati non soltanto per l'adozione delle tecnologie, ma anche per la capacità di conservarne il controllo nel tempo.

La spesa in intelligenza artificiale, infatti, non coincide automaticamente con un investimento in sovranità tecnologica. Può aumentare la produttività rafforzando contemporaneamente la dipendenza da piattaforme esterne. Per contribuire anche all'autonomia industriale deve generare competenze, infrastrutture, strumenti riutilizzabili e possibilità effettive di scelta.

Il criterio che tiene insieme queste dimensioni è quello della reversibilità. Una scelta tecnologica è tanto più sostenibile quanto più un'impresa o un'amministrazione può modificarla senza compromettere il patrimonio informativo e operativo costruito attorno a essa.

Sostituire un modello, cambiare infrastruttura, trasferire dati o ricostruire un servizio su componenti differenti non è soltanto una possibilità tecnica, ma una forma concreta di autonomia economica e strategica. La reversibilità non elimina le dipendenze né rende irrilevante la provenienza delle tecnologie. Permette però di distinguere le interdipendenze governabili da quelle destinate a trasformarsi in vincoli permanenti.

È su questa distinzione che dovrebbe misurarsi una politica della sovranità tecnologica capace di andare oltre le dichiarazioni di principio.

L'apertura dei modelli rappresenta dunque un'opportunità importante, ma il suo valore per l'Italia dipenderà dalle capacità industriali costruite attorno a essa. Non si tratta di competere simbolicamente sul numero di modelli nazionali né di rinunciare ai vantaggi delle tecnologie sviluppate altrove, ma di garantire che il loro utilizzo non comporti la perdita della possibilità di decidere come adottarle, adattarle e sostituirle.

La sovranità, in ultima analisi, non si misura da ciò che si possiede, ma da ciò che si può cambiare senza pagare un prezzo insostenibile. È un criterio meno solenne dei proclami, ma è quello che resiste alla prova dei contratti, dei costi e del tempo.

Fonti e documenti

Parte I - Le radici dell'open source nel software

Free Software Foundation / GNU Project. What is Free Software? Definizione delle quattro libertà fondamentali del software libero e dei principi elaborati da Richard Stallman.

<https://www.gnu.org/philosophy/free-sw.en.html>

Open Source Initiative. The Open Source Definition. Definizione dei criteri che qualificano il software open source, comprese le condizioni relative alla distribuzione, alle modifiche e alle opere derivate.

<https://opensource.org/osd>

Raymond, Eric S. The Cathedral and the Bazaar, 1999. Riflessione sul modello di sviluppo collaborativo del software e sulle differenze rispetto ai processi proprietari.

<https://www.catb.org/~esr/writings/cathedral-bazaar/>

Parte II — Deep learning, community AI e licenze

Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks, 2012. Pubblicazione fondamentale per la diffusione del deep learning applicato al riconoscimento delle immagini.

<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>

Vaswani, Ashish et al. Attention Is All You Need, 2017. Pubblicazione scientifica che introduce l'architettura Transformer.

<https://arxiv.org/abs/1706.03762>

Hugging Face. Transformers Documentation. Documentazione dell'ecosistema di strumenti e librerie per la distribuzione, l'adattamento e l'utilizzo dei modelli.

<https://huggingface.co/docs/transformers/index>

EleutherAI. Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling, 2023. Ricerca sulla trasparenza dei processi di addestramento e sulla riproducibilità dei modelli linguistici.

<https://arxiv.org/abs/2304.01373>

Schuhmann, Christoph et al. LAION-5B: An Open Large-Scale Dataset for Training Next Generation Image-Text Models, 2022. Documentazione del dataset aperto utilizzato nella ricerca sui modelli generativi multimodali.

<https://arxiv.org/abs/2210.08402>

Open Source Initiative. The Open Source AI Definition, versione 1.0, 28 ottobre 2024. Definizione delle condizioni necessarie per qualificare un sistema AI come open source, con particolare riferimento a codice, parametri e informazioni sui dati di addestramento.

<https://opensource.org/ai/open-source-ai-definition>

Open Source Initiative. Open Source AI Definition - Frequently Asked Questions. Approfondimento sulle differenze tra open source AI e semplice disponibilità dei pesi.

<https://opensource.org/ai/faq>

Meta. Llama 3.2 Community License Agreement. Testo contrattuale di riferimento per valutare le condizioni di utilizzo e redistribuzione di una specifica generazione di modelli Llama.

https://www.llama.com/llama3_2/license/

Parte III - La divisione aperto/chiuso: il caso OpenAI

OpenAI. Introducing OpenAI, 11 dicembre 2015. Documento fondativo che descrive missione e impostazione originaria dell'organizzazione.

<https://openai.com/index/introducing-openai/>

OpenAI. OpenAI LP, 11 marzo 2019. Presentazione della struttura capped-profit e delle motivazioni economiche legate alla raccolta di capitale e all'accesso alle infrastrutture di calcolo.

<https://openai.com/index/openai-lp/>

OpenAI. GPT-4 Technical Report, 2023. Documentazione tecnica che illustra capacità, valutazioni e scelta di non divulgare alcuni dettagli del modello, richiamando considerazioni competitive e di sicurezza.

<https://arxiv.org/abs/2303.08774>

OpenAI. GPT-4 System Card, 2023. Analisi dei rischi, delle valutazioni e delle misure di mitigazione associate al rilascio del modello.

<https://cdn.openai.com/papers/gpt-4-system-card.pdf>

Parte IV - Geopolitica dei modelli aperti: Stati Uniti e Cina

Meta. Introducing Meta Llama 3: The Most Capable Openly Available LLM to Date, 18 aprile 2024. Presentazione della strategia di diffusione dei modelli Llama e del relativo ecosistema.

<https://ai.meta.com/blog/meta-ai-llama-3/>

DeepSeek-AI et al. DeepSeek-V3 Technical Report, 2024. Descrizione dell'architettura, dell'addestramento e delle prestazioni dichiarate di DeepSeek-V3, inclusi i dati relativi alle risorse computazionali impiegate.

<https://arxiv.org/abs/2412.19437>

DeepSeek-AI et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, 2025. Documentazione delle tecniche di reinforcement learning e delle capacità di ragionamento della famiglia DeepSeek-R1.

<https://arxiv.org/abs/2501.12948>

Qwen Team / Alibaba Cloud. Qwen2.5 Technical Report, 2024–2025. Documentazione tecnica della famiglia Qwen, delle sue caratteristiche e delle valutazioni pubblicate dagli sviluppatori.

<https://arxiv.org/abs/2412.15115>

Parte V - Nvidia, Nemotron e il controllo dell'infrastruttura

NVIDIA. NVIDIA Launches Family of Open Reasoning AI Models for Developers and Enterprises to Build Agentic AI Platforms, 18 marzo 2025. Presentazione della famiglia Llama Nemotron e del suo posizionamento nell'ecosistema dei modelli aperti.

<https://nvidianews.nvidia.com/news/nvidia-launches-family-of-open-reasoning-ai-models-for-developers-and-enterprises-to-build-agentic-ai-platforms>

NVIDIA. NVIDIA Nemotron Open Model License. Condizioni contrattuali applicabili ai modelli distribuiti sotto questa specifica licenza.

<https://www.nvidia.com/en-us/agreements/enterprise-software/nvidia-nemotron-open-model-license/>

[Comunicato ufficiale del Ministero del commercio cinese del 16 ottobre 2026 e avviso ministeriale n. 48/2026](#)

NVIDIA. CUDA Toolkit Documentation. Documentazione ufficiale dell'ecosistema software per il calcolo accelerato, rilevante per l'analisi delle dipendenze tra hardware, strumenti di sviluppo e infrastrutture AI. <https://docs.nvidia.com/cuda/>

Parte VI - L'Europa e l'Italia: politiche, infrastrutture e iniziative industriali

Commissione europea. Communication on European Tech Sovereignty, accompanied by an EU Open Source Strategy, COM(2026) 503, 3 giugno 2026. Documento programmatico sul rafforzamento dell'autonomia tecnologica europea e sul ruolo dell'open source nella filiera digitale. <https://digital-strategy.ec.europa.eu/en/library/communication-european-tech-sovereignty-accompanied-eu-open-source-strategy>

Commissione europea. The EU Open Source Strategy, 2026. Presentazione delle priorità europee in materia di sviluppo, adozione, interoperabilità e sostenibilità degli ecosistemi aperti. <https://digital-strategy.ec.europa.eu/en/policies/open-source-strategy>

Commissione europea. Commission Selects EUROPA Consortium as the Winner of the Frontier AI Grand Challenge, 19 giugno 2026. Comunicazione ufficiale sull'assegnazione del progetto al consorzio EUROPA guidato da Domyn e sull'obiettivo di sviluppare un modello europeo aperto nelle 24 lingue ufficiali dell'Unione. <https://digital-strategy.ec.europa.eu/en/news/commission-selects-europa-consortium-winner-frontier-ai-grand-challenge-project-build-european-open>

Eclipse Foundation. Eclipse Foundation Launches the Sovereign AI Foundation, 30 settembre 2026. Annuncio della costituzione dell'iniziativa dedicata alla collaborazione e alla governance di progetti di AI sovrana. <https://newsroom.eclipse.org/news/announcements/eclipse-foundation-launches-sovereign-ai-foundation>

Engineering. IS-IA: è IA, ma è sovrana. Engineering lancia l'architettura italiana per un'intelligenza artificiale governabile, aperta e sostenibile, 20 aprile 2026. Presentazione dell'architettura IS-IA e del modello EngGPT 2. <https://www.eng.it/it/news/press-releases/2026/04/is-ia-sovrana-engineering-lancia-architettura-italiana-per-intelligenza-artificiale-governabile-aperta-sostenibile>

Sapienza NLP / FAIR / CINECA. Minerva — Italian Large Language Models. Documentazione del progetto di ricerca sui modelli linguistici per l'italiano. <https://minerva-ai.org/it>

EuroHPC Joint Undertaking. EuroHPC JU Signs Contract to Boost AI Capabilities with IT4Lia AI Factory, 22 aprile 2026. Comunicazione istituzionale sul potenziamento dell'infrastruttura italiana per l'intelligenza artificiale. https://www.eurohpc-ju.europa.eu/eurohpc-ju-signs-contract-boost-ai-capabilities-it4lia-ai-factory-2026-04-22_en

INPS. Determinazione n. 139 del 30 marzo 2026. Documentazione programmatica comprendente il progetto AI Welfare, relativo allo sviluppo di un modello linguistico per il welfare pubblico italiano. https://servizi2.inps.it/servizi/ProvvedimentiFE/provvedimentiDirettoreGenerale/DownloadFile/5399?nomefile=DG2026_139_firmato.pdf

Anthropic. Redeploying Fable 5, 30 giugno 2026, aggiornato il 1° luglio. Ricostruzione delle restrizioni all'accesso ai modelli Fable 5 e Mythos 5 e delle successive condizioni di ripristino. <https://www.anthropic.com/news/redeploying-fable-5>

Nota metodologica

Il presente Paper analizza l'evoluzione dell'open source nell'intelligenza artificiale e le sue implicazioni economiche, industriali e geopolitiche, con particolare attenzione all'Europa e all'Italia. L'obiettivo non è classificare i modelli disponibili né stabilire la superiorità delle tecnologie aperte rispetto a quelle proprietarie, ma individuare le condizioni in cui l'apertura può favorire l'autonomia tecnologica.

L'analisi combina ricostruzione storica, strategie dei principali operatori e valutazione delle dipendenze lungo la filiera AI. Le fonti comprendono documenti istituzionali, studi scientifici, documentazione tecnica, comunicazioni aziendali e analisi indipendenti. Le dichiarazioni dei produttori su prestazioni, costi e benefici sono attribuite alle rispettive fonti e non considerate automaticamente verificate.

Il termine open source è utilizzato nel significato consolidato per il software e, per l'AI, in riferimento alla Open Source AI Definition dell'OSI (2024). Open-weight indica modelli con pesi disponibili secondo specifiche licenze, senza implicare necessariamente l'accesso ai dati e ai processi di addestramento o la piena riproducibilità del sistema.

La sovranità tecnologica è intesa come capacità di governare le dipendenze e preservare alternative praticabili, non come autosufficienza nazionale. Il criterio della reversibilità indica la possibilità di modificare o sostituire tecnologie e fornitori senza compromettere dati, competenze e continuità operativa; costituisce un principio interpretativo, non un indicatore quantitativo standardizzato.

Lo schema costruire, servire e far agire distingue i livelli di dipendenza relativi a modelli, infrastrutture di inferenza, dati e sistemi agentici, senza presumere che la loro apertura garantisca automaticamente sicurezza, interoperabilità o autonomia.

Il confronto tra Stati Uniti, Cina ed Europa è qualitativo e basato su casi selezionati, non statisticamente rappresentativi. Per le iniziative europee e italiane si distinguono, ove possibile, annunci, progetti in sviluppo e capacità effettivamente operative.

Le raccomandazioni rappresentano proposte di indirizzo, non valutazioni economiche o giuridiche validate, e richiedono ulteriori verifiche di fattibilità e compatibilità normativa. L'analisi riflette le informazioni disponibili alla chiusura editoriale. Data la rapida evoluzione del settore, alcuni elementi potrebbero cambiare. La bibliografia finale documenta le fonti utilizzate e consente di distinguerle dalle interpretazioni di Rivista.AI.

Disclaimer

Questa pubblicazione è stata realizzata da Rivista.AI con finalità di analisi, ricerca e posizionamento strategico e intende contribuire al dibattito sull'evoluzione dell'intelligenza artificiale e sui suoi effetti economici, industriali e tecnologici.

Le informazioni e le analisi contenute nel documento si basano su fonti pubbliche, dati istituzionali, letteratura scientifica, report di ricerca e documentazione disponibile alla data di chiusura editoriale. Pur essendo state sottoposte a verifica, tali informazioni possono essere soggette ad aggiornamenti, revisioni o differenti interpretazioni.

Il documento non costituisce parere legale o professionale. Le valutazioni e gli scenari presentati hanno finalità esclusivamente analitiche e non devono essere interpretati come previsioni certe di evoluzioni future.

Eventuali riferimenti ad aziende, piattaforme, prodotti o servizi sono utilizzati a fini esemplificativi e di analisi dell'evoluzione del mercato e non costituiscono endorsement, raccomandazioni né giudizi di merito sui singoli operatori.

Le interpretazioni e le conclusioni espresse nel Paper sono responsabilità editoriale di Rivista.AI.

Nota sull'utilizzo dell'intelligenza artificiale

Nella realizzazione di questo Paper sono stati utilizzati strumenti di intelligenza artificiale generativa come supporto alla realizzazione di elementi grafici come immagini e infografiche.

Rivista.AI, elenco Paper Serie “Sovranità Tecnologica”

“Competitività digitale dell'Unione Europea: semiconduttori, infrastrutture cloud e AI nel contesto geopolitico 2026” - Vol. 01 | giugno 2026

“La sovranità tecnologica nell'era dell'AI: la scommessa francese” - Vol. 02 | luglio 2026

“La sovranità tecnologica europea nell'era dell'intelligenza artificiale. Proposta operativa”
- Vol. 03 | luglio 2026

“Dal modello alla fabbrica: perché la sovranità tecnologica è la sfida decisiva per l'industria italiana” - Vol. 04 | luglio 2026

“La fabbrica invisibile: come l'intelligenza artificiale sta riscrivendo la visibilità dell'industria italiana nel mondo” - Vol. 05 | settembre 2026

“L'effetto mosaico. Come l'intelligenza artificiale ricostruisce il know-how delle imprese e ridisegna i confini della sovranità” - Vol 06 | ottobre 2026