

# Large Language Models: problemi concorrenziali e dinamiche di mercato

di Rossana Arcano, Carlo Cambini, Paolo Lupi,  
Antonio Manganeli, Antonio Perrucci, Giovanni Trotta

SOMMARIO: 1. Big tech e dinamiche competitive. – 2. Problemi concorrenziali a monte e/o a valle della catena del valore. – 3. Il mercato delle applicazioni AI generativa e i FM. – 4. Misure di politica della concorrenza.

## 1. *Big Tech e dinamiche competitive*

La questione più critica e urgente che domina il dibattito di *policy* circa gli assetti della concorrenza nella filiera dell'Intelligenza Artificiale (IA) riguarda l'eventualità che gli *incumbent* digitali, già detentori di un significativo potere di mercato in uno o più segmenti/mercati della catena del valore dell'IA, possano sfruttare strategicamente tale posizione per consolidare ed estendere ulteriormente la loro influenza nel mercato emergente dei *Foundation Models (FM)*. Questo rischio è accentuato dalla natura duale dei FM, che possono essere utilizzati sia come prodotti autonomi che come input per lo sviluppo di applicazioni *downstream*, moltiplicando le occasioni di controllo verticale.

In linea di principio, ciò potrebbe avvenire a più livelli della catena del valore e attraverso una serie di pratiche, tra cui l'integrazione verticale ma anche attraverso *partnership* (esclusive). Il dibattito sulla politica e sul diritto della concorrenza si sofferma sulle grandi imprese tecnologiche (Big Tech), poiché esse operano su più livelli della catena del valore dell'IA. Infatti, come visto nei precedenti capitoli, (i) tutte le Big Tech sono attive o cercano di sviluppare *chip* per l'addestramento dei modelli di IA<sup>1</sup>, anche se il principale fornitore di GPU utilizzate per l'IA è attualmente NVIDIA

---

<sup>1</sup> Oltre a Intel e AMD, Amazon, Apple, Google, Meta e Microsoft stanno sviluppando attivamente *chip* proprietari per l'addestramento dei modelli di intelligenza artificiale. Google è in prima linea

(che comunque vende la maggior parte della propria produzione alle Big Tech); (ii) tre Big Tech (Microsoft, Amazon e Google) sono i principali fornitori di *cloud* nell'UE e in tutto il mondo, con una quota complessiva di circa i 2/3 del fatturato globale, (iii) tutte le Big Tech hanno accesso a un'enorme quantità di dati; e, naturalmente, (iv) tutte sono attive e occupano posizioni di primo piano a livello delle applicazioni *consumer*. Si tratta di una sovrapposizione molto ampia e forse senza precedenti tra diversi livelli della catena del valore e dell'infrastruttura tecnologica, che rende particolarmente rilevante il tema delle possibili estensioni del potere di mercato tra attività infrastrutturali e attività nei mercati applicativi.

Al quadro appena delineato, si aggiunga che nel contesto dello sviluppo dei FM le startup native dell'AI, come OpenAI e Anthropic, hanno ottenuto un vantaggio competitivo iniziale rispetto alle Big Tech, ma ora tutte le Big Tech (v) stanno sviluppando propri modelli di FM e/o hanno stabilito importanti *partnership* con le stesse OpenAI e Anthropic, come nel caso di Microsoft e OpenAI<sup>2</sup>; Google e Anthropic<sup>3</sup>, e Amazon e Anthropic<sup>4</sup>.

Il tema delle *partnership* merita un approfondimento per il loro impatto sullo sviluppo del mercato. Da un lato, le *partnership* possono offrire vantaggi sostanziali alle parti coinvolte, favorendo una maggiore innovazione ed efficienza: gli sviluppatori

---

con le sue *Tensor Processing Unit (TPU)* che fungono da alternativa alle GPU di NVIDIA per l'addestramento dei suoi FM, tra cui Gemini. Inoltre, Qualcomm sta sviluppando *chipset* mobili specializzati nell'IA sufficientemente potenti per eseguire gli FM *on-device*.

<sup>2</sup> Secondo tale accordo: (i) Microsoft ha investito oltre 13 miliardi di dollari in OpenAI; (ii) i modelli GPT di OpenAI (incluso GPT-4) sono disponibili esclusivamente tramite Microsoft Azure; (iii) Microsoft integra i modelli di OpenAI nei suoi prodotti (ad esempio, Copilot in Microsoft 365, GitHub e Windows).

A questo riguardo, nell'aprile 2025, la CMA ha concluso la sua indagine sulla *partnership* tra Microsoft e OpenAI, concentrandosi sui crescenti investimenti e accordi commerciali di Microsoft con OpenAI. La CMA ha concluso che non vi sono prove evidenti che tali dinamiche abbiano portato a un controllo effettivo di Microsoft su OpenAI qualificabile come fusione ai sensi dell'articolo 22, paragrafo 1, dell'*Enterprise Act 2002*. Tuttavia, la CMA ha anche sottolineato che questa decisione non preclude future indagini. Si veda, CMA (2025) *Microsoft Corporation's partnership with OpenAI, Inc. Decision on relevant merger situation*.

<sup>3</sup> Secondo tale accordo: (i) Google ha investito oltre 2 miliardi di dollari in Anthropic (Claude AI); (ii) Anthropic utilizza l'infrastruttura AI di Google Cloud per addestrare ed eseguire i propri modelli; (iii) Google integra Claude nei propri servizi; (iv) i modelli di Anthropic sono disponibili tramite Vertex AI di Google Cloud (servizio di *marketplace* Google AI).

<sup>4</sup> Secondo tale accordo: (i) Amazon si è impegnata a investire 4 miliardi di dollari in Anthropic nel 2023; (ii) i modelli Claude di Anthropic sono ospitati e ottimizzati per AWS Bedrock (il servizio di *marketplace* AI di Amazon); (iii) i *chip* AI di Amazon (Trainium e Inferentia) sono utilizzati per l'addestramento dei modelli di Anthropic; (iv) i clienti AWS ottengono un accesso anticipato ai modelli Claude.

di FM beneficiano delle risorse computazionali, dei finanziamenti e delle infrastrutture delle grandi aziende tecnologiche, mentre queste ultime si assicurano l'accesso anticipato, l'esclusività o l'integrazione di modelli di IA all'avanguardia nei loro ecosistemi<sup>5</sup>. Dall'altro lato, tuttavia, tali collaborazioni presentano anche rischi competitivi, in quanto le aziende consolidate potrebbero sfruttare le *partnership* e gli investimenti per limitare la concorrenza potenziale<sup>6</sup>. Tali dinamiche potrebbero configurare anche fenomeni di *foreclosure* dell'innovazione, disincentivando l'emergere di nuovi entranti e startup innovative che non riescono a ottenere condizioni eque di accesso alle risorse computazionali.

In questo contesto, l'Autorità britannica per la concorrenza e i mercati (CMA) ha già espresso preoccupazione per il rischio che le Big Tech possano limitare la concorrenza tra i FM, sostenendo che «*the growing presence across the foundation models value chain of a small number of incumbent technology firms, which already hold positions of market power in many of today's most important digital markets, could profoundly shape these new markets to the detriment of fair, open and effective competition, ultimately harming businesses and consumers, for example by reducing choice and quality and increasing price*»<sup>7</sup>. L'Autorità per la concorrenza francese (Autorité de la Concurrence) ha fatto eco a queste preoccupazioni, a causa del fatto che «*major digital companies enjoy preferential access to the inputs needed to train and develop foundation models. [...] The vertical integration of certain digital operators and their service ecosystems may give rise to a number of abusive practices*»<sup>8</sup>.

---

<sup>5</sup> Si veda, ad esempio, D. Spulber (2024), *Antitrust and Innovation Competition: Investments and Partnerships in Artificial Intelligence*, in *Artificial Intelligence and Competition Policy*, a cura di T. Schrepel, A. Abbott, dove l'autore suggerisce che queste collaborazioni potrebbero anche fungere da alternative innovative alla completa integrazione verticale.

<sup>6</sup> Queste dinamiche sono descritte anche in un recente rapporto della Federal Trade Commission Office of Technology staff che analizza le *partnership* citate. Cfr. FTC (2025), *Partnerships Between Cloud Service Providers and AI Developers Staff Report on AI Partnerships & Investments 6(b) Study*, 2025. Il rapporto ha rilevato che i grandi fornitori di servizi *cloud* spesso impongono requisiti di spesa per il *cloud* che vincolano gli sviluppatori a ecosistemi specifici, aumentando i costi di passaggio e le barriere all'ingresso, rafforzando così le loro posizioni di potere esistenti sia nei mercati a monte che a valle.

<sup>7</sup> Si veda, CMA, *AI Foundation Models: Technical update report*, 2024.

<sup>8</sup> Autorité de la concurrence (2024), *Parere 24-A-05 sul funzionamento concorrenziale del settore dell'intelligenza artificiale generativa*.

## **2. Problemi concorrenziali a monte e/o a valle della catena del valore**

In tutta la catena del valore riguardante i cosiddetti modelli fondativi, le strategie potenzialmente anticoncorrenziali possono verificarsi sia a monte sia a valle. Finora, le discussioni sulla politica di concorrenza si sono concentrate principalmente sui mercati a monte, dove esistono posizioni di potere di mercato consolidate, che possono essere sfruttate in modo chiaro e diretto attraverso lo *stack* dell'IA, riverberandosi quindi sui mercati dei FM.

Infatti, il potere di mercato a monte, così come eventuali strategie anticoncorrenziali, sono legate al controllo di input produttivi critici per lo sviluppo dei FM: hardware e risorse informatiche, *cloud* e dati. Gli *incumbent* digitali potrebbero quindi limitare la disponibilità e l'accesso a tali input, garantendo un vantaggio competitivo al proprio FM (se sviluppato) o inducendo altri FM a stabilire *partnership* (esclusive) con loro, limitando così la concorrenza.

Il rischio concorrenziale dipende infatti sia (a) dalla presenza delle Big Tech nella maggior parte dei segmenti della catena del valore dell'IA, sia (b) dalla posizione di mercato preminente di cui godono in alcuni stadi della filiera, controllando input critici per i mercati dei FM. Se si va più in dettaglio, nel livello hardware, come già detto, appare chiaro che né le Big Tech né i fornitori di FM occupano una posizione di rilievo nel mercato della produzione di *chip* IA. Un'altra azienda tecnologica, NVIDIA, ha ottenuto un vantaggio del *first mover*, raggiungendo un'elevata importanza e potere di mercato. Più precisamente, NVIDIA attualmente detiene una posizione dominante, poiché la sua GPU per l'addestramento di modelli di intelligenza artificiale ha una quota di mercato superiore al 90%. In questo contesto, il diritto della concorrenza già dispone di strumenti analitici adeguati e di teorie del danno consolidate per sanzionare ed inibire eventuale abuso di posizione dominante.

Peraltro, oltre all'hardware, NVIDIA ha costruito la propria posizione di preminenza e rallentato la concorrenza di altri produttori di *chip* (quali AMD, Intel e gli stessi *hyperscaler*) anche attraverso il suo *framework software* per l'utilizzo delle GPU che consente agli sviluppatori di sfruttare le capacità di elaborazione in parallelo per l'elaborazione generica: un vantaggio fondamentale per le applicazioni di intelligenza artificiale e apprendimento automatico da parte degli sviluppatori. CUDA (Compute Unified Device Architecture), il software proprietario di NVIDIA, è stato progettato specificamente per funzionare in modo perfettamente integrato con l'architettura hardware NVIDIA<sup>9</sup> e vista la sua quota di mercato ed il vantaggio del *first*

---

<sup>9</sup> NVIDIA supporta CUDA con un ricco set di strumenti, librerie e *framework* come cuDNN per il *deep learning*, cuBLAS per l'algebra lineare e TensorRT per l'inferenza di *deep learning* ad alte

*mover* è diventato di fatto lo standard applicativo per l'utilizzo delle GPU<sup>10</sup>. Questa dinamica ha generato potenti effetti di rete<sup>11</sup> e sostanziali barriere all'ingresso per i concorrenti di NVIDIA nel mercato dei *chip* IA.

Per quanto concerne i servizi *cloud*, a parte le grandi aziende digitali e una manciata di aziende con grandi *data centers* interni, come Meta e Samsung, i servizi *cloud* sono il principale mezzo di accesso alla potenza di calcolo necessaria per l'addestramento dei modelli di IA. I Cloud Service Providers (CSP) forniscono agli sviluppatori un'infrastruttura ottimizzata per l'IA e servizi di piattaforma su misura per le loro esigenze, eliminando la necessità di un elevato investimento iniziale nell'infrastruttura IT. Peraltro, il *cloud* funge anche da canale chiave per la distribuzione e l'implementazione dei FM a valle.

Come già richiamato, i tre maggiori fornitori di servizi *cloud* a livello mondiale sono Amazon AWS, Microsoft Azure e Google Cloud Platform, comunemente definiti *hyperscaler*, che sono presenti a tutti i livelli della catena del valore dei servizi *cloud* (IaaS, PaaS, SaaS)<sup>12</sup> e offrono un'ampia varietà di servizi *cloud* su larga scala<sup>13</sup>.

Per quanto riguarda l'ecosistema dell'IA, la preoccupazione maggiore è che i fornitori di servizi *cloud*, grazie alla loro scala, possano sfruttare il potere di mercato detenuto in tale mercato per estenderlo (effetto *leveraging* nella terminologia anglosassone) su altri livelli dell'ecosistema dell'IA.

---

prestazioni. Questo ecosistema ha reso CUDA una parte indispensabile del flusso di lavoro dell'IA, garantendo che il software scritto su CUDA funzioni alle prestazioni ottimali solamente sull'hardware NVIDIA.

<sup>10</sup> *Framework* di programmazione *deep learning* molto popolari come TensorFlow e PyTorch sfruttano CUDA per l'accelerazione GPU. Sia TensorFlow che PyTorch sono *framework* di alto livello che astraggono dalla complessità della programmazione GPU. Quindi, sebbene supportino ed offrano una vasta gamma di piattaforme hardware (tra cui anche AMD e Intel), sono ottimizzati per funzionare su GPU NVIDIA con CUDA per garantire le massime prestazioni.

<sup>11</sup> Sebbene CUDA sia un software proprietario di NVIDIA, l'azienda ha creato un ambiente aperto *open source* in cui sviluppatori, ricercatori e aziende possono contribuire liberamente, condividere conoscenze e ottimizzare le proprie applicazioni. Man mano che sempre più sviluppatori adottano CUDA, l'intero ecosistema diventa più ricco, attraendo a sua volta più sviluppatori.

<sup>12</sup> *Infrastructure-as-a-Service (IaaS)* è il livello base di servizio che comprende l'accesso all'infrastruttura IT. *Platform-as-a-Service (PaaS)* rappresenta un livello intermedio nella catena del valore in cui, rispetto alla configurazione precedente, viene aggiunto il middleware. *Software-as-a-Service (SaaS)* indica il livello più alto della catena del valore e consiste in applicazioni specifiche a cui gli utenti accedono tipicamente tramite un browser web.

<sup>13</sup> Nel 2024, si stima che Amazon AWS e Microsoft Azure avessero quote di mercato rispettivamente del 32% e del 23% nell'Unione Europea, seguiti da Cloud Platform con il 12%.

I dati, come descritto, svolgono un ruolo cruciale nella fase di addestramento, che è richiesta una sola volta per ogni modello, ma può creare colli di bottiglia significativi. Il volume di dati per l'addestramento è un fattore chiave per lo sviluppo di modelli ad alte prestazioni, e la maggior parte di questi dati proviene da fonti disponibili pubblicamente, come le informazioni raccolte sul web. Tuttavia, man mano che i modelli continuano a crescere, il loro addestramento richiede quantità sempre maggiori di dati, sollevando il timore che le fonti pubblicamente accessibili possano diventare insufficienti in un futuro probabilmente prossimo. Questo potrebbe rendere essenziali per le prestazioni dei FM i set di dati proprietari che possono essere controllati da un numero ristretto di attori con posizioni di preminenza nei servizi digitali agli utenti finali. Di conseguenza, i FM sviluppati dalle grandi aziende tecnologiche potrebbero ottenere un vantaggio competitivo grazie all'accesso esclusivo a dati per l'addestramento su larga scala. Ad esempio, Google, Apple e Microsoft, in linea di principio, potrebbero utilizzare tutti i dati generati all'interno dei loro ecosistemi di applicazioni.

I problemi concorrenziali sopra identificati sono reali e concreti e si basano chiaramente sull'ipotesi (derivante dalle esperienze nei confronti delle grandissime piattaforme digitali) che tutti i mercati tecnologici possano rapidamente tendere verso un assetto del tipo “*winner takes all/most*”.

In effetti, le autorità garanti della concorrenza vorrebbero applicare le “lezioni apprese negli ultimi 10-15 anni”<sup>14</sup>, quando nella maggior parte delle circostanze non hanno agito abbastanza velocemente e/o in modo adeguato a evitare che le Big Tech costruissero posizioni consolidate di potere sul mercato.

Pertanto, la domanda principale è se e in quali circostanze le dinamiche competitive all'interno della catena del valore dell'IA potrebbero essere simili a quelle che si sono verificate nel web 2.0, dove le piattaforme digitali molto grandi hanno raggiunto *tipping point* e hanno individualmente creato situazioni di dominio radicato in mercati digitali adiacenti, proteggendosi dalla concorrenza inter-piattaforma.

Per rispondere a questi interrogativi su un piano più generale, sembra utile ricordare che la lezione del passato indica come, difficilmente, il potere di mercato in mercati altamente innovativi possa derivare e radicarsi in modo duraturo grazie al controllo di un input tecnologico a monte, qualora si sia in presenza di dinamiche di innovazione molto forti. Del resto, il caso DeepSeek ha offerto una dimostrazione concreta che smentisce la tesi secondo cui l'accesso a risorse computazionali estese e

---

<sup>14</sup> Osservazioni di Sarah Cardell, CEO della CMA, pronunciate durante il 72o Antitrust Law Spring Meeting, Washington DC, USA, 2024.

i relativi costi costituiscano una barriera all'ingresso insormontabile nel mercato dei FM<sup>15</sup>. Il caso DeepSeek suggerisce che anche attori emergenti, in presenza di innovazioni architetturali e accesso efficiente a infrastrutture pubbliche o condivise, possano entrare nel mercato con modelli competitivi.

Una conferma della vivacità, sia tecnologica sia concorrenziale, del mercato dei FM è il rapido calo dei costi di inferenza, ossia i costi che si devono sostenere per interrogare un modello già addestrato e che sono misurati nei *Large Language Model* (LLM), ma metriche analoghe esistono anche per i modelli multimodali in dollari per milione di *token*<sup>16</sup>. Secondo l'*AI Index Report 2025*, i costi di inferenza a parità di prestazioni sono diminuiti drasticamente: per un modello con prestazioni equivalenti a GPT-3.5, il modello lanciato da OpenAI nel novembre 2022 e su cui si basa la prima versione di ChatGPT, i costi sono scesi da 20 dollari per milione di *token* nel novembre 2022, a soli 0,07 dollari nell'ottobre 2024, con una riduzione di oltre 280 volte in circa 18 mesi<sup>17</sup>.

Va notato che i modelli all'avanguardia sviluppati da aziende come OpenAI, Meta e Anthropic mantengono prezzi più elevati rispetto alle alternative meno avanzate. Questo riflette un posizionamento di mercato differente per i modelli dalle prestazioni superiori, sebbene la velocità con cui i costi si stanno riducendo anche in questo segmento suggerisca che anch'esso è soggetto a forti pressioni competitive.

In aggiunta, si osserva una tendenza crescente verso lo sviluppo di modelli più piccoli ma comunque efficienti che richiedono meno risorse per l'addestramento e per l'impiego, sia in termini di dati sia di potenza di calcolo<sup>18</sup>. Questi modelli, pur fornendo prestazioni generalmente inferiori rispetto ai LLM, in specifici contesti applicativi possono raggiungere risultati competitivi con costi operativi significativamente ridotti.

---

<sup>15</sup> Con l'avvertenza che le affermazioni della startup cinese DeepSeek sui suoi costi non sono state verificate e sono state accolte con un certo scetticismo. Per esempio, Demis Hassabis, CEO di Google DeepMind, ha definito le affermazioni sui costi di DeepSeek «esagerate e un po' fuorvianti». Altri critici sostengono che DeepSeek potrebbe aver perfezionato le tecniche esistenti piuttosto che introdurre metodologie fondamentalmente nuove, che avrebbero un impatto sulla contabilità dei costi.

<sup>16</sup> I *token*, nell'ambito degli LLM sono fondamentalmente le unità di elaborazione del testo, sorta di "atomi" linguistici che il modello utilizza per "comprendere" e generare il testo. Un *token* può essere una parola intera, parte di una parola, un segno di punteggiatura o un simbolo. In linea generale si può stimare il numero di *token* di un testo (sia inglese, sia italiano) moltiplicando il numero di parole per 1,3.

<sup>17</sup> Analoghe riduzioni dei prezzi si riscontrano per modelli capaci di superare *benchmark* più complessi. Secondo le stime di Epoch AI, a seconda delle applicazioni specifiche, i costi di inferenza degli LLM stanno diminuendo a un tasso compreso tra 9 e 900 volte all'anno.

<sup>18</sup> Esempi di FM di dimensioni minori rilasciati di recente sono Gemma 7B di Google, Zephyr 7B di Hugging Face e il "piccolo modello linguistico" Phi-2 di Microsoft.

Peraltro, il passaggio a FM di dimensioni più ridotte sta consentendo l'implementazione di FM a livello locale sui dispositivi (*on-device computing*), tra cui PC, smartphone, dispositivi IoT e sistemi XR<sup>19</sup>, riducendo così la dipendenza da infrastrutture di elaborazione basate su cloud. Questa trasformazione è ulteriormente guidata dalle innovazioni nei chipset specializzati “*AI on the edge*” di aziende come Qualcomm e Intel. In risposta, i produttori di dispositivi hanno iniziato a commercializzare dispositivi dotati di intelligenza artificiale che integrano questi *chip*, aprendo la strada ad applicazioni più indipendenti ed efficienti.

Inoltre, è fondamentale considerare che una quota sostanziale delle vendite di GPU di NVIDIA è specificatamente destinata agli *hyperscaler*, il che conferisce a questi grandi fornitori di *cloud* un significativo contropotere d'acquisto. Peraltro, in ragione del fatto che gli *hyperscaler* continuano a sviluppare i propri *chip*, la loro dipendenza da NVIDIA dovrebbe diminuire nel tempo. Se iniziassero a fornire componenti hardware ad altri sviluppatori di FM, ciò potrebbe introdurre una maggiore pressione competitiva nel mercato. Infatti, riconoscendo i rischi di affidarsi a un unico fornitore di input dominante, appare razionale (e probabile) che le Big Tech andranno a ridurre strategicamente la loro dipendenza da NVIDIA, attraverso una progressiva pressione verso la trasformazione in *commodity* dell'hardware dell'IA, approvvigionandosi da più fornitori o sviluppando internamente alternative di qualità comparabile. Questo approccio mira a impedire a qualsiasi singolo attore di sfruttare il potere di mercato e garantisce una maggiore flessibilità e concorrenza nell'approvvigionamento di hardware per l'IA.

A questo riguardo, anche la preminenza del software applicativo di NVIDIA (CUDA), è sfidata dalle Big Tech che stanno investendo per supportare l'interoperabilità tra i *chip* IA e ridurre la posizione di mercato consolidata di NVIDIA: ad esempio, AWS fornisce Neuron, un kit di sviluppo software, per aiutare i clienti a passare dai *chip* IA di terze parti a quelli di AWS e viceversa.

Per quanto riguarda l'utilizzo dei dati, le grandi aziende tecnologiche possono incontrare limiti nello sfruttare appieno i loro set di dati “proprietary” per ottenere un vantaggio competitivo.

Ad esempio, non ci sono prove evidenti che i modelli Gemini di Google o LLaMA di Meta siano superiori tecnicamente e qualitativamente alla serie GPT di OpenAI o i modelli Claude di Anthropic solo per la quantità di dati. Questo perché il valore unico dei dati proprietari per l'addestramento dei modelli di base rimane incerto, in quanto

---

<sup>19</sup> Che comprendono la Realtà Aumentata, la Realtà Mista e la Realtà Virtuale.

numerose fonti di dati alternative e di terze parti possono fungere da validi sostituti. In molti casi, infatti, i set di dati per l'addestramento e la messa a punto provengono da fonti disponibili pubblicamente, compresi i dati *web-scraped* e i set di dati *open source*. Anche quando si tratta di dati proprietari, spesso è possibile ottenere risultati di formazione simili attraverso *partnership* sui dati, accordi di licenza o l'uso di dati sintetici<sup>20</sup>.

A ciò si aggiunga che in alcuni domini quale quello finanziario modelli “generalisti”, come GPT-4 hanno mostrato prestazioni superiori a quelle dei modelli specializzati, addestrati su dati proprietari di alta qualità. Ciò suggerisce che l'accesso esclusivo a dati proprietari non garantisce automaticamente vantaggi competitivi decisivi, e che l'architettura e la qualità dell'addestramento possono essere fattori più determinanti rispetto alla mera disponibilità di dati esclusivi.

Inoltre, le grandi aziende tecnologiche sono sempre più vincolate nell'uso dei dati che raccolgono dai propri utenti dalle disposizioni normative almeno in Europa come il *Regolamento generale sulla protezione dei dati (GDPR)*<sup>21</sup> e il *Digital Markets Act (DMA)*<sup>22</sup>. Tali regolamenti impongono infatti forti restrizioni sulle modalità di raccolta, elaborazione e sfruttamento dei dati degli utenti, limitando ulteriormente la capacità delle Big Tech di sfruttare i dati come *asset* competitivo. Sarà anche rilevante monitorare l'applicazione delle disposizioni DMA sui servizi “*core platform*” nel contesto dei FM, per evitare che soglie normative troppo generiche non colgano le peculiarità della filiera dell'IA.

Nel complesso, i segmenti a monte della catena del valore dell'IA mostrano una tendenza alla concentrazione del mercato, ma l'innovazione sembra rimanere dinamica a causa della continua competizione conglomerale tra le principali aziende tecnologiche che operano in questi segmenti. Le grandi aziende tecnologiche hanno infatti investito attivamente in quasi tutti i mercati tecnologici legati all'IA, e nessuna sembra aver stabilito un controllo dominante su un segmento specifico. Tuttavia, come già detto, occorre prestare molta attenzione alle *partnership* (e naturalmente alle

---

<sup>20</sup> Ad esempio, OpenAI ha stretto una *partnership* con la casa editrice Axel Springer per dare agli utenti di ChatGPT l'accesso a sintesi in tempo reale dei contenuti di Axel Springer.

<sup>21</sup> Regolamento (UE) 2016/679 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati) [2016] GU L119/1 (GDPR).

<sup>22</sup> Regolamento (UE) 2022/1925 del 14 settembre 2022 relativo a mercati contendibili ed equi nel settore digitale e che modifica le direttive (UE) 2019/1937 e (UE) 2020/1828 (Legge sui mercati digitali).

fusioni) che potrebbero ridurre fortemente la concorrenza e trasformare le strategie di leva concorrenziale in dinamiche di monopolizzazione.

Di conseguenza, si può sostenere che la concorrenza persista in diversi segmenti della catena del valore dell'IA, con le imprese già presenti che si adoperano strategicamente per limitare la capacità dei loro rivali di espandersi e consolidare il potere di mercato. Questa interazione tra grandi aziende tecnologiche favorisce le “interazioni in più mercati” (*multi-market contacts*)<sup>23</sup>, che possono, a loro volta, smorzare l'intensità della “concorrenza per il mercato”<sup>24</sup> e mitigare le dinamiche di tipo “*winner takes it all/most*” che sono state prevalenti nel web 2.0.

La compresenza sul mercato di modelli *open source* e *closed source* è un ulteriore indicatore della dinamicità e vivacità del mercato. Se i modelli *closed source* forniscono prestazioni generalmente superiori, i modelli *open source* offrono trasparenza, personalizzazione e controllo sui dati più elevati. È questa una distinzione non solo tecnica, ma anche concorrenziale: i modelli *open source* possono favorire la condivisione dell'innovazione e abbassare le barriere all'ingresso, fungendo da contrappeso alle tendenze monopolistiche<sup>25</sup>.

Pertanto, le dinamiche competitive a monte della catena del valore dell'IA sembrano differire da quelle osservate nei mercati e servizi digitali “tradizionali”, come i motori di ricerca, i browser web e i social media, e possono portare a risultati di mercato diversi. Mentre i mercati digitali sono stati storicamente soggetti a effetti di *tipping* e monopolizzazione, la catena del valore dell'IA a monte presenta attualmente una struttura più distribuita, con diversi attori principali che operano su più livelli della catena del valore, spesso con sovrapposizioni significative, accanto ad altre imprese specializzate in segmenti specifici.

---

<sup>23</sup> I contatti multi-mercato si riferiscono a quelle situazioni in cui più di due imprese competono contemporaneamente su più mercati del prodotto. La maggior parte degli studi sul contatto multimercato ha analizzato come la sovrapposizione dei mercati crei una “tolleranza reciproca”. Si veda B.D. Bernheim e M. Whinston (1990), *Multimarket Contact and Collusive Behavior*, «The RAND Journal of Economics» 21, no. 1, pp. 1-26. Questo riduce l'intensità della rivalità instaurando dinamiche di cooperazione o “coopetizione”.

<sup>24</sup> La concorrenza per un mercato si riferisce alla definizione di nuovi standard o modelli di business dominanti che tendono a una struttura di mercato monopolistica e di solito è associata al processo di innovazione che porta sul mercato nuove tecnologie di sostituzione. Si veda P. Geroski (2003), *Competition in Markets and Competition for Markets*, «Journal of Industry, Competition and Trade», vol. 3, issue 3.

<sup>25</sup> Tuttavia, il traffico web da e verso i siti dei FM risulta significativamente concentrato, con OpenAI che mantiene oltre il 77% del traffico totale nonostante il proliferare di modelli alternativi, suggerendo che la sola disponibilità di alternative potrebbe non essere sufficiente a garantire un mercato realmente competitivo.

In particolare, nei mercati e nei servizi digitali, recentemente regolati dal DMA, le consolidate posizioni di dominanza delle principali aziende tecnologiche sono state in gran parte sostenute da effetti di *lock-in* del consumatore, spesso sostenuti da “*data feedback loop*” e da effetti di rete diretti<sup>26</sup>, piuttosto che da una superiorità intrinseca in termini di efficienza o qualità derivante dall’accesso esclusivo alle risorse a monte. Allo stesso modo, sembra verosimile e probabile che i rischi concorrenziali più significativi nel panorama dell’IA emergeranno probabilmente a valle, in particolare nella distribuzione e nella diffusione dei FM e nello sviluppo di applicazioni di IA generativa che interagiscono direttamente con gli utenti finali.

Inoltre, dal punto di vista concorrenziale, è fondamentale riconoscere che l’espansione dei sistemi di IA generativa potrebbe avere un impatto dirompente sui mercati attualmente dominati o controllati dalle aziende Big Tech. Ciò è particolarmente vero per il livello delle applicazioni, considerando che l’IA può abilitare nuovi servizi percepiti come sostitutivi (e migliori) dagli utenti finali, sostituendo così i servizi o le piattaforme esistenti, oppure creare nuovi modi di fornire gli stessi servizi “bypassando” quelli esistenti<sup>27</sup>.

Questo potenziale dirompente è generalmente favorevole alla concorrenza, in quanto promuove l’innovazione e l’ingresso nel mercato. Tuttavia, può con tutta probabilità anche incentivare le Big Tech a mettere in atto strategie di *defensive leveraging* per proteggere la propria posizione dominante, introducendo ulteriori complessità competitive. Di conseguenza, i comportamenti anticoncorrenziali nei mercati a valle potrebbero essere anche guidati dagli sforzi per salvaguardare le

---

<sup>26</sup> *Data feedback loop* è un processo (che si auto-rinforza) di miglioramento (oggettivo e/o soggettivo) della qualità, basato sui dati dei clienti. Se il miglioramento della qualità è soggettivo, cioè per un utente specifico, i dati rilevanti dipendono dall’utilizzo del servizio da parte di quel cliente: pertanto, un cliente, utilizzando un servizio, ne migliora la qualità da un punto di vista soggettivo e potrebbe essere soggetto a *lock-in* individuali. Se la qualità è oggettivamente migliorata, cioè per tutti gli utenti, i dati rilevanti dipendono dall’uso complessivo del servizio (e quindi dal numero di utenti): quindi, al di sopra di una certa soglia, può verificarsi una dinamica “winner takes it all” (*lock-in* collettivo). In effetti, un circolo di *feedback* oggettivo è generato dalle esternalità di rete dirette, basate sui dati, in cui il valore per ciascun utente di appartenere a una certa rete dipende dal numero complessivo di utenti della rete e dalla quantità complessiva di dati forniti. Si veda A. Hagiu e J. Wright (2023), *Data-enabled learning, network effects and competitive advantage*, «RAND Journal of Economics» 54 (4), pp. 638-667.

<sup>27</sup> Questo rappresenta uno scenario potenziale in uno dei mercati digitali con un attore dominante da lungo tempo, ovvero i servizi di ricerca. I FM che si aggiornano continuamente effettuando il *crawling* di Internet rappresentano un’alternativa interessante ai motori di ricerca tradizionali come Google. Ad esempio, Perplexity AI sfrutta la ricerca guidata dall’intelligenza artificiale per fornire risposte dirette e basate su citazioni, combinando la ricerca sul web in tempo reale con un’esperienza interattiva di *chatbot*. Gli utenti possono anche affinare le loro richieste in modo conversazionale, rendendo il processo di ricerca più intuitivo. Perplexity non si basa su un proprio modello di base, ma integra diversi modelli preesistenti, tra cui quelli sviluppati da OpenAI e Anthropic.

posizioni di mercato esistenti, con l'obiettivo di mitigare o prevenire l'impatto dirompente dei progressi dell'IA. Le strategie di difesa potrebbero includere anche operazioni di acquisizione preventiva (*killer acquisitions*) o offerte gratuite per limitare l'adozione di soluzioni alternative.

In sintesi, mentre la catena del valore dell'IA a monte presenta caratteristiche economiche distinte rispetto ai mercati digitali tradizionali, questa distinzione non è sempre altrettanto chiara nei segmenti a valle. Sebbene alcune autorità garanti della concorrenza, come l'inglese CMA, abbiano riconosciuto questi rischi a valle, le loro valutazioni sono state finora di livello generale e in gran parte incentrate solo sul generico ruolo di *gatekeeping* delle Big Tech. Sembra invece essenziale distinguere tra i vari mercati a valle e modelli di business per identificare i casi in cui possono emergere effetti di *tipping* e di *lock-in* dei consumatori nei servizi di IA. Inoltre, è fondamentale distinguere tra i diversi modelli di business dell'IA, poiché alcuni di questi problemi potrebbero già essere coperti dal quadro normativo esistente, in *primis* il *Digital Markets Act (DMA)*, che mira a garantire l'equità e la contendibilità del mercato.

### **3. Il mercato delle applicazioni di IA generativa e i FM**

Le imprese a valle nella catena del valore dell'IA possono essere utenti finali dei FM, se questi vengono utilizzati per le loro attività interne, e/o utenti, quando utilizzano i servizi FM come input per fornire servizi di IA generativa ai consumatori. La prima tipologia di imprese usa i modelli internamente per migliorare operazioni, prodotti o processi decisionali, ma non vendono direttamente funzionalità basate sui FM ai clienti. La seconda tipologia integra i FM nei propri prodotti o servizi, mettendo tali funzionalità a disposizione diretta degli utenti finali. Ad esempio, una banca che integra un FM (es. GPT-4) in una *chatbot* usato per assistere i clienti su *home banking*, consulenza finanziaria, richieste di assistenza; oppure una piattaforma per scrittura assistita, riassunti automatici e generazione di idee che integra GPT.

Questa integrazione, come descritto nei precedenti capitoli, può essere fatta in vari modi, principalmente: (i) utilizzano API per accedere a FM ospitati da grandi *provider* di *cloud* (es: GPT-4 di OpenAI tramite Azure; Claude di Anthropic tramite Amazon Bedrock; Gemini di Google tramite Vertex AI); (ii) utilizzando *plug-in* o estensioni (come l'ecosistema di *plug-in* di ChatGPT o Zapier AI, che forniscono componenti software che permettono di integrare le capacità dei FM in app o piattaforme esistenti).

A loro volta, i consumatori possono accedere e utilizzare le applicazioni rivolte all'utente in diversi modi: (i) tramite un sito web o un'applicazione specifica, (ii) utilizzando un prodotto/servizio specifico che è abbinato a un servizio di IA generativa o che lo incorpora; (iii) scegliendo e scaricando un'applicazione di IA generativa autonoma da un app store.

Come evidenziato in precedenza, il mercato a valle della catena del valore dell'IA, dove i FM vengono impiegati e le loro capacità di inferenza sono sfruttate dalle applicazioni generative per gli utenti finali, è particolarmente critico dal punto di vista della contendibilità e dell'equità del mercato. Tuttavia, questo mercato è stato finora oggetto di un'analisi limitata, a causa della sua complessità e della mancanza di una visione chiara di come le grandi imprese tecnologiche potrebbero replicare le dinamiche competitive e di mercato osservate nei mercati digitali tradizionali.

In generale, le grandi aziende tecnologiche, assicurandosi posizioni dominanti nei livelli di applicazioni e contenuti, dove i FM sono utilizzati dagli utenti finali tramite applicazioni di IA generativa, possono influenzare le condizioni di mercato limitando la disponibilità di servizi FM alternativi. Ciò può avvenire attraverso strategie quali il *self-preferencing* dei propri modelli di FM o la formazione di *partnership* esclusive con gli sviluppatori di FM, limitando potenzialmente la concorrenza.

Tuttavia, questa preminenza oligopolistica nel livello applicativo non implica di per sé una leva anticoncorrenziale. Data la diversità delle applicazioni e dei servizi che utilizzano IA generaliste, potrebbe non essere semplice per le imprese dominanti estendere il proprio potere di mercato dal livello applicativo a valle verso il mercato dei FM. Pertanto, nonostante molti importanti mercati e servizi digitali siano dominati dalle Big Tech, sembra improbabile che il potere di mercato di un livello applicativo a valle così composito possa essere esercitato sui FM per stabilire una preminenza di mercato.

Inoltre, a differenza di quanto accade nei mercati digitali "tradizionali", molti studiosi hanno notato che i *data feedback loop* a valle della catena del valore dell'IA potrebbero non essere così forti<sup>28</sup>. In primo luogo, è stato osservato che i dati di *feedback* dell'utente non vengono automaticamente reinseriti nel modello e la maggior parte delle volte sarebbe molto costoso farlo<sup>29</sup>. Inoltre, le applicazioni generative non sempre producono segnali di *feedback* forti e accurati per generare un effetto *feedback* significativo. Ad esempio, il risultato di una *chatbot* potrebbe non ricevere *feedback* o

<sup>28</sup> Si veda: A. Hagiu e J. Wright (2025), *Artificial intelligence and competition policy*, «International Journal of Industrial Organization».

<sup>29</sup> Si veda: CMA (2023), *AI Foundation Models: initial Report*.

un semplice “pollice in su”, che è un *feedback* molto più debole rispetto, ad esempio, alla scelta di uno specifico link/prodotto/feed tra i tanti proposti, come avviene nei servizi digitali “tradizionali”.

Infine, come evidenziato, attualmente esistono più FM di successo e in concorrenza tra loro, il che rende improbabile che i dati degli utenti raccolti da una singola applicazione anche all’interno della vasta base di clienti delle Big Tech possano fornire un vantaggio competitivo significativo nel mercato dei FM. Tuttavia, questo scenario potrebbe cambiare se i dati di *feedback* degli utenti fossero aggregati tra più applicazioni, amplificando potenzialmente il loro impatto sulle prestazioni del modello, come discusso più avanti<sup>30</sup>.

Allo stesso modo, gli effetti di rete indiretti<sup>31</sup> non sembrano essere un fattore importante nel mercato dei FM a valle. A differenza delle tradizionali piattaforme digitali multiversante, dove le scelte degli utenti influenzano gli altri versanti del mercato, le applicazioni di IA generativa costruite sui FM operano in una struttura di mercato lineare. Ciò significa che gli utenti finali scelgono un’applicazione di IA generativa in modo indipendente, senza che la loro scelta influisca direttamente sugli altri utenti o sulle dinamiche di mercato, come avviene invece per i servizi di piattaforma digitale.

Tuttavia, esistono delle rilevanti eccezioni in cui (i) il potere di mercato si può traslare dal livello delle applicazioni verso i FM, direttamente o indirettamente, o (ii) si possono creare “modelli a piattaforma” in grado di creare gli stessi effetti di *lock-in*.

Per quanto riguarda il trasferimento del potere di mercato, questo potrebbe avvenire indirettamente attraverso il *cloud*. Ossia, un fornitore di servizi *cloud* dominante in un mercato di applicativi software veicolati attraverso il *cloud* (SaaS) potrebbe sfruttare questo potere di mercato per accrescere la sua preminenza nei mercati delle infrastrutture *cloud* (IaaS) e quindi, riducendosi la concorrenza nel *cloud*, acquisire una posizione di preminenza anche nello sviluppo dei FM. Un esempio della prima parte di tale dinamica sono le pratiche commerciali inerenti alle licenze software messe in atto da Microsoft, che hanno, secondo l’anglosassone CMA, la capacità e l’incentivo di escludere i concorrenti (anche AWS e Google), rendendo per i clienti dei

---

<sup>30</sup> Si deve notare a questo riguardo che il DMA, ed in specifico il suo art. 5 comma 2 b, vieta ai *gatekeeper* di combinare i dati raccolti su piattaforme diverse per finalità di profilazione o pubblicità mirata, a meno che non venga fornito un consenso libero, informato e specifico dall’utente per quella particolare combinazione di dati.

<sup>31</sup> Gli effetti di rete indiretti (o incrociati) si verificano quando ci sono almeno due gruppi di clienti diversi che sono interdipendenti e l’utilità di almeno un gruppo cresce con la crescita della dimensione dell’altro gruppo o degli altri gruppi.

suoi servizi software meno interessante (sotto un profilo sia tecnico sia commerciale) l'utilizzo di tali prodotti sulle infrastrutture *cloud* dei concorrenti<sup>32</sup>. Proprio contro tali pratiche, Alphabet (Google) ha presentato nel settembre 2024 un reclamo anche alla Commissione europea<sup>33</sup>; inoltre, in novembre 2024, anche la *Federal Trade Commission* (FTC) ha aperto un'ampia indagine antitrust negli Stati Uniti, inviando a Microsoft una estesa richiesta di informazioni<sup>34</sup>.

Per quanto concerne il secondo aspetto, a determinate condizioni, i mercati a valle dell'IA possono essere soggetti a “*platformisation*”, ossia ad assumere caratteristiche simili a quelle di una piattaforma e a replicare i modelli economici e commerciali di quest'ultima. Ciò si verifica in particolare quando le applicazioni e i dispositivi di IA generativa sono integrati commercialmente e tecnicamente in modo da renderli specifici per un FM. In questo scenario, gli utenti possono diventare indirettamente dipendenti da uno specifico FM, non scegliendo direttamente il modello di FM stesso, ma solo indirettamente, in quanto tutte le applicazioni di IA generativa a cui accedono dipendono da quel modello. Questa situazione si può verificare quando un FM è profondamente incorporato in un insieme di servizi o, più significativamente, in un intero ecosistema, limitando la capacità degli utenti finali di utilizzare modelli di IA alternativi al di fuori di quell'ecosistema.

Questa situazione si verifica, ad esempio, quando un FM è integrato con un sistema operativo (OS) o un dispositivo. Tale integrazione può essere di tipo funzionale, quando un'unica azienda sviluppa e fornisce sia il FM, sia il sistema operativo/dispositivo, oppure di tipo commerciale, quando esistono accordi esclusivi tra i produttori di sistemi operativi o di smartphone e gli sviluppatori di FM.

In questo scenario, la catena del valore dei FM, tradizionalmente lineare, inizia a presentare caratteristiche simili a quelle di una piattaforma. Di conseguenza, i distributori di FM e gli sviluppatori di applicazioni di IA generativa che mirano a raggiungere gli utenti finali possono trovarsi costretti o incentivati a fare affidamento sul FM incorporato nel sistema operativo o nel dispositivo. Questa configurazione può portare a esternalità di rete incrociate e a una maggiore probabilità di *tipping* nel mercato, analogamente a quanto si è verificato nei mercati “tradizionali” delle piattaforme digitali.

---

<sup>32</sup> CMA (2025), *Cloud Infrastructure Services Provisional decision*.

<sup>33</sup> Si veda: <https://www.reuters.com/technology/google-files-complaint-eu-over-microsoft-cloud-practices-2024-09-25/>.

<sup>34</sup> Si veda: <https://www.nytimes.com/2024/11/27/technology/microsoft-ftc-antitrust.html>.

In sostanza, le aziende che controllano gli ecosistemi dei sistemi operativi possono acquisire un'influenza significativa sulla diffusione dei FM, limitando potenzialmente la scelta dei consumatori. Dando la priorità ai propri servizi FM attraverso una perfetta integrazione, un'accessibilità preferenziale e/o una maggiore compatibilità, possono creare barriere artificiali che scoraggiano gli utenti dal passare a fornitori FM concorrenti, rafforzando così la loro posizione dominante sul mercato.

Un'altra forma di “*platformisation*” nella catena del valore è strettamente legata ai servizi di *cloud computing*. Se, infatti, da un lato, il *cloud computing* svolge un ruolo cruciale come input a monte per la formazione dei FM, dall'altro lato, funge da canale di distribuzione chiave nel mercato a valle. In particolare, le piattaforme *cloud* fungono da intermediari tra gli sviluppatori di FM, i distributori di FM e gli utenti finali, facilitando sia la distribuzione dei servizi FM, sia l'inferenza.

In effetti, nel mercato a valle, la maggior parte delle aziende si affida al *cloud computing* per il *fine-tuning* del FM e per gestire i servizi di FM e spesso accede a tali servizi attraverso piattaforme configurate come “*marketplace*”, tipicamente gestite dai maggiori fornitori di servizi *cloud* (CSP), ossia Amazon, Microsoft e Google. Numerosi sono gli esempi al riguardo, Microsoft offre i FM di OpenAI attraverso il servizio Azure OpenAI, oltre che all'interno delle sue soluzioni *cloud* aziendali esistenti, tra cui Dynamics e Power Platform; inoltre, le aziende possono distribuire FM di altri fornitori, come Hugging Face, utilizzando l'infrastruttura di Azure. Google fornisce l'accesso ai FM attraverso la propria Google Cloud Platform; la piattaforma IA di Amazon, Bedrock, integra i dati dei clienti archiviati in AWS per il *ne-tuning* di FM pre-addestrati di Amazon, supportando anche modelli di altri fornitori come Anthropic.

Inoltre, Google, AWS e Microsoft non solo forniscono risorse di calcolo essenziali, ma competono direttamente nel mercato a valle dei servizi FM, offrendo le proprie applicazioni di IA rivolte all'utente. I principali fornitori di *cloud* svolgono quindi un ruolo cruciale nel consentire l'accesso degli utenti finali ai servizi FM e nel facilitare il recupero dei dati in tempo reale per l'inferenza.

Attualmente, le piattaforme di *marketplace* di IA promuovono una selezione diversificata di FM, consentendo di scegliere tra una gamma di opzioni. Tuttavia, in questi ambienti, potrebbero emergere strategie di *self-preferencing* da parte dei principali *Cloud Service Providers* (CSP), che potrebbero condurre a: (i) un trattamento preferenziale per i propri FM (proprietary o in *partnership*) rispetto ai modelli concorrenti; (ii) favorire i propri servizi FM a valle rispetto alle alternative di terzi, limitando la concorrenza.

Inoltre, i *data feedback loop* possono avere un impatto molto maggiore sui modelli di intelligenza artificiale generativa integrati nei CSP, poiché questi sistemi beneficiano direttamente delle interazioni continue con gli utenti. Questo vantaggio può consentire loro di perfezionare e personalizzare i modelli in modo più efficiente rispetto agli sviluppatori indipendenti di FM. Poiché i modelli integrati raccolgono dati di inferenza dell'utente in tempo reale su più applicazioni, hanno accesso a un set di dati più ampio e diversificato, che potrebbe rappresentare un vantaggio competitivo non disponibile per gli sviluppatori FM "esterni". Quindi, i modelli di intelligenza artificiale generativa integrati nei CSP possono generare effetti di rete cumulativi, in cui il *feedback* degli utenti viene sfruttato per potenziare i modelli futuri, migliorare le prestazioni e introdurre nuovi servizi<sup>35</sup>.

Queste dinamiche competitive dovrebbero essere considerate come un "effetto ecosistema" piuttosto che come comportamenti isolati del mercato. In particolare, i CSP che controllano anche i sistemi operativi possono espandere in modo significativo la propria posizione di forza lungo la catena del valore dell'IA, accrescendo ulteriormente il potere di mercato.

Da questo punto di vista, l'implementazione dei vari sistemi di FM direttamente sui dispositivi (*on-device AI*)<sup>36</sup> offre diversi vantaggi. Oltre a migliorare le prestazioni, la resilienza e la sicurezza, questo approccio migliora la privacy riducendo la dipendenza dalla raccolta centralizzata dei dati. Di conseguenza, questa evoluzione potrebbe attenuare i problemi di concorrenza legati al ciclo di *feedback* dei dati all'interno degli ecosistemi controllati dai CSP, promuovendo un panorama dell'IA più equilibrato.

#### **4. Misure di politica della concorrenza**

L'identificazione di rischi concorrenziali lungo la catena del valore dell'intelligenza artificiale, che si differenziano sia in termini di intensità, sia di tipologia, riflette la complessità delle dinamiche competitive emergenti ed evidenzia la necessità di considerare un approccio di politica della concorrenza su più livelli.

Il *framework* tradizionale del diritto della concorrenza mantiene un ruolo centrale nell'affrontare eventuali comportamenti anticoncorrenziali lungo alcuni segmenti della

---

<sup>35</sup> Si veda: T. Schrepel, A. Pentland (2024), *Competition between AI foundation models: dynamics and policy recommendations*, «Industrial and Corporate Change», p. 119.

<sup>36</sup> Oppure un approccio ibrido che combina l'intelligenza artificiale su dispositivo e su *cloud*.

catena del valore dell'IA in cui vi sia una posizione dominante: il caso NVIDIA rappresenta un paradigma esemplificativo. L'Articolo 101 TFEU potrebbe invece trovare applicazione nell'analisi delle *partnership* strategiche tra Big Tech e sviluppatori di FM. Le indagini condotte da alcune autorità di concorrenza (soprattutto CMA e FTC) su tali *partnership* hanno rilevato significative preoccupazioni competitive, ma tali analisi erano sviluppate in un contesto di *merger control*. Come dichiarato dalla CMA riguardo la *partnership* Microsoft-OpenAI: «*A finding that the Partnership does not give rise to a relevant merger situation does not constitute a finding that no competition concerns arise from its operation*». L'applicazione dell'Articolo 101 rimane quindi una questione aperta e rilevante per garantire che le dinamiche collaborative nell'ecosistema IA non si trasformino in meccanismi di *foreclosure* anticoncorrenziale che limitino l'innovazione e la concorrenza nel mercato emergente dei *Foundation Models*.

Tuttavia, il diritto della concorrenza presenta limitazioni strutturali, particolarmente nei segmenti dove emergono dinamiche di “piattaforma” quali quelle evidenziate nel paragrafo precedente, che in passato hanno portato all'emanazione del *Digital Markets Act (DMA)*. Questo effetto è prevalente nei mercati *downstream*, ma non è ad esso esclusivo come evidenziato dalla descrizione degli effetti rete e di *lock-in* del *framework* applicativo CUDA. Sarà importante quindi andare a verificare se e come il DMA, che rappresenta il principale strumento regolamentare *ex-ante* per i mercati digitali, possa affrontare i rischi concorrenziali nei segmenti *downstream* della catena del valore dell'IA.

L'approccio più immediato consiste nell'applicazione degli obblighi DMA ai *Core Platform Services (CPS)*<sup>37</sup> esistenti che utilizzano funzionalità di IA. L'applicazione del DMA può tuttavia declinarsi in modo diverso a seconda delle modalità di integrazione di tali funzionalità nei CPS. Quando i FM o gli agenti IA operano come applicazioni *stand-alone* negli ecosistemi dei dispositivi mobili tramite il sistema operativo, si applicano senza problemi gli obblighi di interoperabilità dell'Articolo 6(7) DMA. Questi obblighi garantiscono che i fornitori di servizi IAG

---

<sup>37</sup> Il DMA si concentra sui servizi di piattaforma attualmente più utilizzati dalle imprese e dagli utenti finali, definiti “servizi piattaforma base” (o, più significativamente, in inglese, *core platform services*, CPS), nei quali il legislatore europeo ha individuato per il momento i più evidenti timori di una scarsa contendibilità e di pratiche sleali (art 1.2 DMA). L'elenco completo dei CPS è il seguente: (a) servizi di intermediazione online; (b) motori di ricerca online; (c) servizi di social network online; (d) servizi di piattaforma di condivisione video; (e) servizi di comunicazione interpersonale indipendente dal numero; (f) sistemi operativi; (g) browser web; (h) assistenti virtuali; (i) servizi di *cloud computing*; (j) servizi di pubblicità online. Ai sensi dell'articolo 19, la lista dei CPS può essere integrata da servizi aggiuntivi, da parte della Commissione, a seguito di un'indagine di mercato.

concorrenti possano accedere alle stesse funzionalità hardware-software utilizzate dai servizi del *gatekeeper*, prevenendo forme di *self-preferencing* tecnologico. Quando le funzionalità di IAG sono direttamente integrate in un CPS controllato da un *gatekeeper* designato come un sistema operativo, un assistente virtuale o un browser, anziché offrirle come servizi autonomi si va a creare una situazione di mercato più complessa per l'ambito di applicazione del DMA, che regola i servizi e le strutture di mercato, non le tecnologie in sé. Tuttavia, alcuni obblighi continuano a disciplinare indirettamente le funzionalità IAG integrate, in particolare attraverso disposizioni relative ai dati che si applicano a tutti i CPS pertinenti. Ad esempio, l'articolo 5(2) che proibisce la combinazione di dati personali degli *end-user* attraverso diversi CPS senza consenso esplicito, e l'articolo 6(9) che stabilisce diritti di portabilità dei dati che possono includere i dati di *feedback* generati dalle interazioni con agenti GenAI *embedded*. Quando infine le applicazioni IAG soddisfano autonomamente la definizione di un CPS esistente come motori di ricerca potenziati dall'IA (ad esempio Perplexity.ai) o *marketplace* di servizi IA (ad esempio Amazon Bedrock) in caso di identificazione come *gatekeeper* di tale servizio si applica l'intero set di obblighi DMA, inclusi i divieti di *self-preferencing* dell'articolo 6(5) e le garanzie di accesso non discriminatorio.

Un terzo pilastro della risposta di politica della concorrenza è rappresentato dalle indagini di mercato condotte dalle autorità di concorrenza. Attualmente diverse autorità hanno condotto analisi di mercato per valutare l'esistenza di problemi concorrenziali nei mercati dell'IA: la CMA nel Regno Unito, l'Autorité de la Concurrence in Francia, il JFTC in Giappone ed un numero crescente di autorità sta valutando questa opzione. Esiste tuttavia una grande differenziazione dell'efficacia di queste azioni a seconda dei contesti giuridico-istituzionali. La maggior parte delle altre autorità nazionali di concorrenza può infatti condurre solo *market studies* senza il potere di imporre rimedi vincolanti e quindi usare solamente strumenti di *advocacy*, verso il legislatore, o, al limite, poteri di *moral suasion*. In altri contesti, le autorità di concorrenza hanno poteri "regolatori", con cui possono imporre rimedi comportamentali e/o strutturali in caso constatino dei pervasivi problemi concorrenziali.

Ad esempio, la CMA dispone di forti poteri correttivi, tra cui rimedi strutturali di cessione di attività ed *asset*, attraverso il suo regime di indagine di mercato ai sensi dell'Enterprise Act 2002. Inoltre, ai sensi del *Digital Markets, Competition and Consumers Act 2024*, la CMA può imporre rimedi comportamentali e strutturali nei confronti delle imprese digitali con uno status di mercato strategico (SMS). Anche AGCM ha poteri simili che può esercitare quando, in esito a un'indagine conoscitiva

condotta ai sensi dell'articolo 12, comma 2, della legge 287 del 1990, riscontra problemi concorrenziali che ostacolano o distorcono il corretto funzionamento del mercato con conseguente pregiudizio per i consumatori. L'AGCM può imporre «ogni misura strutturale o comportamentale necessaria e proporzionata, al fine di eliminare le distorsioni della concorrenza». Tali misure non hanno natura sanzionatoria per comportamenti passati ma regolatoria per correggere disfunzioni strutturali in maniera prospettica. I poteri dell'AGCM potrebbero essere particolarmente rilevanti per i mercati dell'intelligenza artificiale, dove emergano rischi di consolidamento e *platformization* che né del diritto antitrust né il DMA possano riuscire a controllare.