

POLITECNICO DI TORINO

Corso di Laurea Magistrale in

INGEGNERIA GESTIONALE

Collegio di Ingegneria Gestionale e della Produzione

Percorso Gestione dell'Innovazione e Imprenditorialità



Tesi di Laurea di II livello

Concorrenza nell'era degli AI Agents:

dinamiche di mercato e implicazioni economiche.

Relatore: Prof. Carlo Cambini

Candidato: Alessandro Albanesi

Anno Accademico 2024-2025

Indice

Indice	3
Introduzione.....	5
Sezione 1 Intelligenza Artificiale e AI Agents.....	8
1.1 Panoramica Generale Intelligenza Artificiale	8
1.1.1 Storia	8
1.1.2 Definizioni e Applicazioni dei modelli di AI	12
1.2 Componenti per lo sviluppo di modelli AI	15
1.2.1 Hardware per lo sviluppo dei modelli AI	15
1.2.2 I dati.....	18
1.2.3 Il Cloud Computing	18
1.3 AI Agents	21
1.3.1 Definizione e panoramica generale degli AI Agents	21
1.3.2 Funzionamento degli AI Agents	22
1.3.3 Multi Agent System e Orchestrazione.....	24
Sezione 2 Value Chain e Main Actors	26
2.1 La catena del valore dell'intelligenza artificiale	26
2.1.1 Hardware – chip	26
2.1.2 Cloud Computing.....	26
2.1.3 Foundation Models e LLM/MLLM.....	27
2.1.4 Orchestrazione agentica	27
2.1.5 Interfaccia	27
2.2 Microsoft.....	28
2.2.1 Microsoft Azure	28
2.2.2 Microsoft Copilot	31
2.2.3 Scelte strategiche	32
2.3 Alphabet - Google	36
2.3.1 Google Cloud Platform.....	36
2.3.2 Google Gemini	37
2.3.3 Scelte Strategiche	38
2.4 Amazon	40
2.4.1 Amazon Web Services.....	41
2.4.2 Anthropic	42
2.4.3 Scelte strategiche	43
2.5 Meta	44
2.5.1 Infrastruttura proprietaria	44
2.5.2 LLAMA.....	45
2.5.3 Scelte strategiche	46
2.6 DeepSeek.....	46
2.7 NVIDIA.....	48
Sezione 3 Opportunità e rischi legati agli AI Agents	49
3.1 Possibili sviluppi degli AI Agents	49
3.1.1 Principali attori attivi nello sviluppo di AI Agents	49
3.1.2 Prospettive future	51
3.1.3 Contesto regolatorio	52

3.1.4 Previsioni di mercato	53
3.2 AI Agents come rischio.....	54
3.2.1 Esternalità di rete	56
3.2.2 Self-preferencing e preinstallazione.....	60
3.2.3 Collusione tacita.....	61
3.2.4 Non neutralità degli agenti intelligenti	63
3.2.5 Dipendenza nell'accesso a input essenziali	64
3.3 AI Agents come opportunità	66
3.3.1 AI Agents come nuovi punti di accesso	67
3.3.2 Minori costi di ricerca	69
3.3.3 Sviluppi open source.....	71
Sezione 4 Conclusioni e prospettive future	74
4.1 Riepilogo di opportunità e rischi degli AI Agents.....	74
4.2 Occasione di apertura o ulteriore consolidamento	78
4.3 Competizione tra le Big Tech	80
4.4 Il contesto normativo	80
Conclusione.....	86
Bibliografia – Sitografia.....	89

Introduzione

Negli ultimi decenni stiamo tutti assistendo a un progresso tecnologico sempre più rapido, talvolta persino superiore alla capacità della mente umana di concepirlo e metterlo in atto. Quella che fino a pochi anni fa chiamavamo intelligenza artificiale ed era considerata soltanto pura immaginazione, oggi è diventata uno strumento concreto che permea la gran parte delle nostre attività quotidiane, spesso senza che ce ne rendiamo conto. Dalla sintesi dei testi, alle diagnosi mediche automatizzate, l'intelligenza artificiale sta certamente rivoluzionando il modo in cui ci interfacciamo con la tecnologia e come essa, in qualche maniera, dà una direzione alla nostra vita, andando ad influire sulle nostre conoscenze e competenze, rafforzandone alcune, creandone di nuove e distruggendone altre.

Quando la tecnologia comincia ad avvicinarsi sempre più alle nostre competenze e conoscenze, significa che il tema si sta facendo delicato e non possiamo permetterci di subire passivamente questo progresso, altrimenti il rischio è che sia la tecnologia a guidare noi e non il contrario.

Nel caso dell'intelligenza artificiale e in particolare nel caso di quelli che oggi vengono chiamati AI Agents, ciò interessa ogni aspetto della vita umana: sociale, culturale, economico, giuridico, lavorativo e certamente anche cognitivo. Quando si cerca di comprendere una nuova tecnologia, ognuno di questi ambiti deve essere considerato oggettivamente e nella sua interezza in modo da comprendere quali siano possibili criticità e riuscire ad agire prima che sia troppo tardi.

Questa tesi, in particolare, vuole osservare e studiare l'aspetto economico, con l'intento di capire quale sia la situazione attuale e in che direzione si stia andando. Laddove qualcuno creda essere più importante ragionare prima su temi sociali, rispondo che l'aspetto economico penetra tutti i settori citati e il motivo è semplice: sebbene una tecnologia nasca come soluzione a certi bisogni, essa poi si sviluppa alimentata da interessi economici tipicamente in capo a certe organizzazioni o aziende. Comprendere questo contesto consente quindi una migliore analisi dei temi sociali, avendo compreso in che direzione stiamo presumibilmente andando.

Il lavoro di questa tesi è strutturato su quattro sezioni che analizzano in ordine sequenziale il tema oggetto di analisi, partendo da una panoramica generale, fino a terminare con la definizione delle implicazioni e dei rischi connessi allo sviluppo degli AI Agents. In particolare:

1. La prima sezione definisce e spiega gli AI Agents. In questa parte viene fornita una panoramica completa sull'intelligenza artificiale in generale, andando a visualizzare l'attuale stato dell'arte. A partire da questa descrizione, si entra poi nel dettaglio a livello di sviluppi tecnologici associati all'emergere degli AI Agents. Da qui, viene

successivamente spiegato come questi strumenti funzionino e quali asset economici, tecnologici e infrastrutturali siano necessari per il loro sviluppo. Si definisce quindi quale sia lo stadio attuale di sviluppo e quali siano le potenzialità future associate a questi sistemi.

2. La seconda sezione analizza il mercato dell'AI. In questa parte, viene analizzata la struttura del mercato, con specifica attenzione a quelli che sono i principali attori operanti nel settore. Vengono quindi identificati i soggetti chiave, scoprendo la loro posizione nella catena del valore. Sulla base dei concetti chiave definiti nella prima sezione, è infatti possibile comprendere in questa parte quali siano i punti di forza e di debolezza associati ad ogni attore.
3. La terza sezione affronta implicazioni, rischi e opportunità associati agli AI Agents e al mercato. In particolare, questa analisi viene fatta su due fronti: lato consumatore e lato imprese. Dal primo lato, vengono affrontati i rischi che incidono in maniera diretta ed immediata sul consumatore, come bias nei sistemi automatizzati, possibili discriminazioni o responsabilità legale associata alle azioni intraprese dagli AI Agents. Dall'altro lato, vengono affrontati i rischi che invece incidono nel consumatore finale nel lungo termine come conseguenza delle scelte di business delle imprese. Tali implicazioni sono figlie di problemi legati all'abuso di posizione dominante, alla presenza di barriere all'ingresso per i nuovi entranti e alla crescente concentrazione del mercato in poche mani.

Successivamente, vengono anche analizzate tutte le possibili opportunità che rendono gli AI Agents delle possibili soluzioni all'eccessiva concentrazione nel mercato digitale.

4. La quarta sezione definisce infine le conclusioni e le prospettive future. In questa parte, vengono riassunti i temi trattati, facendo emergere le problematiche principali e riflettendo sulle evoluzioni del mercato degli AI Agents. Il focus è sulle principali preoccupazioni relative all'impatto di questi strumenti sul potere delle grandi aziende, cercando di comprendere se essi possano rappresentare una valida opportunità per ridurre la concentrazione di potere in mano alle Big Tech o se, al contrario, confermeranno il loro potere, addirittura accentuandolo.

Vengono quindi considerate quali dovrebbero essere le attenzioni e le azioni da parte dei governi e delle autorità regolatorie per garantire il progresso tecnologico, senza però portare a squilibri economici e sociali.

L'obiettivo di questa tesi è quello di permettere al lettore di acquisire conoscenze in questo tema così attuale, in modo da poter sviluppare un pensiero critico ad esso associato.

È questo il motivo per il quale l'elaborato cercherà di essere il più comprensibile possibile anche per gli inesperti nel campo economico trattato, senza però compromettere l'analiticità necessaria per affrontare il tema, anche da un punto di vista tecnico.

Si utilizzerà un linguaggio semplice e accessibile, cercando di definire e spiegare ogni concetto che emerge in maniera chiara e comprensibile per garantire che il lettore, indipendentemente dal suo background, possa seguire il filo del discorso, acquisire una comprensione completa del tema e definire infine la propria idea.

Sezione 1

Intelligenza Artificiale e AI Agents

1.1 Panoramica Generale Intelligenza Artificiale

1.1.1 Storia

Sebbene oggi l'intelligenza artificiale rappresenti oggi uno dei temi più discussi e attuali, le sue origini appartengono ad un passato ben più remoto di quanto si possa immaginare.

Per comprendere l'evoluzione storica dell'AI, è utile partire da un'analisi concettuale del termine stesso, composto da due elementi distinti: *intelligenza* e *artificiale*.

Il termine *artificiale*, secondo la definizione Treccani, fa riferimento a qualcosa che è "*prodotto o ottenuto con l'intervento dell'uomo e mediante procedimenti tecnici*"¹. Ciò implica che qualsiasi forma di intelligenza artificiale sia, innanzitutto, il risultato dell'operare umano il quale uomo progetta e costruisce strumenti caratterizzati da certe funzioni. Queste funzioni, nel nostro caso sono legate al concetto di "intelligenza".

Per comprendere questo termine, una soluzione specifica e adeguata viene fornita dal matematico e logico Alan Turing con il suo articolo intitolato "*Computing Machinery and Intelligence*", pubblicato nel 1950.

Questo articolo deve la sua importanza al tema trattato che, per la prima volta, riguarda il comportamento delle macchine. Turing, infatti, si chiede se le macchine siano in grado di pensare, o meglio, se siano in grado di comportarsi come esseri umani².

Il suo articolo propone un test che prenderà il nome dell'autore (*Test di Turing*) e che si preoccupa di comprendere se per un uomo, l'interrogatore, sia possibile distinguere se le risposte che riceve siano state formulate da una macchina e da un'altra persona.

Laddove l'interrogatore non sia in grado di notare alcuna distinzione, la macchina è da considerarsi intelligente.

Approssimativamente negli stessi anni della pubblicazione di Turing, si introduce e si va sviluppando il concetto di **rete neurale**, strumento che solo successivamente abiliterà lo sviluppo dell'intelligenza artificiale così come la conosciamo oggi.

Come rete neurale, semplificando, bisogna immaginarsi un programma (quindi di fatto un codice che viene eseguito in un computer) che attraverso formule matematiche è in grado di

¹ [https://www.treccani.it/vocabolario/artificiale_\(Sinonimi-e-Contrari\)/](https://www.treccani.it/vocabolario/artificiale_(Sinonimi-e-Contrari)/)

² Turing, A. M. (1950). *Macchine calcolatrici e intelligenza*. <https://disf.org/files/macchine-calcolatrici-e-intelligenza.pdf>

imparare dai dati: più dati vengono forniti, più il sistema è in grado di apportare correzioni e più validi saranno i suoi output.

Ovviamente un sistema così concepito è in grado di funzionare correttamente ed efficacemente solo avendo a disposizione un hardware avanzato e una buona dose di dati.

Questi input cominciano a diventare adeguatamente disponibili solo negli ultimi decenni, in particolar modo con l'avvento di *web services* offerti attraverso computer prima e smartphone poi.

Tutto ciò porta a parlare del concetto di **Deep Learning**. Esso è uno strumento che utilizza le reti neurali addestrate su molti dati (quindi rese sempre più simili al modo di lavorare del cervello umano) per fornire in output nuove soluzioni³.

Su questo strumento si fonda il tema più generale del **Machine Learning** con il quale si identifica *“l'applicazione dell'AI che prevede l'addestramento delle macchine per l'esecuzione di un'attività senza essere appositamente programmate per questo”*⁴.

Si è arrivati quindi a definire dei sistemi che, come immaginava Turing, somigliano molto ad un uomo in quanto capaci di imparare, ragionare e creare output.

Così come sono stati definiti, questi strumenti sembrerebbero rispecchiare la definizione di intelligenza artificiale, almeno per come la intendiamo oggi, ma non basta, si può e si deve introdurre un'ulteriore distinzione fondamentale. Intelligenza artificiale sarebbe infatti troppo generico come concetto e, nella pratica, difficilmente associabile a sistemi concreti.

Per questo, vanno distinte due categorie principali: l'**Intelligenza Artificiale Debole** (ANI - Artificial Narrow Intelligence) e l'**Intelligenza Artificiale Generale** (AGI – Artificial General Intelligence).

- La prima forma di intelligenza artificiale, attualmente la più diffusa e implementata è l'intelligenza artificiale debole.

Questo tipo di intelligenza è stata sviluppata per affrontare problemi specifici, operando con un campo di azione limitato, attraverso l'applicazione di modelli statistici e algoritmi di auto-apprendimento. Sono effettivamente degli strumenti molto specifici e precisi, ma privi della capacità di estendere la loro competenza oltre l'ambito per il quale sono stati progettati.

³ Oracle (s.d.). *Cos'è il deep learning?* Oracle. <https://www.oracle.com/it/artificial-intelligence/machine-learning/what-is-deep-learning/>

⁴ Google Cloud (s.d.). *Cos'è il machine learning?* Google Cloud. <https://cloud.google.com/learn/what-is-machine-learning?hl=it>

Fondamentalmente, l'ANI è capace di svolgere determinate funzioni come il riconoscimento facciale, la classificazione di testi, l'analisi di immagini mediche, la previsione della domanda di mercato o la traduzione automatica, ma non possiede alcuna forma di comprensione generale o di adattamento cognitivo. Manca, infatti, della flessibilità e delle capacità cognitive tipiche dell'intelligenza umana.

Di conseguenza, questi sistemi sono vincolati alla supervisione e all'intervento da parte di esseri umani, sia nella fase di progettazione sia nell'utilizzo effettivo.

Un esempio di ANI sono gli assistenti vocali che tipicamente vengono integrati negli smartphone o in dispositivi come, ad esempio, le casse bluetooth. Sebbene questi strumenti possano simulare una conversazione naturale, la loro comprensione linguistica è in realtà piuttosto limitata a pattern prestabiliti e non sono in grado di cogliere pienamente il significato semantico e contestuale delle interazioni umane.

Di conseguenza, tali sistemi riescono a fornire risposte adeguate solo ad un ristretto numero di input previsti, mentre invece falliscono laddove sia richiesto di gestire ambiguità, ironia, o situazioni non contemplate in fase di progettazione.

- Il secondo tipo di intelligenza, quella generale, è pensata invece come un sistema capace di apprendere in modo autonomo, di prendere le conoscenze acquisite in un campo e applicarle in contesti diversi, di ragionare in modo astratto e di adattarsi a situazioni nuove e complesse per le quali non è stata programmata in maniera predefinita.

Questo tipo di sistemi è in grado di lavorare senza la supervisione né l'intervento umano, garantendo adattabilità a diversi contesti e autonomia nelle decisioni avvicinandosi sempre più alle capacità umane.

Attualmente, nonostante i notevoli progressi nel campo dell'AI, non esistono ancora sistemi che possano essere considerati vere e proprie forme di AGI. Ad oggi la ricerca in questo settore è ancora nelle fasi iniziali, e i principali problemi riguardano non solo l'efficienza degli algoritmi, ma anche questioni legate alla comprensione del contesto.

Nonostante ciò, molte sono le aziende che si stanno muovendo in questa direzione.

Secondo il report *“Artificial General Intelligence Market Report 2025”* di StartUs Insights, sono più di 520 le aziende operanti nel settore AGI e tra queste, 370 sono startup⁵. Questo testimonia il grado innovativo del contesto e l'interesse nella ricerca e nello sviluppo di questa tecnologia.

⁵ StartUs Insights (2024). *Artificial General Intelligence Market Report 2025*. <https://www.startus-insights.com/innovators-guide/artificial-general-intelligence-market-report/>

Non solo, molto interessante è anche il livello di investitori attivi i quali ammontano ad oltre 910, con un valore di mercato di 116 miliardi di dollari entro il 2035. Ciò testimonia la convinzione condivisa che siano molti i possibili sviluppi di business model associati all'AGI.

Ad oggi il valore complessivo degli investimenti dei principali investitori nel mercato dell'intelligenza artificiale generale è maggiore di 9,28 miliardi di dollari.

Di seguito la Figura 1 evidenzia i principali attori e come si ripartiscono gli investimenti.

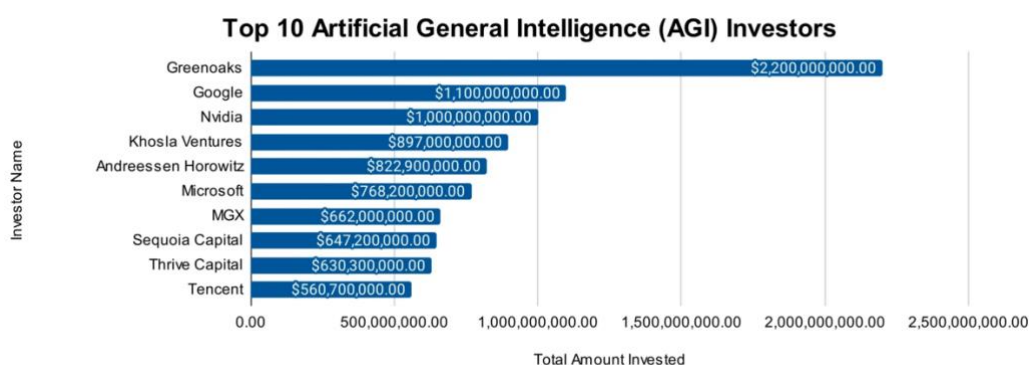


Fig. 1. Artificial General Intelligence Market Report 2025

Per dare un esempio più concreto delle potenzialità di una possibile AGI, possiamo immaginare, una piattaforma installata a bordo di una stazione spaziale, progettata per monitorare in tempo reale diversi parametri, supervisionare lo stato dei sistemi ingegneristici, formulare diagnosi in caso di anomalie e prendere decisioni autonome per consentire la sopravvivenza dell'equipaggio e ottimizzare la riuscita della missione. Un sistema così concepito sarebbe quindi in grado non solo di interpretare grandi moli di dati di diverso tipo, ma anche di apprendere da situazioni non previste, pianificare azioni autonomamente, interagire con esseri umani in linguaggio naturale e modificare il proprio comportamento in base all'evoluzione del contesto in cui opera.

In sintesi, mentre gli attuali modelli di AI mostrano capacità notevoli in compiti ristretti (*narrow*) e predefiniti e sono già comunemente utilizzati in diversi ambiti, lo sviluppo di una vera AGI rappresenta ancora un concetto ideale.

Tra queste due differenziazioni associate all'intelligenza artificiale si potrebbe inserire il protagonista di questa tesi, ovvero gli AI Agents: strumenti molto avanzati che seppur limitati (non hanno una conoscenza illimitata da poter spaziare in ogni ambito), sono dotati di un

impressionante flessibilità che gli permette di prendere decisioni e compiere azioni, anche senza il supporto umano.

Prima di approfondire questo tipo di sistemi, però, è necessario chiarire alcuni concetti associati al funzionamento dell'intelligenza artificiale.

1.1.2 Definizioni e Applicazioni dei modelli di AI

A livello tecnico, questa tesi non vuole approfondire in maniera eccessivamente dettagliata il funzionamento dei sistemi di intelligenza artificiale. Piuttosto, risulta più rilevante ai fini di un'analisi economica e concorrenziale, focalizzarsi solo su quegli aspetti tecnici che risultano critici e determinanti per lo sviluppo di modelli AI e, successivamente, degli AI Agents.

Il primo tema che bisogna affrontare è quello dei **Foundation Model**. Un Foundation Model (FM) è un modello di intelligenza artificiale che è stato addestrato su un'enorme quantità di dati di ogni tipo (testi, immagini, codici, ...), ed è in grado di svolgere diversi compiti, emergendo come uno strumento altamente versatile.

Per rendere più chiaro il concetto, si può immaginare un FM come un cervello digitale che ha letto milioni e milioni di documenti, siti, ha visto immagini e può essere usato per portare a compimento certi obiettivi, anche se non è stato "allenato" specificamente per ognuno di essi. È chiaro quindi da cosa derivi il nome "Foundation": questi strumenti costituiscono la base di molte applicazioni e come conseguenza, anziché creare da zero un nuovo modello ogni volta, vengono utilizzati e riadattati i Foundation Model già addestrati.

Nel loro funzionamento questi modelli di base utilizzano le reti neurali, di cui si è già parlato e in particolare lo fanno attraverso i *Transformer*. Il concetto di *Transformer* è stato introdotto nel 2017 da un team di Google Brain nel paper "*Attention is All You Need*"⁶.

Questo concetto ha rivoluzionato il modo in cui i modelli di AI trattano i dati.

Inizialmente, infatti, i modelli leggevano le parole nell'ordine in cui apparivano: questo approccio rendeva complicato per il modello cogliere i collegamenti tra parole che si trovavano lontane l'una dall'altra nel testo. I *Transformer* rappresentano un'architettura (ovvero un approccio di organizzazione e definizione di come usare le reti neurali) che lavora considerando tutte le parole contemporaneamente e sfrutta un meccanismo chiamato *self-attention*. Questo

⁶ Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017).

https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

meccanismo consiste nel valutare l'importanza delle parole in una sequenza di input in modo da comprenderne meglio le relazioni, ottimizzando così l'uso della potenza di calcolo necessaria per l'addestramento dei modelli⁷.

I Foundation Models, quindi, nella fase di apprendimento “imparano” analizzando molti dati che vengono appresi sotto forma di numeri, tecnicamente chiamati parametri o pesi.

Più parametri possiede un FM, più sarà potente, ma anche più potenza di calcolo richiederà per poter funzionare.

Una volta completata la fase di addestramento del Foundation Model, è possibile adattarlo a svolgere compiti specifici come la comprensione del linguaggio umano (come le chatbot), la scrittura di testi o la creazione di immagini, attraverso un processo chiamato ***fine-tuning***. Questo processo consiste nel “ri-addestrare” il Foundation Model di partenza su un set di dati più specifico per adattarlo meglio ad un certo compito.

Quando i modelli vengono sottoposti a *fine-tuning*, hanno poi la capacità di continuare ad apprendere grazie ai giudizi umani (RLHF – *Reinforcement Learning from Human Feedback*) che permettono loro di perfezionare ulteriormente le proprie risposte.

In conclusione, si può schematizzare come segue il processo di formazione di un Foundation Model:

- Raccolta e controllo dei dati: i FM vengono addestrati su enormi quantità di dati presi dal web, ma questi dati devono essere “puliti” e controllati per evitare problemi di *bias* (ovvero pregiudizi) verso certi temi, violazioni di copyright e di pubblicazione di dati sensibili.
- Addestramento: l'addestramento può essere non supervisionato o auto-supervisionato. Nel primo caso il modello viene “lasciato libero” di scoprire regole o strutture nei dati senza una guida specifica. Nel secondo caso, invece, il modello deve effettuare previsioni a partire da una parte dei dati per scoprire la rappresentazione giusta nel dataset.
- *Fine-tuning*: terminata la fase di addestramento generale, il modello può essere specializzato su compiti specifici.

⁷ IBM (s.d.). *Self-attention: come funziona e perché è importante per l'AI*. IBM Think.
<https://www.ibm.com/think/topics/self-attention>

- RLHF – *Reinforcement Learning from Human Feedback*: in questa fase si confrontano le risposte fornite dal modello con quelle che darebbe un umano e vengono valutate in modo che la macchina riconosca quali siano più vicine al linguaggio umano⁸.

I FM non devono essere confusi con i **Large Language Model (LLM)** che in realtà ne rappresentano un sottoinsieme specifico. Quest'ultimi sono infatti dei modelli focalizzati sul linguaggio naturale: sfruttando le tecniche di fine-tuning vengono addestrati principalmente su grandi quantità di testi (da qui il motivo del termine "*Large*") e sono quindi ottimizzati nel comprendere, generare e interagire in linguaggio umano.

Inoltre, questi modelli sono capaci di apprendimento non supervisionato, ovvero riescono a individuare schemi e regole nei dati senza la necessità di ricevere istruzioni esplicite.

Per via della grande mole di dati richiesta per il loro addestramento, i LLM richiedono un'elevata potenza di calcolo e ciò comporta svantaggi in termini di costi ed efficienza.

Ecco perché nel tempo si sono sviluppati nuovi modelli detti **Small Language Model (SLM)** che a differenza dei primi utilizzano meno dati in fase di addestramento.

Per avere un'idea della differenza di mole di dati utilizzati, IBM indica che gli SLM utilizzano un numero di parametri che varia da pochi milioni a qualche miliardo, mentre i LLM ne utilizzano centinaia di miliardi o addirittura trilioni⁹.

Questa differenza dimensionale si traduce in vantaggi significativi in termini di tempi di addestramento, costi e consumo computazionale.

Tuttavia, la minore quantità di dati e parametri comporta inevitabilmente delle limitazioni: gli SLM tendono a offrire prestazioni inferiori nelle attività più complesse, una minore accuratezza nelle risposte e una maggiore esposizione a bias sistematici.

Per concludere, occorre specificare che ormai i LLM hanno decisamente incrementato le loro capacità, imparando a gestire input di diversi tipi come immagini, video, audio e ovviamente testo¹⁰. Proprio per la loro capacità di lavorare con diverse tipologie di input questi nuovi modelli prendono il nome di **Multimodal Large Language Models (MLLM)**.

⁸ AWS (s.d.). *Cos'è il Reinforcement Learning from Human Feedback (RLHF)?* Amazon Web Services. <https://aws.amazon.com/it/what-is/reinforcement-learning-from-human-feedback/>

⁹ IBM (s.d.). *Small Language Models: cosa sono e perché contano*. IBM Think. <https://www.ibm.com/think/topics/small-language-models>

¹⁰ NVIDIA (s.d.). *Multimodal Large Language Models (MLLMs)*. <https://www.nvidia.com/en-us/glossary/multimodal-large-language-models/>

Questi MLLM sono alla base di molti degli strumenti che utilizziamo quotidianamente, spesso senza che ce ne rendiamo conto. La Figura 2 aiuta a visualizzare e collegare i concetti teorici appena descritti, con i modelli realmente sviluppati e impiegati negli ultimi anni¹¹.

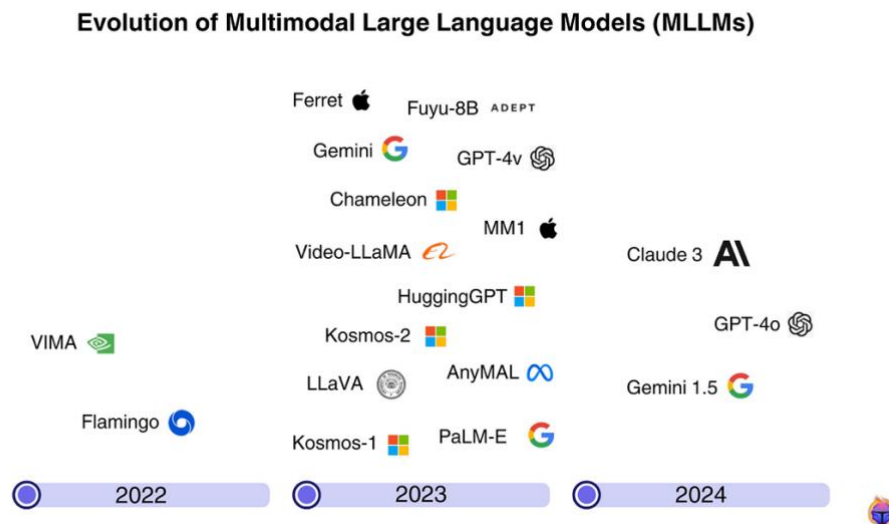


Fig. 2. Some of the Multimodal Large Language Models (MLLMs) developed between 2022 and 2024

1.2 Componenti per lo sviluppo di modelli AI

1.2.1 Hardware per lo sviluppo dei modelli AI

Dopo aver esaminato i modelli che stanno alla base del funzionamento dell'intelligenza artificiale, è importante comprendere quali siano gli strumenti hardware fondamentali per il loro funzionamento e sviluppo. Anche in questo caso l'analisi sarà approfondita sulle componenti fondamentali a livello di catena del valore, tralasciando quelle secondarie.

Possiamo distinguere l'hardware sulla base di tre funzioni:

- 1. Unità di elaborazione:** rappresentano il “motore” di ogni macchina (computer, smartphone, console, ...), si occupano di eseguire calcoli e, più in generale, di trattare i dati. Esistono diversi tipi di unità di elaborazione:

¹¹ Tenyks (2023). *Multimodal Large Language Models (MLLMs) Transforming Computer Vision*. Medium. https://medium.com/@tenyks_blogger/multimodal-large-language-models-mlms-transforming-computer-vision-76d3c5dd267f

- Central Processing Units (CPUs): sono le unità di elaborazione “tradizionali” che si caratterizzano per la loro capacità di gestire calcoli, anche molto complessi, ma in modo sequenziale, uno alla volta.
- Graphics Processing Units (GPUs): nascono nel contesto del gaming, ma oggi giocano un ruolo fondamentale nell’addestramento e il funzionamento dei modelli AI. La loro utilità deriva dalla capacità di elaborare parallelamente un grandissimo numero di dati.

Mentre la CPU è più adatta all’elaborazione di testo semplice e logica, la GPU è in grado di processare efficacemente anche video e immagini.

Nel campo dell’intelligenza artificiale, le GPU sono fortemente preferite rispetto alle CPU perché sono in grado di lavorare molto più velocemente nei processi di addestramento delle reti neurali e quindi dei modelli AI. Questo vantaggio prestazionale comporta però un maggiore consumo energetico rispetto alle CPU¹².

Da questo punto di vista, l’importanza delle GPU nel mercato dell’intelligenza artificiale è confermata dalle sue prospettive di crescita che, come mostrato in figura (Fig. 3), prevedono un aumento da 101,54 miliardi di dollari nel 2025 a circa 1.414,39 miliardi di dollari entro il 2034¹³.

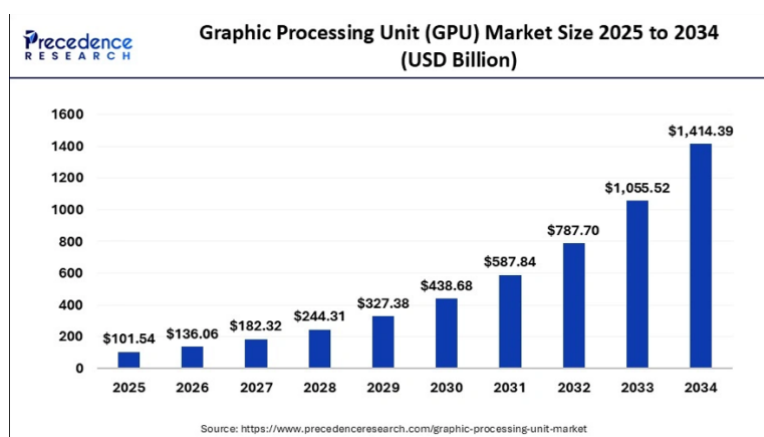


Fig. 1. Graphic Processing Unit (GPU) Market Size, Share and Trends 2025 to 2034¹⁴

¹² Jain, A. (s.d.). *Difference between CPU, GPU, TPU and NPU*. Medium.
<https://medium.com/@abhishekjainindore24/difference-between-cpu-gpu-tpu-and-npu-09fca09f0bb6>

¹³ Precedence Research (2024). *Graphic Processing Unit (GPU) Market – Size, Share and Trends*.
<https://www.precedenceresearch.com/graphic-processing-unit-market>

- Tensor Processing Units (TPUs): sono le unità di elaborazione sviluppate da Google come alternativa alle classiche GPU. Sono specializzate nei lavori di machine learning e deep learning¹⁵.

Da un punto di vista delle prestazioni, le TPU garantiscono risultati sopra la media, ma sono meno flessibili e accessibili rispetto alle GPU. Infatti, non è possibile acquistare TPU sul mercato in quanto Google le offre solo come servizio attraverso la propria piattaforma cloud.

La Figura 4 offre una immediata visualizzazione dei principali attori attivi a livello di unità di elaborazione.

 CPU	 GPU	 TPU
    	    	  

Fig. 2. Processing Unit Producer, Medium, Difference between CPU, GPU, TPU and NPU

- 2. Memoria:** è importante che i dati usati dalle GPU siano conservati in maniera temporanea mentre il modello sta svolgendo più calcoli contemporaneamente. Per questo compito si utilizza la RAM - *Random Access Memory*. Essa “*garantisce che dati e istruzioni siano prontamente disponibili per le unità di elaborazione, consentendo l'esecuzione rapida ed efficiente degli algoritmi di intelligenza artificiale e riducendo i colli di bottiglia nelle operazioni di intelligenza artificiale*”¹⁶.
- 3. Data storage (archiviazione):** se le RAM rappresentano l’archiviazione di memoria volatile (conservata solo temporaneamente), dall’altro lato sono necessari anche strumenti, tipicamente SSD – *Solid State Drive*, che consentano l’archiviazione permanente dei dati. Pur avendo una velocità di accesso inferiore rispetto alla RAM, i SSD permettono di conservare i dati in modo permanente, fino a quando non vengono intenzionalmente eliminati.

¹⁵ Multimodal.dev (s.d.). *What Hardware is Needed for AI?*
<https://www.multimodal.dev/post/what-hardware-is-needed-for-ai>

¹⁶ Intel (s.d.). *AI Hardware: Components and Architecture*. <https://www.intel.com/content/www/us/en/learn/ai-hardware.html>

1.2.2 I dati

I dati rappresentano la base dello sviluppo dei modelli di intelligenza artificiale. Come si è compreso finora, le reti neurali devono lavorare su enormi quantità di informazioni affinché possano apprendere strutture, relazioni e schemi. Non si tratta solo di grandi volumi, ma varietà (a livello semantico e linguistico) e qualità (assenza di errori o bias) sono altrettanto importanti. Da questo punto di vista, avvantaggiate sono tutte quelle aziende che posseggono dati proprietari, raccolti tramite i propri servizi. Questi dati offrono un vantaggio in termini di qualità, ma richiedono risorse economiche e infrastrutturali (data center) significative, infatti, tipicamente sono nelle mani di pochi grandi attori.

Tutta questa mole di dati raccolti, comunemente definita Big Data, sono tipicamente disordinati e poco adatti all'uso immediato nei modelli AI. Ecco perché è necessario un processo fondamentale detto data curation¹⁷. Tale processo si occupa di raccogliere, organizzare, filtrare, descrivere e rendere accessibili i dati affinché possano esser usati per l'addestramento dei modelli.

Dati non curati sarebbero complicati da comprendere per i modelli e questo comporterebbe minori prestazioni da parte di essi.

Occorre infine ricordare che l'utilizzo dei dati non si limita alla fase di addestramento, ma anche a quella di RLHF, ovvero quando il modello è attivo e impara sulla base dei feedback umani. Ecco, quindi, che i dati rappresentano non solo un input tecnico, ma un vero e proprio asset strategico che accompagna i modelli AI in tutte le fasi di vita, dall'addestramento all'utilizzo. Come conferma dell'importanza dei dati da un punto di vista economico e strategico, le previsioni stimano che il mercato globale dei dataset adatti e pronti per l'addestramento dei modelli di intelligenza artificiale valeva circa 2,8 miliardi di dollari nel 2024 e si prevede che raggiungerà i 9,6 miliardi di dollari entro il 2029¹⁸.

1.2.3 Il Cloud Computing

Il cloud computing rappresenta un nodo cruciale in tutta la catena del valore di sistemi AI ed in questo capitolo si cercherà di approfondirne il motivo.

Microsoft Azure definisce il cloud computing come segue:

¹⁷ Lightly (2023). *What is Data Curation and Why is it Important?* <https://www.lightly.ai/blog/what-is-data-curation-and-why-is-it-important>

¹⁸ Markets and Markets (2024). *AI Training Dataset Market by Type, Vertical, and Region – Global Forecast to 2029*. <https://www.marketsandmarkets.com/Market-Reports/ai-training-dataset-market-153819655.html>

“Il cloud computing è la distribuzione di servizi di calcolo, come server, risorse di archiviazione, database, rete, software, analisi e intelligence, tramite Internet ("il cloud"), per offrire la possibilità di innovare rapidamente, risorse flessibili ed economie di scala. Paghi solo per i servizi cloud che usi e risparmi sui costi operativi, esegui l'infrastruttura in modo più efficiente e ridimensioni le risorse in base all'evoluzione delle esigenze aziendali.”

Per meglio comprendere questa definizione e l'importanza del cloud computing può essere utile ricorrere ad un esempio pratico.

Si immagini di voler sviluppare un modello di intelligenza artificiale (per facilità si consideri solo la fase di addestramento), come GPT-3 (lo stesso alla base della famosa chatbot Chat-GPT, sviluppata da Open AI).

Basandosi su alcune stime, l'addestramento di GPT-3 ha richiesto l'utilizzo di 1024 GPU A100 di NVIDIA, per 34 giorni di attività¹⁹. Il costo di acquisto di ciascuna di queste GPU è stimato dagli analisti tra 25.000 e 40.000 dollari²⁰.

Per semplificare l'esempio, si ipotizzi di dover sostenere unicamente il costo per l'acquisto delle GPU, senza considerare invece quelli legati all'energia o alla gestione (che rappresentano comunque un peso rilevante nel costo totale).

Prendendo come riferimento un prezzo medio di 32.500 dollari per GPU, il costo totale da affrontare sarebbe pari a circa 33,3 milioni di dollari (32.500 dollari × 1024 GPU).

Si cambi ora prospettiva e si valuti di affrontare lo stesso investimento, ma affidandosi al cloud computing. Il grafico che segue (Fig. 3) mostra che, il cloud Microsoft Azure nel 2023 ha fornito un'unità di elaborazione grafica modello A100, al prezzo di circa 3,70 dollari all'ora²¹.

¹⁹ Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., Phanishayee, A., & Zaharia, M. (2021). *Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM*. SC21. <https://arxiv.org/pdf/2104.04473>

²⁰ Hawkins Insight (s.d.). *NVIDIA AI GPU Costs and Pricing Strategy Revealed*. <https://www.hawkinsight.com/en/article/nvidia-ai-gpu-costs-and-pricing-strategy-revealed>

²¹ Statista (2023). *Azure GPU Cloud Server Costs Worldwide in 2023*. <https://www.statista.com/statistics/1296926/azure-gpu-cloud-server-costs/>

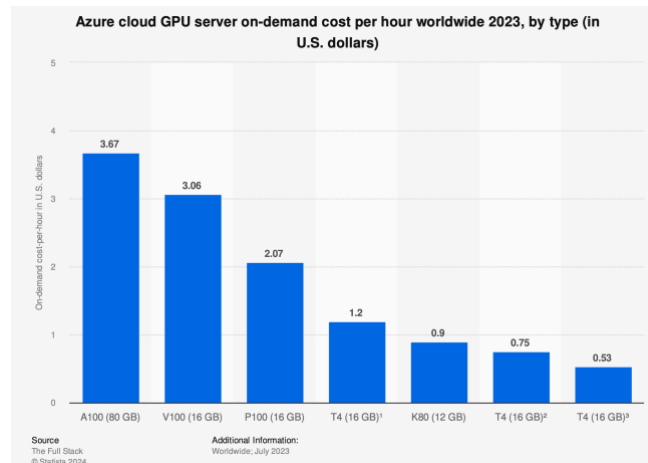


Fig. 3. Azure cloud GPU server on-demand cost per hour worldwide 2023, by type (in U.S. dollars)

Questo costo, che in questo caso comprende anche la gestione associata alle GPU, definisce un investimento complessivo di circa 3,1 milioni di dollari ($3,70 \text{ dollari} \times 24 \text{ ore} \times 34 \text{ giorni} \times 1024 \text{ GPU}$).

Si mostra evidente la differenza di investimento nei due casi: per molte aziende che vogliono sviluppare modelli AI, ma non dispongono di liquidità sufficiente, la soluzione ottimale risiede nel cloud computing.

Ciò a cui si assiste oggi è che solo poche aziende (si entrerà nei prossimi capitoli nel dettaglio) dispongono di sufficiente potenza di calcolo in forma di GPU che vengono raccolte nei data center proprietari; un **data center** è il luogo fisico in cui le aziende conservano computer e apparecchiature hardware. Contiene tutta l'infrastruttura hardware richiesta dai sistemi IT, dunque server, unità di archiviazione dei dati e apparecchiature di rete²².

Ad evidenziare l'importanza del cloud computing, sono le stime che apprezzano tale mercato ad un valore pari a 752,44 miliardi di dollari nel 2024, con una previsione di crescita del 20,4% annuo dal 2025 al 2030, arrivando così ad un valore complessivo di circa 2,4 trilioni di dollari²³. La seguente figura (Fig.) mostra la crescita e la distribuzione geografica globale del mercato nel cloud.

²² AWS (s.d.). *Che cos'è un data center?* Amazon Web Services.

<https://aws.amazon.com/it/what-is/data-center/>

²³ Grand View Research (2024). *Cloud Computing Market Size, Share & Trends Analysis Report 2025–2030*.

<https://www.grandviewresearch.com/industry-analysis/cloud-computing-industry>

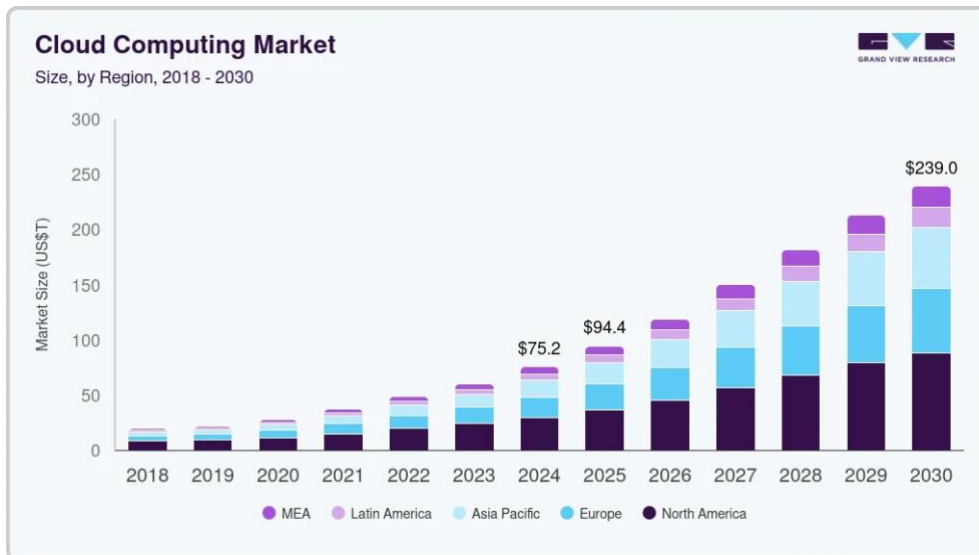


Fig. 1. Cloud Computing Market Size, Share & Trends Analysis Report By Deployment (Public, Private, Hybrid), By Service (IaaS, PaaS, SaaS), By Workload, By Enterprise Size, By End-use, By Region, And Segment Forecasts, 2025 - 2030

1.3 AI Agents

1.3.1 Definizione e panoramica generale degli AI Agents

Compresi i concetti e gli strumenti alla base del funzionamento dell'intelligenza artificiale, è possibile ora affrontare il tema degli AI Agents.

Prima di arrivare ad una definizione tecnica precisa, può essere utile comprendere il contesto attraverso un esempio concreto.

Nel capitolo precedente si è parlato di **Large Language Models – LLM**, definendoli come strumenti in grado di soddisfare molteplici richieste basandosi su dati appresi nella fase di training.

Se si chiedesse ad un qualsiasi LLM di scrivere una mail per fissare un appuntamento, questo strumento funzionerebbe in maniera precisa ed efficace: riceve l'input (la mia domanda), lo elabora in base alle sue conoscenze fornendo in output una risposta (la mail).

Ma cosa accadrebbe se gli si chiedesse quando fissare quell'appuntamento, ovvero quale giorno sarebbe migliore? In questo caso, il sistema non saprebbe rispondere e il motivo è semplice: un LLM, per quanto addestrato possa essere, non ha accesso a informazioni personali come il calendario o gli impegni già pianificati.

Per risolvere questo problema è necessario introdurre il concetto di **AI Workflows**: il sistema deve essere in grado di accedere a strumenti esterni (come l'app calendario, il meteo, ...) in

modo da eseguire flussi operativi per reperire le informazioni personali, analizzarle e successivamente fornire una risposta.

Tuttavia, anche in questi casi, il percorso seguito dal sistema, tecnicamente detto *controllo logico*, è prestabilito da un uomo. Laddove il percorso decisionale del sistema sia predefinito anticipatamente dall'uomo, non si può ancora parlare di AI Agents.

Un **AI Agents**, infatti, è un LLM in grado di decidere autonomamente il percorso da seguire per agire, quali strumenti utilizzare, quali azioni eseguire e in quale ordine, per produrre un output.

Questo comportamento agentic viene descritto dal **Re-Act framework**²⁴, dove Re-Act sta per *Reasoning and Acting*. In questo paradigma, il modello alterna momenti di ragionamento ("*Thought*") e di azione ("*Act*"), ad esempio effettuando ricerche, interrogando un calendario o analizzando dati, fino a generare una risposta finale adeguata.

Da un punto di vista più tecnico gli AI Agents sono quindi “*un sistema o programma in grado di eseguire autonomamente attività per conto di un utente o di un altro sistema progettandone il workflow e utilizzando gli strumenti disponibili*”²⁵.

1.3.2 Funzionamento degli AI Agents

Il funzionamento degli AI Agents è tutt'altro che semplice e si basa sull'integrazione di diverse funzioni che siano in grado di lavorare in maniera fluida, spesso parallelamente.

Per svolgere correttamente un compito, un agente deve innanzitutto comprendere la richiesta dell'utente, inserirla nel contesto appropriato, e quindi analizzare le diverse strategie risolutive possibili. Successivamente, l'agente deve selezionare l'opzione più adatta e procedere alla sua esecuzione. Tuttavia, il processo non termina così: è essenziale che l'agente valuti in modo iterativo gli effetti delle proprie decisioni e, laddove sia necessario, si corregga (Fig. 4).

In questo modo, può apprendere dall'esperienza e affinare le proprie capacità decisionali nel tempo, diventando sempre più efficiente e capace di soddisfare richieste sempre più complesse.

²⁴ Yao, S., Zhao, Y., Yu, D., Zhao, K., Yu, P., & Chen, Y. (2022). *ReAct: Synergizing Reasoning and Acting in Language Models*. arXiv. <https://arxiv.org/abs/2210.03629>

²⁵ IBM (s.d.). *AI Agents: il futuro dell'intelligenza artificiale autonoma*. IBM Think. <https://www.ibm.com/it-it/think/topics/ai-agents>

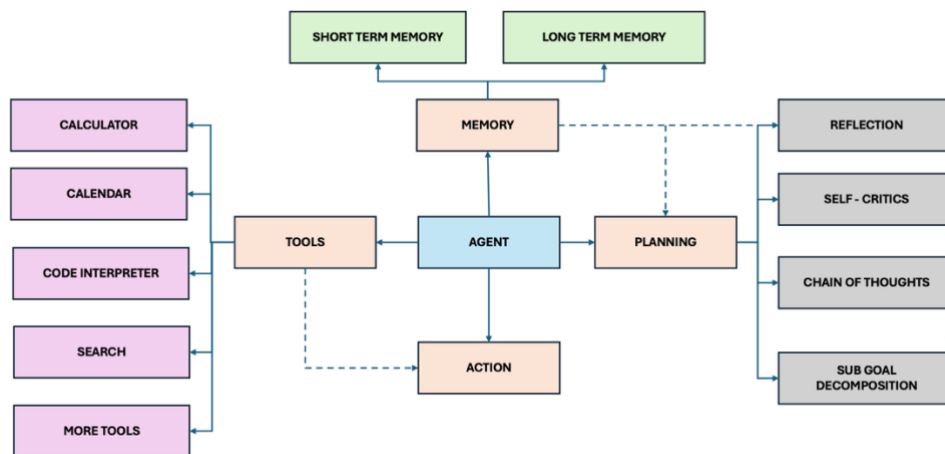


Fig. 2. *What's next for AI agents ft. LangChain's Harrison Chase*²⁶

Si riporta di seguito un esempio di come un AI Agent lavora, andando a considerare le funzionalità tecniche che verranno approfondite successivamente.

1. Richiesta: L'utente richiede di prenotare un pranzo con Luca.
2. Lavoro: l'agente capisce che:
 - serve sapere quando l'utente è libero
 - è necessario sapere chi sia Luca
 - serve inviare un messaggio
3. Memoria: l'agente verifica di conoscere Luca.
4. Tool Calendario: l'agente guarda quando l'utente è libero.
5. LLM: l'agente formula un messaggio in linguaggio naturale.
6. Tool di messaggistica: l'agente invia la notifica a Luca.
7. Fine: l'agente conferma all'utente di aver prenotato ad una certa ora e di aver avvisato Luca.

Come si può notare, alla base del funzionamento dell'agente c'è il **LLM** che funge da “cervello” per l'agente: serve uno strumento per comprendere le richieste e fornire risposte adeguate nel linguaggio umano. Per fare questo le aziende possono affidarsi a FM già sviluppati da terzi, come Open AI, Google, Microsoft, attraverso un'API. Un'**API - Application Programming Interface**, è come un ponte che permette a due software di comunicare tra loro. Permette a un programma (in questo caso all'agente di intelligenza artificiale) di chiedere e ottenere servizi o

²⁶ Chase, H. (2023). *What's next for AI Agents – LangChain*. YouTube.
<https://youtu.be/pBBelpk8hf4>

dati da un altro programma o sistema (LLM), senza dover sapere come esso sia costruito internamente.

In questo processo, l'AI Agent deve eseguire certe azioni e per fare questo si utilizzano strumenti come **LangChain**, che permettono di pianificare (decidere cosa fare prima e dopo), richiamare più modelli o tool (strumenti come le app meteo, calendario, ecc...) e accedere a memoria, al web e ai file.

IBM definisce LangChain come *“un framework di orchestrazione open source per lo sviluppo di applicazioni che utilizzano modelli linguistici di grandi dimensioni”*²⁷.

Per poter agire e imparare autonomamente, l'agente deve necessariamente disporre di memoria e per crearla si usa una **Vector Database**. L'AI utilizza *embedding vettoriali* ossia *“rappresentazioni numeriche di punti dati che esprimono diversi tipi di dati, inclusi dati non matematici come parole o immagini, come un array di numeri che i modelli di machine learning (ML) possono elaborare”*²⁸. Una Vector Database permette di salvare milioni di questi vettori, cercare rapidamente quelli più simili a un nuovo input, aggiornare e organizzare la memoria dell'agente nel tempo.

Nella fase di azione vera e propria l'agente utilizza dei **tool**, ovvero le tradizionali applicazioni: per fissare un appuntamento utilizza il calendario, per inviare una notifica, può usare un servizio di posta elettronica, messaggistica o simili. Infine, ci sarà un'**interfaccia utente - IU** che rappresenta l'ambiente in cui avviene l'interazione tra l'Agente e l'utente.

È cruciale aver chiaro che in tutto questo processo l'utente finale vede esclusivamente ciò che viene mostrato tramite l'interfaccia utente. Egli non si accorge quindi di quali tools l'agente utilizza, né quale percorso ha seguito per portare a termine certi compiti. Nei capitoli successivi sarà evidente quanto ciò possa portare a serie criticità a livello concorrenziale e per l'utilità dell'utente stesso.

1.3.3 Multi Agent System e Orchestrazione

Nei capitoli precedenti si è visto come i Foundation Models, modelli generali, possano essere specializzati in LLM attraverso la fase di *fine tuning*.

Un processo analogo si verifica anche nel contesto agentico, per cui gli AI Agents possono essere specializzati verso funzioni specifiche, diventando così più affidabili ed efficienti nei

²⁷ IBM (s.d.). *LangChain: orchestrazione per modelli linguistici*. IBM Think.

<https://www.ibm.com/it-it/think/topics/langchain>

²⁸ IBM (s.d.). *Vector Embedding: rappresentazione numerica per l'AI*. IBM Think.

<https://www.ibm.com/it-it/think/topics/vector-embedding>

rispettivi ambiti di applicazione. Alcuni agenti, per esempio, possono essere specializzati su processi aziendali, altri su questioni di finance, altri ancora a livello di pianificazione e management. In questo scenario è possibile strutturare insieme di più agenti specializzati che collaborano tra loro: queste strutture prendono il nome di **Multi Agent System – MAS**.

Un MAS, quindi, è una combinazione di più AI Agents che collaborano in un ambiente condiviso per raggiungere obiettivi complessi in maniera più efficace rispetto quanto un singolo agente sarebbe in grado di fare.

Proprio per la natura collaborativa di questi sistemi, diventa cruciale affrontare il tema dell'**orchestrazione degli agenti**.

All'interno di un *Multi-Agent System*, ogni agente svolge un ruolo specifico, spesso altamente specializzato, e per fare in modo che l'intero sistema operi in maniera efficiente è necessario un meccanismo di coordinazione: questo è il compito dell'orchestratore. Si può immaginare l'orchestratore come un agente centrale che supervisiona coordina e gestisce le attività dei singoli agenti, proprio come un direttore d'orchestra.

La sua funzione principale è quella di assicurarsi che l'agente giusto venga attivato nel momento opportuno, in base al tipo di attività da svolgere, allo stato del sistema e agli obiettivi da raggiungere. In questo modo, il sistema nel suo complesso è in grado di far fronte a flussi di lavoro molto complessi, distribuendo i compiti tra gli agenti in base ai loro ruoli.

Sezione 2

Value Chain e Main Actors

2.1 La catena del valore dell'intelligenza artificiale

Prima di andare a definire un'analisi dettagliata associata agli attori attivi nel mercato dell'intelligenza artificiale, è utile considerare la catena del valore. In particolar modo si riprendono i concetti affrontati nel capitolo 1 relativi agli asset necessari per lo sviluppo di modelli di intelligenza artificiale e si relazionano agli aspetti competitivi legati agli attori principali. L'obiettivo è quello di comprendere non solo quali siano le risorse necessarie, ma anche come è strutturato il mercato e quali sono i principali motivi di vantaggio competitivo.

Come si è già visto, cinque sono i segmenti su cui si basa ogni applicazione di intelligenza artificiale: Hardware (chip), Cloud computing, Foundation Models e Modelli Linguistici, Orchestrazione Agentica e Interfaccia Utente.

2.1.1 Hardware – chip

Il primo livello alla base degli sviluppi di intelligenza artificiale è sicuramente la produzione delle unità di calcolo, siano esse GPU, TPU o di altro tipo. Esse rappresentano il motore dei modelli e ne permettono il funzionamento. In questo segmento la concentrazione è elevata, infatti, NVIDIA detiene una posizione quasi di monopolio, nonostante esistano le alternative come le TPU di Google o i chip Trainium e Inferentia di Amazon.

In particolare, il mercato delle GPU, con l'avvento dell'intelligenza artificiale è cresciuto in maniera esponenziale raggiungendo i 125 miliardi di dollari di spesa, trainato da NVIDIA che detiene una posizione dominante con il 92% di market share²⁹.

2.1.2 Cloud Computing

Come il capitolo precedente ha analizzato, il cloud computing rappresenta la modalità più conveniente e diffusa di sviluppo di modelli AI in quanto consente a chi non dispone della liquidità necessaria, di sfruttare una potenza di calcolo che altrimenti sarebbe irraggiungibile.

A questo livello i principali attori sono tre, Amazon Web Services – AWS, Microsoft Azure e Google Cloud Platform.

²⁹ IoT Analytics (2024). *Leading Generative AI Companies to Watch*.
<https://iot-analytics.com/leading-generative-ai-companies/>

In un mercato che vanta una spesa complessiva di 321,3 miliardi di dollari nel 2024 e una crescita costante intorno al 20% annuo³⁰, solo le tre grandi aziende citate contendono oltre il 60% del mercato, con la supremazia netta di Amazon che ne detiene il 33%³¹.

2.1.3 Foundation Models e LLM/MLLM

Sopra il segmento hardware e quello di infrastruttura cloud, si colloca il livello dei Foundation Model e dei Large Language Model. Anche in questo contesto si può osservare che, sebbene siano molte le startup impegnate nello sviluppo di FM e LLM, solo poche raggiungano una quota di mercato considerevole. Tra queste vanno considerate Open AI che primeggia con la maggior market share, pari al 34%, Anthropic segue con il 24%, poi Meta con il 16%, Google DeepMind con il 12%, Mistral il 5% e Cohere il 3%³².

2.1.4 Orchestrazione agentica

I LLM non bastano per lo sviluppo degli AI Agents. Come già visto, occorre anche un sistema di coordinamento che permetta ai modelli di organizzare un flusso di operazioni. In questo livello non si può dire che la struttura di mercato sia definita: in questa fase il mercato dell'orchestrazione è ancora aperto, ma si iniziano a definire le prime supremazie.

LangChain è sicuramente la piattaforma più utilizzata, ma anche altri attori stanno cercando di intervenire come Microsoft con AutoGen³³ o Open AI, Anthropic e Google che invece cercano di integrare la logica di orchestrazione internamente ai loro prodotti.

2.1.5 Interfaccia

L'ultimo livello della value chain è quello con cui il consumatore finale si relaziona direttamente, nonché l'applicazione o, più in generale l'interfaccia stessa.

³⁰ Canalys (2024). *Worldwide Cloud Service Revenues Q4 2024*.

<https://www.canalys.com/newsroom/worldwide-cloud-service-q4-2024>

³¹ Statista (2024). *Worldwide AI funding by quarter*.

<https://www-statista-com.ezproxy.biblio.polito.it/statistics/943151/ai-funding-worldwide-by-quarter/>

³² Medium (2024). *Business and enterprise spending on AI surges 500%*. TechStartups.

<https://techstartups.com/2024/11/21/business-and-enterprise-spending-on-ai-surges-500-to-13-8-billion-as-anthropic-gains-market-share-over-openai/>

³³ Microsoft Research (s.d.). *Project AutoGen*.

<https://www.microsoft.com/en-us/research/project/autogen/>

Da un punto di vista pratico, questa interfaccia è rappresentata dagli AI Agent ed è proprio ciò con cui le aziende possono osservare come l'utente usa il sistema e raccogliere dati con cui migliorarlo in un ciclo continuo che adatta sempre più il prodotto al consumatore finale.

A livello di aziende attive negli sviluppi di AI Agent si entrerà nello specifico nella sezione 3.

2.2 Microsoft

Microsoft è di certo uno degli attori più rilevanti nell'attuale panorama globale dell'intelligenza artificiale. Deve il suo primato nell'attuale contesto economico alla capacità di consolidare la sua posizione in diversi ambiti quali software, infrastrutture, cloud computing e intelligenza artificiale. In questi settori, Microsoft partecipa attraverso la sua piattaforma di cloud computing Microsoft Azure, il suo modello di intelligenza artificiale Copilot e partnership strategiche con alcuni tra i più importanti player nel mondo AI.

2.2.1 Microsoft Azure

Microsoft Azure è la piattaforma attraverso cui Microsoft offre i suoi servizi di cloud computing.

Concretamente parlando, Azure non è altro che una rete globale di oltre 400 data center distribuiti in più di 70 regioni, che permette a chiunque, da aziende a singoli professionisti, di *“utilizzare gli strumenti e i framework che preferiscono per creare, eseguire e gestire applicazioni in ambienti cloud, ibridi ed edge”*³⁴.

Azure, come mostrato nel seguente grafico (Fig. 5), dopo Amazon è la piattaforma preferita nel contesto del cloud computing per via della sua affidabilità, flessibilità e scalabilità che consente ad aziende meno strutturate di adattare le proprie risorse cloud in base alla necessità, pagando esclusivamente per ciò che effettivamente utilizzano.

³⁴ Microsoft (s.d.). *What is Azure?*

<https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-azure>

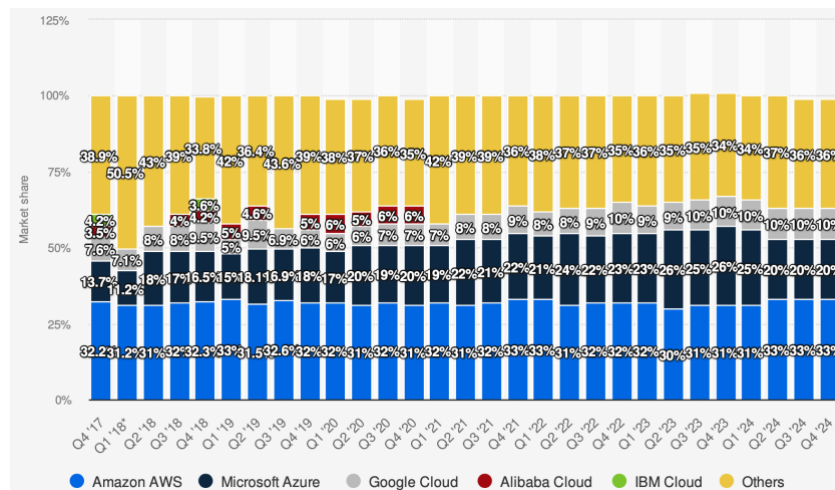


Fig. 3. Cloud infrastructure services vendor market share worldwide from fourth quarter 2017 to fourth quarter 2024³⁵

Nel contesto dell'intelligenza artificiale, Azure offre a Microsoft un vantaggio competitivo cruciale per via di diversi motivi:

1. Potenza di calcolo e scalabilità: l'addestramento di modelli di intelligenza artificiale richiede enormi risorse in termini di calcolo. Azure offre l'infrastruttura necessaria per supportare questi carichi di lavoro, svolgendo quindi un ruolo chiave alla base di tutti gli sviluppi legati all'intelligenza artificiale.
2. Offerta AI-as-a-Service: come visto nei capitoli precedenti, qualunque azienda, anche se piccola, potrebbe sviluppare un AI Agent affidandosi a terzi per certi componenti chiave. Da questo punto di vista, attraverso Azure, Microsoft può offrire servizi di AI preconfigurati, API e strumenti di sviluppo a tutte queste aziende che non dispongono della liquidità necessaria per sviluppare internamente l'infrastruttura adatta ai modelli di AI.
3. Effetti di bundling con l'intero ecosistema aziendale: Microsoft, al di fuori di Azure, dispone di molti altri prodotti adottati da un ingente numero di user, tra questi ci sono Office 365, Dynamics, GitHub, Windows, e altre applicazioni come LinkedIn. Questo permette al colosso tecnologico di integrare il servizio Azure in tutti i prodotti, rendendolo così alla portata di tutti, talvolta senza reale interesse da parte dell'utente finale stesso.
4. Governance e sicurezza: nel trattamento dei dati, uno dei servizi offerto da Azure, cruciale è il tema della sicurezza. La presenza decennale di questa piattaforma la rende

³⁵ Statista (2024). *Worldwide AI funding by quarter*.
<https://www-statista-com.ezproxy.biblio.polito.it/statistics/943151/ai-funding-worldwide-by-quarter/>

una soluzione consolidata e considerata affidabile tanto da essere *“uno dei principali appaltatori governativi che fornisce miliardi di dollari in software e servizi cloud alle agenzie statunitensi, tra cui il dipartimento della Difesa”*³⁶.

Tutto ciò rende Azure una piattaforma particolarmente appetibile per settori altamente regolamentati come finanza, sanità e pubblica amministrazione.

Da questa analisi è ben comprensibile come Azure non rappresenti soltanto una semplice infrastruttura tecnica, bensì un vero e proprio asset strategico attraverso cui Microsoft può consolidare la propria posizione di leadership nel mercato dell'intelligenza artificiale.

Come evidenza di ciò, rilevanti sono due casi antitrust in cui Azure è stata coinvolta:

- Caso 1: nel luglio del 2024 Microsoft ha raggiunto un accordo con CISPE (Cloud Infrastructure Service Provider in Europe, un'associazione europea che riunisce le società del cloud computing) per evitare un'indagine antitrust dell'Ue. CISPE, infatti, accusava Microsoft di limitare la concorrenza nel mercato europeo del cloud per via di licenze discriminatorie: utilizzare software Microsoft su cloud diversi da Azure risultava infatti più costoso e difficile per l'utente finale. Sebbene un accordo sia stato trovato (revisione delle licenze e multa di 20 milioni di euro per il colosso statunitense), il caso è rilevante perché mostra in maniera evidente come Microsoft stia sfruttando la propria posizione dominante nei servizi software (sistema operativo e programmi) per rafforzare il suo cloud³⁷.
- Caso 2: ancora, nel novembre del 2024, il Financial Times³⁸ ha rivelato che la Federal Trade Commission (FTC, Stati Uniti) è in procinto di aprire un'indagine contro Microsoft a causa delle sue pratiche nel cloud Azure. Il sospetto della FTC è che Microsoft stia abusando della posizione dominante nel software per generare lock-in nei clienti verso il cloud Azure. In particolare, c'è il sospetto che Microsoft aumenti le tariffe per chi abbandona Azure, applichi commissioni di uscita elevate e provochi incompatibilità tecnica di Office 365 con i cloud concorrenti³⁹.

³⁶ Il Sole 24 Ore (2024). *Antitrust USA apre un'indagine su Microsoft*.

<https://www.ilsole24ore.com/art/antitrust-usa-apre-un-indagine-microsoft-AGHIsiRB>

³⁷ Start Magazine (2024). *Ecco come Microsoft è scesa a patti con i fornitori cloud in Europa*.

<https://www.startmag.it/innovazione/ecco-come-microsoft-e-scesa-a-patti-con-i-fornitori-cloud-in-europa/>

³⁸ Financial Times (2024). *Microsoft under scrutiny for its role in AI market*.

<https://www.ft.com/content/62f361eb-ce52-47c1-9857-878cfe298d54>

³⁹ Start Magazine (2024). *Stangata Antitrust USA contro il cloud di Microsoft*.

<https://www.startmag.it/innovazione/stangata-antitrust-usa-contro-il-cloud-di-microsoft/>

Entrambi i casi mostrano l'interesse attivo di Microsoft di voler sfruttare la sua posizione dominante per favorire l'adozione e il lock-in verso la sua piattaforma cloud.

In un contesto di intelligenza artificiale, poiché l'accesso a infrastrutture cloud è una condizione fondamentale per lo sviluppo dei modelli, questo rappresenta una seria criticità a livello concorrenziale.

2.2.2 Microsoft Copilot

Microsoft, sfruttando le sue partnership e la sua posizione nel mercato software, ha sviluppato un assistente di intelligenza artificiale chiamato Copilot.

Questo sistema è stato progettato in modo da essere compatibile e direttamente integrato in tutti i prodotti offerti dal colosso americano quali, per esempio, Microsoft 365, GitHub e Window. Concretamente parlando questo sistema funziona come interfaccia tra l'utente e il LLM alla sua base, che altro non è se non GPT-4, sviluppato da Open AI.

Come strumento garantisce un supporto all'utente molto valido per via della sua integrazione con i programmi usati: memorizzando le abitudini e gli interessi dell'utente è in grado di fornire risposte e soluzioni sempre più personalizzate e adeguate:

- Nell'uso di word è in grado di generare testi con un linguaggio adatto, sulla base di documenti precedenti.
- In Teams può riassumere meeting.
- In GitHub può suggerire codici a partire da richieste in linguaggio naturale.⁴⁰

I problemi legati a questo strumento nascono dal momento in cui Microsoft lo integra nei suoi prodotti. In questo modo può infatti aumentare la dipendenza degli utenti dal suo ecosistema e così giustificare un aumento dei prezzi (pratica anticoncorrenziale che prende il nome di *tying*). Allo stesso tempo, basandosi Copilot su LLM eseguiti su cloud Azure, un suo utilizzo rafforza l'uso di Azure stesso, indirizzando clienti e sviluppatori verso l'infrastruttura cloud proprietaria. Oltre a ciò, queste possibilità possono mettere in difficoltà anche gli altri grandi attori come Amazon o Google che non dispongono di così tante piattaforme di lavoro come Microsoft: Word, Excel e Outlook sono strumenti giornalmente utilizzati da una grande mole di utenti e modelli AI concorrenti non hanno la stessa possibilità di integrazione.

⁴⁰ Microsoft (s.d.). *Microsoft Copilot: what it is and how it works*.
<https://www.microsoft.com/en-us/microsoft-copilot/copilot-101/what-is-copilot>

2.2.3 Scelte strategiche

Come reso evidente, Microsoft emerge nel mercato dell'intelligenza artificiale come un attore chiave e si distingue per via di scelte strategiche indirizzate ad un modello orizzontale, basato su partnership e sfruttamento delle proprie infrastrutture cloud.

In questo modo Microsoft riesce ad integrare molte competenze esterne e avere un ecosistema più flessibile e modulare.

Da questo punto di vista, fondamentali per una struttura così fatta sono le partnership con Open AI (che ha dato a Microsoft un vantaggio sostanziale rispetto ai concorrenti) e Inflection AI (che rappresenta un esempio di come Microsoft stia acquisendo talenti in modo da appropriarsi del *know-how* necessario per emergere e mantenere una posizione dominante nel settore AI).

2.2.3.1 Open AI

Open AI viene fondata nel 2015 da un gruppo di investitori, tra cui Elon Musk e Sam Altman, che hanno come obiettivo la realizzazione di un'azienda no-profit che sia in grado di garantire a chiunque sia interessato la possibilità di sviluppare e utilizzare l'intelligenza artificiale a beneficio dell'umanità intera.

Per compiere questa missione però sono necessari ingenti capitali, utili a garantire un'adeguata infrastruttura tecnica e il mantenimento di personale altamente qualificato.

Ecco perché nel 2019, Open AI ha annunciato di passare da ente no-profit ad una struttura ibrida tra una società a scopo di lucro e una no-profit, definita società "capped-profit"⁴¹. In questo modo, come riferisce la società stessa, ciò *“consente di raccogliere capitali di investimento e attrarre dipendenti con un capitale proprio come quello di una startup. Tuttavia, qualsiasi rendimento superiore a tale importo [...] è di proprietà dell'ente no-profit originale di Open AI”*.

Da questo momento, comincia per Open AI un periodo di ristrutturazione molto importante che vede il concretizzarsi di numerose partnership, di cui la più rilevante ed influente è proprio quella con Microsoft.

In prima battuta Microsoft ha effettuato un primo round di investimenti dal valore di un miliardo di dollari, mettendo le basi di quella che sarà poi una duratura collaborazione.

⁴¹ OpenAI (s.d.). *Introducing OpenAI LP*.
<https://openai.com/index/openai-lp/>

Oltre all'investimento monetario, Microsoft ha offerto ad Open AI l'accesso al suo cloud Azure, punto focale nello sviluppo e nell'addestramento dei modelli di deep learning, tra cui GPT-3, attuale GPT-4.

La collaborazione non si esaurisce così, infatti solo nel 2023 avviene la svolta decisiva: Microsoft annuncia un nuovo investimento pluriennale e multimiliardario, stimato in circa 10 miliardi di dollari⁴².

L'accordo decisivo prevede però una serie di clausole non irrilevanti che, da contratto, rimarranno in vigore fino al 2030.

Tra queste, probabilmente, la più importante è quella che riguarda il diritto esclusivo di Microsoft ad integrare la tecnologia Open AI nei propri prodotti⁴³. Come evidenziato nel capitolo precedente, infatti, il modello GPT-4 è integrato in Microsoft Copilot.

Tutto ciò, non impedisce ad Open AI di offrire l'accesso dei propri modelli anche ad altri competitors, ma Microsoft è l'unica azienda a poterli usare in ambito commerciale all'interno dei suoi software.

Un altro punto cruciale del contratto riguarda l'accesso alle API di Open AI che, come affermato dalla Big Tech, sono un'esclusiva di Azure e funzionano su Azure. Concretamente parlando, questo comporta come conseguenza che chiunque voglia utilizzare i modelli come GPT-4 per le proprie applicazioni (o per i propri AI Agents), deve necessariamente farlo attraverso i servizi cloud Azure.

In questo modo Microsoft riesce ad ottenere ricavi sia attraverso le risorse computazionali offerte, sia attraverso la vendita delle licenze API.

Secondo quanto riportato da alcune fonti, Microsoft riceve il 20% dai ricavi generati dalle API di Open AI, percentuale che quest'ultima vuole ridurre al 10% entro il 2030⁴⁴.

Il contratto prevede anche il diritto di prelazione (ROFR, Right of First Refusal) che permette a Microsoft la priorità nell'adozione di qualsiasi tecnologia sviluppata da Open AI, prima che questa possa essere commercializzata ad altri soggetti.

⁴²Semafor (2023). *Microsoft Eyes \$10 Billion Bet on ChatGPT*.

<https://www.semafor.com/article/01/09/2023/microsoft-eyes-10-billion-bet-on-chatgpt>

⁴³ Microsoft (2025). *Microsoft and OpenAI evolve partnership to drive the next phase of AI*.

<https://blogs.microsoft.com/blog/2025/01/21/microsoft-and-openai-evolve-partnership-to-drive-the-next-phase-of-ai/>

⁴⁴ Rivista.ai (2025). *OpenAI taglia la fetta di Microsoft: la nuova partita del potere nell'IA*.

<https://www.rivista.ai/2025/05/07/openai-taglia-la-fetta-di-microsoft-la-nuova-partita-del-potere-nellia/>

Infine, un tema cruciale in questo accordo è legato all'intelligenza artificiale generativa – AGI. Come è emerso negli scorsi capitoli, ad oggi le aziende si stanno muovendo in quella direzione e le stesse Microsoft e Open AI intravedono in essa possibili profitti per 100 miliardi di dollari⁴⁵.

Su ciò grava per Microsoft una clausola contrattuale molto particolare: se Open AI riuscisse a sviluppare una AGI, avrebbe a quel punto la possibilità di limitarne l'accesso a Microsoft.

Considerando le implicazioni economiche e non che una AGI può comportare, questa clausola è stata fissata per evitare conseguenze negative legate all'utilizzo di un'intelligenza artificiale così forte. Tuttavia, alcune fonti affermano che le due aziende stanno lavorando per allentare tale clausola in modo da garantire ulteriori investimenti da parte di Microsoft⁴⁶.

Questo tema si avvicina agli sviluppi di AI Agents da parte di Open AI. Quest'ultima, infatti, ha cominciato a muoversi in questa direzione a partire da inizio 2025 con il lancio di due agenti⁴⁷:

- **Operator**: un agente capace di navigare nel web, cliccare pulsanti, compilare moduli e svolgere funzioni al posto dell'utente.
- **Deep Research**: si occupa di automatizzare la ricerca online; sulla base di una domanda in input, è in grado di cercare fonti, analizzarle, sintetizzarle e produrre report.

Sebbene negli ultimi tempi Open AI stia cercando di allentare i vincoli con il suo maggior investitore, rimane il fatto che molti suoi sviluppi, anche legati agli AI Agents, sono vincolati alla capacità di distribuzione su larga scala offerta dall'ecosistema Microsoft.

Questo significa che anche se Open AI riuscisse a sviluppare potenti AI Agent, l'infrastruttura necessaria per abilitare la loro adozione a livello globale rimane nelle mani di Microsoft.

⁴⁵ Repubblica (2024). *Accordo segreto Microsoft–OpenAI sull'AGI?*

https://www.repubblica.it/tecnologia/dossier/intelligenza-artificiale/2024/12/27/news/microsoft_openai_agi_accordo_segreto_intelligenza_artificiale_generale-423908793/

⁴⁶ Binance (2024). *OpenAI and Microsoft discuss dropping AGI clause to unlock investment.*

<https://www.binance.com/it/square/post/12-06-2024-openai-and-microsoft-discuss-dropping-agi-clause-to-unlock-investment-17201473052113>

⁴⁷ Il Sole 24 Ore (2024b). *OpenAI lancia Operator in Europa: l'AI che naviga e opera per te.*

<https://www.ilsole24ore.com/art/openai-lancia-operator-europa-l-ai-che-naviga-e-opera-web-te-AGAqZMVD>

Questa stretta interdipendenza non è passata inosservata agli occhi delle autorità di regolamentazione della concorrenza. In particolare, la Competition and Markets Authority – CMA, nel 2023 ha avviato un’indagine per comprendere se Microsoft esercitasse o meno un controllo indiretto di Open AI.

Nel 2024 l’indagine è stata poi archiviata con la conclusione che, sebbene Microsoft eserciti un’influenza significativa su Open AI, questa non può ancora essere considerata come una vera e propria fonte di controllo indiretto secondo la legislatura inglese.

Tuttavia, la stessa CMA ha precisato che l’archiviazione dell’indagine “*non costituisce una conclusione secondo cui dal suo funzionamento non derivino preoccupazioni in materia di concorrenza*”⁴⁸.

Questo caso mette in evidenza una preoccupazione legata all’emergere di partnership strategiche tra fornitori di infrastrutture cloud (come si approfondirà, si aggiungono a Microsoft anche Amazon e Google) e aziende impegnate negli sviluppi di AI (Open AI, Anthropic, Mistral AI, ...).

2.2.3.2 Inflection AI

Tra le sue partnership, Microsoft ha stretto legami anche con una delle più promettenti startup americane di intelligenza artificiale, Inflection AI, nota per la sua chatbot Pi (costruita su LLM proprietario).

Stavolta, Microsoft non è intervenuta fornendo infrastruttura o ingenti finanziamenti, ma assumendo delle figure chiave come uno dei due fondatori, Mustafa Suleyman. Oltre ciò ha poi offerto lavoro anche a tutti i dipendenti della startup⁴⁹.

Nello stesso periodo, Inflection AI ha inaspettatamente annunciato di non proseguire la sua attività di sviluppo di modelli AI, ma di trasformarsi in uno studio di ricerca.

⁴⁸ Reuters (2025). *UK says Microsoft-OpenAI partnership does not need anti-trust investigation*. <https://www.reuters.com/technology/uk-says-microsoftopenai-partnership-does-not-need-anti-trust-investigation-2025-03-05/>

⁴⁹ Brian Solis (2024). *AI(n)sights: Microsoft’s Pseudo Acquisition of Inflection AI*. <https://briansolis.com/2024/04/ainsights-microsofts-pseudo-acquisition-of-inflection-ai-and-what-it-means-to-business-customers-and-investors/>

Ciò a cui si è assistito quindi è un'acquisizione velata: assumendo il personale migliore, Microsoft ha di fatto assorbito l'azienda evitando qualsiasi tipo di implicazione legale connessa a problemi concorrenziali.

In tutto ciò, sebbene la Commissione Europea abbia accusato Microsoft, nessuno Stato ha chiesto di indagare sul caso e ciò ha interrotto l'indagine (in considerazione dell'articolo 22 del Regolamento Europeo sulle Concentrazioni)⁵⁰.

Ciò che sicuramente si può notare da questo caso è che Microsoft sta cercando di acquisire *know-how* in ambito AI per ridurre la sua forte dipendenza da Open AI.

2.3 Alphabet - Google

Il secondo grande attore nel contesto dell'intelligenza artificiale è di certo Google. L'approccio adottato da Alphabet, la holding di cui Google fa parte, è particolarmente verticalizzato con l'obiettivo di combinare infrastrutture proprietarie, modelli di intelligenza artificiale sviluppati internamente e prodotti e servizi in grado di abilitare nuovi sviluppi e garantire l'accesso ai dati. Anche in questo capitolo, l'analisi della Big Tech viene strutturata analizzando innanzitutto il servizio di cloud computing offerto attraverso Google Cloud Platform, il modello di AI che è Gemini e la scelta strategica di progettare e produrre internamente i chip (le TPU) oltre a integrare la ricerca attraverso il suo laboratorio specializzato in AI, Google DeepMind.

2.3.1 Google Cloud Platform

Google Cloud Platform rappresenta un asset strategico fondamentale per Alphabet in ambito di intelligenza artificiale. Nel mercato del cloud computing, Google si posiziona come il terzo miglior fornitore dopo Amazon e Microsoft, con una crescita notevole in termini di ricavi che sono passati dai 13 miliardi di dollari del 2020 ai 43 miliardi del 2024 (pari a oltre il 10% del fatturato totale di Google)⁵¹.

Questa crescita è trainata dalla sempre maggiore richiesta di servizi cloud associati all'intelligenza artificiale e uno degli elementi che rende Google Cloud Platform altamente competitivo, è la disponibilità delle Tensor Processing Units.

⁵⁰ European Commission (2024). *Commission opens investigation into Microsoft's practices in the cloud sector.* https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4727

⁵¹ Statista (2024). *Google Cloud public revenue worldwide 2017–2024.* <https://www.statista.com/statistics/478176/google-public-cloud-revenue/>

Questi chip, particolarmente competitivi rispetto alle alternative offerte da NVIDIA in termini di prestazioni e costo, sono accessibili ai clienti solo tramite Google Cloud Platform⁵².

Un altro elemento di vantaggio competitivo per Google Cloud Platform è legato all'accesso ad un enorme quantità di dati, generata da servizi come il Search, YouTube, Maps, Gmail,

Ciò garantisce a Google un vantaggio in termini di addestramento dei modelli AI che l'azienda statunitense può offrire attraverso il suo servizio di cloud. Non solo, Alphabet può sfruttare la sua posizione dominante nel Search e nella pubblicità online per promuovere i propri servizi cloud e AI, limitando quindi la concorrenza.

Non a caso, la FTC ad inizio 2025 si è mossa contro Google avviando indagini per comprendere come l'azienda stia usando i dati e le infrastrutture cloud in un eventuale abuso di posizione dominante⁵³.

C'è poi un discorso legato al leverage effect che Alphabet avrebbe la possibilità di praticare: Google Cloud Platform potrebbe, per esempio, promuovere Gemini (il modello AI proprietario) e questo si tradurrebbe non solo in un rafforzamento nella sua posizione del cloud, ma anche nella fornitura di modelli.

Concettualmente parlando, la logica con cui Google utilizza la sua piattaforma cloud è uguale a quella di Microsoft con Azure: l'azienda sviluppa internamente il suo modello che viene messo a disposizione alle imprese attraverso le API, ma non solo, lo stesso modello viene integrato nelle app mobili, nei motori di ricerca e negli altri servizi offerti.

2.3.2 Google Gemini

Nel mercato dei modelli AI, Gemini rappresenta il tentativo di Google di competere direttamente con i principali modelli disponibili, tra cui quelli di Open AI e Anthropic.

Gemini nasce dal laboratorio di ricerca Google DeepMind e deve la sua particolarità al fatto di essere nato sin da subito sulla base di capacità multimodali: è capace non solo di generare testo, ma di lavorare anche con immagini, codici, video e audio.

Il fatto che Gemini sia stato progettato sin dalla nascita per gestire input multimodali (a differenza degli altri modelli che prima sono stati addestrati per lavorare su testi e poi su altri

⁵² Bytebridge (2023). *GPU and TPU Comparative Analysis Report*. Medium.

<https://bytebridge.medium.com/gpu-and-tpu-comparative-analysis-report-a5268e4f0d2a>

⁵³ Federal Trade Commission – FTC (2025). *FTC Issues Staff Report on AI Partnerships and Investments Study*. <https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study>

tipi di input), lo caratterizza di prestazioni superiori rispetto agli altri competitor, in termini di interpretazione e generazione delle immagini e analisi di contenuti audio.

Un'altra particolarità sta nel fatto che Gemini non è solamente un modello, ma una piattaforma integrata nel sistema Google. Questo emerge in due contesti:

- Gemini come prodotto: si intende il modello AI integrato in Android, Chrome, Gmail e altri prodotti o servizi offerti. Può aiutare nella scrittura delle mail, nel riassumere conversazioni, generare immagini,
- Gemini come piattaforma AI: Gemini è anche accessibile agli sviluppatori attraverso Google Cloud Vortex AI, una piattaforma dove questi hanno la possibilità di personalizzarlo per meglio integrarlo con i dati e le strutture aziendali proprie⁵⁴.

In questa maniera, Alphabet ha la possibilità di ottenere ricavi da due mercati che sono il B2C (i consumatori che utilizzano il modello) e B2B (attraverso i servizi personalizzati offerti sulla piattaforma cloud proprietaria).

Come avviene però per quasi tutte le Big Tech, anche in Google si intravedono problemi anticoncorrenziali legati all'utilizzo dei loro modelli. Bisogna infatti notare che Google integra in maniera predefinita i propri modelli nei dispositivi Android e nelle sue app di Search e questo rappresenta un fenomeno comunemente chiamato *self-preferencing* per cui Google, attraverso la sua piattaforma (Android o Chrome, per esempio), favorisce i propri prodotti (modello Gemini).

2.3.3 Scelte Strategiche

Oltre alla fornitura dei servizi da tutti conosciuti (Search, Gmail, ...) e allo sviluppo di modelli AI, Google ha deciso di strutturarsi nel contesto di intelligenza artificiale in maniera verticale. Per fare questo, da un lato ha acquisito e rafforzato un laboratorio di ricerca che oggi prende il nome di Google DeepMind e dall'altro lato si è impegnata nello sviluppo di chip pensati specificatamente per l'addestramento dei modelli AI, le Tensor Processing Unit.

2.3.3.1 Google DeepMind

Nel contesto della strategia di integrazione portata avanti da Alphabet, un ruolo cruciale è svolto da Google DeepMind. Questo laboratorio di ricerca ha permesso

⁵⁴ Google Cloud (s.d.). *Gemini: l'AI multimodale di Google*.
<https://cloud.google.com/ai/gemini?hl=it>

a Google di collocarsi in una posizione di vantaggio a livello di sviluppo di modelli AI e delle loro applicazioni.

DeepMind non nasce internamente a Google, bensì come startup indipendente fondata da un gruppo di esperti, tra cui Mustafa Suleyman (oggi in Microsoft). L'obiettivo della startup era quello di spaziare nella ricerca associata all'intelligenza artificiale generale. Il primo risultato è l'applicazione del *Deep Reinforcement Learning – DRL*, un'attività svolta in fase di addestramento dei modelli che li implementa nella valutazione e decisione⁵⁵.

DeepMind viene acquisita da Google nel 2014, ma continua a lavorare in maniera indipendente all'interno di Alphabet, fino a quando, nel 2023, viene fusa con il laboratorio di ricerca già attivo in Google, Google Brain.

Questa scelta nasce come conseguenza necessaria al contesto altamente competitivo in cui Microsoft acquisisce sempre più una posizione dominante attraverso la partnership con Open AI.

L'obiettivo di Google DeepMind è quindi quello di rendere competitiva Google nel mercato dell'intelligenza artificiale e il primo grande risultato è rappresentato da Gemini, il modello di Google che nasce sin da subito con la capacità di saper gestire testo, immagini, codici e altri tipi di input (di fatto è un Multinodal Large Language Model - MLLM).

Google DeepMind non si ferma allo sviluppo di MLLM, ma si impegna anche nello sviluppo di AI Agents. In questo campo i risultati sono due:

- **Project Astra:** proposto al mercato nel 2024, Project Astra è una piattaforma in grado di elaborare diversi tipi di input, mantenendo però memoria delle iterazioni avvenute (un importante passo avanti verso la definizione degli AI Agents più affidabili, coerenti e reattivi)⁵⁶.
- **Project Mariner:** è un agente che si concentra nella navigazione nel web. Mariner è in grado di navigare nel web autonomamente, proprio come farebbe un essere umano usando il computer. Per esempio, è in grado di visitare più siti web di più aziende e raccogliere i dati di contatto⁵⁷.

⁵⁵ DeepMind (s.d.). *Deep Reinforcement Learning*.
<https://deepmind.google/discover/blog/deep-reinforcement-learning/>

⁵⁶ DeepMind (s.d.). *Project Astra*.
<https://deepmind.google/models/project-astra/>

⁵⁷ Il Sole 24 Ore – InfoData (2024). *Gemini 2.0, Project Mariner e Deep Research: Google è entrata nell'era degli AI Agent*.

In conclusione, è ben chiaro come DeepMind non rappresenti solamente un asset strategico a supporto di Google, ma un vero e proprio punto di forza per riuscire a guidare gli sviluppi in AI da protagonista.

2.3.3.2 Tensor Processing Unit e Ironwood Chip

A differenza di Microsoft che utilizza le classiche GPU, acquistate e integrate nei propri data center, Alphabet ha maggiormente investito per sviluppare internamente i chip tipicamente usati in ambito AI.

In particolare, ha progettato le Tensor Processing Unit (TPU), chip specifici per accelerare i carichi di lavoro legati al machine learning.

L'obiettivo di Alphabet è quello di ridurre la dipendenza rispetto un fattore chiave (quale le GPU) nello sviluppo dei modelli AI, di cui NVIDIA rappresenta il principale fornitore.

A livello prestazionale, le TPU sono in grado di offrire risultati migliori nelle fasi di inferenza (la fase in cui un modello già addestrato viene utilizzato per generare output). Una TPU v3 è in grado di completare un'inferenza in 1,7 millisecondi, mentre una GPU V100 richiede 3,8 millisecondi. Allo stesso modo le TPU riducono i tempi di addestramento che, per un modello come ResNet-50, richiede 40 minuti se si utilizzano le GPU e 15 se si utilizzano le TPU⁵⁸.

Da un punto di vista economico le GPU sono tipicamente più accessibili: possono essere acquistate o utilizzate attraverso servizi cloud. Per quanto riguarda le TPU, esse sono invece disponibili solo attraverso la piattaforma di Google e hanno costi orari maggiori. Se però si considera anche il tempo impiegato per le operazioni, nel complesso le TPU sono economicamente più convenienti.

2.4 Amazon

Amazon è un altro attore fondamentale nel contesto AI, ma rispetto ad altri concorrenti ha scelto un approccio diverso: piuttosto che investire in un modello proprietario che potesse fronteggiare GPT o Gemini, ha preferito posizionarsi a livello di fornitura infrastrutturale per l'ecosistema AI.

<https://www.infodata.ilsole24ore.com/2024/12/15/gemini-2-0-project-mariner-e-deep-research-google-e-entrata-nellera-degli-ai-agent-datavizandtools/>

⁵⁸ Analytics Insight (s.d.). *Google TPUs vs NVIDIA GPUs for AI Development*. Medium.

<https://medium.com/@analyticsinsight/google-tpus-vs-nvidia-gpus-for-ai-development-3f62fa2392d7>

Come per Microsoft e Google, si analizzerà prima Amazon Web Services – AWS, la divisione cloud dell'azienda nonché punto cruciale nella catena del valore.

Successivamente si proporrà un'analisi della piattaforma Amazon Bedrock e di come questa sia utilizzata per fornire diversi modelli AI tra cui, in particolare, Anthropic.

Sarà infine evidente come l'obiettivo di Amazon sia quello di offrire tutti gli strumenti per costruire soluzioni AI, imponendosi come piattaforma abilitante nello sviluppo dell'AI.

2.4.1 Amazon Web Services

Amazon Web Services – AWS rappresenta il cuore pulsante dell'intero gruppo Amazon. Nel 2024 i ricavi si aggirano intorno ai 100 miliardi di dollari rispetto i 600 miliardi dell'intero gruppo⁵⁹. Di fatto AWS, in funzione dei profitti che è in grado di rendere, rappresenta la maggior fonte di cui Amazon può contare per investimenti futuri legati all'AI.

A livello di mercato cloud, Amazon Web Services detiene il 33% di market share, seguita da Microsoft e Google. Questo si riflette nell'utilizzo da parte della maggior parte delle startup e aziende dei servizi cloud offerti da AWS per l'addestramento e lo sviluppo dei modelli AI. L'azienda ha oltre 25 sedi nel mondo e più di 200 data center utilizzati da attori molto importanti a livello globale, tra cui Netflix, Disney, il Ministero della Giustizia del Regno Unito, Kellogg's, Guardian News and Media e l'Agenzia Spaziale Europea⁶⁰.

L'elevato potere in mano ad AWS non è passato inosservato alle autorità antitrust. Nel Regno Unito, la Competition and Markets Authority – CMA ha infatti avviato un'indagine nei confronti dei principali attori in ambito cloud computing e tra questi AWS.

Ad inizio 2025 la CMA conclude che la preoccupazione maggiore riguarda la concentrazione elevata nel mercato dove AWS e Microsoft detengono le principali quote. Questo potrebbe portare a designare AWS e Microsoft come detentori di uno status di mercato strategico (SMS) nei servizi cloud e ciò permetterebbe alla CMA di attuare politiche molto più restrittive nei confronti delle Big Tech⁶¹.

Anche in Europa, la Commissione Europea ha designato Amazon come gatekeeper, ovvero un'azienda con una tale struttura da dover rispettare obblighi e limitazioni definite nel Digital

⁵⁹ Statista (2024). *Amazon Annual Net Revenue Worldwide (2004–2024)*.

<https://www-statista-com.ezproxy.biblio.polito.it/statistics/266282/annual-net-revenue-of-amazoncom/>

⁶⁰ Statista (2024). *Amazon Web Services Revenue (2013–2024)*.

<https://www-statista-com.ezproxy.biblio.polito.it/statistics/233725/development-of-amazon-web-services-revenue/>

⁶¹ UK Government – Competition and Markets Authority (2024). *Cloud Services Market Investigation – Provisional Findings*.

<https://www.gov.uk/guidance/cloud-services-market-investigation-provisional-findings>

Markets Act – DMA, come ad esempio il divieto di praticare self-preferencing, l'obbligo di interoperabilità e portabilità dei dati e requisiti di trasparenza nelle condizioni e pratiche commerciali⁶².

A basarsi su AWS è Amazon Bedrock, una piattaforma di machine learning utilizzabile per sviluppare applicazioni di intelligenza artificiale, sfruttando modelli di base (FM) già disponibili. In questo modo non è necessario addestrare ogni volta i modelli e ciò permette agli sviluppatori di ridurre i costi. Tra i modelli offerti da Amazon Bedrock ci sono quelli sviluppati da diverse startup, tra cui AI21 Labs, Anthropic, Cohere e Stability AI⁶³.

Attraverso questa piattaforma Amazon vuole quindi offrire la possibilità ai suoi consumatori di poter sperimentare e usare diversi modelli per i loro rispettivi casi specifici, personalizzandoli attraverso addestramenti *fine-tuning* basati sui propri dati.

In realtà, Amazon Bedrock non si limita a fornire solo modelli, ma anche agenti AI attraverso Amazon Bedrock Agents. Amazon definisce i suoi agenti come capaci di utilizzare *“il ragionamento dei modelli di fondazione (FM), delle API e dei dati per suddividere le richieste degli utenti, raccogliere informazioni pertinenti e completare le attività in modo efficiente, consentendo ai team di concentrarsi su lavori di alto valore”*⁶⁴.

2.4.2 Anthropic

In relazione allo sviluppo di modelli AI, sebbene Amazon stia proponendo il suo modello Nova, non si può dire che sia impegnata allo stesso modo di Microsoft e Google.

Piuttosto che sviluppare sin da subito un modello AI, Amazon ha preferito stringere alleanze con diverse startup in AI. Tra queste, la più rilevante è Anthropic, fondata nel 2021 da ex dirigenti di Open AI.

Ad oggi, Amazon ha investito un totale di 8 miliardi di dollari nella startup in intelligenza artificiale attraverso obbligazioni convertibili che nel 2024 sono state in parte convertite in azioni⁶⁵. Nonostante ciò, Amazon mantiene una posizione di minoranza che non le garantisce

⁶² European Commission (2022). *Regolamento (UE) 2022/1925 del Parlamento Europeo e del Consiglio del 14 settembre 2022 (Digital Markets Act)*. EUR-Lex.

<https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32022R1925>

⁶³ TechTarget (s.d.). *Amazon Bedrock (AWS Bedrock) – Definition*.

<https://www.techtarget.com/searchenterpriseai/definition/Amazon-Bedrock-AWS-Bedrock>

⁶⁴ Amazon Web Services (s.d.). *Bedrock Agents*.

<https://aws.amazon.com/it/bedrock/agents/>

⁶⁵ The Guardian (2024). *Amazon Invests Billions in Anthropic Amid AI Race*.

<https://www.theguardian.com/technology/2024/nov/22/amazon-anthropic-ai-investment>

alcun controllo e, conseguentemente, non implica provvedimenti dalle autorità garanti della concorrenza.

Oltre all'investimento monetario, è stato stipulato anche un accordo che prevede l'utilizzo esclusivo da parte di Anthropic dei chip AI proprietari di Amazon, Trainium e Inferentia, e della sua infrastruttura cloud per addestrare e distribuire i propri modelli.

Il più noto tra questi è Claude, oggi disponibile attraverso la piattaforma Amazon Bedrock.

Oltre a ciò, Anthropic partecipa nel legame con Amazon contribuendo nello sviluppo dei chip proprietari Trainium e Inferentia⁶⁶.

Questo legame Amazon-Anthropic può essere paragonabile alle già discusse integrazioni Microsoft-Open AI e Google-Gemini, nonché la risposta di Amazon ad un contesto sempre più competitivo e dinamico.

Il quadro di Anthropic non si esaurisce però così: oltre ad Amazon anche Google ha investito massicciamente nella startup arrivando a detenere una quota del 14%, pur senza esercitare diritti di voto o controllo⁶⁷. Oltre a ciò, Google ha promesso ulteriori investimenti e da parte sua Anthropic ha cominciato ad utilizzare anche i chip e i servizi cloud di Google per l'addestramento dei suoi modelli.

È evidente come entrambe le aziende vogliano adottare un modello di partecipazione che, seppur non dominante, le rafforza dal punto di vista del cloud: Amazon integra i modelli di Anthropic in Amazon Bedrock, mentre Google, seppur senza esclusività, fa sì che Anthropic utilizzi le sue TPU.

2.4.3 Scelte strategiche

Da un punto di vista strategico, Amazon si distingue da Microsoft e da Google per un approccio unico e differente. Invece di impegnarsi nella creazione di un proprio Large Language Model, Amazon preferisce posizionarsi dietro le quinte come fornitore di infrastruttura per l'ecosistema dell'intelligenza artificiale.

Una delle prime azioni in questo senso, è stata lo sviluppo di chip proprietari, quali Trainium e Inferentia⁶⁸: il primo è sviluppato per l'addestramento dei modelli AI, il secondo per ottimizzare

⁶⁶ Reuters (2024). *Anthropic Receives \$4 Billion Investment: Amazon Makes AWS Official Cloud Provider.*
<https://www.reuters.com/technology/artificial-intelligence/anthropic-receives-4-billion-investment-amazon-makes-aws-official-cloud-provider-2024-11-22/>

⁶⁷ AINews (2024). *Il NYT scopre che Google possiede il 14% di Anthropic: le aziende sono davvero avversarie?*
<https://ainews.it/il-nyt-scopre-che-google-possiede-il-14-di-anthropic-le-aziende-sono-davvero-avversarie/>

⁶⁸ Start Magazine (2024). *Amazon sfida Nvidia con i suoi microchip per l'intelligenza artificiale.*
<https://www.startmag.it/innovazione/amazon-microchip-intelligenza-artificiale-nvidia/>

la fase di inferenza. Questi chip rappresentano il tentativo di Amazon di ridurre la sua dipendenza da NVIDIA e rifornisco già molti dei datacenter di Amazon Web Services.

Un'altra strategia legata al posizionamento che l'azienda ha scelto, è quella di offrire non solo un unico modello, ma una pluralità di modelli sviluppati da terze parti, come Anthropic (Claude), Stability AI, Meta (LLaMA), Cohere e altri.

In questo modo i consumatori posso scegliere tra più possibilità, effettuare personalizzazioni, il tutto all'interno di AWS.

Dunque, si può dire che Amazon non vuole offrire un modello unico con il quale imporsi, ma piuttosto gli strumenti necessari (quindi chip, modelli AI, API, Agenti, ...) per aiutare i consumatori a definire autonomamente come strutturare le proprie applicazioni di intelligenza artificiale. Amazon vuole quindi offrire tutta l'infrastruttura alla base degli sviluppi AI.

2.5 Meta

Nel contesto di intelligenza artificiale, un profilo rifilato, ma comunque importante è assunto da Meta che si approccia strategicamente in maniera lontana dalle altre Big Tech.

Da un punto di vista strategico, infatti, l'azienda a cui fa capo Mark Zuckerberg si muove su due assi: da un lato investe in maniera massiccia in infrastrutture cloud e potenza computazionale, dall'altro punta ad offrire un modello open source chiamato LLaMA – Large Language Model Meta AI.

In questo capitolo si affronterà il discorso relativo alle infrastrutture, considerando il fatto che Meta non è da considerarsi un fornitore cloud al pari degli altri concorrenti, ma che comunque utilizza in maniera massiccia data center per il funzionamento delle sue applicazioni e servizi. Si analizzerà il modello LLaMA che, a differenza dei modelli concorrenti, ha la particolarità di essere Open Source, e si cercherà poi di comprendere il motivo di questa scelta strategica.

2.5.1 Infrastruttura proprietaria

L'interesse di Meta nell'intelligenza artificiale è ancor più reso evidente dalle dichiarazioni di Mark Zuckerberg che ad inizio anno ha annunciato un piano di spesa in conto capitale per il 2025 pari a 60-65 miliardi di dollari, una cifra considerevolmente alta se confrontata con i 38-40 miliardi spesi nel 2023⁶⁹.

⁶⁹ HDblog (2024). *Meta, Zuckerberg raddoppia: investimenti in AI per 30 miliardi di dollari.*
<https://www.hdblog.it/mercato/articoli/n606566/meta-zuckerberg-ai-investimenti-2025-miliardi/>

Nello specifico, una quota di questi investimenti sarà destinata all'acquisizione di personale altamente qualificato in AI, mentre la restante quota sarà destinata alla costruzione di data center, fattore chiave non solo nello sviluppo di modelli AI, ma anche per il funzionamento delle principali piattaforme offerte da Meta, quali Facebook, Instagram e WhatsApp.

Sempre Zuckerberg promette di raggiungere entro la fine dell'anno la quota di 1,3 milioni di GPU, un considerevole impegno specie in un contesto di carenza globale di GPU dove tutti i competitor competono per assicurarsi le risorse prodotte da leader come NVIDIA⁷⁰.

Già nel 2023 Meta si era impegnata nella riconversione dei suoi data center per ottimizzarli ai carichi di lavoro richiesti dalle applicazioni di intelligenza artificiale, ma specialmente in questo periodo storico in cui si va sempre più concretizzando la joint venture Stargate, Meta ha compreso che deve accelerare per non rimanere indietro.

2.5.1.1 Stargate

Nel 2025, con l'annuncio del progetto Stargate si è verificata una svolta significativa in ambito AI. Stargate è una joint venture dal valore di 500 miliardi di dollari che nasce con l'obiettivo di dotare gli Stati Uniti della più grande infrastruttura computazionale a livello globale⁷¹.

La società, promossa dall'amministrazione Trump, ha trovato l'appoggio di importanti player quali Open AI, SoftBank, Oracle e MGX. Oltre a partner economici ci sono altri tecnologici tra cui, NVIDIA, Oracle, Arm e Microsoft.

In vista di ciò, Meta ha reagito con i massicci investimenti promessi, ma anche cercando un riavvicinamento con l'amministrazione Trump in considerazione delle vicissitudini passate tra Zuckerberg e il presidente americano⁷².

2.5.2 LLAMA

Meta si è impegnata nello sviluppo dei suoi modelli di intelligenza artificiale LLaMA, ma ha voluto renderli open source. Un modello open source è un sistema che, una volta addestrato, viene reso accessibile a chiunque: tutti possono scaricarlo, scoprirne il funzionamento e

⁷⁰ MSN (2024). *Meta: previsti 1,3 milioni di GPU per l'AI entro la fine del 2025.*

<https://www.msn.com/it-it/money/storie-principali/meta-previsti-1-3-milioni-di-gpu-per-l-ai-entro-la-fine-del-2025/ar-AA1xO86O?ocid=finance-verthp-feeds>

⁷¹ BeBeez (2024). *Data center negli USA: al via il progetto Stargate di OpenAI da 500 miliardi di dollari.*

<https://bebeez.it/rebreakfast/datacenter-negli-usa-al-via-il-progetto-stargate-di-openai-da-500-mld/>
⁷² Bloomberg (2025). *Meta CEO Mark Zuckerberg Bets on Trump — but MAGA Is Skeptical.*

<https://www.bloomberg.com/news/features/2025-05-26/meta-ceo-mark-zuckerberg-bets-on-trump-but-maga-is-skeptical>

modificarlo per adattarlo alle proprie necessità. Tutto si fonda sul rendere esplicito il codice e i dati, permettendo all'intera comunità di contribuire al suo sviluppo.

La scelta di un modello open source è legata ad una duplice considerazione: da un lato si stimola l'adozione del modello e la consapevolezza da parte dei consumatori, dall'altro si accelera il processo di sviluppo perché molti sono gli sviluppatori e i programmatori che ci lavorano.

L'ultima versione di LLaMA, secondo quanto affermato da Chris Cox, responsabile dei prodotti Meta, sarà in grado di sviluppare AI Agents in grado di navigare nel web, completare compiti e supportare il consumatore nei flussi di lavoro⁷³.

C'è da considerare un ultimo aspetto legato alla scelta di offrire un open source ed è legato ai temi antitrust: come sistema aperto LLaMA sarà meno soggetto a controlli in quanto rende più evidenti tutti gli aspetti legati all'interoperabilità, alla concorrenza e all'utilizzo dei dati.

2.5.3 Scelte strategiche

Da un punto di vista strategico, l'attuale condizione di Meta non si può definire chiaramente. Considerando l'aspetto tecnico i modelli LLaMA sono competitivi in termini di prestazioni se confrontati con GPT, Gemini o Claude.

Tuttavia, è evidente come Meta sia più defilata sull'aspetto commerciale e sulla percezione dei consumatori: nonostante l'integrazione di LLaMA nelle piattaforme come WhatsApp, Facebook o Instagram, non si può dire che ciò abbia riscosso successo o interesse nei consumatori: manca una piattaforma conversazionale, paragonabile a Chat GPT o Google Gemini.

Strategicamente, la scelta di un modello open source comporta rischi di questo tipo, ma anche opportunità e tra queste, gli effetti di rete: più sviluppatori utilizzano LLaMA come base per costruire prodotti, maggiore sarà l'influenza di Meta sull'ecosistema IA.

2.6 DeepSeek

Nel gennaio 2025 attira l'attenzione globale una startup cinese specializzata in intelligenza artificiale, DeepSeek. Il rilascio del LLM chiamato DeepSeek R1 è disruptive nel mercato AI, causando uno dei più intensi cali nel mercato americano, specie per NVIDIA, azienda leader

⁷³ CNBC (2025). *Meta product chief: LLaMA 4 will power AI agents.*
<https://www.cnbc.com/2025/03/05/meta-product-chief-says-llama-4-will-power-ai-agents.html>

nella produzione di chip, che nella giornata del 27 gennaio registra una perdita del 17%, corrispondente a circa 600 miliardi di dollari di capitalizzazione di mercato⁷⁴.

Il modello open source DeepSeek R1 è un modello molto efficiente, in grado di offrire prestazioni simili a quelle di modelli americani come Chat GPT o LLaMA, ma a costi molto inferiori. Alcune stime suggeriscono che DeepSeek R1 sia stato dalle 16 alle 33 volte più economico da addestrare rispetto a Chat GPT o1⁷⁵.

Questa efficienza di costo nasce come conseguenza di un'ottimizzazione della fase di inferenza e di una minor dipendenza dalle GPU.

Nonostante i significativi aspetti positivi, DeepSeek è stata accusata di serie limitazioni derivanti dal sistema cinese: si assisteva a censura, allineamento alle decisioni governative e opacità nella gestione dei dati.

Da quel momento molti Stati ne hanno bloccato l'accesso denunciando rischi legati alla sicurezza. In Italia, per esempio, il 30 gennaio, il Garante della Privacy ha bloccato l'accesso alla piattaforma su tutto il territorio nazionale per mancanza di trasparenza nel trattamento dei dati personali⁷⁶.

La situazione per DeepSeek si complica quando il 29 gennaio subisce un attacco informatico con conseguente pubblicazione di dati quali note, documenti tecnici e conversazioni, che aggrava le preoccupazioni associate alla società cinese⁷⁷.

Come già anticipato, il successo di DeepSeek deriva dalla riduzione della dipendenza da GPU, asset sempre più scarso e costoso, e dalla sua ottimizzazione dei processi di inferenza, in grado di funzionare attraverso CPU ARM, come quelle usate nell'architettura Apple⁷⁸.

Tutto ciò ha avuto un impatto disruptive perché se ora i sistemi sono in grado di funzionare con hardware meno impegnativi, sono vanificati tutti gli sforzi basati su data center e GPU, impegnati dagli Stati Uniti (si pensi al progetto Stargate).

Come difesa, gli Stati Uniti si sono mossi limitando l'importazione di GPU NVIDIA verso la Cina, ma ciò si è rilevato più un danno per l'azienda americana che per quella cinese.

⁷⁴ CNBC (2025). *Nvidia loses \$600B in market cap after DeepSeek competition*.
<https://www.cnbc.com/2025/01/27/nvidia-sheds-almost-600-billion-in-market-cap-biggest-drop-ever.html#:~:text=Nvidia%20shares%20plunged%2017%25%20on,from%20Chinese%20AI%20lab%20DeepSeek>.

⁷⁵ Creole Studios (2024). *DeepSeek vs ChatGPT: Cost Comparison*.
<https://www.creolestudios.com/deepseek-vs-chatgpt-cost-comparison/>

⁷⁶ Garante per la protezione dei dati personali (2024). *Documento ufficiale su intelligenza artificiale e privacy*.
<https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10097450>

⁷⁷ The Verge (2024). *DeepSeek breach exposes AI security flaws and user data*.
<https://www.theverge.com/news/603163/deepseek-breach-ai-security-database-exposed>

⁷⁸ DigiTimes (2025). *ARM CPU chips to dominate AI computing landscape*.
<https://www.digitimes.com/news/a20250206VL205/arm-cpu-chips-ai.html>

Il contesto non si esaurisce qui infatti, nell'Aprile 2025, Huawei ha annunciato di essere vicina al rilascio di chip pensati proprio per alimentare l'ultima versione di DeepSeek, preannunciati come eguali a livello prestazionale dei famosi H100 di NVIDIA, ma ad un costo decisamente più contenuto.

2.7 NVIDIA

Nata nel 1993 come produttrice di processori grafici, NVIDIA è diventata una delle più importanti aziende americane nel campo dell'intelligenza artificiale per via della sua leadership nel segmento delle GPU, fondamentali per l'addestramento e l'inferenza dei modelli AI.

Nonostante nel tempo molte aziende come Google, AMD, Intel o Huawei, abbiano cercato di competere con NVIDIA, questa mantiene tutt'oggi una posizione di netto vantaggio per via della sua specializzazione e della sua posizione dominante consolidata, confermata dalla sua capitalizzazione di 2,3 miliardi di dollari⁷⁹, la terza a livello globale.

Questa posizione dominante deriva dalla dipendenza che le Big Tech hanno dei chip utilizzati per applicazioni in AI, e in particolare questo ha permesso a NVIDIA di applicare politiche di prezzo sempre più elevate.

Un contesto di questo tipo non è passato inosservato al Dipartimento di Giustizia degli Stati Uniti che nel 2024 ha avviato un'indagine antitrust nei confronti di NVIDIA volta a comprendere se il dominio dell'azienda sia dovuto a meriti tecnologici o per via di pratiche anticoncorrenziali potenzialmente escludenti per gli attori minori.

Tuttavia, la crescita incontenibile dell'azienda ha subito un fortissimo rallentamento nei primi mesi del 2025, a seguito dell'emergere di DeepSeek, la cui uscita ha determinato un calo giornaliero del 17% delle azioni NVIDIA per i motivi già spiegati nel capitolo precedente⁸⁰.

Ad oggi NVIDIA si ritrova a dover far fronte a due problemi: il primo è legato ai dubbi sorti in merito all'efficienza economica degli investimenti in GPU da parte delle Big Tech statunitensi, il secondo è invece legato alle restrizioni imposte dal governo americano in merito all'esportazione di chip verso la Cina.

⁷⁹ Statista (2024). *Top companies worldwide by market capitalization*.

<https://www-statista-com.ezproxy.biblio.polito.it/statistics/263264/top-companies-in-the-world-by-market-capitalization/>

⁸⁰ Finance Magnates (2025). *The AI Underdog: How DeepSeek Triggered Nvidia's \$560B Plunge*.

<https://www.financemagnates.com/trending/the-ai-underdog-how-deepseek-triggered-nvidias-record-560b-plunge/>

Sezione 3

Opportunità e rischi legati agli AI Agents

3.1 Possibili sviluppi degli AI Agents

3.1.1 Principali attori attivi nello sviluppo di AI Agents

Nel corso degli ultimi due anni, si sta assistendo ad un sempre più crescente interesse nello sviluppo degli AI Agents da parte di numerose aziende e startup.

La più famosa tra le tante, Open AI, è attiva nello sviluppo del suo agente chiamato Operator, in grado di controllare diverse funzionalità del computer.

In particolare, si sta parlando di *Computer-Using Agent - CUA* intendendo uno strumento, come definito dall'azienda stessa, *“addestrato a interagire con le interfacce utente grafiche, i pulsanti, i menu e i campi di testo che le persone vedono su uno schermo, proprio come fanno gli esseri umani”*⁸¹.

Dall'altro lato, anche Anthropic ha implementato il suo modello Claude della capacità di utilizzare il computer e Google sta sviluppando Project Mariner, un agente integrato nel sistema Chrome. Qualcosa di simile è stato proposto in maniera più timida da Amazon che a marzo ha presentato Amazon Nova Act, anch'esso uno strumento capace di utilizzare il computer⁸².

Microsoft sta invece proponendo Copilot Studio, una piattaforma che consente di costruire agenti personalizzati⁸³.

Non sono solo i grandi colossi ad interessarsi all'intelligenza artificiale agentica, ma anche molte startup stanno cercando di emergere proponendo anteprime di agenti attivi in ambiti specifici, tra questi i servizi finanziari, la sanità e l'ottimizzazione di sistemi industriali⁸⁴.

Tra tutte, le applicazioni di agenti più discusse secondo le interviste raccolte da LangChain sono Cursor (un agente che aiuta nella scrittura di codici), Replit (attivo nello sviluppo di software e

⁸¹ OpenAI (s.d.). *Computer-Using Agent*.

<https://openai.com/index/computer-using-agent/>

⁸² Il Sole 24 Ore (2024). *Amazon lancia Nova Act: nuovo agente AI che naviga autonomamente il web*.

<https://www.ilsole24ore.com/art/amazon-lancia-nova-act-nuovo-agente-che-naviga-autonomamente-web-AGfEqFwD>

⁸³ Microsoft (s.d.). *Microsoft Copilot Studio*.

<https://www.microsoft.com/it-it/microsoft-copilot/microsoft-copilot-studio?market=it>

⁸⁴ CB Insights (2024). *AI Agent Market Map*.

<https://www.cbinsights.com/research/ai-agent-market-map/>

applicazioni) e Perplexity (consente di acquistare un prodotto senza uscire dal suo motore di ricerca)⁸⁵.

Nello specifico, il livello di interesse da parte delle piccole startup è evidenziato anche dal crescente livello di investimenti raccolti che è cresciuto da 24 milioni di dollari nel 2020 a 3,8 miliardi di dollari nel 2024.

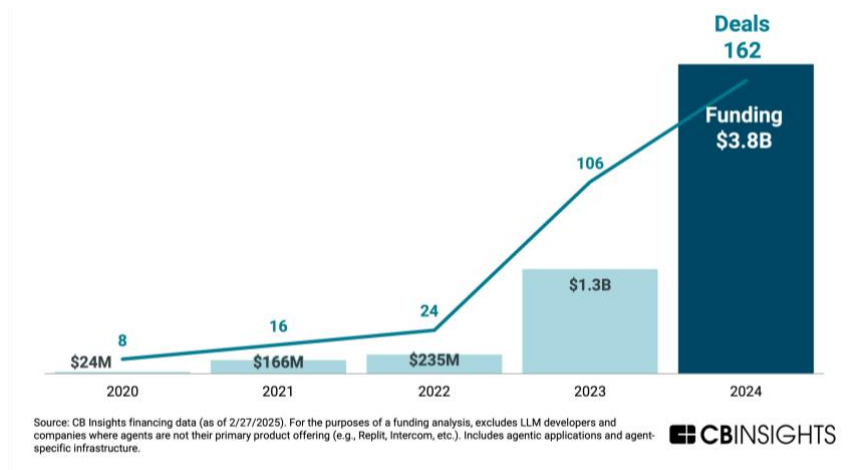


Fig. 1. Disclosed equity deals and funding to AI Agent startups, CBINSIGHTS⁸⁶

È quindi ben chiaro come l'intelligenza artificiale si stia muovendo verso l'agentismo con l'interesse attivo di grandi e piccole imprese.

Sebbene l'interesse sia elevato, non si può dire altrettanto del livello di sviluppo: in questo momento gli agenti disponibili nel mercato sono ancora molto limitati nelle capacità e commettono spesso errori anche banali.

Nonostante ciò, molte aziende non vogliono rimanere indietro, infatti il 51% delle aziende intervistate da LangChain ha affermato di avere già un agente attivo a livello di produzione, mentre il 78% ha in programma di introdurne⁸⁷.

Si nota quindi una diffusione attiva a livello di consumatori finali (non solo sviluppatori e programmatori) sebbene lo stato di sviluppo sia ancora embrionale.

Dati interessanti arrivano anche a livello di venture capital, per cui secondo PitchBook nel corso del 2024 si sono effettuate circa 156 operazioni con un valore medio di circa 3 milioni di dollari

⁸⁵ LangChain (2024). *The State of AI Agents – Use Cases*.
<https://www.langchain.com/stateofaiagents#:~:text=%E2%80%8D%20The%20top%20use%20cases,Similarly%2C%20AI%20agents%20are%20boosting>

⁸⁶ CB Insights (2024). *AI Agent Trends to Watch in 2025*.
<https://www.cbinsights.com/research/ai-agent-trends-to-watch-2025/>

⁸⁷ LangChain, *The State of AI Agents*, cit.

per round⁸⁸. Particolare attenzione deve essere rivolta anche ai movimenti di Y Combinator, un acceleratore americano di startup che nella primavera del 2025 ha investito in 144 aziende di cui 67 sono impegnate nello sviluppo di agenti di intelligenza artificiale⁸⁹.

3.1.2 Prospettive future

Le prospettive future associate allo sviluppo degli AI Agents sono davvero interessanti. Certamente una direzione sarà quella di sviluppare agenti sempre più in grado di pensare e decidere come umani, ma anche di allargare il loro raggio di azione, senza rimanere vincolati a specifici ambiti di applicazione.

Ad oggi, infatti, gli agenti sviluppati sono limitati a determinate applicazioni.

Per esempio, è possibile avere agenti specializzati nella scrittura di codici o agenti adatti a seguire le vendite di un'azienda⁹⁰. Quello che si nota infatti è che molte startup stanno sviluppando agenti specializzati e ciò per via dei vincoli tecnologici ancora non superati. Questo potrebbe portare in futuro a tanti agenti diversi, ognuno per ogni ambiente lavorativo.

Dal lato opposto, altri attori puntano a sviluppare agenti sempre più generali che siano in grado, come affermato da Sam Altman, di ottimizzare tutte quelle attività ripetitive e a basso valore aggiunto⁹¹.

Se si riuscisse a raggiungere lo sviluppo di AI Agents generali, ciò cambierebbe certamente il modo in cui l'utente si interfaccia con strumenti come computer e telefoni mobili. Un agente potrebbe avere accesso a tutte le informazioni e sostituire tutte le tradizionali applicazioni come emerge dal caso di Deutsche Telekom: l'azienda tedesca che ha recentemente presentato un telefono senza applicazioni nel quale l'utilizzatore finale interagisce direttamente con l'agente di intelligenza artificiale che selezionerà autonomamente le applicazioni adatte a soddisfare le sue richieste⁹².

⁸⁸ Fortune (2024). *AI Agent Startups Deal Count Up 81% Year-over-Year*.

<https://fortune.com/2024/09/24/ai-agent-startups-deal-count-up-81-percent-year-over-year-pitchbook/>

⁸⁹ PitchBook (2024). *Y Combinator Is Going All-In on AI Agents*.

<https://pitchbook.com/news/articles/y-combinator-is-going-all-in-on-ai-agents-making-up-nearly-50-of-latest-batch#:~:text=Y%20Combinator%20is%20going%20all,were%20AI%20agent%20companies>

⁹⁰ Toews, R. (2024). *Agents Are the Future of AI: Where Are the Startup Opportunities?* Forbes.

<https://www.forbes.com/sites/robtoews/2024/07/09/agents-are-the-future-of-ai-where-are-the-startup-opportunities/>

⁹¹ Times of India (2025). *Tech Leaders Share 2025 AI Predictions*.

<https://timesofindia.indiatimes.com/technology/tech-news/bill-gates-sam-altman-and-other-tech-leaders-share-2025-predictions/articleshow/117021677.cms>

⁹² Kramer, J., Reggiani, C., & Colangelo, G. (2024). *AI Agents and Ecosystems Contestability*. CERRE – Centre on Regulation in Europe.

<https://cerre.eu/publications/ai-agents-and-ecosystems-contestability/>

Da un punto di vista economico, studi dimostrano che il mercato degli Ai Agent è destinato a crescere rapidamente con un aumento del valore di mercato da 5,1 miliardi di dollari nel 2024 a 47,1 miliardi di dollari entro il 2030. Ciò contribuirebbe all'economia globale con 15,7 trilioni di dollari entro il 2030, con un incremento del PIL mondiale del 26%⁹³.

I settori in cui potrebbero emergere gli Ai Agent nel breve termine sono quelli caratterizzati da azioni ripetitive e facilmente replicabili.

Tra questi, Forbes identifica prevalentemente tre ambiti⁹⁴:

- Assistenza clienti: gli agenti permetterebbero di accelerare la gestione delle richieste dei clienti, molte delle quali sono ripetitive.
Risultati concreti sono già stati ottenuti da Klarna che con l'implementazione dell'intelligenza artificiale agentica ha ottimizzato drasticamente la gestione dell'assistenza clienti svolgendo il lavoro equivalente di 700 agenti a tempo pieno, accelerando i tempi di risposta, lavorando costantemente in più di 35 lingue e garantendo un miglioramento dei profitti pari a 40 milioni di dollari nel 2024⁹⁵.
- Contesto normativo: anche la gestione dei lavori associati alle conformità normative richiedono attività tipicamente ripetibili. In questo ambito due sono le principali aziende attive nello sviluppo di agenti, Norm Ai e Greenlite AI.
- Assistenti personali: questo ambito rappresenta l'applicazione più attesa. Assistenti personali di AI che siano in grado di semplificare la vita quotidiana sono lo sviluppo probabilmente più ambizioso e desiderato. La difficoltà sta nello sviluppare sistemi in grado di gestire una vasta gamma di attività, spesso molto diverse fra di loro, quindi ad alto livello di generalità.

Come già detto, al momento gli agenti sviluppati non sono completamente affidabili ed è questo il motivo per cui tendenzialmente si sta assistendo ad un contesto ibrido in cui il lavoro viene svolto dall'agente, ma revisionato dalla figura umana.

3.1.3 Contesto regolatorio

La crescente adozione dei sistemi agentici solleva dubbi e perplessità sul piano regolatorio, infatti, diversi enti e Stati stanno iniziando ad occuparsi di tale questione. Un esempio è il

⁹³ In AI We Trust (2024). *The Future of AI Agents: Transformative Trends and Predictions 2025–2030*.
<https://inaiwetrust.com/p/the-future-of-ai-agents-transformative-trends-and-predictions-2025-2030>

⁹⁴ Toews, *Agents Are the Future of AI*, cit.

⁹⁵ OpenAI (s.d.). *OpenAI + Klarna*.
<https://openai.com/index/klarna/>

CERRE (Centre on Regulation in Europe) che a fine 2024 ha pubblicato un documento dedicato agli AI Agents in cui affronta le basi da considerare a livello di regolamentazione.

In particolare, si analizzano gli AI Agents evidenziando il rischio associato al fatto che questi sono gli intermediari tra l'utente e i servizi (app e software) e quindi i nuovi gatekeeper nel mercato digitale⁹⁶.

Da questo punto di vista entrano quindi in gioco norme già esistenti come il Digital Markets Act che impone numerosi vincoli ai gatekeeper, ma che necessariamente dovranno essere aggiornate in vista di questi nuovi sistemi. Secondo il rapporto del CERRE, è infatti *“essenziale che i policymaker comprendano le implicazioni tecniche ed economiche degli AI agents”* perché potrebbero alterare significativamente gli equilibri di potere nei mercati digitali.

3.1.4 Previsioni di mercato

Da un punto di vista di mercato, quello degli AI Agents rappresenta il settore con le prospettive di crescita più rilevanti. Diverse stime concordano in un'espansione dai circa 5 miliardi di dollari nel 2024 a oltre 52,62 miliardi nel 2030 con un tasso di crescita annuo composto (CAGR) di circa il 46,3%⁹⁷.

La diffusione è prevista sia a livello aziendale, registrando l'82% delle imprese che pianifica di integrare agenti AI entro i prossimi tre anni⁹⁸, sia a livello consumer, per il quale le statistiche Accenture prevedono per il 2035 un aumento medio della redditività delle aziende del 38% per via dei risparmi in efficienza⁹⁹.

Da un punto di vista geografico, ad oggi il Nord America detiene il 40% del mercato globale degli AI Agents, seguita dall'area Asiatica in crescita con un tasso annuale di quasi il 50%¹⁰⁰.

In conclusione, è evidente come il mercato degli AI Agents si posizioni in un trampolino di lancio, pronto ad una crescita esponenziale nei prossimi anni.

Ad influenzare la diffusione ci saranno certamente questioni tecniche, di sicurezza, ma soprattutto sarà necessaria l'accettazione da parte dei consumatori di questi sistemi in

⁹⁶ Kramer et al., *AI Agents and Ecosystems Contestability*, cit.

⁹⁷ MarketsandMarkets (2024). *AI Agents Market – Global Forecast to 2029*.

<https://www.marketsandmarkets.com/Market-Reports/ai-agents-market-15761548.html>

⁹⁸ Capgemini (2024). *Generative AI in Organizations 2024*.

<https://www.capgemini.com/us-en/insights/research-library/generative-ai-in-organizations-2024/>

⁹⁹ Accenture (2017). *Artificial Intelligence Can Increase Profitability by 38% by 2035*.

<https://newsroom.accenture.com/news/2017/accenture-report-artificial-intelligence-has-potential-to-increase-corporate-profitability-in-16-industries-by-an-average-of-38-percent-by-2035>

¹⁰⁰ In AI We Trust, *The Future of AI Agents*, cit.

considerazione del fatto che cambieranno in modo evidente l'abituale approccio umano alla tecnologia.

Da questo punto di vista, il 2025 rappresenta un anno molto importante in termini di sviluppo e adozione perché sempre più aziende stanno adattandosi a questi nuovi sistemi e sempre maggiore è la consapevolezza da parte dei consumatori.

3.2 AI Agents come rischio

Per quanto possa essere grande l'entusiasmo associato ai possibili sviluppi e utilizzi degli AI Agents, è importante non trascurare tutti quegli aspetti che potrebbero invece incidere negativamente sull'utilità del consumatore finale.

Il rischio, in generale, è che questo tipo di strumenti possano rafforzare esponenzialmente il potere di aziende già consolidate nel contesto economico globale e che esse possano sfruttare tale supremazia per imporre condizioni che massimizzino i loro profitti andando così a discapito dell'utilità degli utenti.

Il fattore alla base di questi rischi risiede nella capacità degli AI Agents di prendere decisioni. Ciò cambia completamente il contesto economico-sociale per cui a decidere non sono più gli utenti, ma dei sistemi digitali. Questo si traduce in una riallocazione del potere contrattuale e in un cambiamento rispetto i tradizionali rapporti di mercato.

Un aspetto interessante correlato a questo cambio di prospettiva viene proposto da Christoph Busch (2017) che nel paper "*Consumer Law for AI Agents*"¹⁰¹ fa notare come il diritto dei consumatori attualmente sia incentrato nell'uomo: esso si basa sul presupposto che le scelte di acquisto vengano prese da esseri umani. Occorre quindi che le autorità di regolamentazione si adattino al nuovo paradigma che si sta imponendo attraverso questi agenti che Bush (2025) definisce in altre parole come "consumatori algoritmici", "clienti macchina" o "custobots"¹⁰² (un ibrido tra cliente e robot).

Un problema correlato a questo contesto consiste nel fatto che questi agenti, pur operando teoricamente nell'interesse dell'utente, sono programmati e serviti da grandi aziende tecnologiche che, come detto, hanno obiettivi diversi dal consumatore: massimizzare

¹⁰¹ Li, C., Zhang, T., Du, Y., et al. (2024). *Foundation Model Competition and Vertical Integration*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5187056

¹⁰² Scheibenreif, A. & Raskino, M., Twigg-Flesner, C., & Howells, G. (forthcoming 2025). *Adapting Consumer Law to New Technologies*, in Brownsword, R. & DiMatteo, L. (Eds.), *The Cambridge Handbook of the Governance of Technology*. Cambridge University Press.

contemporaneamente i profitti aziendali e l'utilità del consumatore non è possibile, infatti, aziende e utenti hanno incentivi divergenti.

Ecco perché è importante analizzare questo disallineamento degli incentivi che sarà la fonte di una serie di criticità che meritano un'attenta analisi al fine di poter poi definire eventuali azioni risolutive, siano esse ex-ante o ex-post.

Tra le possibili criticità, sono state individuate e verranno approfondite in questo capitolo le seguenti:

- Esternalità di rete: è il fenomeno per il quale, il valore di un bene o servizio aumenta all'aumentare degli utilizzatori.

Nel caso degli AI Agents, questo valore è legato agli utenti, ma anche ai dati. Più utenti utilizzano il servizio, più l'agente acquisisce potere e più riesce a fornire soluzioni ottimali per l'utente, alimentando un ciclo continuo.

Allo stesso modo funziona con i dati: più dati un sistema possiede, migliore sarà il suo addestramento e maggiore sarà il numero di clienti, che a loro volta alimentano il ciclo fornendo ulteriori dati.

In entrambi i casi i problemi conseguenti alle esternalità di rete sono tre:

- I mercati con esternalità di rete tendono ad essere maggiormente concentrati, favorendo la logica "*the winner takes it all*" (il vincitore prende tutto).
- Le barriere all'ingresso sono elevate, infatti i *new entrants* avranno difficoltà a competere con *incumbents* che hanno stabilito il loro vantaggio competitivo sulla base utenti (o dati).
- Si verificano effetti di lock-in nell'utente per il quale diventa costoso cambiare piattaforma per via della perdita del valore associato alla rete.
- Self-preferencing: è il meccanismo per il quale le aziende proprietarie, preinstallano i propri agenti nei loro sistemi operativi o ecosistemi chiusi o, più in generale, tendono a proporre i propri servizi piuttosto che quelli concorrenti.
- Collusione tacita: nel contesto dell'intelligenza agentica può verificarsi che diversi agenti, lavorando autonomamente per diversi utenti, possano operare in maniera coordinata, colludendo anche senza un'intenzione esplicita.
- Non neutralità: gli agenti potrebbero non essere sempre neutrali nelle scelte, finendo col massimizzare la propria utilità, piuttosto che quella dell'utente.

- Barriere all'ingresso: sempre maggiori sono gli ostacoli per i *new entrants* a causa del controllo da parte delle Big Tech di risorse essenziali come dati, cloud computing e personale specializzato.

Nel corso di questo capitolo saranno quindi approfonditi tali rischi che incidono negativamente nella contendibilità dei mercati digitali, con gli AI Agents che, anziché aumentare l'utilità dei consumatori e stimolare l'ingresso di nuovi player, rischiano piuttosto di rafforzare il potere degli attuali gatekeepers.

3.2.1 Esternalità di rete

Uno dei rischi più rilevanti associati all'adozione degli AI Agents è rappresentato dal fenomeno delle esternalità di rete.

Da un lato si potrebbe credere che effetti di questo tipo possano favorire un miglioramento dei servizi: se si pensa all'accumulo di dati, si comprende che un maggior numero di dati favorisce agenti meglio addestrati e quindi capaci di rispondere in maniera più adeguata ed efficace alle richieste dell'utente.

Dall'altro lato bisogna però sottolineare che questo meccanismo genera una maggiore adozione dello stesso agente da parte degli utenti e quindi una maggiore raccolta di dati che alimenta il processo di apprendimento e poi di adozione.

Lo stesso funzionamento si verifica con la base utenti che un'azienda può già possedere attraverso, per esempio, una piattaforma o un servizio già offerto. La dimostrazione più evidente è il caso di un'azienda che offre un sistema operativo che conta molti utilizzatori ai quali può essere proposto l'agente di intelligenza artificiale associato al SO stesso. In questo modo l'impresa può sfruttare una base utenti di partenza e questo le garantisce un potere contrattuale maggiore (si entrerà successivamente nel dettaglio) per mezzo del quale l'agente riesce a garantire prezzi più bassi per i suoi utilizzatori. Ciò stimola una maggiore adozione e conseguentemente un potere contrattuale sempre più forte.

In entrambi i casi, quello che va a definirsi è un meccanismo auto-rinforzante che non può che accentuare la concentrazione del mercato, ostacolando l'ingresso di nuovi attori e aumentando l'effetto lock-in negli utenti (per via di elevati switching costs).

Effetti di rete legati ai dati

Come già detto, i dati permettono di sviluppare un circolo vizioso di feedback positivo associato all'adozione dell'agente di intelligenza artificiale. Tuttavia, per abilitare tale circolo occorre avere una grande base dati di partenza.

Questo input chiave rappresenta quindi uno degli asset più strategici che le aziende possono sfruttare come fonte di vantaggio competitivo, difficilmente recuperabile dai competitor.

Secondo il report di Jan Krämer e Friso Bostoen (2024), il problema è reso ancor più evidente e pesante dal fatto che *“alcuni hyperscaler controllano sia i modelli di AI sia l'accesso alle infrastrutture cloud, rendendo più difficile per i rivali costruire prodotti concorrenti senza dipendere da loro”*¹⁰³.

In questo contesto bisogna però considerare un fattore che attenua tale rischio appena descritto. I possibili new entrants, pur non disponendo di una base di dati proprietaria, possono però contare sulla possibilità di acquisirli da data trader o utilizzare i dati gratuitamente disponibili al pubblico.

Inoltre, le innovazioni tecnologiche, come testimoniato dal caso DeepSeek, stanno permettendo l'addestramento dei modelli anche attraverso un minor numero di dati. Oltre a ciò, gli studi¹⁰⁴ dimostrano che molti dati hanno un valore temporaneo perché la loro utilità si erode nel tempo garantendo miglioramenti nelle prestazioni che diventano trascurabili sopra una certa soglia.

In particolare, le performance associate ai modelli di intelligenza artificiale inizialmente migliorano in maniera esponenziale all'aumentare del dataset fino a circa $10^4 - 10^5$ campioni, ma successivamente i miglioramenti si stabilizzano e i costi crescono in maniera evidente¹⁰⁵.

Effetti di rete legati alla base utenti

Molto più fine e articolato è il concetto di effetto di rete che si verifica negli AI Agents a livello di numero utenti.

¹⁰³ Kramer et al., *AI Agents and Ecosystems Contestability*, cit.

¹⁰⁴ CIO (2024). *Big Data Is Not a Barrier to Entry*.

<https://www.cio.com/article/228353/big-data-is-not-a-barrier-to-entry.html#:~:text=Just%20as%20big%20data%20is,free%20data%20available%20to%20them>

¹⁰⁵ Milvus (2024). *What Is the Significance of Dataset Size in ML Performance?*

<https://milvus.io/ai-quick-reference/what-is-the-significance-of-dataset-size-in-machine-learning-model-performance>

Un agente di intelligenza artificiale, se si pensa al contesto di acquisto e vendita di prodotti, funge da intermediario tra l'utente che vuole acquistare un prodotto e il venditore che lo offre.

Nel processo di selezione del venditore da parte dell'agente, entrano in gioco i meccanismi tipici legati al potere contrattuale. Infatti, quando un AI Agent sceglie da quale venditore acquistare un prodotto o un servizio per conto dell'utente, il contesto in cui avviene tale operazione è influenzato da un rapporto di forza tra domanda e offerta, come tipicamente avviene nei mercati tradizionali.

In particolare, il potere contrattuale dell'agente è esprimibile dalla sua capacità di influenzare le condizioni dello scambio, come per esempio prezzo, qualità o tempi di consegna.

A questo punto il problema nasce nel momento in cui l'agente AI rappresenta una grande base di utenti: in questo modo esso acquisisce un potere simile a quello di un grande acquirente.

Di fatto, questa concentrazione di acquisto, come evidenziato da Raz Agranat e Michal Gal (2025) nel documento *"Fueling Concentration: Network Effects and AI Agents"*¹⁰⁶, può trasformare un insieme di consumatori che tipicamente agiscono autonomamente, in un gruppo unico e ciò comporta principalmente due effetti:

- Permette all'agente di negoziare sconti e condizioni agevolate per conto dei suoi utenti;
- Rafforza la posizione di mercato dell'azienda che controlla l'agente, poiché le permette di decidere quali prodotti o servizi mostrare all'utente finale.

In particolare, la piattaforma può escludere o penalizzare tutti quei venditori che non accettano determinate condizioni come, per esempio, il pagamento di elevate commissioni per ottenere visibilità.

Di conseguenza l'agente può orientare le scelte dei consumatori verso venditori che risultano essere più vantaggiosi per la piattaforma stessa e non necessariamente per l'utente (su questo punto, si rimanda al capitolo dedicato alla non neutralità degli agenti).

È importante chiarire che questo fenomeno non può essere inteso come meccanismo di coordinamento tra utenti, ma come semplice conseguenza del fatto che l'agente,

¹⁰⁶ Ganor, A. et al. (2024). *Training Compute as a Barrier to Entry in AI*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5250253

nel suo ruolo di intermediario, raccoglie molte informazioni dei clienti (come bisogni o preferenze) e agisce sulla base di queste per negoziare con l'offerta e garantire all'utilizzatore la soluzione migliore.

Questa descrizione può essere comparata al concetto "*hub and spoke*", un modello che prevede un attore centrale (hub), che coordina le decisioni prese da attori secondari (spoke). Nel caso degli agenti di intelligenza artificiale l'hub è l'agente stesso, mentre gli spoke sono i consumatori.

Tipicamente questo modello è associato alle pratiche collusive tra venditori, ma qui si concretizza in una collusione involontaria che però alla fine genera effetti equivalenti a livello di concentrazione.

Sebbene questo potere degli agenti sembri poter garantire condizioni favorevoli per il consumatore finale, ci sono molti aspetti negativi che devono essere analizzati:

- Se l'AI Agent acquisisse un ruolo sempre più incisivo nel processo di vendita, questo potrebbe costringere i venditori ad adattare le loro logiche di prezzo a quello che è lo standard imposto dall'agente. Ciò si verificherebbe soprattutto nel caso in cui l'agente diventasse l'unico punto di comunicazione tra il venditore e l'utente finale.
- L'agente potrebbe preferire inoltre certi venditori piuttosto che altri, penalizzando quelli che, per esempio, non sono integrati nella piattaforma che fornisce l'agente stesso.
- Il fatto che l'agente possa proporre le soluzioni più economiche porterebbe ad una competizione tra venditori basata prettamente sul prezzo più che sulla qualità e questo andrebbe a discapito della differenziazione e dell'innovazione.

Da questo punto di vista, ci si può però aspettare che, almeno in un primo momento, gli agenti di intelligenza artificiale semplifichino la procedura di scelta ed acquisto di articoli di consumo in cui il cliente tipicamente non fa caso al brand (si pensi a prodotti come lampadine o sacchetti per esempio). Nonostante l'aspetto positivo, dovuto al fatto che il cliente è facilitato nella fase di scelta e riesce a risparmiare tempo e denaro, il problema sorge per le aziende, le quali avranno più difficoltà a fidelizzare il

cliente in un contesto competitivo in cui vengono proposte alternative più economiche, ma altrettanto efficaci¹⁰⁷.

- Raccogliendo le richieste di una base utenti consolidata ed estesa, l'agente diventa sempre più specializzato e adatto per gli utilizzatori e ciò genera in loro lock-in. In questo modo, i venditori non possono permettersi di non essere compatibili con tale agente, altrimenti perderebbero l'accesso ad un vasto gruppo di utenti.

3.2.2 Self-preferencing e preinstallazione

Un altro rischio a livello di concorrenza legato allo sviluppo degli AI Agents riguarda il fenomeno del self-preferencing. Esso consiste nell'integrazione da parte degli sviluppatori degli agenti nella propria piattaforma di prodotti e servizi o comunque, nell'ecosistema proprietario. Concretamente parlando, un esempio è il caso in cui l'agente viene preinstallato nel sistema operativo, nel browser o nei telefoni, escludendo effettivamente la concorrenza o comunque limitandone l'accesso al sistema e la possibile scelta da parte degli utenti.

Da questo punto di vista sono quindi favorite quelle grandi aziende che hanno già ecosistemi consolidati come Microsoft, Google o Amazon che, come emerso nei capitoli precedenti, stanno già integrando gli agenti di intelligenza artificiale all'interno dei loro servizi come sistemi operativi (Windows, Android e iOS), browser (Chrome, Edge e Safari), ambienti di lavoro (Microsoft 365, Google Workspace), ma anche assistenti vocali come Alexa.

Una conseguenza è che diventa l'agente ad interfacciarsi con l'utente, andando così a ridurre l'utilità effettiva delle applicazioni tradizionali o dei motori di ricerca.

Ecco quindi che, in questo modo, l'agente può orientare le richieste degli utenti verso l'utilizzo di applicazioni proprietarie (concetto vicino alla non neutralità degli agenti). A maggior ragione, in ambito di agenti intelligenti il rischio è ancora più evidente poiché essi agiscono autonomamente e ci si aspetta che non sia così semplice e chiaro comprendere quali siano le ragioni di scelta valutate dall'agente e quali i percorsi intrapresi per fornire una soluzione finale all'utente.

Fenomeni di self-preferencing non sono nuovi in ambito antitrust. Un esempio è rappresentato dalla decisione della Commissione Europea che nel 2018 ha accusato Google di self-preferencing nella scelta del motore di ricerca associato a Chrome.

¹⁰⁷Li et al., *Foundation Model Competition*, cit.

In particolare, l'accusa ha sanzionato Google per 4,34 miliardi di euro per aver preinstallato il proprio motore di ricerca, Google Search, e il browser Chrome su tutti i dispositivi Android, minacciando i produttori di non offrire accesso a Play Store in caso di opposizione¹⁰⁸.

In questo modo Google è riuscita ad imporre il suo motore di ricerca a discapito della concorrenza (Bing, DuckDuckGo, Qwant, ...).

Queste pratiche, vietate dal Digital Markets Act (DMA) in quanto configurano un comportamento di self-preferencing da parte dei gatekeeper, sono affrontate in un'ottica di regolazione ex ante. L'intervento dell'autorità non si è limitato a una multa, ma ha imposto a Google l'obbligo di introdurre una schermata iniziale che permettesse all'utente di scegliere autonomamente il motore di ricerca preferito.

La realtà dei fatti dimostra però che *“a causa dello status quo o dei pregiudizi predefiniti degli utenti – di cui Google era ben consapevole – tali restrizioni creavano un vantaggio distributivo a favore di Google Search e Google Chrome che i concorrenti non potevano compensare ricorrendo ad altri canali di distribuzione. Infatti, secondo le inequivocabili conclusioni della Commissione Europea, gli utenti raramente modificano le impostazioni predefinite sui propri dispositivi mobili intelligenti”*¹⁰⁹.

Questo caso è rilevante anche nel contesto degli AI Agents perché mostra che così come Google rafforzava la sua posizione favorendo il suo motore di ricerca, così le aziende che posseggono piattaforme consolidate possono integrare di default l'agente proprietario favorendone l'adozione e il conseguente effetto lock-in nel cliente.

3.2.3 Collusione tacita

Un altro rischio rilevante connesso alla diffusione degli AI Agents riguarda il fatto che essi possano essere in grado di favorire la collusione tacita. Per collusione tacita si intende che più operatori nel mercato si comportano nella stessa maniera (portando quindi ad una riduzione della concorrenza) semplicemente osservandosi l'un l'altro come si comportano e senza alcun accordo esplicito.

Tipicamente un contesto di questo tipo era caratteristico di mercati altamente concentrati, con poche aziende ad operare che potevano facilmente osservarsi.

¹⁰⁸European Commission (2018). *Commission Opens Investigation into Amazon's Data Use*. https://ec.europa.eu/commission/presscorner/detail/en/ip_18_4581

¹⁰⁹Clarke, L. (2024). *RealPage: l'algoritmo che falsa i prezzi degli affitti*. Corriere della Sera. https://www.corriere.it/economia/affitti/24_agosto_24/affitti-sotto-accusa-l-algoritmo-di-realpage-di-thoma-bravo-che-controlla-dardtrace-falsa-i-prezzi-a05ce8ba-baf7-4505-9fb8-b56891ed1x1k.shtml

Già con l'avvento degli algoritmi di *dynamic pricing* il problema ha cominciato a emergere in maniera critica.

Un esempio è il recente caso di RealPage, un'azienda statunitense che ha sviluppato un software che aiuta i proprietari di immobili a decidere un prezzo per gli affitti¹¹⁰. Questo software forniva una soluzione di prezzo basandosi su dati e informazioni relativi agli edifici di ogni zona di interesse.

Il problema è emerso quando gli utilizzatori del servizio erano cresciuti a dismisura e quindi tanti operatori usavano lo stesso strumento. Di conseguenza tutte le decisioni venivano “coordinate” dallo stesso software. Questo, seppur l'appartamento rimanesse sfitto, non consigliava di abbassare il prezzo perché prevedeva che anche gli altri proprietari non avrebbero abbassato i prezzi e così, i profitti collettivi aumentavano.

Di fatto non era stato stipulato alcun accordo tra gli operatori, ma il risultato è stato quello della collusione: i prezzi degli affitti erano più alti, la concorrenza diminuiva e chi cercava un appartamento non aveva altre scelte.

Questo caso mostra come, anche in assenza di accordi espliciti, algoritmi condivisi possono portare ad un contesto di collusione tacita.

Tutto questo può avvenire ancor più facilmente ed efficacemente con l'avvento degli agenti intelligenti: essi sono addestrati per massimizzare l'obiettivo del profitto e non si limitano alla semplice definizione del prezzo, ma prendono decisioni, interagiscono con più attori e gestiscono un'impressionante mole di dati associata a tutti gli utilizzatori.

Secondo alcuni studi¹¹¹, ci sono alcuni fattori che facilitano la collusione tacita, specialmente nei mercati controllati da AI Agents. Tra questi:

- Osservabilità dei dati: gli agenti hanno accesso a prezzi e strategie dei concorrenti (per i quali evidentemente lavorano) e questo li agevola nel controllare ed imitare i comportamenti della concorrenza.
- Interazioni frequenti: nei mercati digitali le decisioni prese dagli agenti sono molte e prese continuamente a gran velocità; questo favorisce l'individuazione di pattern e logiche di scelta.
- Architetture comuni: molti AI Agents potrebbero basarsi su stessi modelli linguistici, API o dati di apprendimento. Ne consegue che se più agenti sono addestrati con le stesse logiche, più probabilmente useranno criteri simili nella scelta dei prezzi e delle strategie.

¹¹⁰ Kramer et al., *AI Agents and Ecosystems Contestability*, cit.

¹¹¹ Kramer et al., *AI Agents and Ecosystems Contestability*, cit.

- Mancanza di soggettività dell'agente: laddove gli agenti comprendano che lavorare in maniera collusiva porta a risultati economicamente convenienti per le aziende, non avranno alcun incentivo nell'agire diversamente: essendo addestrati a massimizzare il beneficio dell'utilizzatore (le aziende), tenderanno a mantenere la collusione attiva.

In tutto ciò, le autorità antitrust avranno seri problemi nel contrastare questo problema per il semplice motivo che esso non è facilmente identificabile.

Le autorità antitrust, per parlare di collusione hanno bisogno di una prova che evidenzia l'atteggiamento coordinato delle aziende, per esempio dei contatti o delle riunioni.

Nel caso degli AI Agents però, nessuno si accorda con nessuno, ma semplicemente gli agenti si adattano al mercato in cui operano; dunque, se apprendono che non abbassare i prezzi porta a soluzioni più vantaggiose, così faranno.

Inoltre, gli agenti agiscono autonomamente sulla base di semplici input da parte delle aziende che li utilizzano. Nonostante ciò, le imprese non sanno come l'agente giunge a certe conclusioni e di fatto non controllano il percorso che esso segue. Di conseguenza per le autorità antitrust risulta complicato anche comprendere a chi attribuire la responsabilità legale di atteggiamenti collusivi emergenti.

Oltretutto le decisioni prese dall'agente non necessariamente devono considerarsi illegali: se l'agente nota che non abbassare il prezzo quando lo fa il concorrente aumenta i profitti, si comporterà in quel modo. Questa però è una scelta del tutto razionale e non necessariamente illegale. Tuttavia, se tutti gli agenti operano in questo modo, il risultato finale è quello collusivo. Il problema di base di questo contesto è che le attuali leggi antitrust non sono adatte a risolvere questi contesti e, anche laddove ci sia l'intenzione di intervenire ex-ante, questo non è del tutto semplice: imporre una regolamentazione troppo stringente potrebbe rallentare il processo innovativo con tutti i benefici che esso potrebbe portare.

3.2.4 Non neutralità degli agenti intelligenti

Come si è avuto modo di osservare, una delle caratteristiche distintive degli AI Agents è la loro autonomia nel prendere decisioni. La domanda che ora sorge spontanea è: queste decisioni sono davvero neutre ed imparziali? I criteri di scelta utilizzati hanno davvero come obiettivo la massimizzazione dell'utilità del consumatore finale?

Sebbene spesso si sia portati a pensare che questi strumenti siano del tutto imparziali e a servizio dell'utente, potrebbe però non essere pienamente così. Bisogna infatti ricordare che essi si

basano su architetture e addestramenti che possono orientarne il comportamento in maniera più o meno evidente.

Studi come quello condotto dal CERRE nel 2024¹¹², affermano che gli agenti non sono neutrali, ma piuttosto tendono a massimizzare l'utilità secondo obbiettivi che sono predefiniti nei modelli alla base, i quali evidenziano interessi specifici e spesso lontani ed incompatibili rispetto quelli dell'utente finale.

Da questo punto di vista il problema sta quindi nell'allineamento degli incentivi: l'agente potrebbe voler ottimizzare il profitto della piattaforma proprietaria, piuttosto che massimizzare il benessere del consumatore.

Un esempio è il caso in cui un agente consiglia dei prodotti sulla base di accordi commerciali, partnership o questioni legate all'integrazione verticale, piuttosto che in base alla qualità o al prezzo più vantaggiosi per l'utente finale.

Questo contesto somiglia molto al già discusso fenomeno di self-preferencing, ma in questo caso si estende a molti più aspetti legati alla capacità decisionale dell'agente.

Anche questo rischio è amplificato dal fatto che l'utente non vede e non sa come funziona internamente l'agente né quali percorsi esso valuta per fornire una soluzione.

La conseguenza più rilevante è legata al fatto che, grazie alla sua capacità di prendere decisioni senza l'intervento diretto dell'utente, è l'agente stesso a stabilire quali prodotti o servizi proporre e quali invece rendere meno visibili.

Il risultato finale legato alla non neutralità degli AI Agents è che essi assumono la figura centrale di decisori di mercato: decidono dal lato della domanda valutando cosa mostrare al cliente finale e agiscono dal lato dell'offerta stabilendo le condizioni da rispettare per poter partecipare al mercato.

In sintesi, la piattaforma che controlla l'agente non si limita ad essere un semplice intermediario tra venditori e acquirenti, ma esercita in essi un'influenza rilevante attraverso atteggiamenti di selezione ed esclusione, ostacolando l'ingresso a nuovi attori e concentrando ancor di più il potere economico nelle mani di pochi preferiti.

3.2.5 Dipendenza nell'accesso a input essenziali

Un elemento fondamentale alla base della competitività del mercato degli AI Agents è il possesso di asset strategici fondamentali per lo sviluppo di questi sistemi, tra questi dati, chip, infrastrutture cloud e software in cui gli agenti vengono eseguiti.

¹¹² Kramer et al., *AI Agents and Ecosystems Contestability*, cit.

Come i capitoli precedenti hanno evidenziato, solo poche aziende sono integrate a tal punto o posseggono questi asset; esempi evidenti sono Amazon, Microsoft e Google.

La criticità non risiede solo nel possesso di input chiave, ma anche nel fatto che, grazie al loro livello di integrazione, queste Big Tech non solo si comportano da fornitori, ma anche da concorrenti, sviluppando ognuna il proprio agente.

Quanto appena detto può essere identificabile come un fenomeno di *double leverage*, per cui un'azienda acquisisce vantaggio competitivo su due livelli e ciò perché controlla sia l'infrastruttura (quindi cloud, dati, GPU) che i prodotti finali (LLM, AI Agents) che funzionano su di essa.

Si identificano così due livelli:

- Primo livello (a monte): gli attori attivi, per esempio nel cloud computing, sanno di possedere un'infrastruttura indispensabile che li rende gli unici fornitori possibili per chiunque voglia sviluppare e offrire AI Agents.
- Secondo livello (a valle): gli stessi attori che forniscono l'infrastruttura cloud, sviluppano anche prodotti basati su essa, basti pensare a Copilot per Microsoft, Gemini su Android, e possono così avvantaggiarsi rispetto ai concorrenti.

In questo modo, le Big Tech possono controllare l'accesso delle risorse chiave e limitarlo o ostacolarlo a concorrenti nel mercato a valle, ottenendo così supremazia nel mercato più vicino al cliente (nonché quello in grado di garantire i maggiori margini).

L'esempio più evidente è rappresentato dalla partnership tra Microsoft e Open AI. Microsoft domina con Azure il mercato del cloud e può offrire condizioni vantaggiose in termini di prezzi e prestazioni a Open AI. Allo stesso tempo però Microsoft ha una posizione rilevante nel mercato degli AI Agents perché può sviluppare sistemi più potenti, a costi minori e, oltretutto, integrarli direttamente nei suoi servizi quali Windows e Office.

In questo senso, la Commissione Europea ha quindi aperto un'indagine sull'accordo Microsoft-Open AI al fine di valutare i possibili effetti a livello di concorrenza nei mercati a valle, in particolare per via dell'uso esclusivo dell'infrastruttura Azure.

Allo stesso modo, la Federal Trade Commission ha richiesto informazioni a Google, Amazon, Microsoft, Anthropic e Open AI per comprendere se l'accesso privilegiato alle risorse strategiche possa creare barriere concorrenziali nel mercato degli AI Agents a valle¹¹³.

¹¹³ Federal Trade Commission – FTC (2024). *FTC Launches Inquiry into Generative AI Investments and Partnerships*.

<https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships>

È bene notare che il vantaggio associato alle risorse non riguarda solamente input come chip o infrastrutture, ma anche talenti utili e necessari per lo sviluppo innovativo degli AI Agents.

In questo senso, bisogna ricordare l'acquisizione del miglior personale di Inflection AI da parte di Microsoft. Anche il personale qualificato, che per motivi di prestigio e condizioni economiche preferisce lavorare in aziende consolidate nel mercato, genera un vantaggio competitivo non irrilevante e limita l'accesso a tali risorse chiave a competitor più piccoli.

Un altro input essenziale per la diffusione di agenti di intelligenza artificiale è rappresentato dalla piattaforma in cui l'agente viene eseguito. Per piattaforma si intende in questo caso i sistemi operativi, i browser o i dispositivi mobili.

Se si immagina che un agente venga integrato sin da subito in un sistema operativo (un esempio è Mariner di Microsoft in Windows) si comprende bene che esso acquisisce una posizione di vantaggio e difficilmente imitabile dai concorrenti. Lavorando sulla piattaforma, oltre che a “crescere” grazie al massiccio utilizzo degli utenti (si veda il concetto di esternalità di rete, discusso nei capitoli precedenti), si impone nell'utilizzo di quest'ultimi che instaurano un rapporto di fiducia che genera elevati costi di switching.

Ovviamente un vantaggio di questo tipo è possibile solo per quelle aziende che possiedono piattaforme significativamente utilizzate, come Amazon, Google, Microsoft o Meta.

In conclusione, la dipendenza delle piccole aziende per input essenziali si traduce in barriere all'ingresso che riguardano aspetti:

- Tecnici: quindi l'accesso a infrastrutture cloud e software proprietari;
- Economici: le piccole aziende che non possono sfruttare sin da subito economie di scala si ritrovano a sostenere costi elevatissimi;
- Comportamentali: la proprietà degli ambienti in cui far lavorare Ai Agent genera lock-in nei clienti verso gli agenti preinstallati nella piattaforma da essi già utilizzata.

3.3 AI Agents come opportunità

Ora che il capitolo precedente ha affrontato i rischi concorrenziali legati agli sviluppi degli AI Agents, è bene anche comprendere quali siano le opportunità che sistemi di questo tipo offrono per stimolare la concorrenza e ridurre così il potere accentrato nelle mani delle Big Tech.

In questo capitolo si approfondiranno quindi quali siano le varie possibilità e i contesti in cui gli AI Agents possono favorire una maggior contendibilità dei mercati.

In particolare, si affronterà tale analisi cercando di comprendere innanzitutto come gli AI Agents possano diventare i nuovi punti di accesso ai servizi digitali, andando così ad indebolire il ruolo dei rinomati gatekeeper.

Successivamente si cercherà di comprendere in che modo questi strumenti possono aumentare la scelta dei consumatori riducendo i loro costi di ricerca, gli switching costs e facendo così in modo che le aziende competano più apertamente.

Infine, si approfondirà come i nuovi entranti nel mercato possono affidarsi a soluzioni open source per competere attivamente con le attuali aziende detentrici dei più importanti asset strategici.

3.3.1 AI Agents come nuovi punti di accesso

Una delle più grandi opportunità che gli AI Agents possono offrire, è legata al cambiamento dei punti di accesso ai servizi digitali, contesto che può diminuire drasticamente il potere degli attuali gatekeeper.

Ad oggi gli utenti possono utilizzare diversi servizi come applicazioni o strumenti di ricerca attraverso strumenti (sistemi operativi, app store, o motori di ricerca) che sono prevalentemente controllati da poche Big Tech.

In un contesto di questo tipo gli AI Agents permettono di rompere questa struttura: essi potrebbero intervenire come interfaccia unica con cui l'utilizzatore finale interagisce e questo evita ai clienti di dover passare per una singola app, o un motore di ricerca.

Un esempio concreto di questa situazione è quello del già citato telefono *app-less* di Deutsche Telekom. Con uno strumento del genere l'utilizzatore finale può porre domande e richieste all'AI Agent e sarà esso stesso (né l'utente né il sistema operativo del telefono) a decidere quali applicazioni sottostanti utilizzare per risolvere la richiesta.

Può essere utile in questo caso riportare un esempio per rendere più evidente il funzionamento: si immagini che l'utente richieda di acquistare un generico prodotto; a questo punto l'agente può scegliere il servizio di prenotazione o l'applicazione migliore; quest'ultima potrebbe quindi essere un'applicazione di proprietà di un grande colosso tecnologico (si pensi ad Amazon per l'acquisto di un prodotto), ma anche quella di una piccola startup (magari un e-commerce specializzato nei prodotti come quello richiesto).

Ne risulta quindi evidente che il potere di indirizzare la domanda in una direzione o nell'altra risiede nell'agente.

Ecco che questo contesto apre molte opportunità in termini di concorrenza perché l'utente non è più vincolato a scelte predefinite magari dal proprietario della piattaforma (come il sistema operativo che si sta utilizzando), ma può spaziare attraverso l'agente ad una miriade di servizi che potrebbero in alcuni casi essere addirittura non conosciuti dell'utente stesso.

Dunque, ben si capisce come queste innovazioni siano altamente disruptive e offrano la possibilità di influenzare e cambiare drasticamente l'attuale paradigma tecnologico.

Gli stessi Bostoen & Krämer (2024) nel documento *"AI Agents and Ecosystems Contestability"*¹¹⁴ affermano:

"questi cambiamenti consentono nuove modalità di comunicazione, risoluzione dei problemi e creatività, portando spesso all'emergere di attori completamente nuovi nel settore e catalizzando cambiamenti radicali nel design, nella funzionalità e nell'utilizzo dei dispositivi".

Un esempio evidente di questa possibilità si sta verificando già ora con l'avvento dei servizi offerti attraverso i Large Language Model come GPT.

Di fatto, la maggioranza degli utenti ad oggi non utilizza più Google Search per cercare informazioni, ma piuttosto si affida a strumenti di intelligenza artificiale ed il motivo è ben citato da Alessandro De Vita nel blog *"AI vs Search Engines — Are We Witnessing the End of Google?"*¹¹⁵ in cui afferma che *"a differenza dei tradizionali motori di ricerca che forniscono agli utenti un elenco di link da consultare, questi modelli di intelligenza artificiale forniscono risposte immediate e dirette in un formato conversazionale"*.

Oltre a ciò, gli agenti intelligenti potrebbero anche agire in un altro contesto. Potrebbero infatti essere in grado di cambiare il modo in cui le aziende riescono trovare nuovi clienti. Attualmente questo obiettivo viene perseguito attraverso metodi come la Search Engine Optimization – SEO o la pubblicità online.

Ad oggi, infatti, un'azienda per farsi conoscere deve tentare di comparire tra i primi risultati di una ricerca Google o utilizzare inserzioni su piattaforme come Facebook o Instagram. Questi strumenti, però, sono nelle mani delle solite poche grandi aziende che di fatto decidono chi mostrare e quanto costa essere mostrati.

Gli sviluppi negli AI Agent però possono suggerire all'utente prodotti sulla base delle sue preferenze o bisogni senza passare da un motore di ricerca né da una pubblicità.

¹¹⁴ Kramer et al., *AI Agents and Ecosystems Contestability*, cit.

¹¹⁵ Devita, A. (2024). *AI vs Search Engines: Are We Witnessing the End of Google?* Medium.
<https://medium.com/@devitaalexandro/ai-vs-search-engines-are-we-witnessing-the-end-of-google-b8d6f1667fe6>

Per esempio, si potrebbe richiedere un consiglio per un prodotto tipico da regalare e l'agente potrebbe consigliare una piccola azienda che produce prodotti tipici. Esso potrebbe quindi farlo anche se questa non compare tra i primi risultati su Google o anche se non ha fatto pubblicità: la scelta deriva solo da una ricerca dell'AI Agent e della sua valutazione in funzione della pertinenza tra le caratteristiche dell'azienda e la richiesta dell'utente.

Dunque, anziché dipendere da Google Search o dall'advertising nelle piattaforme più grandi, le aziende possono essere raggiunte dagli utenti in funzione della loro capacità di “piacere” all'AI Agent.

Se questo contesto rappresenta una grandissima opportunità per cambiare il paradigma tecnologico attuale, è doveroso però ricordare, come emerso nei capitoli precedenti, che le Big Tech si stanno già muovendo con sforzi significativi verso lo sviluppo degli agenti.

Il motivo è ora chiaro: se si comprende che chi possiede l'agente possiede anche il punto di accesso ai servizi digitali, allora è logico che le aziende si stiano muovendo per accaparrarsi la possibilità di imporsi come agente adottato dalla massa.

In definitiva, tutto dipenderà da chi fornirà questi AI Agents. In uno scenario ottimistico la situazione descritta avverrà solo se i nuovi agenti nasceranno da startup o da new entrants.

In uno scenario pessimistico, invece, se gli agenti vengono forniti dagli attuali gatekeeper, questo rappresenterebbe solamente uno “spostamento interno”: per Google, ad esempio, si passerebbe dall'utilizzo di Google Search, all'utilizzo del suo agente proprietario, ma comunque alla base della ricerca rimarrebbe lo stesso attore, Google.

Da questo punto di vista, la regolamentazione giocherà un ruolo fondamentale, ma, nonostante ciò, rimane il fatto che gli AI Agents offrono un'opportunità che può e deve essere sfruttata per abilitare l'ingresso di nuovi attori, limitando la supremazia di quelli già attivi.

3.3.2 Minori costi di ricerca

Un altro fattore rilevante che agisce positivamente a livello di concorrenza è la possibilità che gli AI Agents offrono di aumentare la scelta offerta ai consumatori e ridurre così i costi di ricerca associati.

Ad oggi, infatti, la scelta di una piattaforma piuttosto che un'altra da parte dell'utente è spesso vincolata al costo, in termini di tempo, di ricerca e valutazione dell'alternativa più adatta. Ad esempio, un consumatore può scegliere di comprare un certo tipo di articoli esclusivamente su una piattaforma come Amazon per questioni di abitudine o di *status quo* (per cui si preferisce optare per una soluzione che già si conosce).

Inoltre, all'interno delle piattaforme stesso entrano in gioco meccanismi che rendono più costosa la ricerca dell'utente e lo spingono ad accontentarsi dei primi risultati che gli vengono mostrati, che tipicamente potrebbero non essere i più economici o adatti.

Un esempio in questo caso è proprio quello di Amazon che non consente di ordinare i prodotti per prezzo/unità, aumentando i costi di ricerca per chi vuole trovare l'articolo più conveniente¹¹⁶.

Da questo punto di vista, gli AI Agents offrono un punto di svolta perché come strumenti digitali altamente intelligenti ed efficaci in termini di ricerca e selezione, permetterebbero di confrontare in pochi secondi migliaia di siti web, in funzione della variabile di interesse del consumatore come, per esempio, il prezzo.

Un agente intelligente può anche effettuare un confronto più efficace basandosi su una vasta gamma di informazioni di diverso tipo come recensioni o acquisti recenti.

In tutto questo l'utente finale partecipa come semplice osservatore. Bush (2025) in *"Consumer Law for AI Agents"*¹¹⁷ afferma esattamente: *"Questi problemi, però, potrebbero essere mitigati grazie agli AI agent. Mentre per noi umani è semplicemente troppo difficile e lungo cercare ovunque il prodotto più adatto, gli AI agent possono farlo senza fatica. Inoltre, questi "Custobot" (un tipo di AI agent focalizzato sui clienti) saranno in grado di usare una quantità molto più vasta di dati per confrontare i prodotti: pensa alle recensioni dei clienti, ai forum di utenti, ai test sui prodotti, e ai contenuti online. Questo potrebbe aiutare i consumatori a ridurre le attuali asimmetrie informative"*.

Lo stesso documento evidenzia anche come gli AI Agents risolvano sotto questo punto di vista anche molti problemi associati all'asimmetria informativa (un contesto di inefficienza in cui in un processo di acquisto un soggetto ha meno informazioni rispetto l'altro).

Uno tra questi sono i dark pattern presenti in molti siti web: le piattaforme offrono interfacce volutamente progettate per indirizzare il comportamento degli utenti, come potrebbe essere, per esempio, il posizionare il pulsante "annulla" o "disiscriviti" in punti scarsamente visibili o accessibili (per scoraggiare l'azione dell'utente).

Queste strategie sarebbero annullate con l'utilizzo degli AI Agents perché si presterebbero come strumenti oggettivi di scelta ed estremamente precisi nella ricerca di informazioni e dati utili per indirizzare l'acquisto in funzione della variabile di interesse.

¹¹⁶Hastings Science & Technology Law Journal (2024). *U.S. Algorithmic Regulation and Competitive Harm*. https://repository.uclawsf.edu/cgi/viewcontent.cgi?article=1139&context=hastings_science_technology_law_journal

¹¹⁷ Li et al., *Foundation Model Competition*, cit.

In aggiunta, gli AI Agent risolverebbero il già citato problema legato alle SEO: i clienti, per questioni di bias tendono tipicamente a selezionare le prime soluzioni offerte nella ricerca o proposte dall'e-commerce. Un agente non è influenzabile da questo meccanismo ma può ordinare i prodotti o le soluzioni in funzione del reale interesse dell'utente finale.

Ecco, dunque, che proprio in questi mesi si sta assistendo all'avvento dell'AI Agent Optimization – AAO come sostituto al tradizionale SEO. Di fatto, mentre ad oggi l'obiettivo dei venditori era quello di “piacere” a Google, per esempio, per essere mostrati come prima soluzione, in futuro la prospettiva cambierà. Occorrerà infatti proporre prodotti che siano interessanti dal punto di vista dell'AI Agent il quale analizzerà fattori come dati, schede di prodotto ben compilate, recensioni positive, questioni legate alla logistica, politiche di reso, ecc¹¹⁸.

Le conseguenze a livello concorrenziale sono evidentemente molto positive: le aziende comprendono che non potranno più fare affidamento su bias cognitivi e questioni di status quo, ma piuttosto potranno competere solo su qualità e prezzo.

Ciò rappresenta un importante stimolo concorrenziale che premia le migliori aziende a prescindere dalla notorietà del brand o dalla grandezza.

In conclusione, si passa da un mercato in cui si è abituati a vedere la vittoria di chi controlla il canale di distribuzione o la visibilità, ad un contesto in cui vince chi offre la soluzione ottimale in termini di criteri di interesse valutati dall'utente per mezzo dell'agente.

Il potere delle Big Tech che, come emerso dai capitoli precedenti, è fondato proprio sul controllo della domanda (si è visto come Google o Amazon decidano di mostrare prima e dare priorità a certi prodotti solo sulla base di fattori soggettivi) cala drasticamente.

Anche in questo caso, però, occorre che sia soddisfatto il requisito di neutralità dell'agente e che quindi esso sia orientato agli interessi dell'utente e non della piattaforma che lo ha sviluppato.

3.3.3 Sviluppi open source

L'ultimo fattore che si vuole analizzare come opportunità offerta dagli AI Agents nella riduzione del potere associato alle Big Tech è legato al fatto che essi possono essere sviluppati su piattaforme open source.

¹¹⁸ Agenda Digitale (2024). *AI Agent Optimization: come fare eCommerce con gli assistenti AI*. <https://www.agendadigitale.eu/mercati-digitali/e-commerce/ai-agent-optimization-come-fare-e-commerce-con-gli-assistenti-ai/>

In particolare, lo sviluppo di agenti basati su una piattaforma aperta, rappresenta un'opportunità che agisce a livello di interoperabilità degli agenti.

Tipicamente, infatti, le Big Tech basano il loro vantaggio competitivo su ecosistemi chiusi che impediscono l'interoperabilità e impongono ai concorrenti condizioni severe per far parte della piattaforma. Oltre al già citato caso Google-Android, un esempio di questo contesto è fornito da Apple che offre un ecosistema altamente chiuso: nei sistemi Apple (iPhone o Mac, per esempio) gli utenti non possono installare applicazioni da fonti esterne e sono quindi vincolati all'utilizzo dell'App Store in cui ogni app deve essere approvata da Apple che impone una commissione fino al 30% su tutte le transazioni¹¹⁹.

Ciò a cui stiamo assistendo oggi è una corsa verso entrambe le direzioni: i colossi tech (Google, Microsoft, Amazon e Open AI) investono massicciamente in agenti da integrare nelle loro piattaforme chiuse, mentre Meta e molte altre startup rilasciano continuamente modelli e anteprime di agenti aperti¹²⁰.

In questo modo, oltre che accelerare lo sviluppo (perché si lavora in maniera "condivisa") dell'intelligenza agentica, si riducono drasticamente le barriere d'ingresso per tutti coloro che vogliono sviluppare il proprio AI Agent. Tra i vari strumenti open source disponibili per lo sviluppo degli agenti c'è, per esempio, il già discusso framework LangChain, utile per l'orchestrazione tra gli agenti.

Un contesto open source è una valida soluzione per la creazione di un ecosistema interoperabile, ideale per lo sviluppo della concorrenza. Un esempio di seguito può aiutare a comprendere più chiaramente il contesto.

Si immagini che ogni grande azienda, come Google o Microsoft, sviluppi i propri agenti di intelligenza artificiale in modo che essi siano in grado di operare solo ed esclusivamente con i propri servizi. Un AI Agent di Google, quindi, sarebbe in grado di funzionare solo con Gmail, Google Maps, ecc., ma non con altri servizi offerti da Microsoft (Outlook, Teams, ...).

Questo rappresenta ciò che si intende per ecosistema chiuso e ciò, si comprende bene, è un grande limite alla concorrenza.

In questo contesto, costruire un ecosistema di AI Agents aperto, dove il sistema è in grado di interagire con diversi servizi forniti anche da altre piattaforme, evita che una sola azienda

¹¹⁹ DDay.it (2024). *Apple cambia ancora l'App Store in Europa: il nuovo sistema di commissioni è un puzzle.*
<https://www.dday.it/redazione/53504/apple-cambia-ancora-lapp-store-in-europa-il-nuovo-sistema-di-commissioni-e-un-puzzle>

¹²⁰ HCLTech (2024). *The Race to Define Agentic AI Standards: Open Source vs Big Tech.*
<https://www.hcltech.com/trends-and-insights/race-define-agentic-ai-standards-open-source-vs-big-tech#:~:text=A%20global%20race%20is%20underway,of%20this%20new%20AI%20era.&text=Agentic%20AI%20isn't%20just,and%20adapt%20based%20on%20feedback.>

diventi l'unico punto di accesso per usare un certo agente e soprattutto permette agli utenti di scegliere liberamente le migliori soluzioni più adatte alle loro esigenze, senza limitarsi all'ecosistema proprietario associato all'agente.

Si riassumono quindi di seguito i vantaggi che un contesto open source legato agli agenti porta con sé:

- L'open source permette a tutte le piccole aziende (startup, PMI, centri di ricerca, ...) di accedere ad una tecnologia in via di sviluppo, ma senza dover investire un ingente capitale (cosa che invece le Big Tech possono fare senza problemi). Si pensi ad una piccola azienda che per sviluppare il proprio agente, può decidere di basarlo su modelli open source come LLaMA di Meta e utilizzare dataset aperti per addestrarlo, piuttosto che affidarsi ad API a pagamento fornite per esempio da Microsoft. Ciò abbassa le barriere all'ingresso in termini di costi iniziali: più attori possono partecipare, più aumenta la concorrenza e lo sviluppo innovativo.
- Contesti open source favoriscono, come già detto, l'interoperabilità, fondamentale per evitare fattori di lock-in: gli utenti (o sviluppatori) non sono vincolati ad un ecosistema fisso, ma possono combinare diverse componenti (il servizio di ricerca Google, il servizio di mail Microsoft, ...). In questo modo non c'è un'unica azienda che domina ad ogni livello fornendo tutte le applicazioni per ogni ambito di utilizzo.

È bene notare che il contesto open source non è solamente un'ipotesi, ma una realtà concreta che le Big Tech hanno compreso e con la quale comprendono di doversi confrontare necessariamente.

Un esempio è la reazione di alcune grandi aziende come Microsoft che ha proposto un software open source chiamato Autogen che permette a più agenti di cooperare (orchestrazione)¹²¹. L'interesse di Microsoft è ovviamente quello di proporre, attraverso una piattaforma aperta, il proprio standard con il conseguente obiettivo di imporre i propri agenti come dominant design (ovvero un'architettura che si conferma a livello globale e che offre vantaggi economici e di mercato all'azienda proprietaria per un lasso di tempo prolungato).

In conclusione, un ecosistema agentic aperto è decisamente auspicabile da un punto di vista concorrenziale e i presupposti presentano una possibilità concreta che deve essere sfruttata per evitare la persistenza di un'eccessiva concentrazione di potere.

¹²¹ Microsoft Research (s.d.). *Project AutoGen*.
<https://www.microsoft.com/en-us/research/project/autogen/>

Sezione 4

Conclusioni e prospettive future

4.1 Riepilogo di opportunità e rischi degli AI Agents

Gli AI Agents rappresentano ad oggi una delle innovazioni con il maggior potenziale disruptive a cui il mondo tecnologico abbia mai assistito.

Nel corso dei capitoli precedenti si è affrontato questo tema e in particolare come questi strumenti possono avere un duplice impatto: da un lato possono offrire delle opportunità senza precedenti, ma dall'altro potrebbero generare seri rischi e criticità.

A livello di opportunità, è emerso come gli AI Agents possano ridurre le barriere di ingresso: andandosi ad imporre come unica interfaccia con il consumatore finale, essi possono evitare all'utente di passare attraverso canali oggi detenuti da pochi gatekeeper come, ad esempio, motori di ricerca, applicazioni e piattaforme. Ciò significa che un utente potrebbe interfacciarsi con l'agente senza rimanere confinato ad un unico ecosistema.

Un esempio, citato nei capitoli precedenti, è quello di un agente integrato nel telefono app-less, in cui l'utente si preoccupa esclusivamente di formulare una richiesta (per esempio, "acquista un certo prodotto") e sarà poi l'agente a scegliere quale applicazione o servizio utilizzare per rispondere alla richiesta, che potrebbe essere una app proprietaria di una Big Tech, ma anche quella di una piccola startup.

Uno scenario di questo tipo apre contesti di maggior concorrenza: il potere decisionale si sposta infatti sull'agente, il quale non è vincolato ad una piattaforma chiusa (può tecnicamente sfruttare qualunque applicazione, sia essa sviluppata da Microsoft, piuttosto che da Google o da altri ancora).

In altre parole, un ecosistema aperto di AI Agents, capace di sfruttare diversi servizi di ogni tipo, impedirebbe ad una singola azienda di essere l'unico punto di accesso ai servizi digitali e, così facendo, aumenterebbe la libertà di scelta del consumatore finale, favorendo la concorrenza.

Come visto, si sta già assistendo ad uno scenario molto simile in cui moltissimi utenti stanno già sostituendo la tradizionale ricerca effettuata su Google Search, con strumenti di AI i quali, in maniera molto più precisa e personalizzata, offrono risposte dirette, piuttosto che una lista di link da consultare.

Un secondo fattore di opportunità riguarda la riduzione dei costi di ricerca e delle asimmetrie informative. Gli AI Agents possono analizzare e consultare un'elevatissima mole di dati,

comparando diversi fattori (come prezzo, recensioni o caratteristiche di prodotto) cosa che tipicamente un essere umano non compie in fase di acquisto.

Ciò ha una conseguenza importante: il consumatore, attraverso l'agente, potrebbe rapidamente trovare l'offerta più adatta alle sue esigenze senza doversi limitare ad una piattaforma specifica (come potrebbe essere Amazon per l'acquisto di beni generici).

Questo concetto è tutt'altro che banale perché l'utente è fortemente influenzato da bias e fattori di status quo, per cui tende a preferire le prime soluzioni offerte dalla piattaforma (che tipicamente non sono le più adatte) e ad utilizzare piattaforme conosciute.

Un agente neutrale potrebbe quindi "liberare" il consumatore da tali bias, permettendogli di consultare un numero molto elevato di opzioni disponibili che verranno selezionate in maniera molto più accurata sulla base dei criteri di preferenza impostati dal consumatore.

Infine, l'ultima opportunità analizzata è stata quella legata allo sviluppo degli AI Agents su piattaforme open source. L'open source offre infatti la possibilità di rendere l'intelligenza artificiale agentic alla portata di tutti: anche piccole imprese, startup e centri di ricerca possono sfruttare modelli aperti (come LLaMA offerto da Meta) e dataset pubblici gratuiti per sviluppare i propri agenti senza dover investire ingenti capitali o dipendere da servi a pagamento forniti dalle Big Tech (per esempio, le API).

In questo modo le barriere d'ingresso vengono drasticamente ridotte e si favorisce un mercato in cui possono partecipare molti *new entrants*, fattore che stimola la concorrenza e riduce la concentrazione.

Quando si fa riferimento all'ingresso di nuovi entranti, questo è un contesto reale e seriamente incisivo. L'esempio lampante in questo caso è offerto dalla startup cinese DeepSeek che ha dimostrato come innovazioni tecnologiche radicali possano arrivare anche da nuovi entranti: l'azienda cinese ha infatti sviluppato un modello linguistico che, nonostante alcuni limiti e controversie, si è dimostrato altamente efficiente ed in grado di mettere in seria difficoltà anche i modelli sviluppati dalle consolidate Big Tech.

Inoltre, il concetto open source è legato a quello di interoperabilità tra agenti e servizi: se ogni colosso tecnologico sviluppasse agenti funzionanti solo ed esclusivamente all'interno del proprio ecosistema (si pensi all'agente di Google, in grado di funzionare esclusivamente su Android e con le applicazioni fornite da Google stessa), l'utente sarebbe vincolato nelle scelte e ciò, oltre che a ridurre le possibili soluzioni offerte al consumatore, aumenta anche la concentrazione di mercato.

Di contro, la possibilità di un ecosistema aperto, offrirebbe un agente in grado di interagire con servizi o prodotti offerti anche da diverse piattaforme e ciò permetterebbe agli utenti di combinare più soluzioni in funzione delle proprie esigenze.

Accanto a queste opportunità, non mancano però anche seri rischi e problematiche che meritano un'attenzione particolare.

Innanzitutto, c'è il pericolo che gli AI Agents rafforzino tutte quelle dinamiche di concentrazione a cui i mercati digitali stanno già assistendo, andando ad amplificare il potere di chi ad oggi ha già un vantaggio competitivo.

Nei capitoli precedenti si è visto come gli agenti tendono a generare effetti di rete che, nel loro caso, sono amplificati da due lati. Le grandi piattaforme dispongono infatti di una grande mole di dati e una base utenti estesa che alimentano tali effetti, generando cicli auto-rinforzanti per l'agente.

Ad esempio, una Big Tech come potrebbe essere Microsoft, può facilmente preinstallare il proprio agente nel suo sistema operativo e, così facendo, esso si imporrebbe come strumento agentico alla base di milioni di dispositivi e utenti, assicurandosi una massima adozione sin dal primo giorno e avviando meccanismi di lock-in nei consumatori finali.

Questo fenomeno di self-preferencing è molto più pericoloso nel contesto degli AI Agents rispetto ai mercati tradizionali: l'agente agisce in autonomia e in modo spesso non chiaro, per cui l'utente non sempre sarebbe in grado di rendersi conto se le soluzioni proposte sono davvero le più adatte o, piuttosto, influenzate da fattori e preferenze associate all'azienda fornitrice dell'agente stesso.

Di fatto, l'agente è fondato su un modello che tende a massimizzare il profitto della piattaforma piuttosto che l'utilità del consumatore finale, per questo, esso finirà per consigliare i prodotti sulla base di accordi commerciali, partnership o fattori di integrazione, anziché per qualità e convenienza per il cliente.

Dunque, la piattaforma che controlla l'agente acquisisce un enorme potere di selezione ed esclusione del mercato: dal lato della domanda decide cosa l'utente può vedere, mentre dal lato dell'offerta stabilisce quali debbano essere le condizioni affinché un venditore possa essere mostrato (finendo per preferire alcuni ed escluderne altri).

Un altro rischio discusso è quello della collusione tacita. Se molte aziende finissero per affidarsi allo stesso agente o ad agenti basati sullo stesso LLM di sviluppo, queste potrebbero arrivare a prendere decisioni strategiche allineate, se non uguali, riducendo così la competizione e generando un contesto di collusione non esplicita.

Si è visto il caso RealPage e come, allo stesso modo, potrebbero instaurarsi condizioni collusive di fatto se in futuro molti venditori utilizzassero agenti con logiche simili per gestire scelte strategiche come quelle relative ai prezzi. Tutto ciò sarebbe complicato da individuare, ma soprattutto da regolare con gli attuali strumenti regolativi antitrust.

Per ultimo, ma non per importanza, si è affrontato il problema legato all'accesso agli input essenziali per lo sviluppo e il funzionamento degli AI Agents. Ciò richiede infatti risorse strategiche come enormi quantità di dati, chip (GPU), infrastrutture cloud e piattaforme software, che solo poi attori a livello globale possiedono e controllano direttamente.

Colossi come Amazon, Microsoft e Google sono altamente integrati su tutta la filiera o hanno accordi strategici con aziende attive in sviluppi di intelligenza artificiale agentica: esse forniscono spesso il cloud a terzi, ma allo stesso sviluppano i propri agenti che lavorano sulla base di quelle stesse infrastrutture fornite ai competitor.

Si assiste così ad un duplice vantaggio competitivo: da un lato le imprese possono garantirsi condizioni vantaggiose in termini di accesso a risorse negandole o limitandole a potenziali concorrenti, dall'altro lato possono integrare i loro agenti sulle loro piattaforme (SO, software o dispositivi) che contano già un significativo numero di utenti.

Un tale “doppio controllo” può facilmente bloccare qualunque tentativo di ingresso nel mercato per agenti di new entrants: chi vuole partecipare deve necessariamente appoggiarsi ad infrastrutture delle Big Tech (non essendoci alternative), andando così di fatto a finanziare i propri competitor nel mercato degli agenti. L'alternativa considerata è stata quella di sistemi open source che però, ad oggi, non offrono prestazioni confrontabili con quelle dei modelli proprietari.

Nel corso di questa sezione si cercherà di proporre un'analisi complessiva del contesto, tenendo in considerazione questo panorama ambivalente.

Da un lato gli AI Agents si presentano come strumenti altamente disruptive in grado di offrire la possibilità di ribilanciare i mercati digitali; dall'altro, l'attuale realtà mostra una situazione molto delicata in cui le Big Tech sono decisamente in vantaggio nella corsa agli agenti e dispongono di tutti gli strumenti necessari per garantirne un loro sviluppo altamente efficiente. Nei prossimi paragrafi si analizzerà nel dettaglio se e come gli AI Agents potranno effettivamente ridurre la concentrazione di potere nelle mani dei principali attori o se, al contrario, finiranno per confermarne la supremazia.

Da questo punto di vista si cercherà di comprendere quali misure siano necessarie affinché il progresso innovativo-tecnologico sia garantito, mantenendo al contempo livelli accettabili di concorrenza.

4.2 Occasione di apertura o ulteriore consolidamento

In questo capitolo si tenterà di offrire una risposta alla domanda focale di questa analisi, ovvero se gli AI Agents saranno davvero in grado di ridistribuire il potere che ad oggi è concentrato quasi completamente nelle mani delle Big Tech.

In altre parole, gli agenti intelligenti favoriranno un mercato più aperto e competitivo o accentueranno il potere delle Big Tech, spostando solo il contesto di competizione (dal motore di ricerca all'agente, per esempio)?

L'analisi svolta in questa tesi sembra suggerire che entrambi gli esiti sono plausibili, ma non allo stesso modo probabili. Da un lato, gli scenari ottimistici aprono alla speranza di mercati ricchi di attori e indebolimento dei poteri monopolistici. Dall'altro lato, uno scenario pessimistico, decisamente più in linea con le dinamiche attuali, vede le Big Tech confermarsi nella loro posizione dominante mantenendo, o addirittura rafforzando, la loro supremazia.

È ancora presto per definire in quale direzione si muoverà il mercato, ma le analisi finora affrontate sembrano confermare in maniera abbastanza netta che le grandi imprese tecnologiche hanno ben compreso il potenziale strategico degli AI Agents e stanno muovendosi con ingenti investimenti per dominarne lo sviluppo. Il motivo è semplice: chi controlla l'agente, controlla l'unico punto di accesso ai servizi digitali, quindi la relazione diretta con il cliente finale.

Non solo, per le Big Tech farsi trovare impreparate potrebbe significare perdere la propria posizione dominante a vantaggio di nuovi player.

Ecco perché, come emerso nei capitoli precedenti, stiamo assistendo ad un'attenzione sempre maggiore dei colossi tecnologici per gli agenti intelligenti. Ognuno di loro sta sviluppando la propria soluzione e alcune anteprime vengono già integrate nelle piattaforme e nei servizi già offerti.

Laddove questi sviluppi proseguano con successo, il rischio è che si avveri lo scenario pessimistico: saranno attivi nel mercato un agente per ogni attore (Google, Microsoft, Amazon, e così via), ciascuno profondamente integrato nell'ecosistema proprietario del rispettivo fornitore.

Dal lato della domanda, la maggior parte degli utenti utilizzerà l'agente già fornito di default sul proprio dispositivo (o piattaforma di servizio) per i motivi di inerzia cognitiva o convenienza. Questo è ciò che si verifica già oggi con molte applicazioni preinstallate nei dispositivi (si pensi alle app orologio, meteo o mappe).

In questo contesto, i benefici che gli AI Agents sarebbero in grado portare verrebbero in buona parte annullati: per esempio, Google manterrebbe comunque il controllo dell'informazione

perché l'utente, seppur non passando direttamente per Google Search, continua ad usare uno strumento comunque appartenente all'ecosistema Google, il che non intacca la sua posizione di partenza.

Il risultato sarebbe che le Big Tech confermano il loro potere e anzi, lo rafforzano, considerando che gli AI Agents offrono maggiori possibilità di lock-in rispetto alle tradizionali applicazioni (per via della loro elevata personalizzazione e adattabilità al cliente).

Nel caso in cui nemmeno l'interoperabilità sia garantita efficacemente, non si assisterebbe più nemmeno a quella minima concorrenza che ad oggi ancora si può osservare, se si pensa a Google, come motore di ricerca, in contrapposizione a Bing.

Di fatto, se ogni ecosistema ha il suo agente esclusivo, la concorrenza si svolge solo tra ecosistemi integrati (come per esempio oggi, iOS e Android) e quindi in un mercato altamente concentrato.

A confermare ulteriormente la possibilità dello scenario pessimistico ci sono fattori come le esternalità di rete che favoriscono un contesto del tipo *the winner takes it all*. Concretamente parlando, i colossi tecnologici possono sussidiare i propri agenti (quindi offrirli gratuitamente o abbinati ad altri servizi) pur rimettendoci, ma garantendosi così una rapida diffusione a discapito di agenti concorrenti più piccoli, impossibilitati a competere nei prezzi e monetizzare per via delle scarse risorse (economiche, ma anche in termini di input come dati, utenti, infrastrutture).

Ciò non è tutto, proprio in termini di risorse le Big Tech potrebbero ostacolare l'utilizzo del proprio cloud da parte di sviluppatori di agenti concorrenti, o praticare prezzi svantaggiosi.

Si pensi al caso in cui, in futuro, tutti i dispositivi Android integreranno l'AI Agent di Google e per installarne uno diverso servano operazioni complesse. La conseguenza sarà che solo pochi consumatori si sforzeranno di cambiare, confermando all'agente di Google la posizione dominante.

In conclusione, dunque, la bilancia pende maggiormente verso l'ipotesi che gli AI Agents non saranno direttamente in grado di abbattere lo strapotere delle Big Tech, a meno che non intervengano fattori esterni come autorità antitrust, ma anche la consapevolezza dei consumatori, a riequilibrare la situazione.

Prima di affrontare il tema di quali leve abbiano a disposizione le autorità antitrust per evitare questo esito, si introduce un'interessante considerazione per cui, se anche il potere degli AI Agents dovesse consolidarsi nei soliti pochi attori, una conseguenza che rimane comunque positiva è che, se prima le Big Tech si spartivano i mercati in diversi settori, con l'avvento dell'intelligenza artificiale agentica, saranno costrette a competere fra di loro in ogni mercato.

4.3 Competizione tra le Big Tech

Attualmente i colossi tecnologici dominano prevalentemente su servizi distinti: Google detiene il mercato dell'informazione, Amazon domina nell'e-commerce e Microsoft è forte nel software. Per questo motivo ad oggi non si assiste ad una vera e propria competizione diretta tra questi attori nei loro settori dominati.

Lo sviluppo degli AI Agents, anche se avvenisse per mano di questi attori, potrebbe andare a rompere un contesto di questo tipo in cui ognuno si preoccupa prevalentemente del suo mercato. Se Microsoft, Google e Amazon offrissero ognuno il proprio AI Agent, essi potrebbero trovarsi a competere l'uno contro l'altro per spartirsi gli utenti.

Si riporta un esempio per rendere più chiaro questo possibile contesto. Un agente intelligente potrebbe occuparsi di cercare informazioni, effettuare acquisti online o prenotare su piattaforme diverse, indipendentemente dal proprietario. In questo scenario ciascuna Big Tech offrirebbe agenti diversi per gli stessi bisogni del cliente finale. La conseguenza è che così facendo, tali aziende si ritroverebbero in diretta competizione sugli stessi servizi.

In altre parole, anziché ragionare autonomamente ognuno per i propri servizi (come stanno facendo ora), le Big Tech potrebbero dover far fronte ad una competizione ad armi pari: se Amazon, Google e Microsoft dovessero sfidarsi sulle stesse caratteristiche di un agente, sarebbero incentivate ad innovare e a differenziare i propri agenti per accaparrarsi più utenti.

4.4 Il contesto normativo

Considerando le importanti implicazioni associate allo sviluppo degli AI Agents, governi ed autorità antitrust avranno la seria responsabilità di strutturare un contesto che massimizzi tutti i benefici associati all'innovazione e al contempo eviti i possibili scenari di squilibrio di mercato. L'intelligenza artificiale si muove sempre più velocemente ed affidarsi ad una regolamentazione ex-post (come le sanzioni a fatto compiuto), rischia di non essere sufficiente, considerato che gli AI Agents potrebbero facilmente consolidare il potere di singoli attori creando contesti difficili da intaccare a posteriori.

Numerosi sono gli esperti che concordano sul fatto che siano necessari interventi ex-ante, ispirati ai principi di equità e contestabilità, come quelli introdotti dal Digital Markets Act¹²². Ciò vuole dire che, invece di aspettare che i colossi tecnologici diventino troppo potenti e danneggino l'assetto concorrenziale del mercato in maniera irrimediabile, occorre stabilire delle

¹²² Colangelo, G. (2024). *Competitive Risks of Generative AI and the Digital Markets Act*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5242028

regole chiare, applicabili sin da subito per prevenire comportamenti anticoncorrenziali come quelli emersi nei capitoli precedenti.

Questi interventi hanno quindi l'obiettivo di garantire che il mercato rimanga aperto a possibili new entrants, quindi competitivo per tutti gli attori, anche i più piccoli e autonomi.

Di seguito vengono proposte alcune raccomandazioni a livello di policy, basate sulle criticità emerse nel corso di questa analisi.

Promuovere l'interoperabilità e gli standard

Per evitare la divisione del mercato tra pochi attori i quali propongono ognuno i propri ecosistemi chiusi, le autorità dovrebbero stimolare, o se necessario imporre, la definizione di standard comuni che permettano agli AI Agents di differenti fornitori di interagire e comunicare con servizi e piattaforme diverse.

Da questo punto di vista, potrebbero essere requisiti di compatibilità API o di condivisione di protocolli di funzionamento: ad esempio, un agente dovrebbe essere in grado di operare, con le stesse possibilità e senza restrizioni, su un sistema operativo di qualunque tipo o fornitore.

In ambito europeo, il Digital Markets Act prevede già degli obblighi a livello di interoperabilità tra servizi.

Un altro esempio è l'obbligo per i gatekeeper di permettere l'interoperabilità dei propri AI Agents con servizi e hardware di terze parti: se un'azienda come Google ha un assistente già integrato su Android, essa deve permettere all'utilizzatore di poter utilizzare altri agenti, anche se sviluppati da altri fornitori, sullo stesso sistema operativo.

Dunque, un ecosistema aperto deve garantire all'utente la possibilità di scegliere liberamente l'agente che preferisce, indipendentemente dal dispositivo o dalla piattaforma utilizzata. Allo stesso modo un fornitore di servizi digitali deve poter essere raggiungibile tramite qualunque agente l'utente abbia scelto.

In questo modo si eviterebbe il rischio che gli utenti restino bloccati dentro ecosistemi chiusi controllati da una sola azienda.

Di fatto, l'interoperabilità abbasserebbe le barriere d'ingresso e impedirebbe la creazione di monopoli basati sul lock-in dell'utente.

Divieto di self-preferencing e obbligo di neutralità

Per garantire un contesto adeguatamente concorrenziale, occorre anche impedire agli agenti di agire per conto di interessi diversi da quelli dell'utente, finendo per favorire i servizi della piattaforma proprietaria o partner commerciali.

Occorre quindi stabilire regole a livello di comportamento imparziale degli AI Agents, soprattutto per quelli sviluppati e offerti da quelle imprese con un maggior potere di mercato.

Un agente dovrebbe, per esempio, essere obbligato a non discriminare servizi o prodotti ugualmente performanti in base alla loro provenienza e a non dare tantomeno preferenza ai propri. Per esempio, se l'agente sviluppato da Amazon consiglia esclusivamente prodotti venduti sulla sua piattaforma anche se esisterebbero soluzioni altrettanto adatte altrove, ciò costituirebbe la violazione del principio di neutralità.

Oltre al divieto in sé, occorre però imporre trasparenza sugli algoritmi di ranking e selezione utilizzati dagli agenti: gli utenti, come le autorità, devono essere in grado di poter conoscere i criteri generali con cui l'agente prende decisioni. Questo non è semplice, considerata la complessità di funzionamento associata ai LLM alla base del funzionamento degli agenti, ma permetterebbe di impedire favoritismi.

Già oggi il Digital Markets Act proibisce ai gatekeeper la pratica di auto preferenza dei propri servizi.

Allo stesso modo, così come accaduto nel caso Google¹²³, la Commissione Europea dovrebbe obbligare gli attori dominanti a livello di sistemi operativi o piattaforme su cui un AI Agent può girare, a presentare una schermata di scelta dell'agente durante la fase di configurazione del dispositivo.

In questo modo, si eviterebbe il contesto di lock-in per l'utilizzatore (per via di bias cognitivi e fattori di status quo) verso l'agente preinstallato.

Garanzia di portabilità dei dati

Come i capitoli precedenti hanno dimostrato, i dati sono il carburante degli AI Agents e tale risorsa è detenuta in maniera privilegiata da poche aziende a livello

¹²³ LegalTech Italia (2023). *DMA – Digital Markets Act: la nuova normativa dell'Unione Europea destinata a ridisegnare il panorama digitale europeo (e non solo)*.
<https://www.legaltechitalia.eu/dma-digital-markets-act-la-nuova-normativa-dellunione-europea-destinata-a-ridisegnare-il-panorama-digitale-europeo-e-non-solo/>

globale. Per evitare che solamente tali aziende siano in grado di addestrare agenti efficaci, occorrono delle norme che garantiscano una corretta ed equa circolazione dei dati, tenendo ovviamente conto del rispetto della privacy.

L'obbligo più importante riguarda di certo la portabilità dei dati degli utenti, per cui un consumatore dovrebbe poter liberamente trasferire al nuovo agente di un concorrente tutto quel set di dati relativo a preferenze, acquisti, profilo, e così via, che aveva accumulato nell'agente precedente.

In questo modo sarebbe possibile evitare il lock-in degli utenti dovuto alla necessità e la comodità di conservare i propri dati.

Inoltre, le autorità dovrebbero permettere un equo accesso a questa risorsa, ad esempio, assicurando a qualunque attore attivo nello sviluppo di AI Agents almeno i dati di interesse generali come le recensioni, le informazioni di prodotti e i cataloghi.

Ciò va necessariamente bilanciato con la tutela della privacy e della proprietà dei dati, ma fattori come l'open data (dati accessibili a chiunque, solitamente messi a disposizione da enti pubblici) o il data sharing (quell'attività attraverso cui un soggetto mette a disposizione di altri i propri dati, nel rispetto di certe regole, magari a pagamento) potrebbero diminuire il vantaggio competitivo dei pochi attori che oggi detengono la gran parte dei dati dei consumatori¹²⁴.

Eque possibilità di accesso ad input essenziali

Come visto, il controllo di infrastrutture chiave come cloud, chip, dati o sistemi operativi, da parte delle Big Tech può concretizzarsi in comportamenti escludenti verso possibili concorrenti nel mercato degli agenti.

Le autorità antitrust devono attentamente osservare il contesto relativo alle fusioni, partnership o clausole di contratti in questi ambiti.

Ad esempio, accordi di lungo periodo che legano un'azienda fornitrice di cloud con una sviluppatrice di LLM, devono essere tenuti sott'occhio dalle autorità perché mettono a rischio la libera concorrenza.

Da questo punto di vista, le autorità potrebbero valutare l'utilizzo di rimedi strutturali o regole di separazione tra infrastrutture essenziali e servizi che si basano su esse.

¹²⁴OECD (2024). *Artificial Intelligence, Data and Competition*. OECD Publishing.
https://www.oecd.org/en/publications/artificial-intelligence-data-and-competition_e7e88884-en.html

La storia conosce già eventi di questo tipo, come quello che ha visto protagonista l'azienda americana AT&T, la quale deteneva il monopolio delle telecomunicazioni negli Stati Uniti. Essa controllava sia i servizi telefonici, che le infrastrutture (come le linee telefoniche), che la produzione di apparecchiature. Tutto ciò le permetteva di escludere concorrenti e dominare l'intero mercato. Le autorità avviaron una causa antitrust contro AT&T che alla fine subì lo smembramento¹²⁵.

Certamente questo tipo di azioni è radicale, ma garantisce l'interruzione di monopoli e l'apertura del mercato.

Nel contesto degli AI Agents, il rischio verificatosi con AT&T riguarda diverse aziende come Amazon (la quale rappresenta il maggior fornitore di cloud, ma anche uno sviluppatore di agenti con Amazon BedRock) o Microsoft, per via del suo accordo con Open AI.

Diversi studi¹²⁶ fanno notare come questi casi appena citati rappresentino potenziali conflitti di interesse in cui il fornitore di infrastruttura potrebbe ostacolare l'accesso a soggetti rivali.

Un'alternativa all'azione più drastica dello smembramento di un'azienda potrebbe essere quella di impedire all'azienda protagonista, di non poter decidere autonomamente le condizioni per i propri agenti o più in generale per il mercato a valle.

Una o l'altra azione sono strettamente necessarie perché mercati concorrenziali come quello degli AI Agents, non possono nascere in presenza di forti barriere infrastrutturali.

Per garantire la competizione è necessario che il contesto di concorrenza sia neutrale e ciò può anche richiedere che chi fornisce le infrastrutture chiave non sia anche attore in gioco.

In conclusione, considerata la portata e la natura dell'oggetto di regolamentazione, occorrerà necessariamente una cooperazione a livello globale tra i regolatori: l'Europa con il DMA e l'AI Act, gli Stati Uniti con le iniziative della FTC e gli altri Paesi dovranno condividere considerazioni e approcci per evitare situazioni non definite ed incerte.

¹²⁵ Ezrachi, A., & Stucke, M. E. (2012). *The Curious Case of Competition and Quality*. The Antitrust Bulletin, 57(1), 191–231. JSTOR.

<https://www.jstor.org/stable/41623144>

¹²⁶ Kramer et al., *AI Agents and Ecosystems Contestability*, cit.

L'attenzione che deve essere posta, però, non deve essere solo alla definizione delle norme, ma anche e soprattutto, nel non frenare l'innovazione degli AI Agents, che anzi può portare benefici enormi, anche a livello concorrenziale.

In altre parole, è necessario che le autorità traccino dei paletti che delimitino il percorso che lo sviluppo degli AI Agents può seguire, andando così ad indirizzarlo verso un contesto che massimizzi i vantaggi collettivi (in termini di efficienza, concorrenza e nuovi servizi) e minimizzi gli effetti collaterali (come nuovi monopoli, inefficienze di mercato e per il consumatore finale).

Conclusione

Questa tesi ha cercato di mettere in evidenza luci ed ombre associate allo sviluppo di una delle tecnologie più disruptive a cui il mondo si sta preparando.

Dall'analisi emerge un quadro sfumato, ma tendenzialmente pessimistico: se le autorità dovessero fallire nei loro interventi, lo scenario più plausibile sarebbe quello di un ulteriore accentramento del potere nelle mani dei colossi tecnologici.

Se da un lato quest'ultimi stano facendo già ora il possibile per imporre i propri AI Agents, dall'altro esistono alcune possibilità per un mercato futuro più aperto e contendibile: gli stessi AI Agents possiedono tutte le caratteristiche necessarie per rompere monopoli e permettere l'ingresso di nuovi player.

È bene anche notare che, sebbene sia complicato, il progresso tecnologico e l'equilibrio socio-economico non sono necessariamente incompatibili. Al contrario, uno sviluppo della concorrenza nel mercato digitale può maggiormente stimolare l'innovazione per via della presenza di più individui attivi nel creare nuovi AI Agents.

Un'attenzione vuole essere data ad un aspetto non prettamente tecnico, ma non per questo meno importante nel contesto socio-economico associato agli sviluppi degli AI Agents.

È bene notare che, accanto all'intervento delle autorità preposte alla regolamentazione, c'è un altro attore chiave dal cui comportamento dipenderà il futuro degli AI Agents: il consumatore finale, cioè tutti noi.

Questo fattore è di cruciale importanza perché, di fatto, è proprio il consumatore ad avere il potere, più di ogni altra autorità di regolamentazione, di accettare o rifiutare le soluzioni proposte dal mercato.

Non basta dunque affidarsi esclusivamente alle azioni governative, ma occorre anche che ci sia una crescente consapevolezza degli utenti del contesto tecnologico ed economico in cui gli AI Agents operano.

Non solo, in tutto questo vale anche una questione morale: la corretta informazione e la piena consapevolezza di ciò con cui si ha a che fare riduce la probabilità di subire passivamente atteggiamenti opportunistici messi in atto dalle Big Tech.

Se si pensa alla storia passata è facile notare come spesso i consumatori abbiano beneficiato di grandi comodità in cambio di una minima, se non assente, trasparenza da parte delle aziende.

Si pensi ai social network. Questi servizi, comunemente accettati e condivisi dalla massa, sono sempre stati considerati gratuiti, ma la realtà era ben diversa. A pagare è stato sempre e

comunque il consumatore finale e il mezzo di pagamento erano proprio i suoi dati personali. Tutto questo ha avuto implicazioni che solo oggi comprendiamo a pieno, proprio come emerso anche in questa analisi.

Nel caso degli AI Agents la posta in gioco è ancora maggiore: ogni consumatore si prepara a delegare decisioni quotidiane a degli strumenti che ad oggi si presentano in maniera poco trasparente e che vengono forniti da aziende sempre più avidi di potere. Questa volta non ci si può permettere di subire passivamente questo progresso. Come anche scritto nell'introduzione, il rischio è che sia la tecnologia a guidare noi, piuttosto che il contrario.

Diventa quindi indispensabile che il consumatore sia pienamente consapevole del contesto in cui finirà per trovarsi. È importante, anzi cruciale, chiedersi chi abbia sviluppato l'agente, quali obiettivi potrebbe avere (oltre al semplice soddisfare la propria richiesta), quali dati stia utilizzando e, soprattutto, con quali potenziali bias.

Questa consapevolezza deve essere promossa anche dalle organizzazioni pubbliche, ma è importante che ciascuno, nel suo piccolo, presti attenzione agli strumenti che utilizza.

Tutto ciò può sembrare un ostacolo alla comodità che tali agenti possono offrire, ma un consumatore consapevole è anche un consumatore difficile da ingannare, che sa usare l'agente come strumento al suo servizio e non viceversa.

Concretamente, un utente potrebbe utilizzare un agente per ottenere una prima scrematura delle opzioni di acquisto, ma poi verificare autonomamente solo certe alternative.

Allo stesso tempo potrebbe utilizzare più di un agente per confrontare i risultati ottenuti. Infine, l'utente potrebbe anche indirizzare l'agente chiedendo di motivare le scelte prese.

Di certo, non tutti i consumatori diventeranno esperti, ma anche una maggioranza potrebbe costringere, o almeno incentivare, le aziende a rendere i propri agenti più trasparenti ed effettivamente orientati all'utente.

C'è anche un aspetto legato alla responsabilità sociale: sapere che dietro un agente che offre numerose comodità potrebbero nascondersi meccanismi nocivi per la concorrenza, dovrebbe spingere ognuno di noi a sostenere tutte quelle politiche e quelle aziende che invece preferiscono un approccio più etico.

Concretamente parlando, se, ad esempio, si scoprisse che un certo agente privilegia prevalentemente le proprie offerte e servizi, la reazione dei consumatori (che dovrebbe essere quella di affidarsi ad un nuovo agente) sarebbe sicuramente il metodo correttivo più efficace che punisce direttamente l'azienda colpevole.

È chiaro quindi come la consapevolezza sia un elemento imprescindibile in un mercato equilibrato dal punto di vista concorrenziale, in grado di aiutare le autorità a “far rispettare le regole”.

In conclusione, si mostra evidente che gli AI Agents rappresentano oggi una sfida e un’opportunità. La direzione che il mercato prenderà dipende però dalle decisioni e dalle scelte prese oggi.

La speranza è che, grazie ad una combinazione di azioni governative efficaci e consapevolezza del consumatore, si riesca ad indirizzare questa innovazione tanto disruptive verso un futuro in cui la tecnologia serva davvero l’uomo e non lo domini.

Il contesto è decisamente delicato, ma questo è l’equilibrio a cui occorre ambire: un mondo in cui gli AI Agents migliorano la vita dei consumatori finali, senza per questo concentrare la ricchezza e il potere nelle mani di pochi.

La partita comincia adesso e il finale, tutt’altro che scritto, dipenderà dalle mosse che verranno messe in campo fin d’ora.

Bibliografia – Sitografia

[https://www.treccani.it/vocabolario/artificiale_\(Sinonimi-e-Contrari\)/](https://www.treccani.it/vocabolario/artificiale_(Sinonimi-e-Contrari)/)

Turing, A. M. (1950). Macchine calcolatrici e intelligenza.

<https://disf.org/files/macchine-calcolatrici-e-intelligenza.pdf>

Oracle (s.d.). Cos'è il deep learning? Oracle.

<https://www.oracle.com/it/artificial-intelligence/machine-learning/what-is-deep-learning/>

Google Cloud (s.d.). Cos'è il machine learning? Google Cloud.

<https://cloud.google.com/learn/what-is-machine-learning?hl=it>

StartUs Insights (2024). Artificial General Intelligence Market Report 2025.

<https://www.startus-insights.com/innovators-guide/artificial-general-intelligence-market-report/>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017).

https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

IBM (s.d.). Self-attention: come funziona e perché è importante per l'AI. IBM Think.

<https://www.ibm.com/think/topics/self-attention>

AWS (s.d.). Cos'è il Reinforcement Learning from Human Feedback (RLHF)? Amazon Web Services.

<https://aws.amazon.com/it/what-is/reinforcement-learning-from-human-feedback/>

IBM (s.d.). Small Language Models: cosa sono e perché contano. IBM Think.

<https://www.ibm.com/think/topics/small-language-models>

NVIDIA (s.d.). Multimodal Large Language Models (MLLMs).

<https://www.nvidia.com/en-us/glossary/multimodal-large-language-models/>

Tenyks (2023). Multimodal Large Language Models (MLLMs) Transforming Computer Vision. Medium.

https://medium.com/@tenyks_blogger/multimodal-large-language-models-mlllms-transforming-computer-vision-76d3c5dd267f

Jain, A. (s.d.). Difference between CPU, GPU, TPU and NPU. Medium.

<https://medium.com/@abhishekjainindore24/difference-between-cpu-gpu-tpu-and-npu-09fca09f0bb6>

Precedence Research (2024). Graphic Processing Unit (GPU) Market – Size, Share and Trends.

<https://www.precedenceresearch.com/graphic-processing-unit-market>

Multimodal.dev (s.d.). What Hardware is Needed for AI?
<https://www.multimodal.dev/post/what-hardware-is-needed-for-ai>

Intel (s.d.). AI Hardware: Components and Architecture.
<https://www.intel.com/content/www/us/en/learn/ai-hardware.html>

Lightly (2023). What is Data Curation and Why is it Important?
<https://www.lightly.ai/blog/what-is-data-curation-and-why-is-it-important>

Markets and Markets (2024). AI Training Dataset Market by Type, Vertical, and Region – Global Forecast to 2029.
<https://www.marketsandmarkets.com/Market-Reports/ai-training-dataset-market-153819655.html>

Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., Phanishayee, A., & Zaharia, M. (2021). Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM. SC21.
<https://arxiv.org/pdf/2104.04473>

Hawkins Insight (s.d.). NVIDIA AI GPU Costs and Pricing Strategy Revealed.
<https://www.hawkinsight.com/en/article/nvidia-ai-gpu-costs-and-pricing-strategy-revealed>

Statista (2023). Azure GPU Cloud Server Costs Worldwide in 2023.
<https://www.statista.com/statistics/1296926/azure-gpu-cloud-server-costs/>

AWS (s.d.). Che cos'è un data center? Amazon Web Services.
<https://aws.amazon.com/it/what-is/data-center/>

Grand View Research (2024). Cloud Computing Market Size, Share & Trends Analysis Report 2025–2030.
<https://www.grandviewresearch.com/industry-analysis/cloud-computing-industry>

Yao, S., Zhao, Y., Yu, D., Zhao, K., Yu, P., & Chen, Y. (2022). ReAct: Synergizing Reasoning and Acting in Language Models. arXiv.
<https://arxiv.org/abs/2210.03629>

IBM (s.d.). AI Agents: il futuro dell'intelligenza artificiale autonoma. IBM Think.
<https://www.ibm.com/it-it/think/topics/ai-agents>

Chase, H. (2023). What's next for AI Agents – LangChain. YouTube.
<https://youtu.be/pBBelpk8hf4>

IBM (s.d.). LangChain: orchestrazione per modelli linguistici. IBM Think.
<https://www.ibm.com/it-it/think/topics/langchain>

IBM (s.d.). Vector Embedding: rappresentazione numerica per l'AI. IBM Think.
<https://www.ibm.com/it-it/think/topics/vector-embedding>

IoT Analytics (2024). Leading Generative AI Companies to Watch.

<https://iot-analytics.com/leading-generative-ai-companies/>

Canalys (2024). Worldwide Cloud Service Revenues Q4 2024.

<https://www.canalys.com/newsroom/worldwide-cloud-service-q4-2024>

Statista (2024). Worldwide AI funding by quarter.

<https://www-statista-com.ezproxy.biblio.polito.it/statistics/943151/ai-funding-worldwide-by-quarter/>

Medium (2024). Business and enterprise spending on AI surges 500%. TechStartups.

<https://techstartups.com/2024/11/21/business-and-enterprise-spending-on-ai-surges-500-to-13-8-billion-as-anthropic-gains-market-share-over-openai/>

Microsoft Research (s.d.). Project AutoGen.

<https://www.microsoft.com/en-us/research/project/autogen/>

Microsoft (s.d.). What is Azure?

<https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-azure>

Statista (2024). Worldwide AI funding by quarter.

<https://www-statista-com.ezproxy.biblio.polito.it/statistics/943151/ai-funding-worldwide-by-quarter/>

Il Sole 24 Ore (2024). Antitrust USA apre un'indagine su Microsoft.

<https://www.ilsole24ore.com/art/antitrust-usa-apre-un-indagine-microsoft-AGHIsRB>

Start Magazine (2024). Ecco come Microsoft è scesa a patti con i fornitori cloud in Europa.

<https://www.startmag.it/innovazione/ecco-come-microsoft-e-scesa-a-patti-con-i-fornitori-cloud-in-europa/>

Financial Times (2024). Microsoft under scrutiny for its role in AI market.

<https://www.ft.com/content/62f361eb-ce52-47c1-9857-878cfe298d54>

Start Magazine (2024). Stangata Antitrust USA contro il cloud di Microsoft.

<https://www.startmag.it/innovazione/stangata-antitrust-usa-contro-il-cloud-di-microsoft/>

Microsoft (s.d.). Microsoft Copilot: what it is and how it works.

<https://www.microsoft.com/en-us/microsoft-copilot/copilot-101/what-is-copilot>

OpenAI (s.d.). Introducing OpenAI LP.

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

Semafor (2023). Microsoft Eyes \$10 Billion Bet on ChatGPT.

<https://www.semafor.com/article/01/09/2023/microsoft-eyes-10-billion-bet-on-chatgpt>

Microsoft (2025). Microsoft and OpenAI evolve partnership to drive the next phase of AI.

<https://blogs.microsoft.com/blog/2025/01/21/microsoft-and-openai-evolve-partnership-to-drive-the-next-phase-of-ai/>

Rivista.ai (2025). OpenAI taglia la fetta di Microsoft: la nuova partita del potere nell'IA.
<https://www.rivista.ai/2025/05/07/openai-taglia-la-fetta-di-microsoft-la-nuova-partita-del-potere-nellia/>

Repubblica (2024). Accordo segreto Microsoft–OpenAI sull'AGI?
https://www.repubblica.it/tecnologia/dossier/intelligenza-artificiale/2024/12/27/news/microsoft_openai_agi_accordo_segreto_intelligenza_artificiale_generale-423908793/

Binance (2024). OpenAI and Microsoft discuss dropping AGI clause to unlock investment.
<https://www.binance.com/it/square/post/12-06-2024-openai-and-microsoft-discuss-dropping-agi-clause-to-unlock-investment-17201473052113>

Il Sole 24 Ore (2024b). OpenAI lancia Operator in Europa: l'AI che naviga e opera per te.
<https://www.ilsole24ore.com/art/openai-lancia-operator-europa-l-ai-che-naviga-e-opera-web-te-AGAqZMVD>

Reuters (2025). UK says Microsoft-OpenAI partnership does not need anti-trust investigation.
<https://www.reuters.com/technology/uk-says-microsoftopenai-partnership-does-not-need-anti-trust-investigation-2025-03-05/>

Brian Solis (2024). AI(n)sights: Microsoft's Pseudo Acquisition of Inflection AI.
<https://briansolis.com/2024/04/ainsights-microsofts-pseudo-acquisition-of-inflection-ai-and-what-it-means-to-business-customers-and-investors/>

European Commission (2024). Commission opens investigation into Microsoft's practices in the cloud sector.
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4727

Statista (2024). Google Cloud public revenue worldwide 2017–2024.
<https://www.statista.com/statistics/478176/google-public-cloud-revenue/>

Bytebridge (2023). GPU and TPU Comparative Analysis Report. Medium.
<https://bytebridge.medium.com/gpu-and-tpu-comparative-analysis-report-a5268e4f0d2a>

Federal Trade Commission – FTC (2025). FTC Issues Staff Report on AI Partnerships and Investments Study.
<https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study>

Google Cloud (s.d.). Gemini: l'AI multimodale di Google.
<https://cloud.google.com/ai/gemini?hl=it>

DeepMind (s.d.). Deep Reinforcement Learning.
<https://deepmind.google/discover/blog/deep-reinforcement-learning/>

DeepMind (s.d.). Project Astra.
<https://deepmind.google/models/project-astra/>

Il Sole 24 Ore – InfoData (2024). Gemini 2.0, Project Mariner e Deep Research: Google è entrata nell'era degli AI Agent.

<https://www.infodata.ilsole24ore.com/2024/12/15/gemini-2-0-project-mariner-e-deep-research-google-e-entrata-nellera-degli-ai-agent-datavizandtools/>

Analytics Insight (s.d.). Google TPUs vs NVIDIA GPUs for AI Development. Medium.

<https://medium.com/@analyticsinsight/google-tpus-vs-nvidia-gpus-for-ai-development-3f62fa2392d7>

Statista (2024). Amazon Annual Net Revenue Worldwide (2004–2024).

<https://www-statista-com.ezproxy.biblio.polito.it/statistics/266282/annual-net-revenue-of-amazoncom/>

Statista (2024). Amazon Web Services Revenue (2013–2024).

<https://www-statista-com.ezproxy.biblio.polito.it/statistics/233725/development-of-amazon-web-services-revenue/>

UK Government – Competition and Markets Authority (2024). Cloud Services Market Investigation – Provisional Findings.

<https://www.gov.uk/guidance/cloud-services-market-investigation-provisional-findings>

European Commission (2022). Regolamento (UE) 2022/1925 del Parlamento Europeo e del Consiglio del 14 settembre 2022 (Digital Markets Act). EUR-Lex.

<https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32022R1925>

TechTarget (s.d.). Amazon Bedrock (AWS Bedrock) – Definition.

<https://www.techtarget.com/searchenterpriseai/definition/Amazon-Bedrock-AWS-Bedrock>

Amazon Web Services (s.d.). Bedrock Agents.

<https://aws.amazon.com/it/bedrock/agents/>

The Guardian (2024). Amazon Invests Billions in Anthropic Amid AI Race.

<https://www.theguardian.com/technology/2024/nov/22/amazon-anthropic-ai-investment>

Reuters (2024). Anthropic Receives \$4 Billion Investment: Amazon Makes AWS Official Cloud Provider.

<https://www.reuters.com/technology/artificial-intelligence/anthropic-receives-4-billion-investment-amazon-makes-aws-official-cloud-provider-2024-11-22/>

AINews (2024). Il NYT scopre che Google possiede il 14% di Anthropic: le aziende sono davvero avversarie?

<https://ainews.it/il-nyt-scopre-che-google-possiede-il-14-di-anthropic-le-aziende-sono-davvero-avversarie/>

Start Magazine (2024). Amazon sfida Nvidia con i suoi microchip per l'intelligenza artificiale.

<https://www.startmag.it/innovazione/amazon-microchip-intelligenza-artificiale-nvidia/>

HDblog (2024). Meta, Zuckerberg raddoppia: investimenti in AI per 30 miliardi di dollari.
<https://www.hdblog.it/mercato/articoli/n606566/meta-zuckerberg-ai-investimenti-2025-miliardi/>

MSN (2024). Meta: previsti 1,3 milioni di GPU per l'AI entro la fine del 2025.
<https://www.msn.com/it-it/money/storie-principali/meta-previsti-1-3-milioni-di-gpu-per-l-ai-entro-la-fine-del-2025/ar-AA1xO86O?ocid=finance-verthp-feeds>

BeBeez (2024). Data center negli USA: al via il progetto Stargate di OpenAI da 500 miliardi di dollari.
<https://bebeez.it/rebreakfast/datacenter-negli-usa-al-via-il-progetto-stargate-di-openai-da-500-mld/>

Bloomberg (2025). Meta CEO Mark Zuckerberg Bets on Trump — but MAGA Is Skeptical.
<https://www.bloomberg.com/news/features/2025-05-26/meta-ceo-mark-zuckerberg-bets-on-trump-but-maga-is-skeptical>

CNBC (2025). Meta product chief: LLaMA 4 will power AI agents.
<https://www.cnn.com/2025/03/05/meta-product-chief-says-llama-4-will-power-ai-agents.html>

CNBC (2025). Nvidia loses \$600B in market cap after DeepSeek competition.
<https://www.cnn.com/2025/01/27/nvidia-sheds-almost-600-billion-in-market-cap-biggest-drop-ever.html#:~:text=Nvidia%20shares%20plunged%2017%25%20on,from%20Chinese%20AI%20lab%20DeepSeek.>

Creole Studios (2024). DeepSeek vs ChatGPT: Cost Comparison.
<https://www.creolestudios.com/deepseek-vs-chatgpt-cost-comparison/>

Garante per la protezione dei dati personali (2024). Documento ufficiale su intelligenza artificiale e privacy.
<https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10097450>

The Verge (2024). DeepSeek breach exposes AI security flaws and user data.
<https://www.theverge.com/news/603163/deepseek-breach-ai-security-database-exposed>

DigiTimes (2025). ARM CPU chips to dominate AI computing landscape.
<https://www.digitimes.com/news/a20250206VL205/arm-cpu-chips-ai.html>

Statista (2024). Top companies worldwide by market capitalization.
<https://www-statista-com.ezproxy.biblio.polito.it/statistics/263264/top-companies-in-the-world-by-market-capitalization/>

Finance Magnates (2025). The AI Underdog: How DeepSeek Triggered Nvidia's \$560B Plunge.
<https://www.financemagnates.com/trending/the-ai-underdog-how-deepseek-triggered-nvidias-record-560b-plunge/>

OpenAI (s.d.). Computer-Using Agent.
<https://openai.com/index/computer-using-agent/>

Il Sole 24 Ore (2024). Amazon lancia Nova Act: nuovo agente AI che naviga autonomamente il web.

<https://www.ilsole24ore.com/art/amazon-lancia-nova-act-nuovo-agente-che-naviga-autonomamente-web-AGfEqFwD>

Microsoft (s.d.). Microsoft Copilot Studio.
<https://www.microsoft.com/it-it/microsoft-copilot/microsoft-copilot-studio?market=it>

CB Insights (2024). AI Agent Market Map.
<https://www.cbinsights.com/research/ai-agent-market-map/>

LangChain (2024). The State of AI Agents – Use Cases.
<https://www.langchain.com/stateofaiagents#:~:text=%E2%80%8D%20The%20top%20use%20cases,Similarly%2C%20AI%20agents%20are%20boosting>

CB Insights (2024). AI Agent Trends to Watch in 2025.
<https://www.cbinsights.com/research/ai-agent-trends-to-watch-2025/>

LangChain, The State of AI Agents, cit.

Fortune (2024). AI Agent Startups Deal Count Up 81% Year-over-Year.
<https://fortune.com/2024/09/24/ai-agent-startups-deal-count-up-81-percent-year-over-year-pitchbook/>

PitchBook (2024). Y Combinator Is Going All-In on AI Agents.
<https://pitchbook.com/news/articles/y-combinator-is-going-all-in-on-ai-agents-making-up-nearly-50-of-latest-batch#:~:text=Y%20Combinator%20is%20going%20all,were%20AI%20agent%20companies>

Toews, R. (2024). Agents Are the Future of AI: Where Are the Startup Opportunities? Forbes.
<https://www.forbes.com/sites/robtoews/2024/07/09/agents-are-the-future-of-ai-where-are-the-startup-opportunities/>

Times of India (2025). Tech Leaders Share 2025 AI Predictions.
<https://timesofindia.indiatimes.com/technology/tech-news/bill-gates-sam-altman-and-other-tech-leaders-share-2025-predictions/articleshow/117021677.cms>

Kramer, J., Reggiani, C., & Colangelo, G. (2024). AI Agents and Ecosystems Contestability. CERRE – Centre on Regulation in Europe.
<https://cerre.eu/publications/ai-agents-and-ecosystems-contestability/>

In AI We Trust (2024). The Future of AI Agents: Transformative Trends and Predictions 2025–2030.
<https://inaiwetrust.com/p/the-future-of-ai-agents-transformative-trends-and-predictions-2025-2030>

Toews, Agents Are the Future of AI, cit.

OpenAI (s.d.). OpenAI + Klarna.

<https://openai.com/index/klarna/>

Kramer et al., AI Agents and Ecosystems Contestability, cit.

MarketsandMarkets (2024). AI Agents Market – Global Forecast to 2029.

<https://www.marketsandmarkets.com/Market-Reports/ai-agents-market-15761548.html>

Capgemini (2024). Generative AI in Organizations 2024.

<https://www.capgemini.com/us-en/insights/research-library/generative-ai-in-organizations-2024/>

Accenture (2017). Artificial Intelligence Can Increase Profitability by 38% by 2035.

<https://newsroom.accenture.com/news/2017/accenture-report-artificial-intelligence-has-potential-to-increase-corporate-profitability-in-16-industries-by-an-average-of-38-percent-by-2035>

In AI We Trust, The Future of AI Agents, cit.

Li, C., Zhang, T., Du, Y., et al. (2024). Foundation Model Competition and Vertical Integration. SSRN.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5187056

Scheibeneif, A. & Raskino, M., Twigg-Flesner, C., & Howells, G. (forthcoming 2025).

Adapting Consumer Law to New Technologies, in Brownsword, R. & DiMatteo, L. (Eds.), The Cambridge Handbook of the Governance of Technology. Cambridge University Press.

Kramer et al., AI Agents and Ecosystems Contestability, cit.

CIO (2024). Big Data Is Not a Barrier to Entry.

<https://www.cio.com/article/228353/big-data-is-not-a-barrier-to-entry.html#:~:text=Just%20as%20big%20data%20is,free%20data%20available%20to%20the m>

Milvus (2024). What Is the Significance of Dataset Size in ML Performance?

<https://milvus.io/ai-quick-reference/what-is-the-significance-of-dataset-size-in-machine-learning-model-performance>

Ganor, A. et al. (2024). Training Compute as a Barrier to Entry in AI. SSRN.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5250253

Li et al., Foundation Model Competition, cit.

European Commission (2018). Commission Opens Investigation into Amazon's Data Use.

https://ec.europa.eu/commission/presscorner/detail/en/ip_18_4581

Clarke, L. (2024). RealPage: l'algoritmo che falsa i prezzi degli affitti. Corriere della Sera.
https://www.corriere.it/economia/affitti/24_agosto_24/affitti-sotto-accusa-l-algoritmo-di-realpage-di-thoma-bravo-che-controlla-dardtrace-falsa-i-prezzi-a05ce8ba-baf7-4505-9fb8-b56891ed1x1k.shtml

Kramer et al., AI Agents and Ecosystems Contestability, cit.

Kramer et al., AI Agents and Ecosystems Contestability, cit.

Kramer et al., AI Agents and Ecosystems Contestability, cit.

Federal Trade Commission – FTC (2024). FTC Launches Inquiry into Generative AI Investments and Partnerships.
<https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships>

Kramer et al., AI Agents and Ecosystems Contestability, cit.

Devita, A. (2024). AI vs Search Engines: Are We Witnessing the End of Google? Medium.
<https://medium.com/@devitaalejandro/ai-vs-search-engines-are-we-witnessing-the-end-of-google-b8d6f1667fe6>

Hastings Science & Technology Law Journal (2024). U.S. Algorithmic Regulation and Competitive Harm.
https://repository.uclawsf.edu/cgi/viewcontent.cgi?article=1139&context=hastings_science_technology_law_journal

Li et al., Foundation Model Competition, cit.

Agenda Digitale (2024). AI Agent Optimization: come fare eCommerce con gli assistenti AI.
<https://www.agendadigitale.eu/mercati-digitali/e-commerce/ai-agent-optimization-come-fare-e-commerce-con-gli-assistenti-ai/>

DDay.it (2024). Apple cambia ancora l'App Store in Europa: il nuovo sistema di commissioni è un puzzle.
<https://www.dday.it/redazione/53504/apple-cambia-ancora-lapp-store-in-europa-il-nuovo-sistema-di-commissioni-e-un-puzzle>

HCLTech (2024). The Race to Define Agentic AI Standards: Open Source vs Big Tech.
<https://www.hcltech.com/trends-and-insights/race-define-agentic-ai-standards-open-source-vs-big-tech#:~:text=A%20global%20race%20is%20underway,of%20this%20new%20AI%20era.&text=Agentic%20AI%20isn't%20just,and%20adapt%20based%20on%20feedback.>

Microsoft Research (s.d.). Project AutoGen.
<https://www.microsoft.com/en-us/research/project/autogen/>

Colangelo, G. (2024). Competitive Risks of Generative AI and the Digital Markets Act. SSRN.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5242028

LegalTech Italia (2023). DMA – Digital Markets Act: la nuova normativa dell’Unione Europea destinata a ridisegnare il panorama digitale europeo (e non solo).

<https://www.legaltechitalia.eu/dma-digital-markets-act-la-nuova-normativa-dellunione-europea-destinata-a-ridisegnare-il-panorama-digitale-europeo-e-non-solo/>

OECD (2024). Artificial Intelligence, Data and Competition. OECD Publishing.

https://www.oecd.org/en/publications/artificial-intelligence-data-and-competition_e7e88884-en.html

Ezrachi, A., & Stucke, M. E. (2012). The Curious Case of Competition and Quality. *The Antitrust Bulletin*, 57(1), 191–231. JSTOR.

<https://www.jstor.org/stable/41623144>

Kramer et al., AI Agents and Ecosystems Contestability, cit.