

# Le strategie di Business nel settore dell'Intelligenza Artificiale, Il caso Amazon

Prof. Luigi Nasta

---

RELATORE

Matr. 266571

---

CANDIDATO



## INDICE

<b>CAPITOLO 1: INTRODUZIONE</b> .....	1
<b>CAPITOLO 2: LE STRATEGIE DI BUSINESS NELLA TEORIA</b> .....	3
2.1 Leadership di costo .....	5
2.2 Differenziazione .....	7
2.3 Focalizzazione.....	10
2.4 Sfide e rischi legati alle strategie.....	11
<b>CAPITOLO 3: STRATEGIE BUSINESS APPLICATE AL SETTORE DELL'INTELLIGENZA ARTIFICIALE</b> .....	13
3.1 Cos'è l'IA? .....	13
3.2 Analisi del settore dell'IA .....	18
3.3 Leadership di costo nel settore dell'IA.....	37
3.4 differenziazione nel settore dell'IA.....	41
3.5 Focus nel settore dell'IA .....	48
<b>CAPITOLO 4: IL CASO AMAZON</b> .....	54
4.1 Panoramica azienda Amazon .....	54
4.2 In che modo l'IA ha contribuito ad aumentare l'efficienza di processi ed operazioni in Amazon e suoi clienti .....	63
4.3 Innovazione tecnologica e strategie di business.....	68
<b>CAPITOLO 5: CONCLUSIONE</b> .....	71
<b>BIBLIOGRAFIA</b> .....	73
<b>SITOGRAFIA</b> .....	74

## **CAPITOLO 1: INTRODUZIONE**

L'intelligenza artificiale (IA) sta velocemente trasformando il mondo in cui viviamo. Auto a guida autonoma e sistemi di raccomandazione personalizzati testimoniano che l'IA sta già avendo un impatto significativo su diversi settori, e il suo potenziale di innovazione è solamente all'inizio. In questo contesto, è di massima importanza comprendere come le aziende possono usufruire dell'IA per ottenere un vantaggio competitivo e posizionarsi in modo strategico in un mercato altamente mutevole.

L'obiettivo di questa tesi è quello di osservare in che modo le strategie di business sono applicabili al settore dell'IA, concentrandosi in particolar modo su: leadership di costo, differenziazione e focalizzazione. Nel lavoro si cercherà di esaminare come le suddette strategie possono essere applicate dalle società per ottenere vantaggi in termini di costi e valore aggiunto.

La tesi è articolata in cinque capitoli. Il primo capitolo introduce l'argomento e definisce quali sono gli obiettivi e la struttura della ricerca. Il secondo capitolo fornisce una spiegazione delle strategie di business dal punto di vista teorico, esaminando i concetti di leadership di costo, differenziazione e focalizzazione. Il terzo capitolo illustra in che modo queste strategie siano applicabili al settore dell'IA, effettuando una valutazione delle principali dinamiche competitive e opportunità di innovazione. All'interno del quarto capitolo viene preso in esame il caso di studio di Amazon, nel quale si osserverà come l'azienda ha utilizzato l'IA per migliorare le sue operazioni ed ottenere un vantaggio competitivo. Da ultimo, il quinto capitolo conclude la tesi, riassumendo i risultati principali ed offrendo spunti per future ricerche.

Il settore dell'IA sarà analizzato attraverso l'uso di strumenti strategici come l'analisi PESTEL e il modello delle 5 forze di Porter. Questi modelli consentiranno l'identificazione delle principali forze competitive che caratterizzano il settore, dei fattori esterni che ne provocano l'evoluzione e delle opportunità e minacce che le società devono affrontare.

Il lavoro sarà incentrato in particolare sull'IA cosiddetta "debole" o "ristretta", cioè quella concepita per svolgere compiti come: il riconoscimento vocale e la traduzione automatica.

La motivazione risiede nel fatto che l'IA suddetta rappresenta la forma più diffusa ed utilizzata attualmente, con un impatto significativo su diversi settori.

La trattazione ha come scopo precipuo quello di fornire una visione d'insieme di come le strategie di business siano applicabili al settore in questione. Attraverso l'analisi teorica e l'esame di casi concreti, ci si propone di fornire spunti utili alle aziende che intendono sfruttare l'IA in modo da acquisire un vantaggio competitivo e sostenibile. Questa ricerca potrà contribuire ad una migliore comprensione delle dinamiche del settore e delle opportunità di innovazione offerte dall'IA.

## **CAPITOLO 2: LE STRATEGIE DI BUSINESS NELLA TEORIA**

La finalità delle strategie di business è quella di consentire una crescita organica dell'impresa, ossia crescita di fatturato e quota di mercato, mantenendo un margine operativo stabile o in crescita. Esse si contrappongono agli approcci che prevedono una crescita di fatturato tramite semplice aggiunta di nuove linee di prodotto, l'individuazione di nuove opportunità di vendita in microaree del mercato (c.d. approccio granulare) e la vendita di servizi complementari.

L'approccio strategico alle attività d'impresa ha cominciato a svilupparsi negli Stati Uniti a cavallo tra la fine dell'800 e l'inizio del '900. A fronte dello sviluppo di mercati sempre più ampi ed aziende sempre più grandi e complesse, emerse infatti la necessità di definire approcci per poter influenzare attivamente tali mercati. Un ulteriore impulso allo sviluppo del pensiero strategico applicato alle attività aziendali fu dato dallo scoppio della Seconda guerra mondiale, che richiese "l'allocazione di risorse scarse attraverso l'intero sistema economico" statunitense (Ghemawat, 2002, pag. 39).

Nel secondo dopoguerra, una molteplicità di istituzioni economiche ed accademiche, in particolare Boston Consulting Group, McKinsey e la Harvard Business School, intraprese l'elaborazione di una serie di teorie e strumenti pratici alla base della moderna strategia competitiva aziendale, che sostiene la necessità da parte delle aziende di riconoscere il proprio ambiente competitivo al fine di poterlo influenzare. Tra le principali teorie sviluppate si ricordano:

- La SWOT analysis, proposta negli anni '60 da Kenneth Andrews della Harvard Business School al fine di identificare i punti di forza e debolezza di un'azienda rispetto alle opportunità e alle minacce del mercato di appartenenza;
- Il concetto delle curve di apprendimento, su cui si basò la competenza distintiva di BCG nei primi anni della sua attività, secondo cui le aziende beneficiano di costi inferiori a seguito della permanenza in un dato settore grazie ad una maggiore esperienza, tramite economie di scala e innovazione;
- La matrice BCG per descrivere i business di un'azienda in base al loro tasso di crescita e alla quota di mercato da essa detenuta, distinguendo tra Cash Cows, Dogs, Stars e Question marks.

	<b>Quota mkt Alta</b>	<b>Quota mkt Bassa</b>
<b>Crescita Alta</b>	Star	Question Mark
<b>Crescita Bassa</b>	Cash Cow	Dog

Tale approccio fu il primo di una serie di strategie per la gestione del portafoglio delle attività d'impresa, tra cui si può menzionare la matrice General Electric-McKinsey per combinare l'attrattività di un settore e la posizione competitiva dell'azienda;

- Gli studi di Industrial Organization di Joe Bain, volti a comprendere i fattori che determinano la differente profittabilità di diversi settori industriali;
- L'analisi delle 5 forze di Michael Porter, finalizzata a comprendere come l'ambiente competitivo in un dato mercato sia influenzato da:
  - Competitor del settore;
  - Fornitori (a monte);
  - Acquirenti (a valle);
  - Nuovi entranti;
  - Prodotti sostituiti;
- L'analisi dei player della catena del valore elaborata da Brandenburger e Nalebuff, che individuarono anch'essi il ruolo fondamentale di concorrenti, clienti, fornitori e produttori di beni complementari.

A partire dagli anni '70, in particolare, furono analizzati più approfonditamente i driver di costo aziendali, al fine di ridurre le spese tramite l'ottenimento di economie di scala. Il focus si ampliò poi dai costi per considerare il punto di vista e la prospettiva dei consumatori, in quanto apparve chiaro che per molte categorie di beni non era sufficiente offrire il prodotto più economico: la scelta di differenziare la propria offerta poteva rivelarsi vincente per un'azienda, laddove a fronte di maggiori costi di differenziazione fosse stata in grado di applicare prezzi superiori ai concorrenti. Il vantaggio competitivo dell'impresa può essere dunque conseguito attraverso due capacità di base:

- la capacità di commercializzare un prodotto/servizio analogo a quello dei concorrenti ad un costo più basso;

- la capacità di offrire un prodotto/servizio con caratteristiche che lo distinguono da quelle dei competitors, per le quali il cliente è disposto a pagare un prezzo più alto.

Secondo Michael Porter, queste alternative sono potenzialmente applicabili ad ogni singola attività di un'impresa, cosicché è possibile ottenere vantaggi di costo o differenziazione nella gestione delle risorse umane, dell'innovazione, nel marketing, etc.

In virtù di quanto sopra, sono possibili tre alternative strategiche (Porter, 1985):

- leadership di costo
- differenziazione
- focus

## **2.1 Leadership di costo**

Per leadership di costo si intende la capacità dell'azienda di aumentare la propria efficienza, in modo da produrre a costi inferiori rispetto ai rivali. L'azienda può quindi ridurre il prezzo al di sotto di quello dei concorrenti pur mantenendo un margine positivo. In questo modo riesce ad aumentare la propria quota di mercato e a conseguire economie di scala accumulando, altresì, esperienza. Si crea, quindi, un circuito virtuoso tra aumento dell'efficienza, incremento del margine operativo e aumento dei volumi venduti. Il processo avviene tuttavia con alcune frizioni: per esempio, il consumatore non si accorge subito della diminuzione di prezzo, oppure l'azienda può non avere la capacità produttiva per soddisfare la maggiore domanda potenziale (Caroli, 2021).

Per sfruttare il vantaggio di costo, la riduzione di prezzo deve essere, quindi, percepibile e l'aumento delle vendite dell'azienda non deve essere tale da causare diseconomie di scala. Anche qualora venga mantenuto un prezzo di livello pari a quello dei competitors ma a costi minori, si otterranno ovviamente margini superiori.

Il processo per ottenere vantaggio di costo si articola in tre fasi:

- scomposizione dei costi per le singole attività della catena del valore;
- confronto dei costi con quelli dei concorrenti;
- individuazione e studio delle determinanti del livello dei costi (driver).

Si procede, quindi, a definire la strategia per conseguire la posizione di vantaggio rispetto ai competitors.

Per quanto riguarda le determinanti dei costi, esse si dividono in due categorie: fattori relativi all'attività specifica e fattori relativi ai legami tra l'attività specifica e le altre attività delle catene del valore dei vari business dell'impresa (Caroli, 2021).

Nella prima categoria rientrano:

- **economie di scala:** esse determinano la riduzione del costo medio all'aumentare della quantità prodotta;
- **economie di estensione:** consistono nella riduzione dei costi medi dovuta alla maggiore estensione dell'attività dell'impresa che consente un maggiore assorbimento delle spese generali trasversali;
- **economie di apprendimento:** all'aumentare della produzione cumulata (nel tempo) si riducono i costi medi;
- **grado di utilizzazione della capacità produttiva:** può essere considerato come una delle determinanti delle economie di scala, che conseguono anche alla più completa utilizzazione dei fattori di produzione;
- **innovazione dei processi produttivi:** può consentire la produzione di beni e servizi pressoché simili a costi inferiori;
- **innovazione di prodotto e materie prime:** l'innovazione del prodotto dal punto di vista dell'assemblaggio e l'utilizzo di materiali meno costosi ha una rilevante influenza sui costi;
- **localizzazione geografica delle attività:** la vicinanza delle attività produttive alle fonti di approvvigionamento ed ai mercati di sbocco ha anch'essa un notevole impatto sui costi;
- **gestione verticale della filiera:** un importante differenziale di costo è inoltre costituito dalla posizione negoziale rispetto a fornitori e distributori;
- **efficienza generale:** l'efficienza complessiva delle attività della catena del valore dell'impresa può costituire un ulteriore fattore determinante il vantaggio di costo rispetto ai concorrenti.

Per quanto riguarda la seconda categoria, essa è rappresentata dalle connessioni tra le diverse attività della catena del valore. I costi sono influenzati dall'efficienza con cui materiali e semilavorati sono movimentati tra le varie fasi del processo operativo e dall'efficacia con cui le attività sono connesse.

Per acquisire il vantaggio di costo l'azienda può agire sulle singole determinanti di costo o riconfigurare l'intera catena del valore modificando la propria organizzazione e il modo in cui svolge le sue attività. Il suddetto cambiamento organizzativo può essere attuato in quattro differenti modalità:

- **outsourcing**: esternalizzazione di alcune attività della catena di valore, generalmente quelle in cui l'impresa ha uno svantaggio di costo;
- **reingegnerizzazione dei processi produttivi**: riorganizzazione delle attività produttive in modo da evitare o ridurre alcune voci di costo;
- **razionalizzazione dell'insieme delle unità produttive**: concentrazione delle attività in un numero piuttosto limitato di stabilimenti posizionati strategicamente e tra loro integrati;
- **modificazione della posizione nella filiera produttiva**: integrazione verticale a monte o a valle. Si tratta dell'opzione opposta all'outsourcing.

## 2.2 Differenziazione

Il vantaggio della differenziazione consiste in una maggiore disponibilità del consumatore a pagare un prezzo più elevato oppure, praticando un limitato aumento di prezzo e migliorando la qualità del prodotto, in un aumento dei volumi venduti (essendo maggiore il valore percepito) (Caroli, 2021).

Per acquistare un vantaggio competitivo tramite differenziazione, il prodotto di un'azienda deve presentare le caratteristiche di:

- **unicità**: possesso di caratteristiche tali da distinguerlo in modo netto da quelli dei concorrenti;
- **valore degli elementi di unicità**: essi devono rivestire un valore significativo per i consumatori (minori costi per svolgere un'attività o migliori prestazioni ottenute da una determinata attività);

- **percezione dell'unicità e del valore da parte dei consumatori:** i clienti devono essere in grado di percepire l'incremento di valore del prodotto;
- **sostenibilità economica:** il cliente deve essere disposto a pagare un prezzo più alto di quello praticato dai concorrenti. Tale prezzo deve essere superiore ai maggiori costi sostenuti dall'impresa per la differenziazione dell'offerta.

Il miglioramento del rapporto valore offerto/prezzo, generato dalla differenziazione aumenterà la soddisfazione ed il grado di fidelizzazione del cliente.

La differenziazione si può realizzare intervenendo sulle componenti materiali, immateriali o relazionali dell'offerta.

Per quanto riguarda gli aspetti materiali del prodotto/servizio, è possibile agire in relazione ai seguenti elementi:

- contenuto tecnologico all'avanguardia;
- performance (ad esempio la velocità del processore di un pc);
- affidabilità;
- caratteristiche e qualità dei materiali utilizzati;
- integrabilità con altri prodotti;
- ampiezza della gamma di cui il prodotto è parte;
- grado di innovatività;
- design.

Eventuali componenti aggiuntive dell'offerta possono ampliare i contenuti offerti dal prodotto. Tra esse rientrano i cosiddetti fattori di segnalazione, che consentono al consumatore una più facile valutazione del valore dell'offerta. Ne è un esempio la garanzia concessa per un lungo termine e molto ampia, la quale costituisce un indice di affidabilità e durata del prodotto e rafforza la percezione da parte del consumatore dell'unicità dello stesso.

Ulteriori fattori di segnalazione sono costituiti, per esempio, dagli investimenti in ricerca e sviluppo, dall'intensità della comunicazione e dal prezzo. Inoltre, è opportuno inviare segnali al cliente non solo prima o durante l'acquisto ma anche successivamente ad esso in modo che questi sia ulteriormente convinto della bontà della scelta e sia disposto a riconfermarla in futuro.

Dal punto di vista delle componenti immateriali, cioè intangibili del valore di un bene, si rileva che esse influenzano lo schema cognitivo del consumatore, cioè il modo in cui questi si pone nei confronti del prodotto.

Tra di esse riveste importanza fondamentale in primo luogo la marca, che rende il prodotto riconoscibile e consente di svilupparne la reputazione. Quest'ultima è un fattore intangibile di differenziazione molto importante in quanto contribuisce a caratterizzare l'identità del prodotto percepita dal consumatore.

Vi sono, poi, ulteriori fattori intangibili di differenziazione costituiti dai valori che il prodotto o l'impresa si prefiggono di diffondere, valori nei quali il cliente si identifica e che soddisfano la sua esigenza di aderire e diffondere un determinato sistema di idee.

La differenziazione può, infine, basarsi sulla relazione tra l'impresa ed il suo cliente, con particolare riferimento ai seguenti aspetti:

- **facilità di acquisto del prodotto:** la facilità nell'acquisto e nel contatto con il venditore può rappresentare un elemento decisivo nella scelta del consumatore;
- **ambiente in cui il cliente viene in contatto con il fornitore ed il prodotto:** il contesto in cui avviene l'acquisto può aggiungere valore al prodotto, influenzando l'immagine percepita, il posizionamento strategico del prodotto e i costi che l'acquirente sostiene per l'acquisto;
- **condizioni che facilitano l'utilizzazione del prodotto:** esse dipendono innanzitutto dalle caratteristiche intrinseche del prodotto ma possono anche essere migliorate mediante servizi di assistenza resi dal produttore che rivestono una particolare importanza nel caso dei beni industriali e dei prodotti tecnologicamente complessi. Essi migliorano le condizioni di utilizzo del prodotto, aumentano l'utilità tratta dal consumatore, diminuiscono i costi di utilizzazione e aumentano i benefici intangibili del prodotto.

In ogni caso, gli interventi posti in atto per la differenziazione di un prodotto devono essere tali da garantire l'integrità dello stesso.

Si distinguono un'integrità interna ed un'integrità esterna. La prima consiste nella congruenza tra la struttura fisica del prodotto e le funzioni che esso deve svolgere, la

seconda si sostanzia nella coerenza tra componenti tangibili ed intangibili del prodotto e le aspettative ed esigenze del cliente target (Clark e Fujimoto, 1991).

### **2.3 Focalizzazione**

La focalizzazione consiste nella ricerca di vantaggio di costo o differenziazione in un'area molto ristretta del mercato (nicchia). Data la limitata estensione del mercato a cui si rivolge, si tratta di una strategia più adatta per piccole imprese che hanno a disposizione risorse limitate (Caroli, 2021).

I vantaggi di tale strategia consistono:

- nella possibilità di beneficiare di una maggiore forza competitiva a parità di risorse;
- nella maggiore specializzazione delle risorse e delle conoscenze con conseguente maggiore possibilità di conseguire una posizione di vantaggio nella propria area di business;
- nella minore pressione competitiva esercitata dalle grandi imprese, che hanno minore attenzione verso le aree di business di piccola dimensione.

A fronte dei vantaggi sopra descritti, la focalizzazione presenta alcuni rischi.

L'impresa focalizzata è legata all'andamento della sua nicchia e ne segue il ciclo di vita. Pertanto, quando la nicchia si avvia verso il declino, l'impresa perde le opportunità di sviluppo e subisce la contrazione dei volumi di attività.

Inoltre, l'area di mercato in cui l'azienda si focalizza potrebbe essere troppo limitata in termini di valore potenziale o essere interessata da una competizione troppo intensa e quindi non essere sostenibile dal punto di vista economico.

L'azienda che sceglie di focalizzarsi potrebbe non valutare correttamente le potenzialità economiche dell'area prescelta, rischiando di indirizzare la propria attività in una direzione non redditizia.

Infine, l'area di nicchia potrebbe comunque attrarre l'attenzione di qualche grande impresa, quando la domanda nei mercati più rilevanti è in fase di contrazione.

## 2.4 Sfide e rischi legati alle strategie

Le varie strategie aziendali, seppure correttamente elaborate ed applicate, comportano necessariamente dei rischi che ne possono compromettere il successo nel lungo periodo se non vengono fronteggiati con la dovuta preparazione e consapevolezza. La leadership di costo, la differenziazione e la focalizzazione presentano specifiche sfide che derivano dalla natura stessa di queste strategie e dall'ambiente competitivo in cui operano.

La leadership di costo ha lo scopo di ottenere un vantaggio competitivo tramite costi inferiori rispetto a quelli delle aziende concorrenti ma comporta il concreto rischio costituito dalla "guerra dei prezzi". Quando più aziende adottano questa strategia, il risultato è spesso una competizione feroce sul prezzo, che riduce i margini di profitto per tutti i giocatori coinvolti. Questo fenomeno non solo può erodere i ricavi, ma può anche spingere alcune aziende a scendere sotto la soglia di sostenibilità economica pur di rimanere competitive. Inoltre, l'attenzione quasi esclusiva alla riduzione dei costi può portare a compromessi sulla qualità del prodotto o servizio, rischiando di danneggiare la reputazione aziendale e alienare i clienti che cercano un valore aggiunto. La dipendenza da economie di scala e processi standardizzati rende questa strategia vulnerabile anche a cambiamenti improvvisi nei costi delle materie prime o nelle condizioni di mercato, che possono rapidamente annullare i benefici di costo precedentemente ottenuti (Il Fatto Quotidiano, 2023) (FasterCapital, 2024).

La strategia di differenziazione si basa sulla creazione di un valore che deve essere percepito dai clienti come unico. Pertanto, affronta una sfida cruciale nel mantenimento di tale percezione. In mercati fortemente mutevoli e competitivi vi è il rischio che gli elementi di differenziazione del prodotto introdotti da un'azienda vengano velocemente imitati dai suoi competitor. Ciò riduce il vantaggio competitivo di chi ha messo in pratica tale strategia. Inoltre, il costo dell'innovazione e della comunicazione per mantenere una posizione differenziata può essere elevato, e non sempre i clienti sono disposti a pagare un prezzo premium per un valore che percepiscono come marginale. La stessa fedeltà del cliente, sulla quale solitamente si basa questa strategia, può essere compromessa da offerte concorrenti nelle quali le caratteristiche innovative si accompagnano ad un prezzo minore. Inoltre, le aziende devono essere sempre aggiornate sulle reali esigenze del

mercato in modo da non commettere l'errore di considerare unico ciò che non è più percepito come tale dai consumatori (FasterCapital, 2024).

La focalizzazione, che si concentra su un segmento di mercato ben definito, presenta criticità specifiche legate ad una eccessiva dipendenza dalla nicchia scelta. Questo approccio, nonostante possa essere efficace nell'assicurare una forte posizione competitiva all'interno del segmento target, può al contempo rendere l'azienda particolarmente esposta ai cambiamenti nelle preferenze dei consumatori, alle condizioni economiche o all'evoluzione del settore. Ad esempio, se un'azienda è specializzata in un prodotto o servizio destinato a un segmento di mercato che improvvisamente subisce un declino o diventa scarsamente rilevante, le possibilità di adattamento possono essere limitate. La focalizzazione è anche soggetta alla minaccia di grandi concorrenti che, percependo l'attrattiva della nicchia, possono investire maggiori risorse per conquistare la stessa fetta di mercato, superando l'azienda focalizzata grazie alla loro scala e capacità (FasterCapital, 2024).

Le criticità sopra esaminate mettono in evidenza come nessuna strategia sia esente da rischi. La leadership di costo richiede un'attenta gestione dell'efficienza operativa per evitare di cadere nella trappola della guerra sui prezzi. Parallelamente, la differenziazione necessita di una costante innovazione e allineamento con le aspettative dei clienti. In ultimo, la focalizzazione richiede un'elevata flessibilità per potersi adattare ai cambiamenti di una nicchia di mercato spesso imprevedibile.

## **CAPITOLO 3: STRATEGIE BUSINESS APPLICATE AL SETTORE DELL'INTELLIGENZA ARTIFICIALE**

### **3.1 Cos'è l'IA?**

La nostra specie non sarebbe progredita molto senza la meccanizzazione. Dalla ruota che ha rivoluzionato l'agricoltura, alla vite che ha tenuto insieme progetti di costruzione sempre più complessi, fino alle linee di assemblaggio robotizzate odierne, le macchine hanno reso possibile la vita così come la conosciamo. Eppure, nonostante la loro apparentemente infinita utilità, gli umani hanno da tempo paura delle macchine, più specificamente della possibilità che un giorno esse possano acquisire intelligenza umana e agire autonomamente. Tuttavia, tendiamo a vedere la possibilità di macchine senzienti anche con attrazione. Questa curiosità ha aiutato a trasformare la fantascienza in scienza reale.

L'IA è la capacità di una macchina di svolgere funzioni cognitive che associamo alla mente umana, come percepire, ragionare, apprendere, interagire con l'ambiente, risolvere problemi e persino esercitare creatività. Probabilmente tutti hanno già interagito almeno una volta con l'IA senza rendersene conto: assistenti vocali come Siri e Alexa si basano sulla tecnologia IA, così come alcuni chatbot di assistenza clienti che appaiono per aiutarti a navigare nei siti web. Nel 2024, la maggior parte dei ricercatori e dei professionisti dell'IA, così come la maggior parte delle informazioni legate all'IA, si sono concentrati sui progressi dell'IA generativa (Gen AI), una tecnologia che può creare testi, immagini, video e altri contenuti originali. Per comprendere appieno l'IA generativa, è importante innanzitutto capire le tecnologie su cui si basano gli strumenti che essa utilizza: il machine learning (ML) e il deep learning.

Un modo semplice per pensare all'IA è considerarla come una serie di concetti derivati, emersi nell'arco di oltre 70 anni:

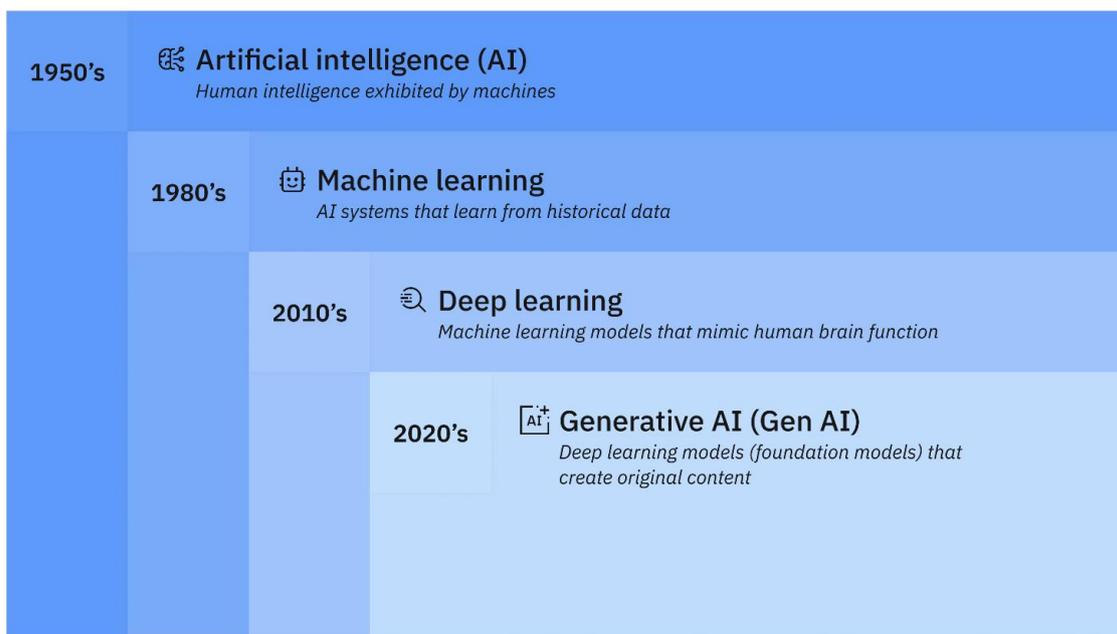


Figura 1: i processi che rappresentano l'evoluzione dell'intelligenza artificiale (Stryker, Kavlakoglu, 2024).

Come si può notare nella figura 1, il machine learning è posizionato immediatamente sotto l'IA; esso implica la creazione di modelli addestrando un algoritmo ad effettuare previsioni o prendere decisioni basate sui dati. Include una vasta gamma di tecniche che permettono ai computer di apprendere dai dati e fare inferenze basate su questi, senza essere esplicitamente programmati per compiti specifici. Esistono molti tipi di tecniche o algoritmi di machine learning, tra cui: linear regression, logistic regression, decision trees, random forest, support vector machines (SVMs), k-nearest neighbor (KNN), clustering, e altri ancora. Ciascuno di questi approcci è adatto a tipi diversi di problemi e dati (Chui, Hazan, Roberts, Singla, Smaje, Sukharevsky, Yee, Zemmell, 2023).

Uno dei tipi di algoritmo di machine learning più popolari è chiamato rete neurale (o rete neurale artificiale). Le reti neurali sono modellate sulla struttura e funzione del cervello umano. Una rete neurale è composta da strati di nodi interconnessi (analoghi ai neuroni) che lavorano insieme per elaborare e analizzare dati complessi. Le reti neurali sono particolarmente adatte ai compiti che coinvolgono l'identificazione di schemi complessi e relazioni in grandi quantità di dati. La forma più semplice di apprendimento automatico è chiamata supervised learning (apprendimento supervisionato), che coinvolge l'uso di set di dati etichettati per addestrare algoritmi a classificare i dati o a prevedere i risultati in

modo accurato. Nell'apprendimento supervisionato, gli esseri umani associano ogni esempio di addestramento con un'etichetta di output. L'obiettivo è che il modello impari la corrispondenza tra input e output nei dati di addestramento, così da poter prevedere le etichette di nuovi dati non visti prima (Stryker et al., 2024).

Una volta appreso il concetto di machine learning si può approfondire il significato di deep learning. Quest'ultimo è una sottocategoria del machine learning che utilizza reti neurali multilivello, chiamate deep neural networks (reti neurali profonde), che simulano in modo più accurato la complessa capacità decisionale del cervello umano. Le reti neurali profonde includono un input layer (livello di input), almeno tre (ma di solito centinaia) di hidden layers (livelli nascosti), e un output layer (livello di output), a differenza delle reti neurali utilizzate nei modelli di machine learning classici, che generalmente hanno solo uno o due livelli nascosti (Chui et al., 2023).

## Deep Neural Network

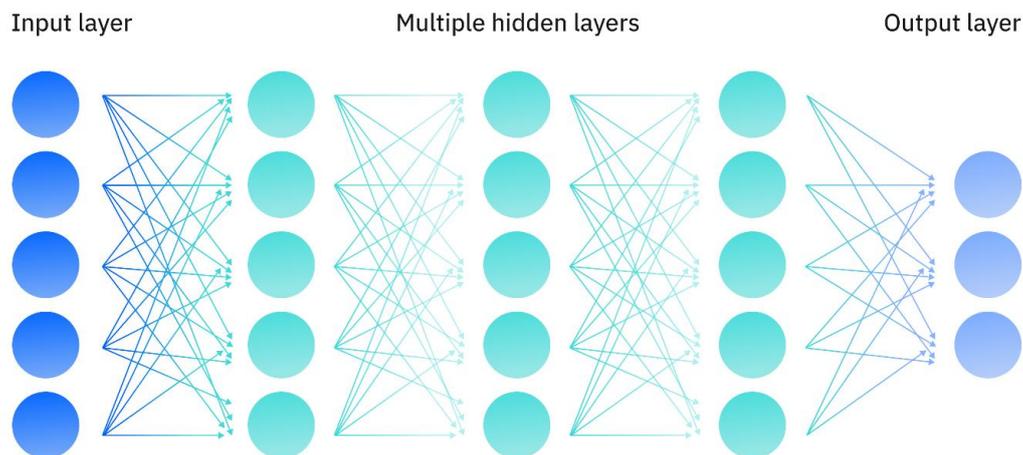


Figura 2: rappresentazione della rete neurale profonda che costituisce il deep learning (Stryker et al., 2024).

Questi strati multipli permettono l'apprendimento non supervisionato: possono automatizzare l'estrazione di caratteristiche da grandi set di dati non etichettati e non strutturati, e fare previsioni autonome su cosa rappresentano i dati. Poiché il deep learning non richiede l'intervento umano, consente di effettuare il machine learning su vasta scala. È particolarmente adatto al natural language processing (NLP), ovvero la capacità del

computer o dispositivo elettronico di riconoscere, comprendere e generare testo sotto forma di linguaggio umano, alla computer vision e ad altri compiti che coinvolgono l'identificazione rapida e accurata di schemi complessi e relazioni in grandi quantità di dati. Oggi, molte applicazioni di intelligenza artificiale (IA) nella nostra vita quotidiana sono alimentate da qualche forma di deep learning (Stryker et al., 2024) (Stryker, Holdsworth, 2024).

Il deep learning abilita anche:

- Semi-supervised learning (Apprendimento semi-supervisionato), che combina apprendimento supervisionato e non supervisionato utilizzando sia dati etichettati che non etichettati per addestrare modelli di IA per compiti di classificazione e regressione.
- Self-supervised learning (Apprendimento auto-supervisionato), che genera etichette implicite dai dati non strutturati, anziché affidarsi a set di dati etichettati per ottenere segnali di supervisione.
- Reinforcement learning (Apprendimento per rinforzo), che apprende tramite tentativi, errori e funzioni di ricompensa anziché estrarre informazioni da schemi nascosti.
- Transfer learning (apprendimento per trasferimento), in cui la conoscenza acquisita attraverso un compito o set di dati viene utilizzata per migliorare le prestazioni del modello su un altro compito correlato o su un diverso set di dati.

Si può ora analizzare il concetto di IA generativa, a volte chiamata "Gen AI". Si riferisce a modelli di deep learning che possono creare contenuti complessi e originali come: testi lunghi, immagini di alta qualità, video o audio realistici e altro ancora in risposta a un input o richiesta dell'utente. Ad alti livelli, i modelli generativi codificano una rappresentazione semplificata dei loro dati di addestramento e poi utilizzano questa rappresentazione per creare nuovi lavori che sono simili, ma non identici, ai dati originali.

I modelli generativi sono stati utilizzati per anni in statistica per analizzare dati numerici. Tuttavia, nell'ultimo decennio, si sono evoluti per analizzare e generare tipi di dati più complessi. Questa evoluzione ha coinciso con l'emergere di tre tipi di modelli di deep learning sofisticati:

- Variational Autoencoders o VAE, introdotti nel 2013, che hanno permesso ai modelli di generare più variazioni di contenuti in risposta a un input o istruzione.
- Modelli di diffusione, apparsi per la prima volta nel 2014, che aggiungono "rumore" alle immagini fino a renderle irriconoscibili, e poi rimuovono il rumore per generare immagini originali in risposta agli input.
- Transformers (chiamati anche modelli transformer), che sono addestrati su dati sequenziati per generare sequenze estese di contenuti (come parole in frasi, forme in un'immagine, fotogrammi di un video o comandi nel codice software). I transformer sono al centro della maggior parte degli strumenti di IA generativa di grande successo odierni, tra cui ChatGPT e GPT-4, Copilot, BERT, Bard e Midjourney.

Tramite chatbots e assistenti virtuali gestiti mediante IA le aziende possono anche gestire le richieste dei clienti, i ticket di supporto ed altro. Questi utilizzano NLP (natural language processing) e IA generativa per rispondere a domande frequenti dei clienti, come lo stato dell'ordine, le caratteristiche dei prodotti e le policy sui resi. I chatbot e gli assistenti virtuali offrono supporto 24/7 e risposte più veloci alle domande della sezione FAQ, riducendo il carico di lavoro per gli operatori umani e migliorando l'esperienza complessiva del cliente.

Le piattaforme di reclutamento del personale gestite tramite l'IA possono rendere più veloce il processo di assunzione esaminando i curriculum, abbinando i candidati ai profili richiesti e persino svolgendo colloqui preliminari attraverso l'analisi dei video. Questi e altri strumenti consentono di ridurre radicalmente la grande mole di lavoro amministrativo necessaria per la gestione di un elevato numero di candidati. Possono inoltre accorciare i tempi di risposta e il tempo di assunzione, migliorando l'esperienza dei candidati, sia che ottengano il lavoro o meno.

L'intelligenza artificiale è classificabile, a seconda del suo grado di complessità, in due categorie: "Weak AI" e "Strong AI".

L'Intelligenza Artificiale Ristretta (Artificial Narrow Intelligence), anche conosciuta come Weak AI (Intelligenza Artificiale Debole), è l'unico tipo di IA esistente oggi. Qualsiasi altra forma di IA è ancora teorica. L'IA ristretta può essere addestrata per

svolgere un singolo compito o un compito ristretto, spesso in modo molto più rapido e preciso rispetto alla mente umana. Tuttavia, non è in grado di operare al di fuori del compito specificato. Si concentra su un sottoinsieme limitato di abilità cognitive e fa progressi solo in quel campo. Esempi di Weak AI includono Siri, Alexa di Amazon e IBM Watson®. Anche ChatGPT di OpenAI è considerato una forma di IA debole, poiché è limitato al singolo compito della conversazione basata su testo (IBM, 2024) (Chui et al., 2023).

Se i ricercatori riuscissero a sviluppare la Strong AI (Intelligenza Artificiale Forte), la macchina avrebbe un'intelligenza pari a quella umana; avrebbe una coscienza propria, con la capacità di risolvere problemi, apprendere e pianificare il futuro. L'obiettivo della Strong AI è creare macchine intelligenti indistinguibili dalla mente umana. Tuttavia, come un bambino, una macchina dotata di Strong AI dovrebbe apprendere attraverso input ed esperienze, progredendo e migliorando costantemente le sue abilità nel tempo. Sebbene i ricercatori di IA, sia nel mondo accademico che nel settore privato, siano impegnati nella creazione della Strong AI, oggi essa esiste solo come concetto teorico (IBM, 2024).

Dunque, l'Intelligenza Artificiale, capace di replicare funzioni cognitive umane come l'apprendimento, il ragionamento e la creatività, è ormai parte della vita quotidiana. Tecnologie come Siri e Alexa, basate su IA debole, ne sono esempi tangibili, mentre la ricerca più recente si concentra sull'IA generativa, capace di produrre contenuti originali. Come già detto, la Strong IA, che rappresenterebbe un'intelligenza comparabile a quella umana rimane, allo stato attuale, un concetto ancora teorico. In questa analisi, quindi, l'attenzione sarà rivolta esclusivamente alla Weak AI, esaminandone le applicazioni pratiche e i limiti, evitando le speculazioni teoriche riguardanti la Strong AI.

### **3.2 Analisi del settore dell'IA**

Dal lato delle dinamiche del mercato, l'industria dell'intelligenza artificiale è caratterizzata da una stringente concorrenza tra i grandi operatori globali quali Google, Amazon, Microsoft, nonché tra numerose start-up innovative. Grazie all'alta rilevanza strategica dell'IA per il futuro economico e tecnologico, il settore attrae considerevoli investimenti in ricerca e sviluppo sia da parte dei governi che delle aziende. Il modello di analisi PESTEL e quello delle 5 forze di Porter sono utili strumenti per interpretare le dinamiche favorevoli e ostili in questo settore in forte evoluzione.

Questi strumenti consentono di esplorare le forze competitive che modellano il mercato, di identificare i fattori esterni (politici, economici, sociali e tecnologici) che influenzano lo sviluppo del settore e di valutare i punti di forza e debolezza delle imprese coinvolte. Lo studio di questi aspetti consente di avere una visione globale delle opportunità e dei rischi che l'industria dell'IA deve fronteggiare, fornendo alle aziende gli strumenti per comprendere come posizionarsi strategicamente in questo scenario in rapida evoluzione. Sarà inoltre di grande importanza osservare come le strategie di business possano essere sfruttate dalle imprese in questo settore per ottenere un vantaggio competitivo duraturo.

La figura 3 illustra l'andamento degli investimenti aziendali globali nell'IA dal 2013 al 2023, inclusi fusioni e acquisizioni, partecipazioni di minoranza, investimenti privati e offerte pubbliche. Per due anni consecutivi, gli investimenti aziendali globali nell'IA hanno registrato un calo. Nel 2023, il totale degli investimenti è sceso a 189,2 miliardi di dollari, con una diminuzione di circa il 20% rispetto al 2022. Nonostante una lieve riduzione negli investimenti privati, il calo più significativo si è verificato nelle fusioni e acquisizioni, che sono diminuite del 31,2% rispetto all'anno precedente. Tuttavia, negli ultimi dieci anni, gli investimenti legati all'IA sono aumentati di tredici volte (Maslej, Fattorini, Perrault, Parli, Reuel, Brynjolfsson, Etchemendy, Ligett, Lyons, Manyika, Niebles, Shoham, Wald, Clark, 2024).

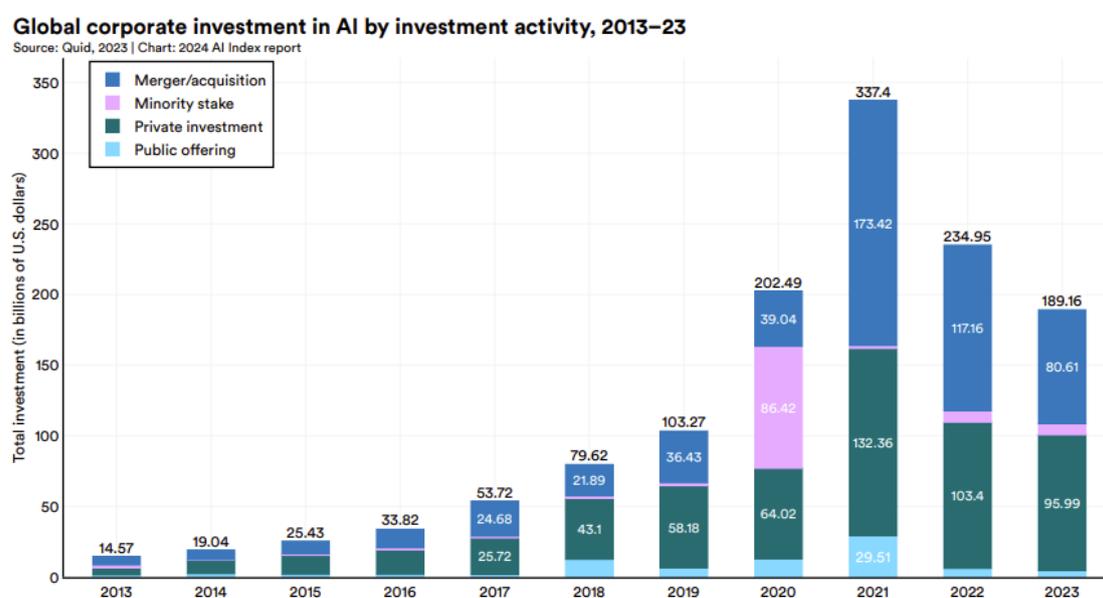


Figura 3: ammontare degli investimenti totale da parte delle aziende nell'IA dal 2013 al 2023 (Maslej et al., 2024).

La figura 4 mostra come gli investimenti privati globali nell'IA sono diminuiti per due anni consecutivi (2022-2023). Tuttavia, la diminuzione rispetto al 2022 è stata modesta (-7,2%) e inferiore al calo osservato tra il 2021 e il 2022. Nonostante i recenti cali, gli investimenti privati nell'IA a livello globale sono cresciuti significativamente nell'ultimo decennio.

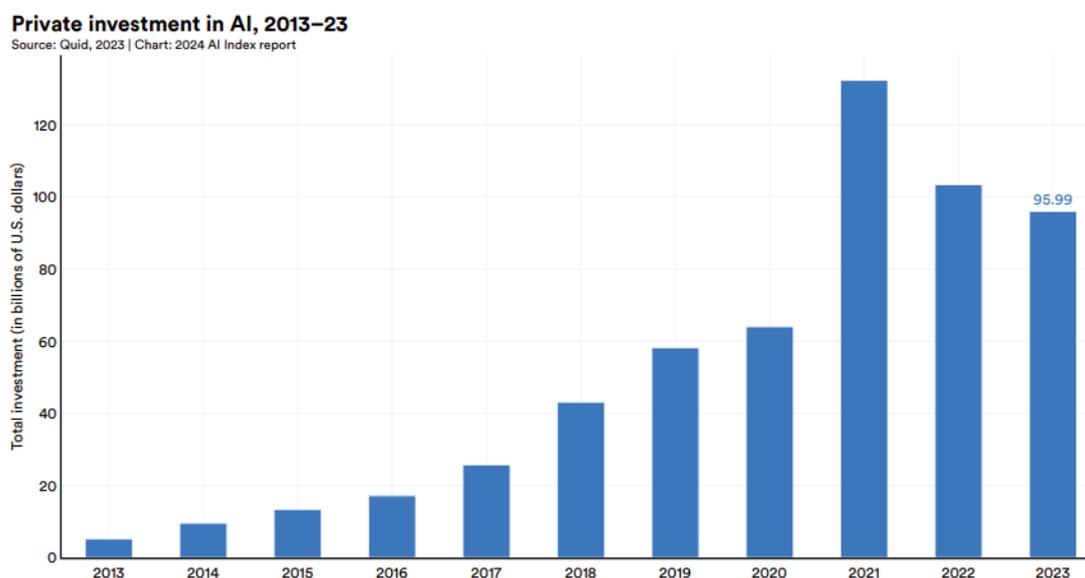


Figura 4: ammontare degli investimenti privati nell'IA dal 2013 al 2023 (Maslej et al., 2024).

Sebbene gli investimenti privati complessivi nell'IA siano diminuiti nel 2023, i finanziamenti per l'IA generativa sono aumentati drasticamente (Figura 5). Nel 2023, il settore ha attirato 25,2 miliardi di dollari, quasi nove volte l'investimento del 2022 e circa 30 volte la cifra del 2019. Inoltre, l'IA generativa ha rappresentato più di un quarto di tutti gli investimenti privati legati all'IA nel 2023.

### Private investment in generative AI, 2019–23

Source: Quid, 2023 | Chart: 2024 AI Index report

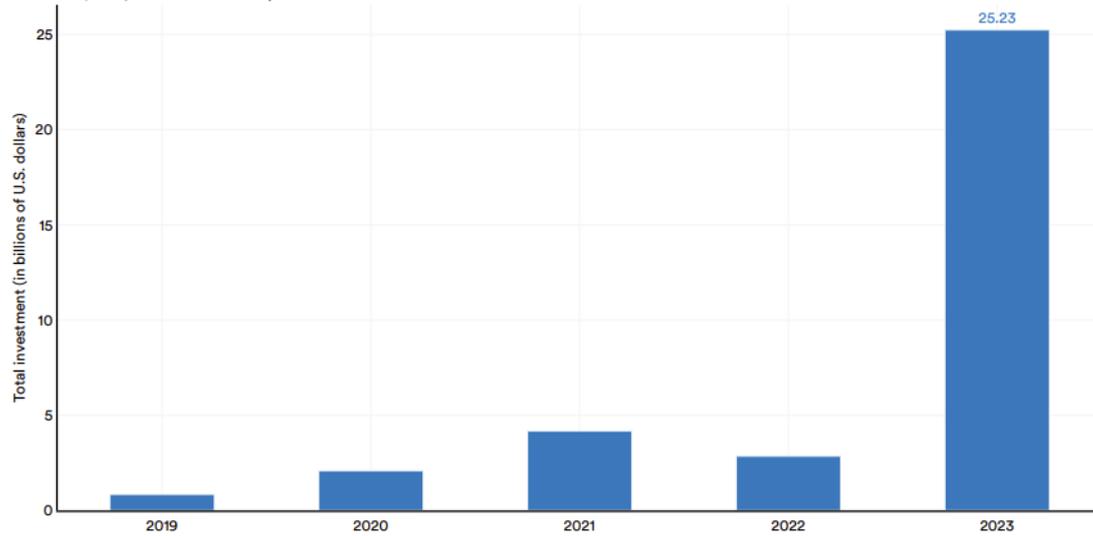


Figura 5: ammontare degli investimenti privati nell'IA generativa dal 2019 al 2023 (Maslej et al., 2024).

Interessante è il fatto che il numero di nuove aziende fondate nel settore dell'IA sia aumentato a 1.812, con un incremento del 40,6% rispetto all'anno precedente (Figura 6).

### Number of newly funded AI companies in the world, 2013–23

Source: Quid, 2023 | Chart: 2024 AI Index report

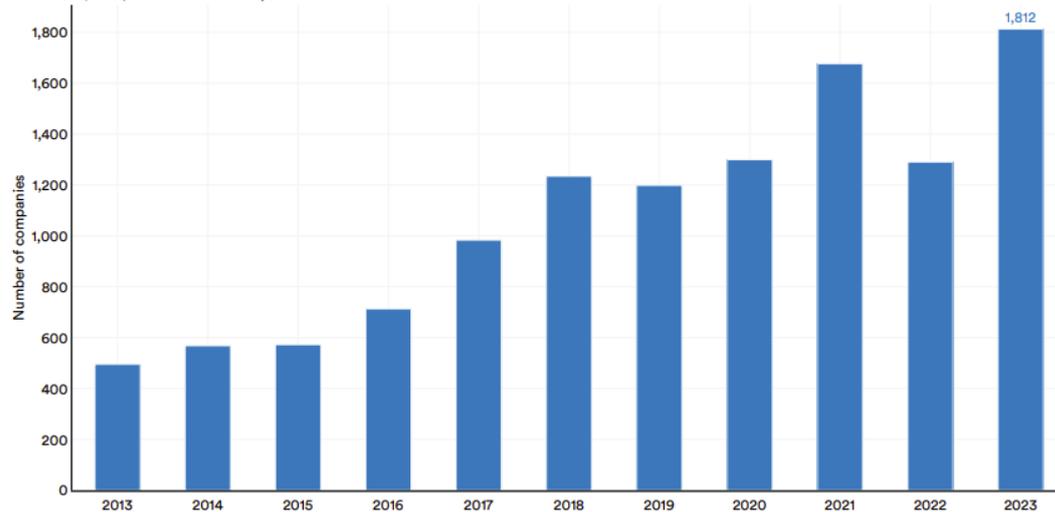


Figura 6: numero di nuove imprese di IA fondate nel mondo dal 2013 al 2023 (Maslej et al., 2024).

Gli Stati Uniti guidano la classifica degli investimenti privati per area geografica con 335,2 miliardi di dollari investiti, seguiti dalla Cina con 103,7 miliardi di dollari e dal Regno Unito con 22,3 miliardi di dollari (figura 7).

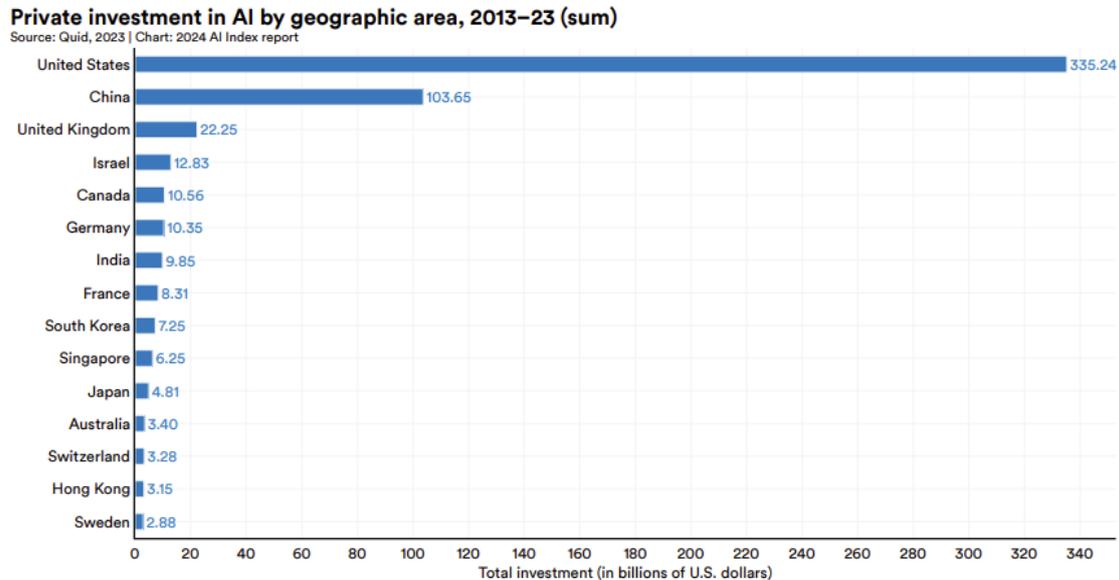


Figura 7: somma degli investimenti privati in IA per area geografica dal 2013 al 2023 (Maslej et al., 2024).

Il contesto ampio in cui l'impresa opera ne influenza in modo rilevante l'evoluzione e i risultati potenziali perché determina un insieme di vincoli e opportunità di cui essa non può non tenere conto nel delineare il suo comportamento strategico e in alcuni casi anche operativo. Non è raro che la crisi di un'azienda derivi proprio dal non aver saputo adeguarsi efficacemente ai cambiamenti significativi e repentini del contesto ampio. Le imprese devono considerare questi fattori come dati che salvo casi eccezionali non possono essere influenzati. Essi non sono quindi l'oggetto della strategia aziendale, ma determinanti dello scenario in cui questa deve essere pensata e attuata (Caroli, 2021).

Le numerose condizioni e gli attori del contesto ampio rilevante per le imprese possono essere raggruppate in sei ambiti memorizzabili con l'acronimo PESTEL: politico-istituzionale, economico, socio-culturale, tecnologico, ecologico, legale-amministrativo. La rilevanza di ciascuno di questi ambiti e delle specifiche condizioni o attori raggruppati al suo interno è ovviamente diversa in relazione al settore e allo specifico business in cui l'azienda opera (Caroli, 2021).

# PESTEL

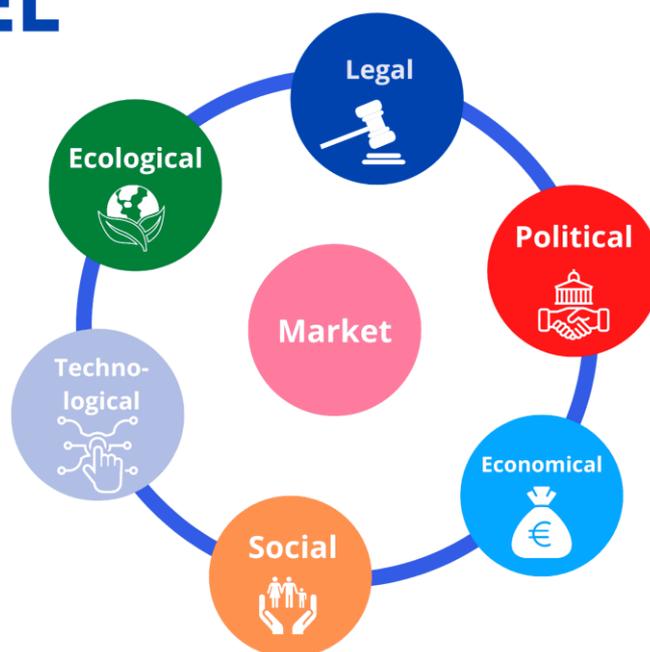


Figura 8: modello componenti del modello PESTEL (Waalaxy, 2024)

Applicando questo modello al settore dell'IA emerge un complesso quadro di attori ed elementi coinvolti a livello globale:

## 1. Fattori Politici

**Politiche di investimento negli USA:** Ogni dicembre, il National Science and Technology Council rilascia un rapporto sul budget per la ricerca e lo sviluppo dell'IA nel settore pubblico. Secondo il rapporto del 2023, per l'anno fiscale 2023, le agenzie governative statunitensi hanno speso in totale 1,8 miliardi di dollari in ricerca e sviluppo nell'IA. I fondi spesi per la R & D nell'IA hanno continuato a crescere dal 2018, triplicando il valore di quell'anno. Per l'anno fiscale 2024, è stato richiesto un budget superiore a 1,9 miliardi di dollari (Maslej et al., 2024) (Khan, Shoaib, Arledge, 2024).

**Politiche di investimento nell'UE:** nel maggio 2021 è stato approvato ufficialmente il programma "Europa Digitale". È stato annunciato un cospicuo finanziamento di 67.5 milioni di euro destinato a finanziare progetti volti ad accelerare e rafforzare l'adozione delle tecnologie di intelligenza artificiale in Europa. Ciò comprende il sostegno alle imprese e alle PMI nella loro preparazione al rispetto della legge sull'IA (News Article, 2024).

**Politiche di investimento in Cina:** In Cina sono di grande rilevanza i “Government Guidance Funds”, ovvero fondi di investimento privati controllati dal governo cinese. Essi vengono utilizzati per iniettare capitale nei settori strategici prioritari, come l'intelligenza artificiale (IA) e altre tecnologie emergenti. Entro il 2022, le autorità cinesi avevano istituito 2.107 Guidance Funds, con un obiettivo registrato di 1,86 trilioni di dollari. Tuttavia, secondo un rapporto di Zero2IPO del 2023, questi fondi avevano raccolto solo un totale di 940 miliardi di dollari da fonti pubbliche e private (Luong, Konaev, 2023)

**Politiche di investimento in Corea del Sud:** Nel 2024, la Corea del Sud ha adottato un piano di investimenti da 710,2 miliardi di won, pari a circa 527 milioni di dollari, in 69 settori per stimolare l'innovazione guidata dall'intelligenza artificiale nella vita quotidiana, nell'industria e nei servizi amministrativi del governo. Un rapporto del governo ha affermato che l'integrazione dell'IA potrebbe avere un impatto economico di 310 trilioni di won all'anno entro il 2026 (The Korea Times, 2024).

## **2. Fattori Economici**

**Rivoluzione o bolla speculativa?:** Il valore delle aziende legate all'intelligenza artificiale (IA) è aumentato vertiginosamente a causa della paura di rimanere indietro (FOMO - fear of missing out), nonostante vi siano pochi segni di un profitto imminente per queste tecnologie. Ad esempio, OpenAI, creatore di ChatGPT, ha raccolto 6,6 miliardi di dollari in un round di finanziamento, portando la sua valutazione a 157 miliardi di dollari e sollevando preoccupazioni su una possibile bolla dell'IA. Gli analisti come Rob Enderle avvertono che siamo in una fase di "bolla", in cui le aziende si affrettano a implementare l'IA come parte della trasformazione digitale, ma molte persone non comprendono appieno la tecnologia. Secondo Enderle, ci vorranno ancora alcuni anni prima che la situazione si stabilizzi. Inoltre, questi avverte che la cosiddetta intelligenza artificiale generale, quella che simula pienamente il pensiero umano, non arriverà prima del 2030 (Chapman, 2024) (Jorzik, Klein, Kanbach, Kraus, 2024).

Infine, anche se i giganti del settore tecnologico, come Amazon, Google, Meta e Microsoft, stanno investendo enormi somme nell'IA, ci sono dubbi sui ritorni economici a breve termine. Ad esempio, OpenAI dovrebbe perdere 5 miliardi di dollari quest'anno, nonostante preveda di raggiungere ricavi esponenziali pari a 100 miliardi di dollari entro

il 2029. Secondo l'analista Carolina Milanese, anche se il modello di business dell'IA generativa è difficile da sostenere ora, con il tempo diventerà parte integrante della nostra vita (Chapman, 2024) (Liang, Luo, Shao, Shi, 2022).

**Costi di addestramento dei modelli:** i costi che le aziende devono sostenere per addestrare i modelli di punta sono sempre più elevati. Questo vale specialmente per la categoria chiamata “foundation models”. Questi ultimi sono infatti addestrati utilizzando vastissimi dataset. Degli esempi di foundation models sono: GPT 4, Claude 3 e Llama 2. Queste IA hanno delle capacità davvero straordinarie e per questo motivo sono sempre più utilizzate in situazioni reali (Alawamleh, Shammass, Alawamleh, Ismail, 2024).

Come mostrato in figura 9, nel 2017 il modello originale Transformer, che ha introdotto l'architettura che è alla base praticamente di ogni moderno LLM, ha avuto un costo di circa 900 \$ per la formazione. RoBERTa Large, rilasciato nel 2019, che ha ottenuto risultati all'avanguardia su molti benchmark di comprensione canonici come SQuAD e GLUE, ha avuto un costo di circa 160.000 \$ per la formazione. Facendo un salto al 2023, i costi di formazione per GPT-4 di OpenAI e Gemini Ultra di Google sono stimati rispettivamente in circa \$78 milioni e \$191 milioni (Maslej et al., 2024).

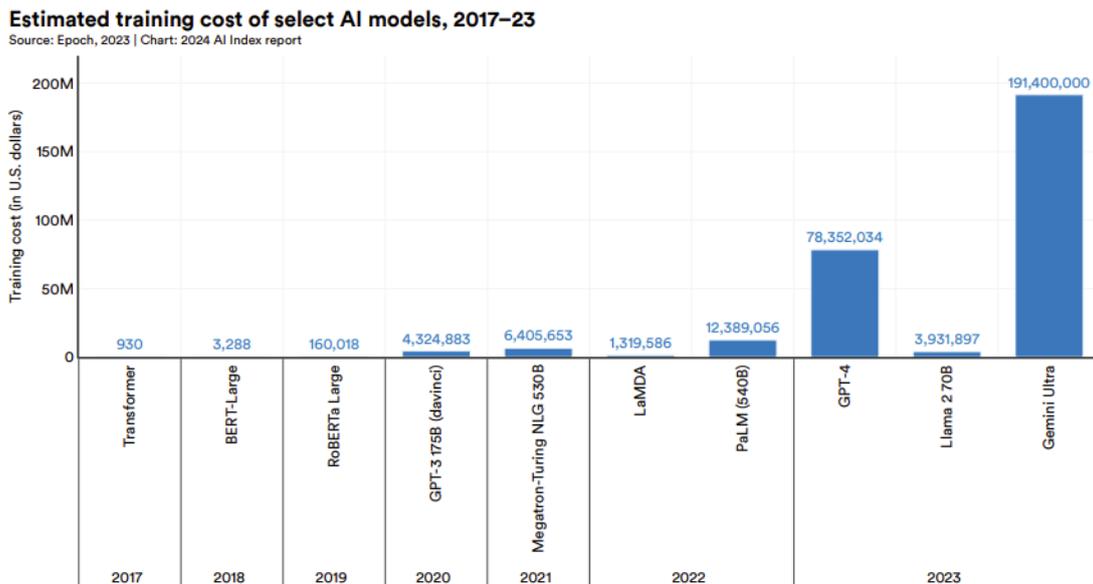


Figura 9: stima dei costi di addestramento dei modelli IA selezionati dal 2017 al 2023 (Maslej et al., 2024).

### 3. Fattori Sociali

**Creazione di posti di lavoro:** I dati reperiti da Lightcast mostrano trend chiave lungo i quali le offerte di lavoro che richiedono competenze in IA sono variate nel corso degli ultimi 10 anni. Nel 2023, i paesi con la percentuale più elevata di offerte di lavoro per profili con competenze nel campo dell'IA sono rappresentati dagli Stati Uniti, con l'1,6%, dalla Spagna, 1,4% e dalla Svezia, 1,3%. Nel suddetto anno, tuttavia, la percentuale registrata negli Stati Uniti, (1,6% come sopra riportato), risulta diminuita rispetto a quella del 2022, che era pari al 2,0% e riflette un calo generale più diffuso.

La figura 10 evidenzia la percentuale di annunci di lavoro con richiesta di competenze nell'IA. Il grafico mostra chiaramente che il fenomeno dell'IA ha indubbiamente creato una domanda di profili specializzati con competenze specifiche in questo campo. Tuttavia, i numeri mostrano che l'offerta di posti di lavoro per risorse con competenze specifiche di IA è stata risibile nel periodo esaminato (Maslej et al., 2024).

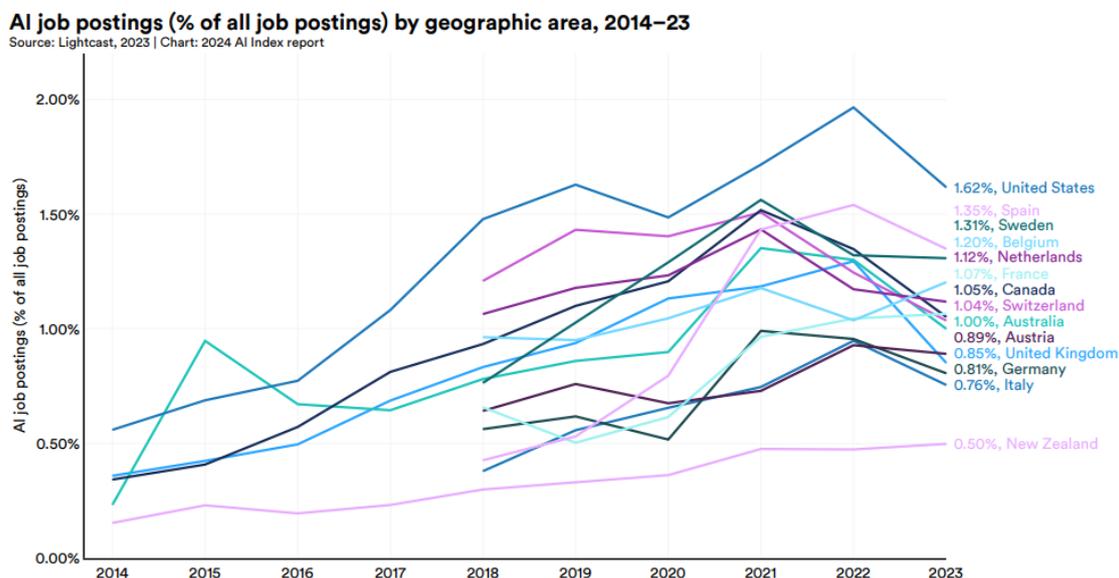


Figura 10: percentuale di annunci di lavoro (sul totale degli annunci) che richiedono competenze nell'IA per area geografica dal 2014 al 2023 (Maslej et al., 2024).

La figura 11 confronta le 10 competenze specializzate più richieste negli annunci di lavoro legati all'IA nel 2023 rispetto a quelle relative al periodo dal 2011 al 2013. Su scala assoluta, la domanda per quasi tutte le competenze specializzate è aumentata nell'ultimo decennio, con un incremento significativo della popolarità di Python, evidenziando la sua

ascesa come linguaggio di programmazione preferito nell'ambito dell'IA (Maslej et al., 2024).

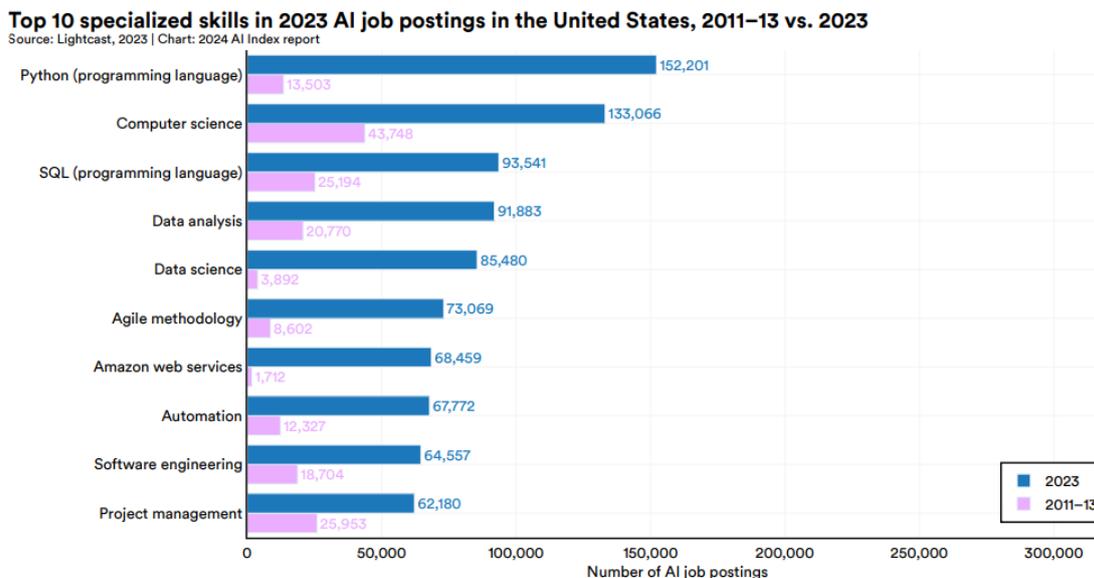


Figura 11: le dieci skill più richieste nelle offerte di lavoro riguardanti l'IA, confronto tra i dati del periodo 2011-2013 e i dati del 2023 (Maslej et al., 2024).

**Distruzione di posti di lavoro:** il rapido sviluppo di tecnologie generative come ChatGPT o Gemini Pro ha suscitato preoccupazioni significative riguardo alla possibile distruzione di posti di lavoro. Secondo un rapporto di Goldman Sachs queste piattaforme potrebbero mettere a rischio fino a 300 milioni di posti di lavoro in tutto il mondo nei prossimi dieci anni. Negli Stati Uniti, circa il 66% delle occupazioni potrebbe essere vulnerabile all'automazione parziale o completa, con fino al 50% delle attività lavorative in alcuni settori che potrebbero essere sostituite. In particolare, il World Economic Forum prevede che entro il 2025, circa 85 milioni di posti di lavoro potrebbero essere persi a causa della divisione del lavoro tra esseri umani e macchine (Stropoli, 2023).

Un altro problema chiave degli Stati Uniti è il sistema fiscale. Attualmente esso incentiva in misura maggiore l'automazione dei lavori rispetto all'assunzione di nuovo personale, con tassi impositivi effettivi sul lavoro che vanno dal 25,5 al 33,5%, contro il 10% sugli investimenti nel capitale tecnologico. Per impedire un'ondata di sostituzioni di posti di lavoro, secondo alcuni economisti, le policy fiscali dovrebbero favorire l'impiego di AI come integrazione piuttosto che come sostituto del lavoro umano (Stropoli, 2023).

**Accettazione e fiducia:** esistono molti esempi tratti dal mondo reale che spiegano come mai l'intelligenza artificiale non sia ancora riuscita a conquistare la piena fiducia dei consumatori, soprattutto per quanto riguarda la trasparenza e i rischi legati al suo utilizzo. Due delle principali problematiche connesse alla fiducia verso l'IA sono:

- **Diffusione di immagini fake:** a gennaio 2024, sono apparse immagini esplicite generate dall'intelligenza artificiale, presumibilmente raffiguranti la cantante Taylor Swift, sulla piattaforma X. Queste immagini, che hanno accumulato 45 milioni di visualizzazioni in 17 ore prima di essere rimosse, dimostrano come i modelli di IA generativa possano creare immagini false, anche in assenza di tali immagini nei dataset originali. Sebbene esistano filtri per prevenire la creazione di tali contenuti, questi vengono spesso aggirati con facilità (Gibson, 2024).
- **Preoccupazioni riguardanti la privacy:** molti chatbot raccolgono una quantità significativa di informazioni personali dai loro utenti, tra cui dati sensibili come dettagli sulla salute o informazioni finanziarie. Una revisione condotta dalla Mozilla Foundation ha rivelato che 11 chatbot romantici esaminati raccolgono informazioni personali in eccesso, che possono essere facilmente abusate. Ad esempio, la politica sulla privacy di Crushon.AI consente di raccogliere informazioni riguardanti la salute sessuale e altre informazioni personali delicate, sollevando preoccupazioni riguardo le pratiche di gestione dei dati da parte di questi servizi (Mozilla, 2024).

Dunque, le persone sono giustamente diffidenti nei confronti di queste tecnologie, soprattutto quando i meccanismi di protezione dei dati e le misure di sicurezza sembrano inadeguati o facilmente aggirabili. Per ottenere una maggiore fiducia da parte del pubblico sarà necessario: migliorare la trasparenza dei processi, implementare filtri più efficaci e garantire una gestione etica dei dati personali.

#### **4. Fattori Tecnologici**

**Abbondanza di dati:** La figura 12 mostra come la quantità totale di dati creati, acquisiti, copiati e consumati a livello globale abbia avuto un rapido aumento, raggiungendo i 47 zettabyte nel 2020. Nei prossimi 10 anni, fino al 2035, si prevede che la creazione di dati globali crescerà a fino oltre 2142 zettabyte. Nel 2020, la quantità di dati creati e replicati ha raggiunto un nuovo picco massimo, la crescita si è rivelata superiore alle aspettative,

a causa dall'aumento della domanda dovuta alla pandemia di COVID-19, poiché più persone hanno lavorato e studiato da casa, utilizzando anche più frequentemente opzioni di intrattenimento domestico (Taylor, 2023).

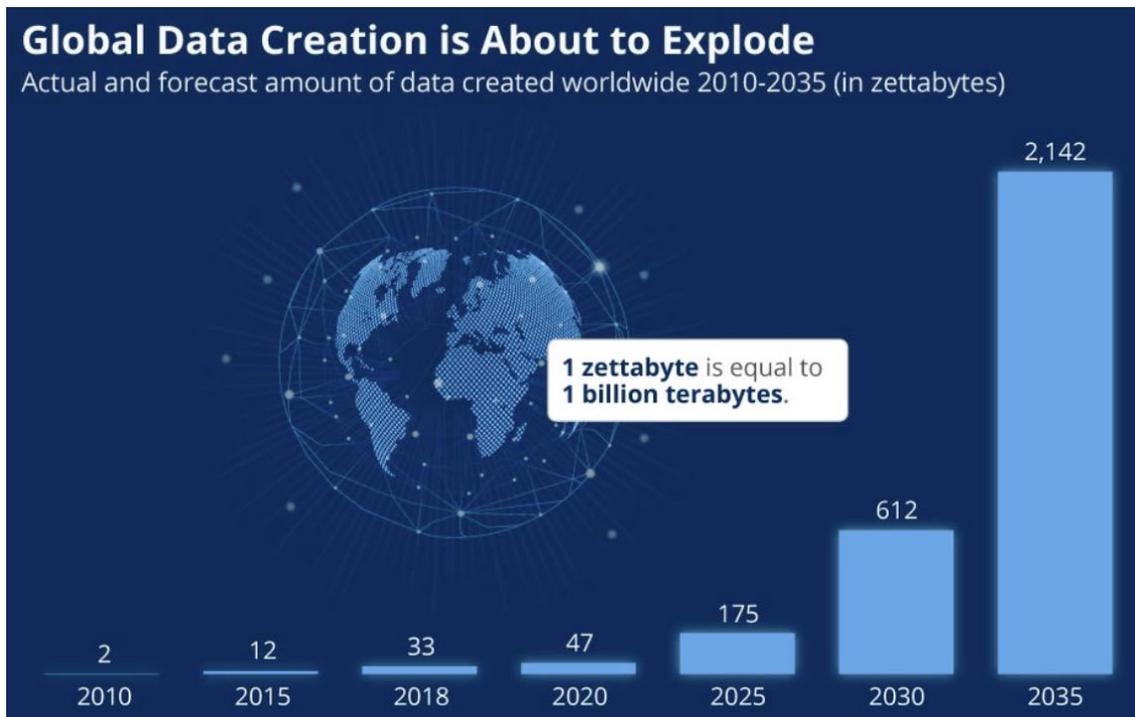


Figura 12: volume di dati creati (in zettabytes) dal 2010 al 2025, con previsione fino al 2035 (statista, 2023).

L'enorme crescita della quantità di dati prodotti a livello globale rende decisamente più accessibile e potente l'uso dell'intelligenza artificiale. I sistemi di machine learning, deep learning e le reti neurali, infatti, dipendono fortemente da grandi quantità di dati per apprendere e per migliorare le loro prestazioni. Senza un'ampia base di dati, questi modelli non riuscirebbero a identificare schemi, fare previsioni accurate o affrontare problemi complessi. Modelli come GPT-4, che sono stati addestrati su dataset di dimensioni enormi (fino a 100 trilioni di parametri), riescono a offrire risposte più precise e pertinenti proprio grazie a questa vasta mole di informazioni (knight, 2024).

**Infrastrutture e dispositivi hardware a supporto dell'IA:** solo una piccola percentuale di questi nuovi dati viene conservata: per esempio, solo il 2% dei dati prodotti e consumati nel 2020 è stato salvato e mantenuto fino al 2021. In linea con la forte crescita del volume dei dati, si prevede un aumento della base installata di capacità di archiviazione, con un tasso di crescita annuale del 272% durante il periodo di previsione dal 2020 al 2025. Nel

2020, la base installata di capacità di archiviazione ha raggiunto i 6,7 zetabyte (Taylor, 2023).

Come si evince dalla figura 13, la dimensione del mercato globale dei semiconduttori è stata valutata a 544,78 miliardi di USD nel 2023. Secondo le previsioni, raggiungerà circa 1.137,57 miliardi di USD entro il 2033, con un incremento previsto ad un tasso di crescita annuo del 7,64% durante il periodo di previsione dal 2024 al 2033 (Precedence research, 2024) (Batra, Jacobson, Madhav, Queirolo Santhanam, 2018).

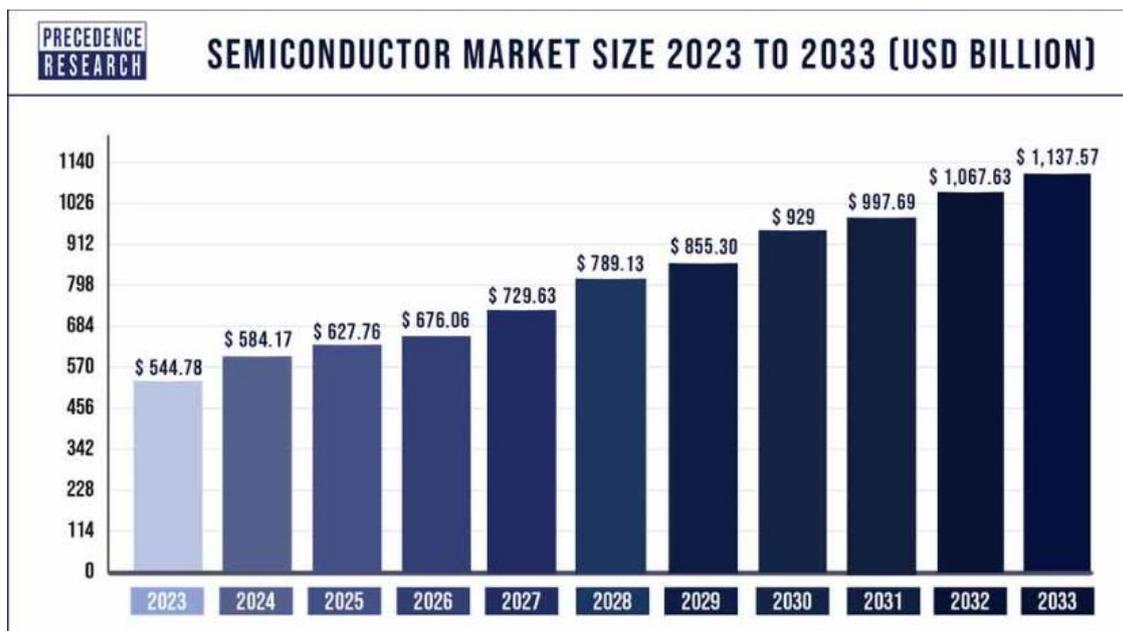


Figura 13: previsioni sulle dimensioni del mercato dei semiconduttori dal 2023 al 2033 (Precedence research, 2024).

## 5. Fattori Ecologici

**Impatto energetico:** l'elaborazione di grandi quantità di dati e il cloud computing, quando eseguiti con l'ausilio dell'IA, richiedono notevoli risorse energetiche per l'alimentazione dei data center. Si prevede, perciò, che entro il 2030 l'ICT e l'IA possano rappresentare fino al 20% della domanda globale di elettricità. Questo comporterebbe un elevato rischio ecologico, qualora l'energia utilizzata non provenisse da fonti rinnovabili e comprometterebbe il raggiungimento degli obiettivi di sostenibilità ambientale (Vinuesa, Azizpour, Leite, Balaam, Dignum, Domisch, Felländer, Langhans, Tegmark and Nerini, 2020).

**Controlli sull'ambiente:** l'IA dà anche la possibilità di monitorare e proteggere le aree naturali, disponendo di algoritmi in grado di rilevare l'inquinamento marino o prevenire la desertificazione. Inoltre, l'elaborazione di immagini satellitari mediante reti neurali e sistemi di apprendimento automatico consente di migliorare la gestione e la conservazione del territorio, preservando la biodiversità e di combattere la deforestazione (Vinuesa et al., 2020).

## 6. Fattori Legali

Negli ultimi anni, il numero di regolamentazioni riguardanti l'intelligenza artificiale (IA) è cresciuto notevolmente, riflettendo l'importanza sempre maggiore di questa tecnologia e la necessità di garantire il suo sviluppo in modo sicuro ed etico. In questo contesto, approfondiremo di seguito le principali normative ed approcci adottati da Stati Uniti, Europa e Cina, tre delle principali potenze mondiali che stanno plasmando il futuro dell'IA con strategie normative differenti.

**Regolamentazioni sull'IA negli USA:** come si evince dalla figura 14, il numero di regolamentazioni legate all'intelligenza artificiale è aumentato significativamente, sia nel 2023 che nei cinque anni precedenti. Nel 2023, c'erano 25 regolamentazioni relative all'IA, un aumento notevole rispetto al 2016, anno in cui ve ne era soltanto una. Dal 2022 al 2023, il numero totale di regolamentazioni relative all'IA è cresciuto del 56,3%.

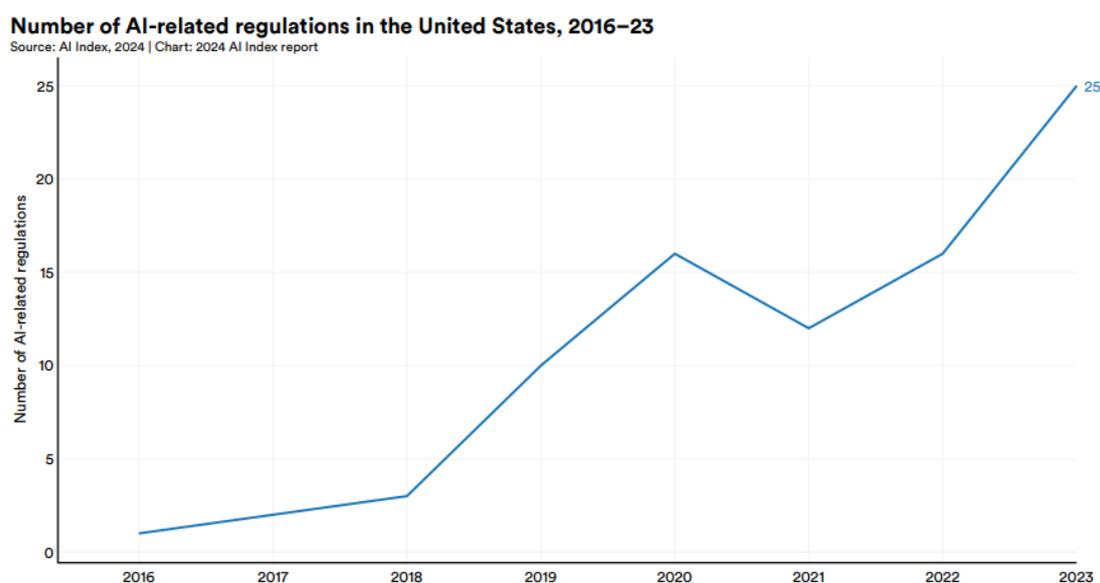


Figura 14: numero di regolamentazioni sull'IA negli USA dal 2016 al 2023 (Maslej et al., 2024).

Tuttavia, negli Stati Uniti non esiste attualmente una legislazione federale completa che regoli in modo specifico l'uso o lo sviluppo dell'intelligenza artificiale (IA). Esistono solamente leggi federali che toccano l'IA in vari contesti, sebbene in modo limitato. Ne è un esempio il National AI Initiative Act del 2020, il quale è stato aggiornato nel 2023 e mira ad espandere la ricerca e lo sviluppo nell'ambito dell'IA. Questa legge ha anche istituito l'Ufficio Nazionale per l'Iniziativa sull'Intelligenza Artificiale, incaricato di coordinare e attuare la strategia nazionale americana sull'IA (Cataleta, Grillo, 2024).

**Regolamentazioni sull'IA in Europa:** Come si può osservare dalla figura 17, l'UE è di certo una delle aree più avanzate in materia, come indicato dal numero di regolamenti sui relativi temi emessi tra il 2017 e il 2023. In particolare, l'AI Act rappresenta una pietra miliare dei principi regolatori delle tecnologie IA. Esso classifica i sistemi AI in base ai rischi; per quelli ad “alto rischio” richiede che siano soddisfatti i requisiti di trasparenza, supervisione umana e sicurezza informatica. In caso di violazione, sono previste pene e sanzioni fino a 35 milioni di euro o al 7% del fatturato annuo globale dell'azienda. Queste normative comportano una significativa modifica dei piani di molte imprese appartenenti al settore, poiché i requisiti legali influenzeranno direttamente le modalità di sviluppo ed implementazione delle soluzioni AI (AI Act, 2024).

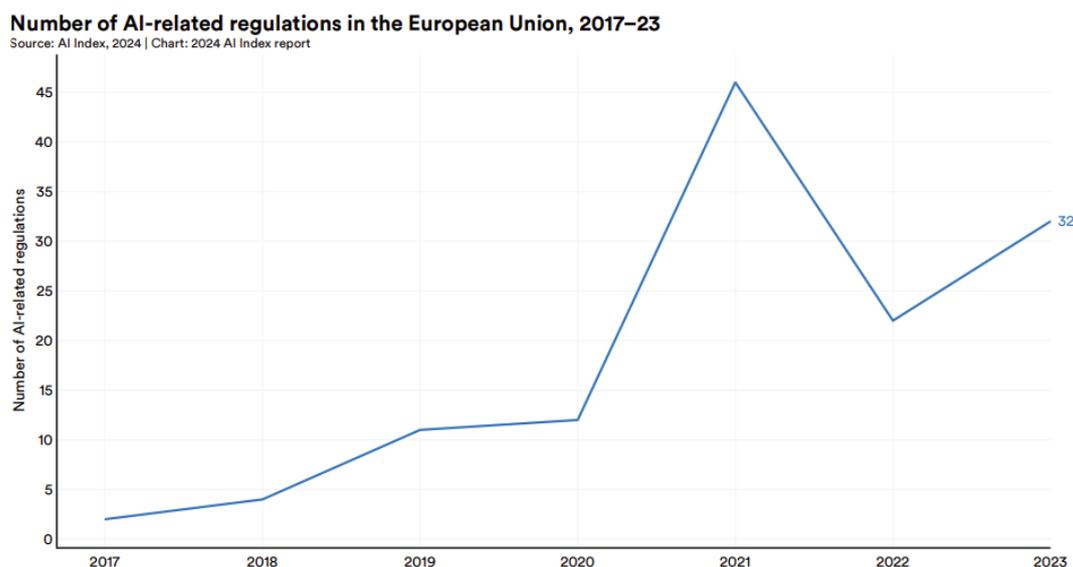


Figura 15: numero di regolamentazioni sull'IA nell'UE dal 2017 al 2023 (Maslej et al., 2024).

Il regolamento proibisce, inoltre, alcuni usi specifici dell'IA, come il social scoring (punteggi sociali), il riconoscimento facciale non autorizzato in spazi pubblici e l'uso di

IA per manipolare comportamenti o riconoscere emozioni in contesti lavorativi e scolastici. Tali limitazioni sono state introdotte per proteggere i diritti fondamentali dei cittadini (AI Act, 2024).

**Regolamentazioni sull'IA in Cina:** Inizialmente, la Cina ha adottato un approccio orientato alla promozione dell'innovazione e alla creazione di un'industria competitiva a livello globale, senza introdurre normative restrittive. Questo periodo è stato caratterizzato dal “Piano per lo Sviluppo dell'IA di Nuova Generazione” (2017), che ha posto le basi per le future politiche IA e ha sottolineato la necessità di integrare profondamente internet, big data e IA con l'economia. Dal 2020 il Governo cinese ha previsto standard nazionali non obbligatori, riguardanti le tecnologie quali gli algoritmi e l'apprendimento automatico. Il documento “Linee Guida per la Costruzione del Sistema di Standard IA della Nuova Generazione”, datato 2020, ha tracciato dunque le guidelines per la definizione degli standard tecnici. A seguito dell'intervento, in primo luogo, c'è stata una maggiore vigilanza governativa; tuttavia, la Cina non ha ancora elaborato una struttura legislativa completa con la quale poter esercitare una regolazione diretta dell'IA.

Dal 2022, lo Stato ha iniziato a promulgare regolamenti obbligatori specifici per le tecnologie IA. Tra le normative più rilevanti figurano le “Disposizioni per la Gestione degli Algoritmi di Raccomandazione” (2021), le “Disposizioni sulla Sintesi Profonda” (2022) e le “Misure per la Gestione dei Servizi di Intelligenza Artificiale Generativa” (2023), che stabiliscono norme precise per settori come gli algoritmi di raccomandazione e l'IA generativa (Gong, Qu, Dorwart, 2024).

Questa analisi PESTEL evidenzia come il settore dell'IA sia fortemente influenzato da una combinazione di fattori politici, economici, sociali, tecnologici, ambientali e legali. Mentre l'IA offre enormi opportunità, deve affrontare sfide significative, come la regolamentazione, la fiducia del pubblico e l'impatto ambientale, tutte questioni che modellano il futuro di questo settore in rapida crescita.

Dopo aver analizzato l'ampio contesto con il modello sovrastante si può focalizzare l'attenzione sull'ambiente competitivo. Quest'ultimo è costituito dall'insieme di attori con i quali si stabiliscono interazioni dirette e ripetute con frequenza significativa. L'attributo “competitivo” non indica quindi semplicemente l'ambiente dove l'impresa compete, ma, quello dove operano gli attori e si manifestano le forze che interagiscono in

modo più intenso con essa, contrastando o favorendo il raggiungimento dei suoi obiettivi (Caroli, 2021).

Per comprendere le condizioni dell'ambiente competitivo e la dinamica delle interazioni tra le aziende al suo interno occorre approfondire la configurazione delle così dette "forze competitive". Un modello ormai ampiamente consolidato per assolvere a questo compito è quello delle cinque forze competitive indicate da Michael Porter (Caroli, 2021).

## Il modello di Porter

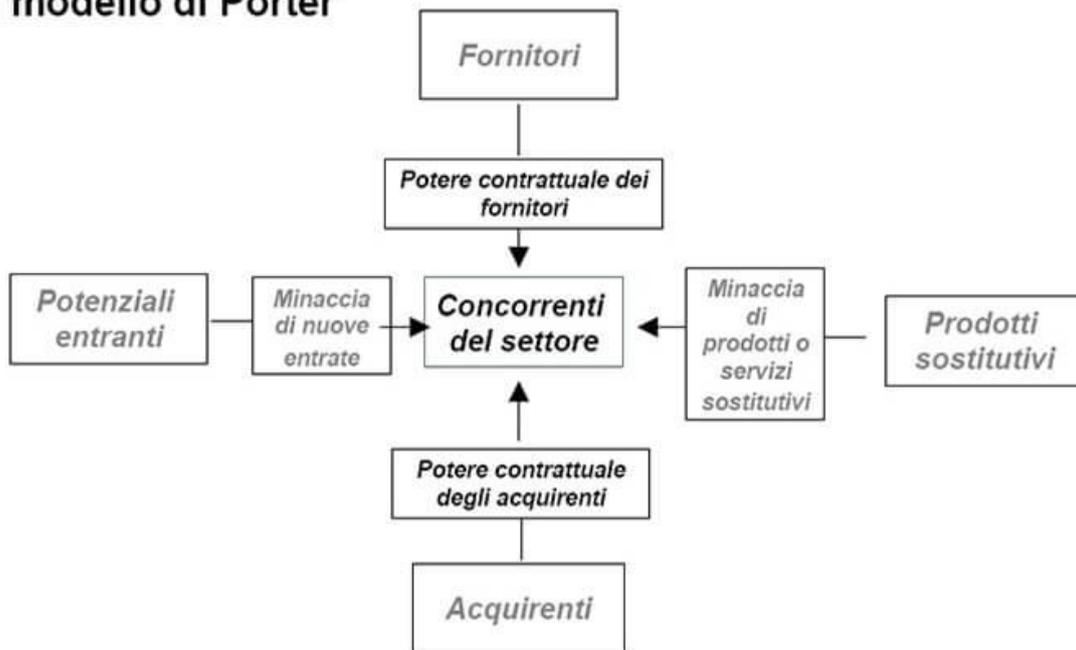


Figura 16: modello di Porter (Spinosi Marketing, 2022)

### 1. Minaccia di nuovi entranti

Il settore dell'intelligenza artificiale presenta barriere d'ingresso elevate. Lo sviluppo dell'IA richiede, infatti, ingenti investimenti, accesso a grandi quantità di dati e capacità tecniche avanzate. Sono poche le aziende che dispongono dei mezzi e delle competenze necessarie per accedere a questa tecnologia (ad esempio: Google, Microsoft e Meta, Amazon, ecc...). Inoltre l'Unione Europea con il già citato AI Act, ha introdotto regolamentazioni molto rigorose, soprattutto per i settori ad alto rischio, che costituiscono ulteriori barriere, di carattere legale, che scoraggiano l'ingresso nel settore (Martens, 2024) (AI Act, 2024).

## **2. Potere contrattuale dei fornitori**

Le aziende che sviluppano IA, in particolare quelle che utilizzano i modelli di machine learning e deep learning, richiedono grandi quantità di dati per l'addestramento dei propri modelli. Tuttavia, l'acquisizione di dati e la creazione di database di alta qualità sono processi costosi e complessi. Di conseguenza, le aziende sono molto dipendenti dalle società che offrono questi dati, le quali possiedono competenze e infrastrutture adeguate alla raccolta di dati su larga scala ed alla loro organizzazione. Ad esempio Nexdata offre enormi datasets pronti all'uso: ad esempio 200.000 ore di dati vocali, 800TB di dati video. Questa abbondanza ed eterogeneità dei dati consente alle aziende di aumentare l'accuratezza dei propri modelli IA, di accelerarne l'addestramento e lo sviluppo delle applicazioni (Kelly, 2024).

## **3. Potere contrattuale dei clienti**

Poiché i clienti hanno aspettative sempre più elevate, le imprese che adottano l'IA devono garantire esperienze personalizzate e prestazioni di livello superiore. Ad esempio, Booking sfrutta l'IA per capire il tipo di viaggio che meglio si adatta a determinate richieste dei clienti anche quando queste risultano molto vaghe. Usando le capacità di computer vision, raccomandazione e comprensione del linguaggio naturale, Booking aiuta a capire meglio ciò che i clienti desiderano. Per esempio, la funzione "Interests" personalizza la homepage degli utenti in base ai loro interessi e alle loro passioni, migliorando l'engagement dei clienti (Katariya, 2019).

## **4. Minaccia di prodotti sostitutivi**

Le imprese che utilizzano già l'IA dovrebbero affrontare tecnologie emergenti come: quantum computing e blockchain. Il quantum computing è un settore all'avanguardia dell'informatica e sfrutta le proprietà della meccanica quantistica per risolvere problemi che vanno al di là della portata dei computer e dei supercomputer di oggi (Schneider, Smalley, 2024). Ha le potenzialità per superare l'IA in quanto a capacità computazionale; ciò potrebbe consentire la risoluzione di problemi complessi in tempi molto ristretti rispetto ai computer standard. Dunque, settori come: la crittografia, la ricerca medica e l'ottimizzazione dei sistemi industriali potrebbero trasformarsi radicalmente. Tuttavia, il quantum computing è ancora in fase di sviluppo, e ci sono sfide tecniche, come la

correzione degli errori e la riduzione del rumore, che devono essere affrontate prima che possa essere adottato su larga scala (Philo, 2024) (AI Tech Park, 2023).

## **5. Rivalità tra concorrenti**

La competizione nel settore dell'intelligenza artificiale (IA) è molto intensa. Per quanto riguarda i chip per l'intelligenza artificiale, Nvidia domina il mercato con una quota stimata tra il 70% e il 85%. I chip sono utilizzati nell'addestramento e nel deployment di modelli simili a quelli di OpenAI. Nvidia detiene una posizione stabile sul mercato grazie ai suoi GPU H100 e alla piattaforma software CUDA che le conferiscono un vantaggio competitivo sul mercato a causa delle prestazioni superiori. Tuttavia, l'azienda compete con altre società che cercano di ottenere una quota di mercato con chip più economici e specializzati, tra cui AMD, con il chip Instinct MI300X, e Intel, con il Gaudi 3. La competizione è alimentata dal fatto che le GPU (Graphics Processing Unit), circuiti elettronici progettati per accelerare la grafica e l'elaborazione delle immagini su una varietà di dispositivi, inclusi schede video, schede madri, telefoni cellulari e computer personali (PC) (Flinders, Smalley, 2024), realizzate da Nvidia, pur essendo dominanti hanno un costo elevato (circa 30.000 dollari a chip), spingendo i clienti a cercare alternative più economiche. Diverse aziende, tra cui D-Matrix e Cerebras Systems, stanno sviluppando nuove architetture che potrebbero ridurre i costi e migliorare l'efficienza per alcune applicazioni specifiche di IA. Inoltre, grandi provider di cloud come Google, Amazon e Microsoft stanno sviluppando chip personalizzati per i loro data center, allo scopo di ridurre la dipendenza da Nvidia. Anche startup stanno entrando nel mercato dei chip per IA, con investimenti significativi da parte di venture capital (6 miliardi di dollari nel 2023), cercando di sviluppare soluzioni innovative che competano con le offerte consolidate di Nvidia (Leswing, 2024).

Un altro fattore rilevante in questo settore è costituito dai vantaggi del first mover. Nel settore dell'IA esistono varie barriere all'entrata in grado di proteggere i primi entranti dai concorrenti futuri. Ad esempio, il costo proibitivo per addestrare modelli di linguaggio su larga scala (LLM), che può raggiungere fino a 10 milioni di dollari per modello. Tuttavia, questi costi potrebbero ridursi con hardware più economico e algoritmi più efficienti. Inoltre, i dataset utilizzati per addestrare i modelli non sembrano rappresentare un'adeguata barriera per le aziende che utilizzano dati pubblici (Spaul, 2023).

La vera barriera potrebbe quindi risiedere nell'inerzia dei clienti: una volta che gli utenti si abituano a utilizzare un assistente IA specifico, sarà difficile convincerli a cambiare, a meno che il nuovo concorrente non offra vantaggi evidenti. Proprio come nei mercati dei motori di ricerca e degli smart speaker, le aziende che si affermano per prime potrebbero mantenere il loro vantaggio competitivo (Spaul 2023) (Costa Climent, Haftor, Staniewski, 2024).

Il settore dell'IA è quindi caratterizzato da barriere all'ingresso elevate e da una forte concentrazione del potere nelle mani di poche grandi aziende. Per i nuovi entranti, trovare una nicchia o sviluppare una strategia di partnership potrebbe essere l'unico modo per avere successo in un settore così competitivo e regolamentato.

Nei successivi paragrafi, si procederà ad esaminare il modo in cui le aziende possono servirsi delle tre strategie competitive per collocarsi in modo efficace in questo scenario altamente mutevole.

### **3.3 Leadership di costo nel settore dell'IA**

Le aziende che operano nel settore dell'IA possono adottare diverse strategie per ottenere vantaggi di costo ed un conseguente vantaggio competitivo rispetto ai loro concorrenti.

Un modo per ridurre i costi è sfruttare le piattaforme di cloud computing. Il cloud computing permette l'accesso su richiesta a risorse di elaborazione, come server (fisici o virtuali), spazio di archiviazione dati, strumenti di rete, software e strumenti di intelligenza artificiale, attraverso Internet con un modello di pagamento a consumo. (Stephanie Susnjara, Ian Smalley, 2024)

Mentre l'IT "on premise" richiede a un'azienda di possedere e gestire direttamente server e data center per l'elaborazione dei dati, il cloud computing offre molti vantaggi a seconda dei servizi selezionati:

- **Risparmio sui costi operativi:** Con il cloud, le aziende possono evitare o ridurre in modo significativo i costi relativi all'acquisto, installazione e gestione di hardware complesso come i mainframe e le infrastrutture fisiche. Invece, pagano solo per le risorse che utilizzano quando ne hanno bisogno, eliminando così costi di capitale inutili. (Stephanie Susnjara, Ian Smalley, 2024)

- **Elasticità e provisioning automatico:** Il cloud computing offre un alto livello di elasticità e gestione autonoma delle risorse. Invece di sovradimensionare la capacità (che poi resta inutilizzata in periodi di bassa attività) per gestire picchi di domanda sporadici, le risorse possono essere scalate in base alle necessità in tempo reale. (Stephanie Susnjara, Ian Smalley, 2024)

Questi vantaggi rendono il cloud computing una soluzione ideale per le aziende che vogliono essere più agili e ottimizzare i costi, specialmente nel contesto della crescita dell'intelligenza artificiale e dei dati.

Un'altra pratica che può essere utilizzata dalle imprese per ridurre i propri costi consiste nel dare in outsourcing determinate attività fondamentali per i modelli di machine learning e quindi per lo sviluppo delle IA, ma che possono anche essere costose sia dal punto di vista delle risorse che del tempo. Un esempio è l'attività di "data labeling". Essa consiste nell'identificare dei dati non elaborati (ad esempio immagini, file di testo, video) e quindi aggiungere una o più etichette a tali dati per specificarne il contesto, consentendo ai modelli di machine learning di effettuare previsioni accurate. (IBM 2024)

A tal fine, le aziende possono collaborare con università e altri enti di ricerca per ridurre i costi legati all'ottenimento di dati o allo sviluppo di tecnologie di intelligenza artificiale (IA). Queste collaborazioni offrono vari vantaggi:

- Le università e gli enti di ricerca dispongono di esperti accademici e di dottorandi con competenze avanzate in settori specifici dell'IA, come il machine learning, la computer vision e l'elaborazione del linguaggio naturale. Collaborare con loro può ridurre i costi di sviluppo, poiché i ricercatori lavorano spesso a progetti innovativi come parte del loro percorso formativo o di ricerca (La Repubblica, 2024).
- Molti governi offrono incentivi per la ricerca accademica congiunta con il settore privato, come sovvenzioni o crediti d'imposta. Ad esempio, l'Unione Europea finanzia programmi come Horizon Europe che promuovono la collaborazione tra aziende e università per sviluppare soluzioni di IA. IBM, per esempio, ha creato il programma IBM Academic Initiative, collaborando con università di tutto il mondo per fornire accesso a risorse cloud e IA per lo sviluppo di progetti innovativi (La Repubblica, 2024).

Infine, un'altra possibilità per ridurre i costi è quella della model compression. Questa pratica comprende un insieme di tecniche di compressione che riducono le dimensioni e la complessità dei modelli di deep learning per renderli più efficienti senza comprometterne troppo le prestazioni. Ciò è indispensabile per rendere i modelli adatti a dispositivi con risorse limitate, come gli smartphone, e ridurre i costi di calcolo e di memoria. In tal modo è anche possibile eseguire operazioni di addestramento più rapidamente e con un minor impiego di risorse, con una conseguente riduzione dei costi di memorizzazione (DataDeep, 2022).

I due principali procedimenti utilizzati per la riduzione dei modelli sono:

- **Pruning:** questo metodo consiste nel rimuovere i pesi e le connessioni meno rilevanti in una rete neurale. In una rete neurale completamente addestrata, alcuni dei collegamenti tra i neuroni hanno pesi molto piccoli, che non contribuiscono significativamente all'output del modello. Eliminando questi collegamenti, si riduce il numero di parametri e quindi la complessità del modello. Un modello pruned richiede meno risorse di calcolo e memoria durante l'inferenza, riducendo anche il tempo di elaborazione (DataDeep 2022).
- **Quantizzazione:** riduce la precisione con cui i pesi e le attivazioni del modello sono rappresentati. I modelli di deep learning utilizzano comunemente valori in virgola mobile a 32 bit per rappresentare i pesi. La quantizzazione riduce questa precisione, utilizzando ad esempio interi a 8 bit, il che riduce la quantità di memoria necessaria per archiviare il modello. Ridurre la precisione può comportare una lieve perdita di accuratezza, ma in molti casi questa riduzione è trascurabile rispetto ai vantaggi in termini di velocità e risparmio di risorse (DataDeep 2022).

Ad esempio Google ha usato la model compression per migliorare le performance di "Google Assistant" sui dispositivi mobili. L'azienda è riuscita nel ridurre le dimensioni del modello di riconoscimento vocale di Google Assistant da 100GB a 0,5GB senza perdere precisione. Intel è riuscita a migliorare l'efficienza energetica dei suoi processori per l'elaborazione di reti neurali (DataDeep 2022).

In sintesi, pruning e quantizzazione sono tecniche essenziali nella model compression, rendendo i modelli di IA più leggeri e efficienti, e quindi più adatti all'implementazione in dispositivi con risorse limitate e a basso consumo energetico.

Una volta ottenuto il vantaggio di costo le imprese devono avere la capacità di sfruttarlo. Questo vale sia per le aziende “produttrici” di IA sia per aziende che operano a valle e che sfruttano l'IA per aumentare l'efficienza dei propri processi produttivi o distributivi.

Le aziende che producono IA spesso reinvestono i vantaggi di costo ottenuti per alimentare ulteriormente ricerca e sviluppo (R&D). Inoltre, in molti casi queste aziende usano la stessa IA per le attività di R&D, così facendo possono accelerare l'innovazione e mantenere un vantaggio competitivo nel lungo periodo (Berteletti, Borden, 2024).

Secondo McKinsey, le aziende che reinvestono in R&D e adottano l'IA per accelerare lo sviluppo dei propri prodotti possono migliorare la produttività fino al 50% e ridurre i tempi di lancio sul mercato fino al 40%. L'intelligenza artificiale, integrata nei processi di simulazione e progettazione, permette alle aziende di esplorare soluzioni innovative in meno tempo e con meno risorse, sfruttando al massimo le economie di scala generate. Inoltre, le aziende con una forte cultura dell'innovazione, alimentata da investimenti in IA e tecnologie avanzate, tendono a superare i loro concorrenti in termini di crescita dei ricavi e dei profitti, con un incremento medio del 16% nel ritorno sul capitale investito (Banholzer, Gartner, LaBerge, 2023).

Nel caso di Google, ad esempio, l'impresa ha reinvestito quanto ottenuto grazie ai vantaggi di costo per migliorare ulteriormente la sua IA e i suoi prodotti e servizi. Google ha sviluppato i Tensor Processing Units (TPU), chip progettati per addestrare i modelli di intelligenza artificiale dell'azienda. L'investimento ha infatti permesso di velocizzare ulteriormente le operazioni legate all'IA, riducendone i costi complessivi e migliorandone i sistemi di apprendimento automatico. Ad esempio, quelli utilizzati per prodotti chiave come il motore di ricerca e Google Assistant (Berteletti e Borden, 2024).

Per quanto riguarda invece le aziende che operano a valle e che sfruttano l'IA per ottimizzare i propri processi produttivi o distributivi si può fare riferimento ad Amazon. Quest'ultima, dopo aver ridotto i propri costi grazie all'ottimizzazione delle operazioni tramite l'IA (come l'automazione nei magazzini e l'ottimizzazione delle rotte di consegna),

ha scelto, da più di venti anni, di applicare prezzi bassi. Amazon è infatti fermamente convinta che questa pratica aumenti la fiducia dei clienti, motivandoli così a tornare (Amazon, 2023).

Un'analisi indipendente condotta da Profitero ha confermato che per sette anni consecutivi, i prezzi di Amazon sono stati i più bassi tra tutti i principali rivenditori online degli Stati Uniti. Nel 2023, essi sono stati in media più bassi del 16% rispetto a quelli dei concorrenti. Amazon continua a lavorare per confrontare ed adattare i propri prezzi a quelli dei concorrenti, consolidando la sua reputazione come il luogo ideale per trovare prezzi bassi. (Black, 2023)

### **3.4 differenziazione nel settore dell'IA**

Questo paragrafo si limiterà all'analisi delle Gen AI (intelligenza artificiale generativa) principali e non metterà a confronto tutte le soluzioni di IA disponibili al fine di ridurre la gamma alle alternative più avanzate e rilevanti per il settore. GPT-4 Turbo, Claude 3 Opus e Gemini 1.5 Pro sono tra le Gen AI più innovative, create per far fronte ad una vasta gamma di applicazioni e soddisfare esigenze nel mercato. Questi modelli sono stati presi in considerazione specialmente per la capacità di adattarsi a molti contesti diversi e complessi e per le loro funzionalità avanzate che li rendono i leader del settore.

In un panorama competitivo e in rapida evoluzione come quello dell'intelligenza artificiale, la strategia di differenziazione è un elemento di vitale importanza per tutte le aziende che cercano di ottenere o mantenere il proprio vantaggio competitivo. Con l'espansione del mercato dell'IA e l'aumento della concorrenza, le aziende devono distinguersi offrendo soluzioni uniche che creano valore per i clienti. Le prestazioni tecnologiche sono sicuramente uno dei punti fondamentali in cui le aziende di IA possono agire con la strategia di differenziazione. Nell'ambito dell'IA generativa si può osservare come Gemini 1.5 Pro di Google sia sicuramente la Gen AI più avanzata.

#### **Differenziazione sulle performance tecniche**

Gemini 1.5 Pro è un modello multimodale, progettato per gestire lunghe sequenze di dati attraverso diversi formati, inclusi testo, video e audio. Rispetto a GPT-4 Turbo, Gemini 1.5 Pro offre miglioramenti significativi nella gestione dei contesti lunghi, supportando fino a 10 milioni di token, a fronte dei 128.000 token di GPT-4 nella sua variante Turbo.

Osservando il Technical Performance Report di Gemini ultra e Gemini 1.5 si può notare una netta superiorità anche rispetto a Claude 3 Opus (di Anthropic). Infatti, Claude 3 Opus supporta al massimo un milione di token, limite che come riportato in precedenza è ampiamente superato da Gemini 1.5 Pro. Quest'ultimo registra un tasso di richiamo del 99,7% anche con un 1 milione di token e del 99,2% fino a 10 milioni di token. Inoltre, Gemini 1.5 Pro è nativamente multimodale, il che significa che può elaborare simultaneamente testo, video e audio, offrendo prestazioni eccellenti in tutte queste modalità, cosa che manca a Claude 3 Opus. Questo rende Gemini 1.5 Pro molto più versatile in compiti che richiedono l'integrazione di dati provenienti da diverse fonti (Gemini Team Google, 2024).

Un token è una sequenza di caratteri che ha un significato semantico per un modello di intelligenza artificiale. Quando un modello di IA elabora un testo, non lavora direttamente con le parole nel loro formato originale. Queste vengono scomposte in unità più piccole chiamate token, che possono essere intere parole o anche parti di parole. Il processo di tokenizzazione è il passaggio in cui il testo, come quello del prompt fornito, viene suddiviso in questi token. Ogni parola o porzione di parola viene convertita in uno o più token, che rappresentano una forma leggibile per il modello di IA. I modelli di linguaggio come Gemini, utilizzando i token per processare e rispondere alle domande, funzionano considerando qualsiasi sequenza di token come un'unica unità di significato. Ad esempio, "intelligenza" potrebbe essere frammentato in più token, mentre parole corte come "IA" potrebbero formare un token unico. Questa è una caratteristica fondamentale dei modelli; è ciò che consente a Gemini e agli altri modelli di elaborare e rispondere in maniera corretta al testo inserito dall'utente (IBM, 2024). Gemini dunque, supportando più token, è in grado di comprendere testi dieci volte più lunghi rispetto a Claude 3 Opus e circa 78 volte più lunghi rispetto a GPT-4 Turbo (Gemini Team Google, 2024).

I miglioramenti significativi mostrati nella figura 17, ottenuti in soli tre mesi, evidenziano chiaramente come Google abbia investito notevolmente nello sviluppo di Gemini 1.5 Pro per garantire un vantaggio competitivo e una chiara differenziazione dal punto di vista delle prestazioni tecniche nell'intelligenza artificiale. Questi progressi spaziano dalla risoluzione di problemi matematici e scientifici complessi fino alla generazione di codici e alla comprensione del linguaggio multimodale; essi sottolineano l'impegno dell'azienda

nell'innovazione tecnologica per mantenere e rafforzare la propria leadership nel mercato dell'IA generativa.

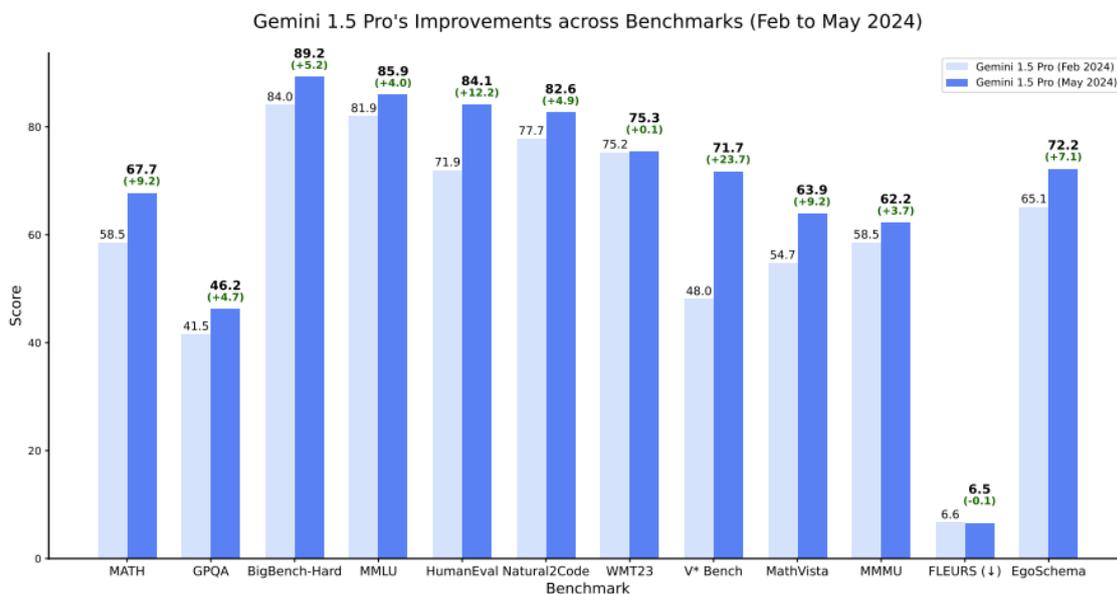


Figura 17: miglioramenti di Gemini 1.5 Pro nei vari benchmark da Febbraio a Maggio 2024 (Gemini Team Google, 2024).

Successivamente, per esaminare ulteriormente le avanzate capacità di Gemini 1.5 Pro, si può utilizzare il test del richiamo in contesti lunghi facendo affidamento sulla valutazione recentemente introdotta del compito "needle-in-a-haystack" (ago nel pagliaio) (Kamradt, 2023), che verifica la capacità di un modello di recuperare un testo (l'ago) inserito in varie posizioni all'interno di una sequenza (il pagliaio). Seguendo studi precedenti (Dhinakaran, 2024), sono stati utilizzati una serie di saggi concatenati e ripetuti scritti da Paul Graham per raggiungere la lunghezza del contesto desiderata. L'ago viene inserito a intervalli spaziali lineari dall'inizio alla fine del contesto, dove l'ago è una frase come "Il numero magico della città {city} è: {numero}", successivamente si chiede al modello di restituire il numero magico per una specifica città. Il sistema di coordinate può essere descritto come segue: l'asse orizzontale x è la lunghezza totale del testo (pagliaio) mentre l'asse verticale y è la posizione del dato di riferimento "ago" in termini di profondità nel pagliaio. Così, il 100% indica che l'informazione è posta alla fine del testo, lo 0% indica l'inizio.

Il test serve a capire quanto bene il modello riesce a trovare l'informazione inserita in punti diversi del testo, indipendentemente dalla lunghezza del contesto. Come si può

vedere nella Figura 18, Gemini 1.5 Pro ottiene un richiamo del 100% fino a 530.000 token e oltre il 99,7% fino a 1 milione di token. Questo compito dimostra chiaramente come Gemini 1.5 Pro sia in grado di recuperare in modo affidabile informazioni da documenti lunghi fino a 1 milione di token. A scopo di confronto, sono riportati anche i risultati di GPT-4 Turbo fino alla lunghezza massima supportata dalla sua API di 128.000 token (Gemini Team Google, 2024).

Gemini 1.5 Pro ha dimostrato, al termine di una significativa serie di valutazioni nel compito "needle in a haystack", di essere capace di estrarre informazioni da testi che superano la lunghezza di un milione di token con un'accuratezza del 99,2%. La figura 18 mostra come performano Gemini 1.5 Pro e GPT-4 Turbo nel compito di trovare "l'ago nel pagliaio" testuale. Le celle verdi indicano che il modello ha recuperato con successo il numero segreto, le celle grigie rappresentano errori dell'API (Application Programming Interface), un insieme di regole o protocolli che permette alle applicazioni software di comunicare tra loro per scambiarsi dati, funzionalità e caratteristiche (Goodwin, 2024). Le celle rosse indicano che la risposta del modello non conteneva il numero segreto. La riga superiore mostra i risultati di Gemini 1.5 Pro, con un contesto da 1.000 a 1 milione di token (a sinistra) e da 1 a 10 milioni di token (a destra). La riga inferiore presenta i risultati per GPT-4 Turbo, con un contesto fino alla lunghezza massima supportata di 128.000 token (Gemini Team Google, 2024).

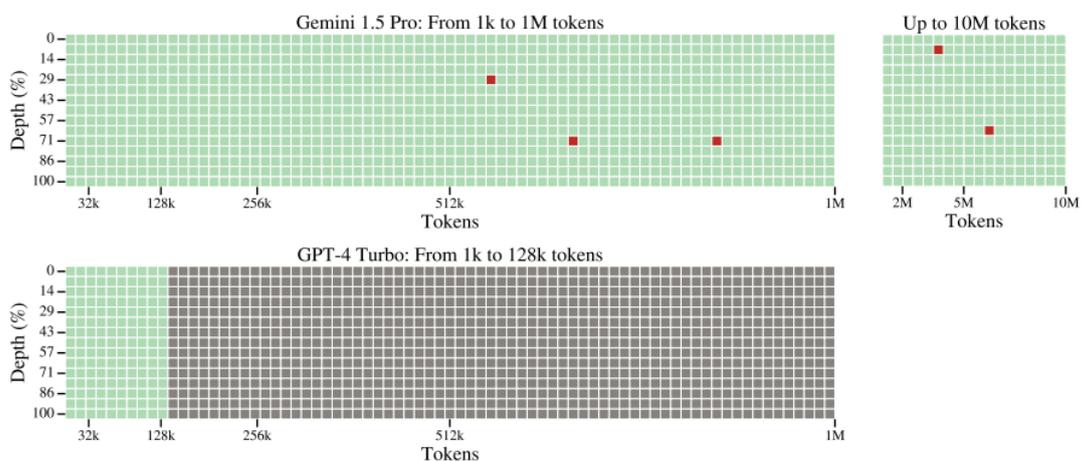


Figura 18: come Gemini 1.5 Pro e GPT 4 Turbo hanno risposto al test del richiamo in contesti lunghi (Gemini Team Google, 2024)

Come si può osservare nella figura 19, Gemini 1.0 è riuscita a superare GPT 4 Turbo nell'MMLU (Massive Multitask Language Understanding), ovvero un benchmark progettato per valutare la conoscenza generale e la capacità di risolvere problemi di un modello IA su 57 materie diverse. Queste includono campi delle scienze STEM come matematica e fisica, così come discipline umanistiche e sociali come storia, diritto, medicina ed etica. Il test comprende domande che spaziano dalle nozioni di base fino a problemi più complessi, che richiedono capacità di ragionamento e analisi (Google Deepmind, 2024).

L'MMLU viene usato per valutare la capacità di linguaggio dei modelli di IA e la loro abilità nel generalizzare le conoscenze acquisite applicandole in altri contesti. Questo benchmark è di particolare interesse in quanto contempla sia domande di livello scolastico sia domande più tecniche e specialistiche. In questo modo si riesce ad elaborare una valutazione completa del modello nel gestire un vasto spettro di argomenti e sfide accademiche.

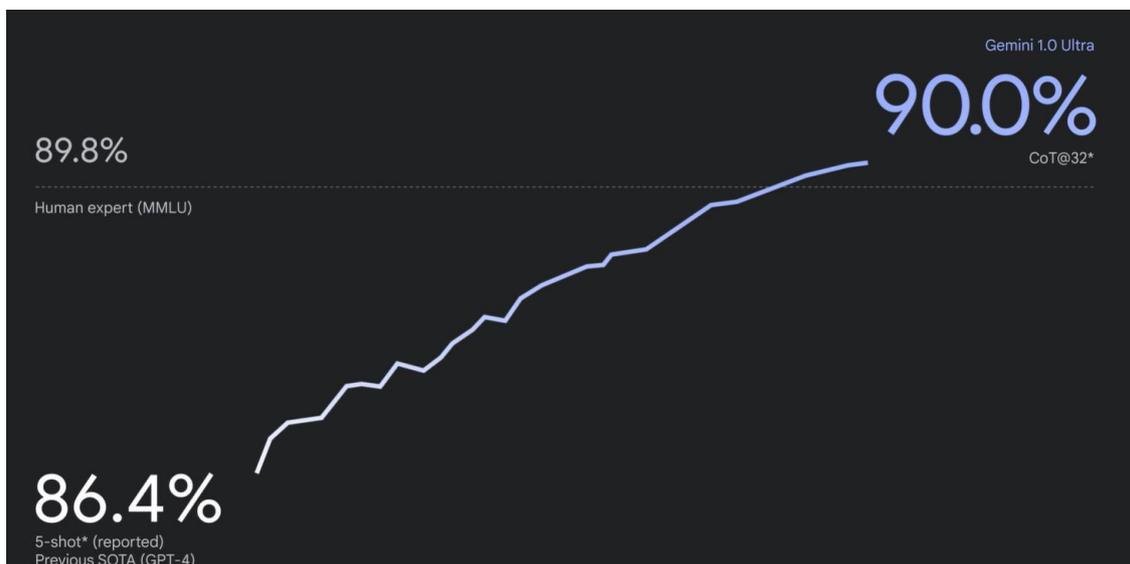


Figura 19: gemini 1.0 (versione precedente rispetto a gemini 1,5 pro) è riuscita a superare GPT 4 e, per la prima, volta anche un esperto umano nell'MMLU (Massive Multitask Language Understanding) (Google Deepmind, 2024).

Quindi, Gemini Ultra costituisce un esempio di come la differenziazione delle prestazioni tecnologiche rappresenti un'ottima soluzione per conseguire un margine competitivo nel campo dell'IA. Come si è visto, la piattaforma si distingue dai concorrenti GPT 4 Turbo e Claude 3 Opus in quanto consente di gestire lunghe sequenze di dati (in formati

multimodali) e contesti. Inoltre, straordinarie capacità di risoluzione di problemi sono attestate dai risultati eccezionali nei benchmark come l'MMLU, ragione per cui il modello viene preferito da aziende e professionisti con elevate aspettative. Questa esclusiva caratteristica rende Gemini Ultra la scelta più idonea per tutti coloro che ricercano il massimo nel campo dell'innovazione, dell'efficienza e della performance analitica.

### **Differenziazione sul prezzo**

Dal punto di vista dei prezzi Claude 3 Opus costituisce una soluzione economica migliore rispetto a GPT 4 Turbo in termini di rapporto qualità prezzo. Entrambi i modelli hanno un costo di €20 al mese; tuttavia, Claude 3 Opus è in grado di gestire contesti di lunghezza superiore rispetto a quelli di GPT 4 Turbo. Per quanto riguarda Gemini 1.5 Pro, le tariffe non sono attualmente note. Il modello è disponibile principalmente attraverso infrastrutture cloud e servizi aziendali, il che rende più difficile una valutazione diretta del costo per gli utenti finali (OpenAI, 2024) (Anthropic, 2024).

Rytr costituisce una soluzione più economica rispetto a Claude 3 Opus e GPT 4 Turbo, con una tariffa mensile di €7,50. Ovviamente non è paragonabile a nessuno dei modelli già citati in termini di performance ma rappresenta una valida alternativa per gli utenti che accettano compromessi in termini di prestazioni e possibilità di elaborazione, non necessitando delle capacità avanzate dei modelli più costosi ma cercano comunque uno strumento utile per la generazione di contenuti (Rytr, 2024).

La differenziazione di Rytr si basa sull'offerta di una soluzione più economica, sebbene questo comporti prestazioni inferiori e una minore capacità di elaborazione rispetto a modelli più potenti. Questa scelta sul mercato consente a Rytr di acquisire utenti sensibili ai prezzi, disposti a rinunciare a caratteristiche aggiuntive per un prezzo inferiore, il che dimostra come un prodotto possa competere sul prezzo e non sulla qualità.

Da ultimo, è molto importante precisare che tutte le piattaforme di IA già menzionate offrono la possibilità di provare delle loro versioni gratuite e depotenziate nelle performance (come GPT 3.5). Con questa strategia si permette ai clienti di testare il prodotto senza dover sostenere alcuno tipo di costo. Tuttavia, queste versioni presentano dei limiti sostanziali: ad esempio non possono elaborare immagini o audio, né possono analizzare file. Questo significa che i clienti saranno poi incentivati ad abbonarsi per

sbloccare tutte le funzioni aggiuntive. Questo approccio consente alle aziende di far conoscere i loro prodotti e, una volta che l'utente percepisce i limiti della versione gratuita, lo incentiva a scegliere un abbonamento a pagamento per ottenere accesso completo alle funzionalità più avanzate.

### **Differenziazione sulla facilità d'uso**

GPT-4 Turbo riunisce prontezza d'accesso e facilità di utilizzo sulla piattaforma ChatGPT. L'interfaccia è user-friendly, è accessibile a tutti e garantisce a qualsiasi persona la possibilità di sfruttarne le funzionalità. Semplificata ulteriormente dalle numerose integrazioni con applicazioni e software standard quali Microsoft Office, risulta perfetta per il fruitore che desidera impiegare subito l'intelligenza artificiale nel proprio lavoro, accrescendo il rendimento senza dover apprendere modalità complesse e nuove. L'accesso a GPT-4 Turbo tramite abbonamento ChatGPT Plus consente inoltre di passare facilmente da account personali a team, facilitando il lavoro collaborativo (Kommunicate, 2024).

Anche la strategia di differenziazione di Claude 3 Opus ha previsto la realizzazione di un'interfaccia chiara e pulita. Questa piattaforma è in grado di rendere al meglio in termini di coerenza nelle risposte, soprattutto considerando grandi volumi di informazioni. Tuttavia, non possiede un livello di integrazione con applicazioni e software al pari di GPT 4 Turbo per cui ne risulta meno immediato l'utilizzo nel lavoro. Inoltre, la piattaforma pone enfasi sulla sicurezza nel trattamento delle informazioni basandosi sui principi di "Constitutional AI" (Kommunicate, 2024).

Gemini 1.5 Pro, la soluzione di Google, è il modello più potente tra i tre. Tuttavia, la sua complessità lo rende meno immediato da usare rispetto a GPT-4 Turbo e Claude 3. Pur avendo un'interfaccia moderna e ben progettata secondo i principi del design di Google, Gemini richiede una maggiore competenza per essere sfruttato appieno. Questo modello infatti è più adatto a professionisti e aziende che necessitano di capacità multimodali avanzate. Proprio come Claude 3 Opus, Gemini 1.5 Pro non possiede ancora un elevato livello di integrazione con applicazioni di uso comune come Microsoft Office. Questo problema, se trascurato, potrebbe limitare la sua accessibilità solo a chi è già inserito nell'ecosistema Google Cloud. Quindi, il modello in questione risulta essere meno adatto

per chi cerca un'esperienza immediata e semplice in contesti di lavoro quotidiano (Kommunicate, 2024).

Nel complesso, GPT-4 Turbo è la scelta ideale per chi cerca semplicità e integrazioni ampiamente diffuse, Claude 3 Opus è perfetto per chi necessita di risposte complesse e sicure, mentre Gemini 1.5 Pro eccelle nelle applicazioni multimodali, ma richiede maggiori competenze per l'utilizzo completo.

### **3.5 Focus nel settore dell'IA**

Finora sono state prese in considerazione tre aziende (OpenAI, Anthropic e Google) che adottano una strategia di mercato ampia. Ciò significa che non si focalizzano su un particolare segmento ma sulla totalità del mercato dell'IA generativa. I tre modelli già descritti possono essere infatti utilizzati per una vastissima gamma di applicazioni: casi d'uso aziendali, assistenza clienti, analisi di grandi quantità di dati e creazione di contenuti. Ciò significa che il target è costituito da molti settori anche profondamente diversi l'uno dall'altro.

Al contrario, esistono varie aziende che decidono di dedicarsi solamente ad una certa nicchia di mercato. DeepL ne è un esempio: essa si concentra infatti sul settore della traduzione automatica ed offre servizi incredibilmente raffinati per professionisti con esigenze estremamente elevate in termini di naturalezza della traduzione. Un'altra azienda altamente specializzata è Jasper AI. Essa è specializzata (con la sua IA chiamata Jasper) nella creazione di contenuti di marketing ed offre strumenti dedicati ai copywriter e marketer per la creazione di testi pubblicitari. Queste aziende si differenziano per il loro approccio mirato, servendo specifiche esigenze di mercato piuttosto che puntare a una base d'utenza generalista.

#### **DeepL**

DeepL utilizza una propria rete neurale, una tecnologia che imita vagamente il funzionamento del cervello umano, e che, secondo l'azienda, le dà un vantaggio rispetto ai concorrenti. Nonostante Google Translate abbia un bacino di utenza (oltre un miliardo di installazioni) largamente superiore rispetto a quello di DeepL, essa è riuscita ad ottenere ingenti finanziamenti ed investimenti che l'hanno portata ad essere valutata due

miliardi di dollari dopo una raccolta di 300 milioni di dollari nel maggio scorso (Magee, 2024).

La startup, con sede a Colonia, è riuscita a conquistare molte aziende multinazionali come clienti, affermandosi con oltre 100.000 clienti aziendali, tra cui Nikkei, Fujitsu, Deutsche Bahn, e metà delle aziende di Fortune 500. Gli usi variano dalla traduzione di messaggi di assistenza clienti al miglioramento delle comunicazioni interne tra le sedi globali (Magee, 2024).



Figura 20: la figura mostra che le traduzioni di DeepL sono 1,3 volte più preferite rispetto a Google Translate, 1,7 volte più preferite rispetto a GPT 4 e 2,3 volte più preferite rispetto a Microsoft (DeepL, 2024).

La strategia adottata da DeepL le ha quindi consentito di offrire servizi progettati specificatamente per le esigenze di traduzione di utenti aziendali e professionisti, ponendosi come principale obiettivo il rispetto di standard qualitativi molto elevati. Questa strategia è stata inoltre combinata con una strategia di differenziazione sul prezzo e sulla qualità (come mostrato in figura 20). Tutto ciò le ha permesso di distinguersi dai suoi principali competitor.

Un chiaro risultato della differenziazione si può notare nei confronti di Systran, ovvero un'altra intelligenza artificiale generativa specializzata nella traduzione automatica. Sebbene anche Systran offra una versione avanzata del servizio, il costo della sua opzione

più avanzata, Systran Pro Premium, è di 31,99 € al mese, un prezzo più basso rispetto ai 49,99 € al mese della versione DeepL Ultimate. Tuttavia, DeepL giustifica il costo più elevato con prestazioni superiori, offrendo traduzioni di qualità significativamente migliore rispetto a Systran. Le traduzioni di DeepL sono apprezzate soprattutto per la loro precisione e capacità di preservare il contesto e le sfumature linguistiche anche in testi complessi come documenti tecnici e legali (DeepL, 2024) (Systran, 2024).

Riassumendo, la strategia di DeepL è essenzialmente una combinazione di focus e differenziazione. Infatti, offre un prodotto di traduzione automatico premium, in grado di giustificare il prezzo maggiore rispetto a quello applicato dai competitor. Così facendo, si posiziona come prima scelta delle aziende che esigono qualità e precisione nelle traduzioni (Byron, 2024).

## **Jasper**

Un'altra azienda che ha adottato strategie di focus su una nicchia è Jasper AI (creatrice dell'IA "Jasper"). in questo caso la nicchia riguarda principalmente campagne e attività di marketing. Questa piattaforma è stata progettata specificamente per rispondere alle esigenze dei team di marketing aziendale, offrendo strumenti e funzionalità che facilitano la creazione, l'analisi e l'ottimizzazione di contenuti marketing su larga scala. Nel confronto seguente, si osserveranno le differenze tra ChatGPT e Jasper, analizzando le tecniche usate da quest'ultima per specializzarsi nelle attività di marketing.

ChatGPT è stato creato per una vasta gamma di scopi e utenti generici, dagli sviluppatori agli studenti, dai professionisti della finanza agli appassionati di viaggi. Tuttavia, le aziende e i loro team di marketing hanno bisogno di una soluzione di intelligenza artificiale che offra più dei casi d'uso generici che ChatGPT fornisce. Hanno bisogno di una soluzione costruita per loro.

Il motore AI proprietario di Jasper garantisce accesso a vari modelli linguistici avanzati da OpenAI, Google, Anthropic e altri, contando in totale oltre una dozzina di modelli. Questo approccio di interoperabilità permette di selezionare la combinazione ottimale di modelli per specifici casi d'uso o settori. Tale strategia riduce la dipendenza da una singola fonte, contrastando i problemi di disponibilità che possono affliggere piattaforme come ChatGPT, che dipendono unicamente dai modelli di OpenAI. Jasper mantiene così un

uptime garantito del 99,95%. Jasper è in grado di utilizzare i punti di forza dei vari modelli per fornire i migliori risultati possibili. L'eccessiva dipendenza di ChatGPT dai suoi modelli GPT comporta che i suoi risultati sono più limitati in termini di portata ed efficienza (Jasper AI, 2024).

Sia Jasper che ChatGPT si impegnano per rispettare le politiche di sicurezza dei dati e degli accessi, conformandosi, per esempio, ai protocolli SOC2 e l'SSO. Tuttavia, attraverso approfondimenti in materia di sicurezza, si può constatare che Jasper è la scelta migliore per i team di marketing aziendale. Come anche ChatGPT, Jasper crittografa tutti i dati in transito tramite i protocolli TLS 1.2 e AES-256. In tal modo, vi è la garanzia che tutte le informazioni sensibili siano protette in ogni circostanza e condizione, sia in fase di trasmissione che di archiviazione. Una delle migliori garanzie di sicurezza offerte da Jasper ai clienti è la possibilità di creare spazi privati per diversi progetti, controllando le fonti dei dati e i diritti di pubblicazione. Pertanto, la tutela della privacy è sistematicamente garantita (Jasper AI, 2024).

Anche se ChatGPT offre un piano Bug Bounty per la divulgazione responsabile delle vulnerabilità, Jasper fa un passo in più in termini di sicurezza servendosi di una terza parte indipendente per i test di penetrazione annuali della rete e delle applicazioni. Ciò consente di individuare e risolvere in modo proattivo qualsiasi potenziale vulnerabilità prima che possa diventare un problema. Sia Jasper che ChatGPT sono conformi alle normative GDPR e CCPA. Il programma di sicurezza di Jasper è anche conforme ai principi dei servizi fiduciari dell'AICPA, dimostrando il suo impegno per il rispetto delle best practice del settore. Jasper mostra anche i report SOC2 per fornire prove del suo impegno sulla sicurezza (Jasper AI, 2024)

In breve, mentre sia Jasper che ChatGPT danno priorità alla sicurezza dei dati e offrono crittografia, controllo degli accessi e conformità alle normative, le caratteristiche aggiuntive di Jasper lo rendono la scelta ideale per i team di marketing aziendali. Dal controllo amministrativo agli spazi privati e a un team di sicurezza dedicato, Jasper offre un approccio alla sicurezza completo, solido e su misura che lo distingue veramente.

ChatGPT può disporre di funzionalità aziendali che consentono ai membri del team di produrre contenuti su larga scala; tuttavia, non esistono funzionalità dedicate alla gestione di tali contenuti. Le funzionalità Spaces e Team Acceleration di Jasper offrono un

vantaggio significativo, riducendo le comunicazioni errate e migliorando la trasparenza. Con gli Spaces di Jasper, i team di marketing possono controllare l'accesso e l'utilizzo da parte di tutti i membri del team e di eventuali fornitori di soluzioni esterne. È possibile impostare spazi privati per diversi progetti, controllare le fonti di dati e gestire i diritti di pubblicazione. Di conseguenza, tutte le operazioni all'interno del copilota Jasper si svolgono senza problemi, riducendo le incomprensioni e i disguidi (Jasper AI, 2024).

I grandi team di marketing spesso gestiscono più campagne contemporaneamente. Ogni campagna può coinvolgere diversi membri del team e collaboratori esterni. Creando spazi separati per ogni progetto, i team leader possono garantire che tutti abbiano accesso alle informazioni e agli strumenti giusti. Questo non solo riduce il rischio di errori di comunicazione, ma assicura anche che le informazioni sensibili siano condivise solo con le parti interessate. Le funzioni di accelerazione dei team di Jasper sono dunque progettate per migliorare il coordinamento tra i team di marketing. Queste funzioni consentono ai team di pianificare progetti e campagne con roadmap dettagliate di compiti e incarichi. Il pregio di questo sistema è che i processi di pianificazione e creazione sono integrati, in modo che i membri del team non perdano tempo a riferire sui progressi o ad aggiornare uno strumento di gestione dei progetti esterno (Jasper, 2024).

Ad esempio, un responsabile marketing che intenda monitorare i progressi del lancio di un nuovo prodotto, può utilizzare le funzioni di Team Acceleration per averne una visione completa e aggiornata. Questo migliora l'informazione e consente anche di intervenire tempestivamente quando necessario. Queste funzioni migliorano la comunicazione ed aiutano i team di marketing a operare in modo più efficiente ed efficace. ChatGPT non possiede queste funzionalità.

Nel presente capitolo si è esaminata l'intelligenza artificiale da diverse prospettive, delineando prima di tutto cosa sia l'IA e quali siano le sue principali applicazioni e potenzialità. Successivamente, si è discusso di come le strategie di leadership di costo possano consentire alle aziende di offrire soluzioni di IA a prezzi competitivi, estendendo l'accesso e la diffusione delle tecnologie basate su IA.

Infine, lo studio di differenziazione e focus ha evidenziato, da un lato che unicità e innovazione dell'intelligenza artificiale generano un notevole vantaggio sul mercato attraendo clienti e creando valore aggiunto, dall'altro che concentrarsi su una nicchia del

mercato (come la traduzione o il marketing) consente alle aziende di ottimizzare l'uso delle proprie risorse migliorando la qualità e l'efficacia delle soluzioni offerte. Questi approcci, se ben integrati, possono non solo accelerare il progresso tecnologico ma anche favorire una maggiore adozione dell'IA in diversi ambiti industriali e sociali.

## CAPITOLO 4: IL CASO AMAZON

### 4.1 Panoramica azienda Amazon

Amazon è una delle società più grandi ed influenti al mondo, oltre che uno dei simboli dello spirito imprenditoriale e della digitalizzazione. Fondata da Jeff Bezos nel 1994 a Seattle, Washington, inizialmente come semplice libreria online, nel tempo si è evoluta fino a divenire una mega corporation che domina settori come l'e-commerce, l'intrattenimento digitale, l'intelligenza artificiale e via dicendo. Si tratta di una realtà estremamente diversificata, le cui attività possono essere classificate in diverse categorie:

**E-commerce e Marketplace:** Il business core di Amazon si fonda sul suo sito e-commerce in cui vende direttamente al consumatore finale o agisce anche come intermediario per la vendita di merci tra consumatori e venditori terzi. Con il servizio "Fulfillment by Amazon" migliaia di società ricorrono alla rete logistica di Amazon per ottenere visibilità e vendere i propri prodotti in tutto il mondo.

**Abbonamenti e Servizi Prime:** Il servizio Amazon Prime (con più di 200 milioni di abbonati) consente spedizioni gratuite, accesso a Prime Video e Prime Music e altri servizi premium. Si tratta di una strategia che ha aumentato drasticamente la fedeltà degli utenti e costituisce una fonte di entrate ricorrenti (HDBlog, 2021).

**Amazon Web Services:** Il marchio è leader mondiale nel campo del cloud computing offrendo servizi di storage, machine learning, data analysis, nonché infrastrutture per aziende di ogni dimensione. L'attività genera una percentuale rilevante dei profitti.

**Media e Intrattenimento:** lo streaming video Amazon è paragonabile a quello di titani del settore quali Netflix e Disney+; Amazon Music è una piattaforma di streaming musicale; Amazon, tramite la divisione Amazon Studios, produce inoltre film e serie TV (come The Boys e Gli anelli del potere), alcuni dei quali hanno riscosso un grande successo.

**Dispositivi tecnologici e AI:** L'azienda sviluppa altresì dispositivi quali Kindle, Fire TV e altri dispositivi elettronici che si integrano con Alexa, il suo assistente digitale.

**Servizi di pagamento:** Amazon dispone anche di un servizio di pagamento digitale, chiamato Amazon Pay.

Quanto a dimensioni e impatto globale, Amazon è un colosso in termini di: risorse economiche, dipendenti e tecnologie utilizzate.

### Risorse economiche

Come mostra il grafico in Figura 21 Amazon fino al 2023 ha avuto un'impressionante crescita delle vendite nette a livello internazionale. Le aree geografiche includono Stati Uniti, Germania, Regno Unito, Giappone e "Resto del mondo". Le vendite nette dell'azienda sono aumentate da circa 89 miliardi di dollari nel 2014 a oltre 600 miliardi di dollari nel 2023, ciò mostra in che modo Amazon abbia consolidato la sua leadership globale nel settore dell'e-commerce.

Un altro importante punto di svolta nell'evoluzione della società può essere osservato a partire dal 2020. In questo anno, infatti, si è verificata un'impennata delle vendite nette, che sono passate da 280 miliardi nel 2019 a 386 nel 2020. Questo fenomeno è certamente attribuibile all'impatto della pandemia di COVID-19, che ha costretto i consumatori a imparare con urgenza a servirsi dell'e-commerce come fornitore quotidiano dei prodotti di cui avevano bisogno. È qui che Amazon ha dimostrato di essere in grado di adattarsi alla situazione secondo le necessità del momento.

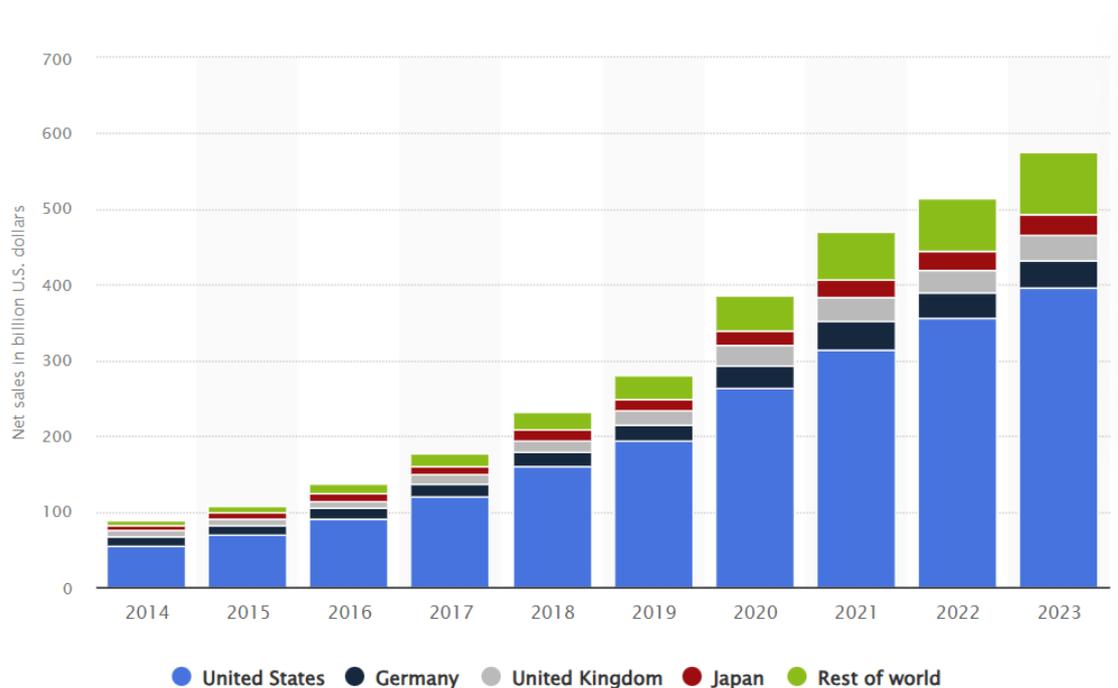


Figura 21: vendite nette annuali di Amazon nei principali mercati selezionati dal 2014 al 2023 (Statista, 2024)

Nel grafico in figura 22 è illustrato l'andamento delle vendite nette di Amazon per un periodo di un anno (dal quarto trimestre del 2022 al quarto trimestre del 2023), sia dal punto di vista geografico che settoriale delle entrate. Una delle principali informazioni è l'incremento del 12% annuo (Y/Y) delle vendite, esse passano da 513,983 miliardi di dollari nel Q4 2022 a 574,785 miliardi di dollari nel Q4 2023. Questo aumento riflette una rilevante espansione delle operazioni aziendali, sostenuta probabilmente da vari fattori tra cui: l'aumento della domanda dei consumatori, l'espansione in mercati internazionali e la continua crescita del segmento AWS.

Il grafico a torta mostra che il Nord America genera il 61% delle vendite totali (352,828 miliardi di dollari), ed è quindi il principale mercato dell'azienda. Questo risultato evidenzia l'importanza della forte base di clienti negli USA e in Canada. I mercati internazionali contribuiscono per il 23% delle vendite (131,200 miliardi di dollari) dimostrando il ruolo strategico dell'espansione globale, nonostante la crescita in queste regioni si trovi spesso ad affrontare sfide come differenze normative e culturali. L'ultima porzione del grafico è ricoperta da AWS, che genera il 16% delle vendite totali (90,757 miliardi di dollari) e rimane dunque un pilastro fondamentale del modello di business aziendale.

L'andamento positivo delle vendite evidenziato dal grafico riflette la capacità dell'azienda di adattarsi a un ambiente di mercato complesso e competitivo, capitalizzando su strategie mirate di crescita regionale, diversificazione dei ricavi e investimenti tecnologici.

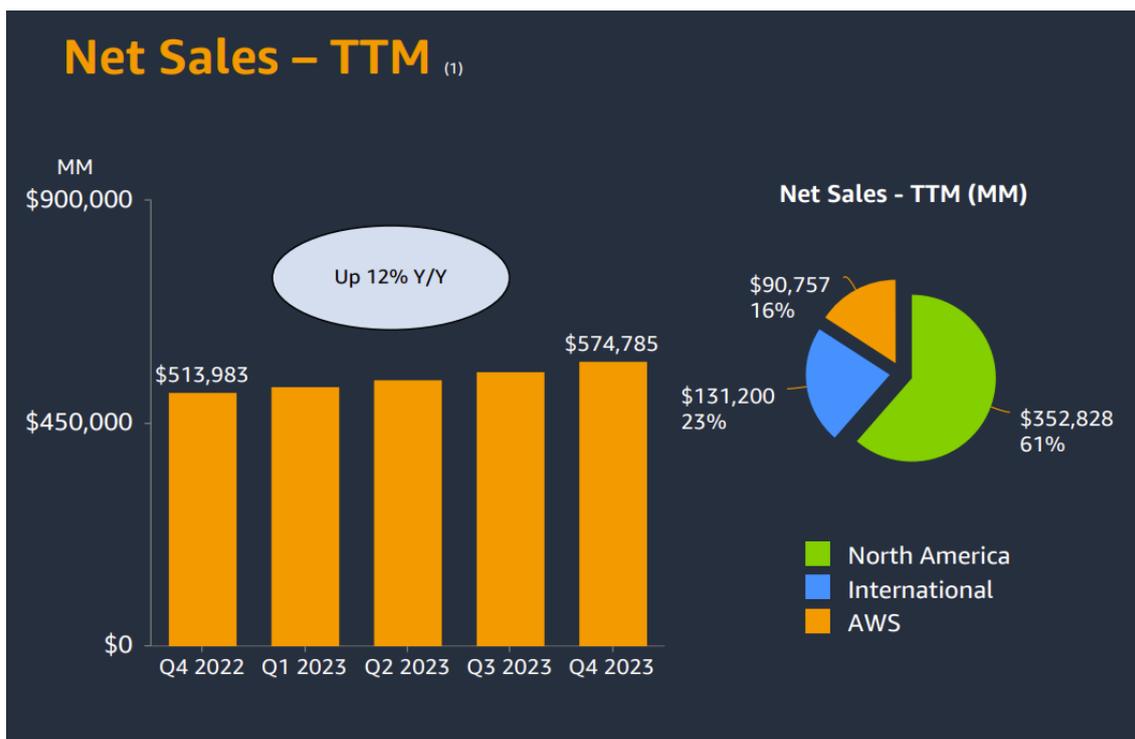


Figura 22: andamento delle vendite dal Q4 2022 al Q4 2023 e loro distribuzione geografica (Amazon Team, 2023).

Nel grafico in figura 23 è rappresentato l'andamento dell'utile operativo per un periodo di un anno (dal quarto trimestre del 2022 al quarto trimestre del 2023). In questo periodo si può notare una crescita consistente: l'utile è infatti aumentato da 2,737 miliardi di dollari nel Q4 2022 fino a 13,209 miliardi di dollari nel Q4 2023. Ciò si traduce in un sorprendente aumento del 383% (Y/Y). Il margine operativo, a sua volta, è passato da  $2,737/513,983=0,53\%$  nel Q4 2022 a  $13,209/574,785=2,30\%$  nel Q4 2023.

Dal grafico si può osservare una crescita costante in ogni trimestre. Si può quindi presumere l'introduzione di fattori che hanno portato ad una gestione più efficiente dei costi, una maggiore ottimizzazione dei processi oppure un incremento delle entrate in segmenti di grande rilevanza per l'azienda (ad esempio AWS, il segmento ecommerce oppure i ricavi pubblicitari).

L'importanza di questo risultato risiede nella capacità dell'azienda di non solo generare ricavi, ma di trasformarli in un significativo miglioramento della redditività, un segno di forte salute finanziaria e operativa. L'aumento netto, costante e accelerato verso il Q4 2023, dimostra che l'azienda è riuscita a consolidare la sua posizione di mercato, ottimizzando le sue operazioni e ampliando il margine operativo in modo considerevole.

Questo trend, soprattutto in un contesto economico globale incerto, rappresenta una testimonianza della resilienza e della forza strategica dell'organizzazione.

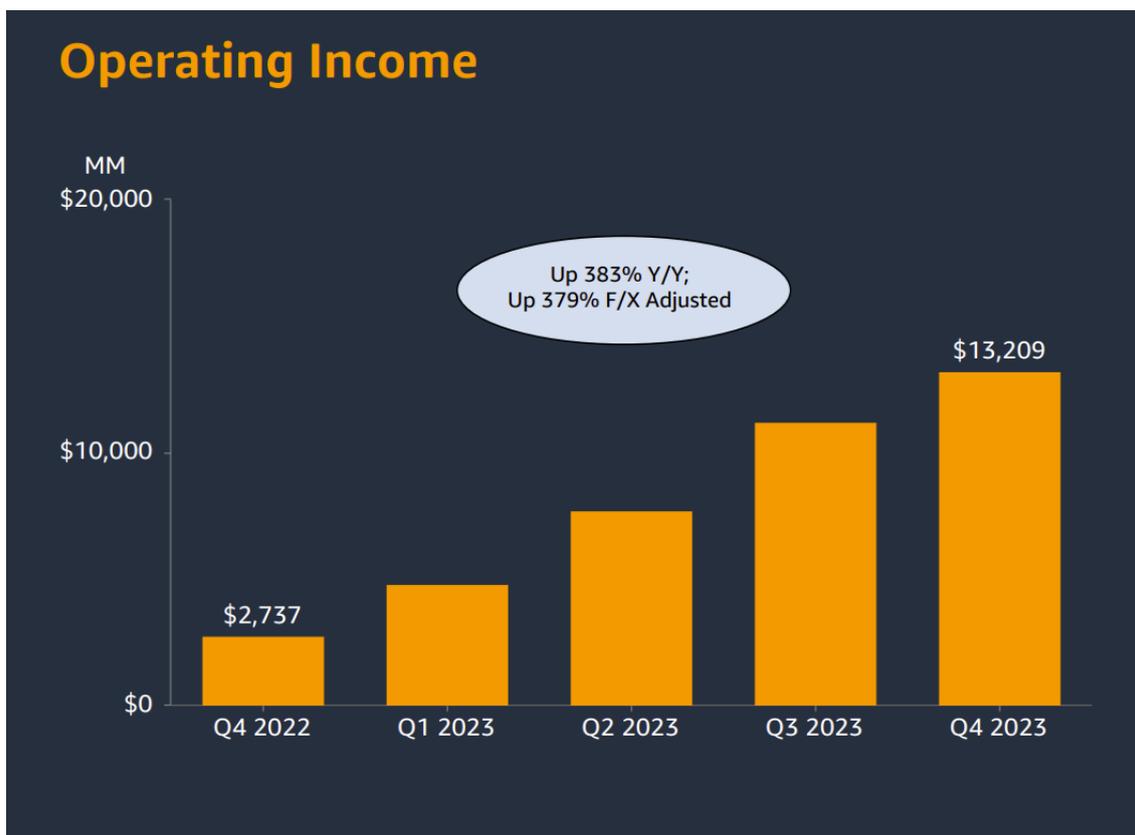


Figura 23: andamento dell'utile operativo dal Q4 2022 al Q4 2023 (Amazon Team, 2023).

Procedendo con un'analisi del reddito netto (figura 24), si può osservare che dal Q4 2022 al Q4 2023 è stata registrata una crescita notevole. Nel Q4 2022 è stata registrata una perdita di 2,722 miliardi di dollari, sinonimo di un momento di difficoltà finanziaria della società; ciò deriva probabilmente da fattori esterni, come il contesto economico globale incerto e svalutazioni di asset, tra cui un impatto significativo derivante dall'investimento in Rivian Automotive.

Trimestre dopo trimestre sono stati registrati dei miglioramenti crescenti, già nel Q1 2023 il reddito è tornato ad essere positivo, fino a raggiungere quota 30,425 miliardi di dollari nel Q4 2023. Una tale crescita in un tempo così ristretto è stata possibile grazie all'implementazione da parte di Amazon di strategie vincenti per fronteggiare le difficoltà precedenti. Degli esempi possono essere: l'espansione nei mercati internazionali ed una più efficace diversificazione delle entrate. Inoltre, nel grafico viene anche riportato come il miglioramento sia stato influenzato anche da variazioni contabili riguardanti

investimenti: nel Q4 2023 Amazon ha registrato un guadagno “pre tasse” di 800 milioni di dollari grazie alla rivalutazione dell’investimento in Rivian Automotiv, rispetto a una perdita di valutazione di 12,7 miliardi di dollari nel Q4 2022.

Questo grafico mostra che Amazon è una società in grado di gestire momenti di crisi adattandosi rapidamente a un mercato ormai sempre più dinamico. L’incremento costante evidenziato non rappresenta soltanto un successo operativo ma anche una prova della fiducia che investitori e consumatori continuano a riporre nell’azienda.

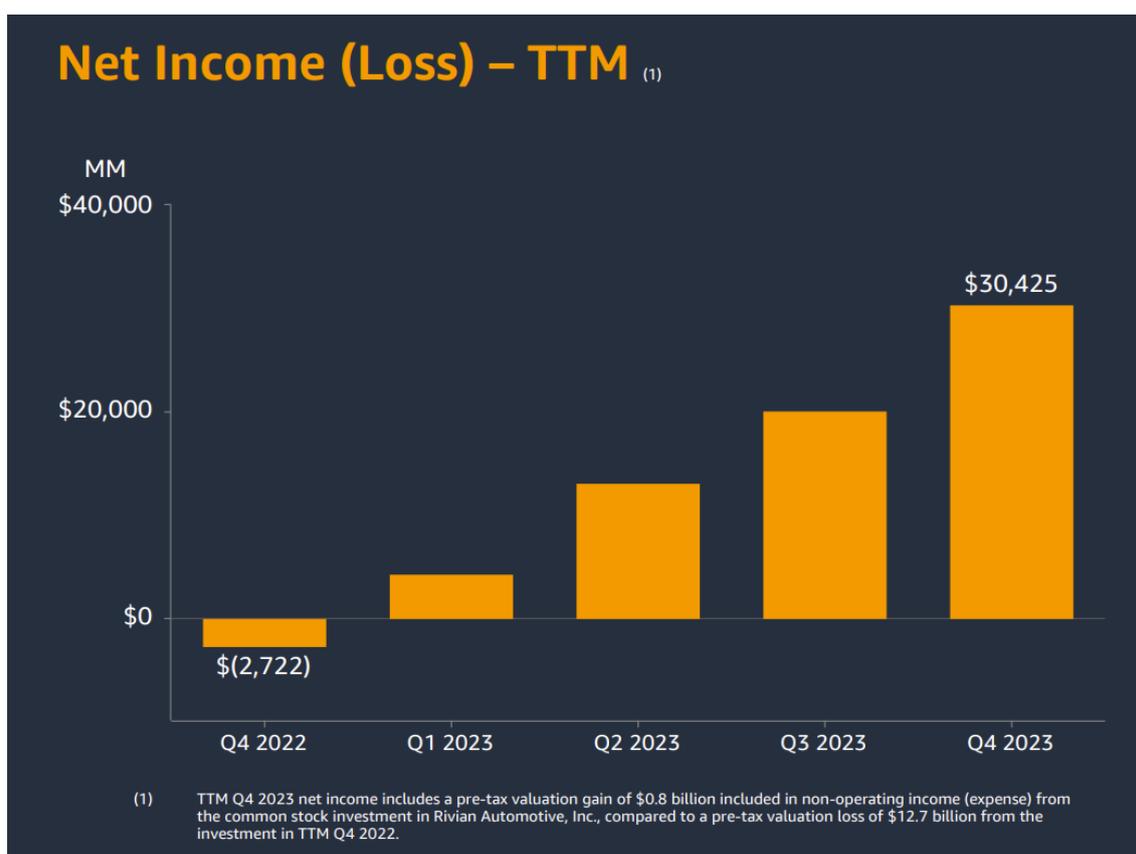


Figura 24: andamento del reddito netto dal Q4 2022 al Q4 2023 (Amazon Team, 2023)

### Dipendenti:

Amazon è uno dei maggiori datori di lavoro privati a livello mondiale. Infatti occupa 1,5 milioni di dipendenti in tutto il mondo. Le sedi dell'azienda sono dislocate in numerose regioni (tra cui Africa, Asia, Europa, America Latina, Medio Oriente e Nord America). Ciò riflette la portata globale della sua attività. Lo sviluppo professionale, l'inclusione e la sicurezza dei dipendenti sono colonne portanti della sua strategia.

## Tecnologie

Amazon è, inoltre, una delle aziende tecnologiche più innovative al mondo. Già dai primi anni 2000 ricorre all'intelligenza artificiale nelle sue attività. Utilizza algoritmi di machine learning per migliorare le raccomandazioni dei prodotti e ottimizzare la gestione dell'inventario. La società ha realizzato un ecosistema tecnologico che ha ridisegnato le aspettative dei consumatori e rivoluzionato interi settori: ecommerce, cloud computing, logistica e assistenza clienti.

Nel grafico (figura 25) è evidenziata la distribuzione percentuale di venditori e brand operanti sulla piattaforma Amazon che fanno uso dell'intelligenza artificiale in varie attività (aggiornato a Gennaio 2024).

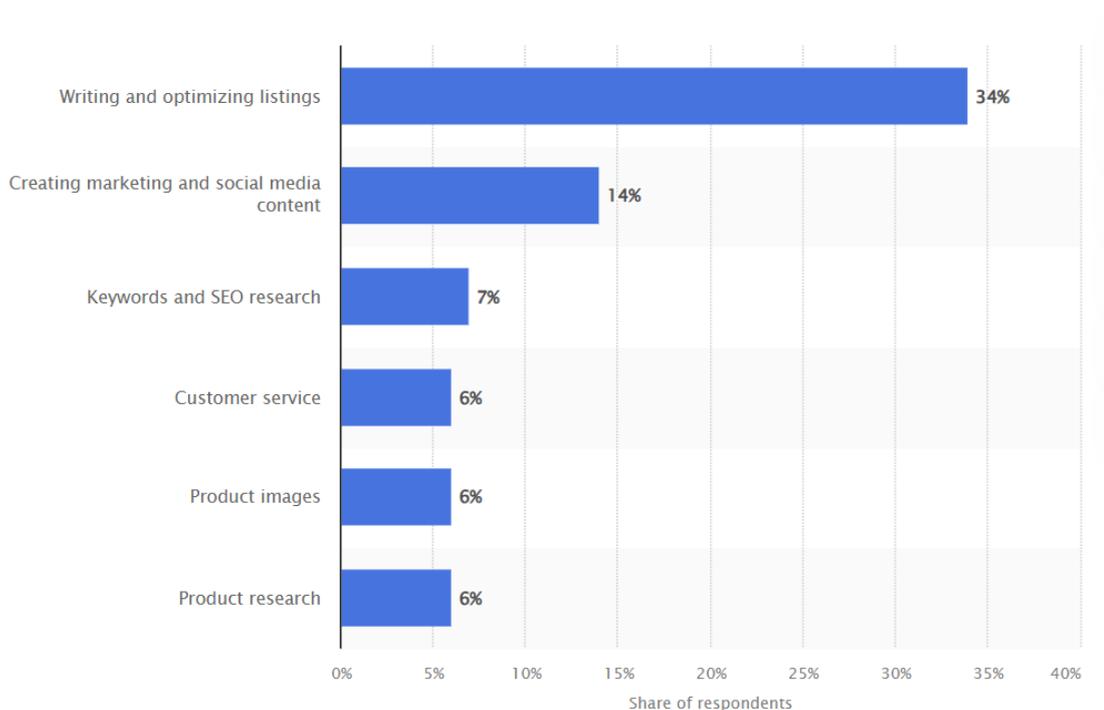


Figura 25: Quota di venditori e marchi su Amazon che utilizzano l'intelligenza artificiale (IA) a gennaio 2024 (Statista, 2024).

Dall'esame dei dati si riscontra che l'IA è prevalentemente utilizzata per la scrittura e l'ottimizzazione delle inserzioni dei prodotti (come indicato dal 34% dei rispondenti). Questo risultato evidenzia l'elevata importanza della presentazione dei prodotti in un contesto competitivo come quello di Amazon, dove descrizioni chiare e pertinenti possono migliorare significativamente la visibilità e le vendite.

Seguono, con il 14%, le attività di creazione di contenuti per il marketing e i social media. Questo dato mostra come i venditori ritengano fondamentale l'uso dell'IA per creare contenuti efficaci ed accattivanti, con conseguente aumento della capacità di attirare, coinvolgere e mantenere la propria clientela.

Una percentuale più bassa, pari al 7%, si registra nella ricerca di parole chiave e SEO (Search Engine Optimization). Pur essendo questa un'area strategica per la rilevabilità dei prodotti nei motori di ricerca e sul marketplace, la percentuale relativamente bassa porta a ritenere che molti venditori ricorrano ancora a strategie SEO tradizionali oppure non sfruttino a pieno le potenzialità dell'intelligenza artificiale in questo ambito.

Seguono, poi, con una percentuale pari al 6% altre applicazioni come: il servizio clienti, le immagini dei prodotti e la ricerca di prodotti. Il ricorso all'IA, in supporto al servizio clienti, per esempio con l'utilizzo di chatbot o assistenti virtuali, sebbene possa migliorare l'efficienza delle relazioni con i clienti, sembra essere ancora meno prioritario rispetto alle attività direttamente connesse alla vendita e al marketing. Allo stesso modo rimane ancora limitato l'uso dell'IA per ottimizzare immagini di prodotto o condurre ricerche sui prodotti, con la prospettiva, però, di una futura crescita all'aumentare dell'accessibilità degli strumenti di IA.

I dati del grafico dimostrano, quindi, che in Amazon l'IA è utilizzata principalmente per migliorare le attività di presentazione e promozione dei prodotti mentre in altri settori come il servizio clienti e la ricerca di prodotto vi sono elevati margini di crescita, tenuto conto che gli strumenti di IA diventeranno più accessibili.

Amazon analizza, tramite algoritmi avanzati di machine learning, una grande mole di dati provenienti dai consumatori, quali abitudini d'acquisto, cronologia di navigazione, recensioni, valutazioni dei prodotti e comportamenti di utenti simili. Tali algoritmi si basano su tecniche come il filtraggio collaborativo e l'elaborazione del contesto e consentono di prevedere quali prodotti saranno di maggiore interesse per un determinato cliente. Secondo un rapporto del MIT Technology Review, le indicazioni personalizzate di Amazon generano fino al 35% del totale delle vendite della piattaforma, dal che si comprende l'importanza di questa tecnologia per il successo dell'azienda (Caballar, Stryker, 2024) (Shopify, 2023).

L'applicazione di questa tecnologia si riscontra nel modulo "Chi ha acquistato questo articolo ha anche acquistato..." oppure nelle sezioni "Consigliati per te", le quali usano i dati dei consumatori per fornire suggerimenti personalizzati. In questo modo l'esperienza del cliente è migliore riducendosi notevolmente il tempo necessario per trovare i prodotti desiderati e, nello stesso tempo, aumentano il valore medio degli ordini e la fidelizzazione, poiché i clienti percepiscono un valore aggiunto nella personalizzazione (Blendee, 2022).

Un altro strumento fondamentale delle tecnologie di IA utilizzate da Amazon è l'elaborazione del linguaggio naturale (NLP). Con essa il sistema riesce a comprendere il significato delle ricerche effettuate dai clienti e fornisce risultati più precisi. Ad esempio, se un cliente cerca "scarpe da corsa in montagna impermeabili", il motore di ricerca analizza le parole chiave, il contesto della frase e la cronologia di navigazione del cliente per offrire risultati mirati che rispecchiano la specificità della richiesta (Caballar, Stryker, 2024) (Amazon Team, 2023).

Amazon, inoltre, incrocia i dati della ricerca con le tendenze del mercato e con le recensioni dei prodotti per mostrare risultati ancora più accurati. L'interpretazione dell'NLP accresce, dunque, l'efficienza della ricerca, la gradevolezza dell'esperienza e il tasso di conversione. Si stima che l'accuratezza della ricerca e la pertinenza dei risultati abbiano migliorato le vendite del 20-30%, grazie alla capacità del sistema di connettere i clienti con i prodotti desiderati in modo rapido ed intuitivo (The Verge, 2023).

In conclusione, Amazon riesce ad eccellere grazie alla capacità di analizzare enormi volumi di dati ogni giorno; la società processa petabyte di dati provenienti dalle sue interazioni con i clienti, dagli acquisti e dalle recensioni. Grazie a questi dati si possono aggiornare gli algoritmi di ricerca, i quali riusciranno sempre a fornire ciò di cui i consumatori hanno bisogno. Ad esempio, se un cliente acquista un set di pentole, il sistema suggerirà anche altri articoli di cucina, migliorando l'esperienza d'acquisto ed aumentando il valore del carrello (Bitter, 2024).

Nel prossimo paragrafo saranno illustrati i modi in cui effettivamente l'impiego dell'IA da parte di Amazon aumenta l'efficienza dei processi.

## **4.2 In che modo l'IA ha contribuito ad aumentare l'efficienza di processi ed operazioni in Amazon e suoi clienti**

L'intelligenza artificiale (IA) ha contribuito ampiamente a migliorare l'efficienza operativa di Amazon riducendo i costi di vari settori dell'azienda. Questi sono alcuni modi con cui l'IA ha contribuito a alla riduzione dei costi:

- **Riduzione degli imballaggi;**
- **Automazione del servizio clienti;**
- **Gestione del magazzino.**

### **Riduzione degli imballaggi**

Amazon ha sviluppato un avanzato modello di IA (il Package Decision Engine) in grado di ottimizzare l'uso di imballaggi, incrementando così sia l'efficienza dell'azienda sia la sua sostenibilità. Grazie a questo modello si riesce a stabilire quale sia il tipo di imballaggio più appropriato per centinaia di milioni di articoli venduti. In questo modo si riesce a tagliare in misura significativa il consumo di materiale e si garantisce che i prodotti rimangano protetti durante le spedizioni. Dal 2015, il Package Decision Engine, insieme ad altre innovazioni di imballaggio, ha evitato l'uso di oltre 2 milioni di tonnellate di materiale da imballaggio, un risultato che sottolinea l'impegno dell'azienda verso pratiche sostenibili (Amazon, 2024).

Tutto ciò è reso possibile da una serie di fasi di raccolta dati effettuate su ogni articolo. Una volta che i prodotti arrivano nei centri di distribuzione Amazon, vengono identificati all'interno di un tunnel di computer vision. Vengono quindi rilevate le dimensioni, eventuali difetti e la presenza o meno di un imballaggio protettivo. Inoltre, il modello utilizza l'elaborazione del linguaggio naturale (NLP) per analizzare i dati testuali associati al prodotto come: nome, descrizione, prezzo e dimensioni. Tramite le recensioni online e i report sul centro resi di Amazon anche i feedback dei clienti vengono integrati (Amazon, 2024). Una volta processate queste informazioni, il sistema assegna un punteggio che indica la miglior opzione di imballaggio per il prodotto (in termini di efficienza e sicurezza). Ad esempio, per un articolo robusto come una coperta, il modello potrebbe suggerire una borsa senza imbottitura, mentre per un set di piatti fragili, potrebbe consigliare una scatola più robusta. Il modello memorizza le decisioni prese, in modo da

migliorare quelle future. L'addestramento del Package Decision Engine ha richiesto l'analisi di milioni di spedizioni campione, considerando sia i prodotti consegnati con successo sia quelli danneggiati. Il modello ha compreso, ad esempio, che gli articoli con parole chiave come "generi alimentari" o "gres" nella descrizione richiedono una maggiore protezione rispetto ai prodotti con termini come "multipack" o "borsa", che indicano una maggiore robustezza intrinseca. Questo apprendimento continuo consente al Package Decision Engine di adattarsi a nuovi tipi di prodotti, mercati regionali e lingue. Il modello è talmente efficace che Amazon ha iniziato a distribuirlo a livello globale. Attualmente è ampiamente utilizzato nei centri di distribuzione in Nord America ed Europa, con un'espansione in corso in paesi come India, Australia e Giappone (Amazon, 2024).

Esso rappresenta, quindi, uno dei modi in cui Amazon si sta servendo dell'IA sia per aumentare l'efficienza delle operazioni logistiche che per raggiungere gli obiettivi legati alla sostenibilità. Grazie al Package Decision Engine le spedizioni saranno: meno costose, più sicure e più sostenibili.

### **Automazione del servizio clienti**

L'uso di chatbot, assistenti virtuali e sistemi di risposta vocale interattiva (IVR), basati sull'intelligenza artificiale, ha mutato notevolmente le modalità con cui le aziende interagiscono con i consumatori.

In questo modo si aumenta la soddisfazione dei clienti e si riducono i costi operativi, ottenendo un vantaggio competitivo in mercati sempre più impegnativi. Con la potenza dei grandi modelli linguistici (LLM) come quelli forniti da Amazon Web Services, le aziende possono implementare soluzioni conversazionali dinamiche e personalizzate, facili da integrare nei contact center. La novità principale è la capacità di sostenere conversazioni sofisticate in un tono naturale. I chatbot basati sull'intelligenza artificiale generativa utilizzano l'NLP per capire l'intento del cliente, mantenere il livello di conversazione e fornire risposte adeguate basate su fonti di conoscenza attendibili. Non solo migliorano la qualità del servizio, ma creano anche un'esperienza più coinvolgente e soddisfacente per il cliente, che ottiene velocemente risposte alle sue richieste. Le soluzioni self-service intelligenti consentono ai clienti di risolvere i problemi 24 ore su 24, 7 giorni su 7 tramite canali come: telefono, Web, messaggistica ed e-mail. Questo

approccio riduce drasticamente i tempi di attesa e offre ai clienti il controllo diretto sul processo di risoluzione (Amazon, 2024).

Da una prospettiva aziendale, l'implementazione di assistenti virtuali e chatbot porta vantaggi tangibili. Grazie all'automatizzazione di attività ripetitive e standardizzate, come rispondere a domande frequenti o gestire richieste comuni, le aziende possono ridurre il personale dedicato alle attività di routine. Ciò comporta la possibilità di dedicare un maggior numero di risorse a questioni più complesse che richiedono competenze specifiche e intervento umano, aumentando la produttività (Amazon, 2024).

Quindi, con il supporto di piattaforme avanzate come AWS, le aziende possono trasformare il modo in cui interagiscono con i clienti, ponendo le basi per una crescita sostenibile e competitiva.

### **Gestione del magazzino**

Amazon ha completamente rivoluzionato il settore della logistica e delle spedizioni utilizzando nei suoi magazzini e centri di smistamento tecnologie avanzate, robotica ed intelligenza artificiale. Queste innovazioni hanno comportato una sensibile riduzione del tempo intercorrente tra l'ordine e la spedizione e notevoli miglioramenti nel campo della sostenibilità, dell'efficienza operativa e della puntualità della consegna. Come evidenziato da Jacopo Bruna, membro del team operativo di Amazon Shipping, con l'introduzione di automazione e nuove tecnologie l'azienda ha trasformato i propri centri di smistamento in hub operativi all'avanguardia. Bruna afferma che il processo di innovazione è costante allo scopo di perfezionare le operazioni di prelievo, imballaggio e spedizione mediante l'inserimento di strumenti basati sull'IA (Amazon, 2024).

I tempi di evasione degli ordini si riducono notevolmente grazie agli algoritmi di raccolta. Questi, dopo un'analisi degli articoli in magazzino, danno la precedenza a quelli da consegnare con urgenza. Inoltre, il sistema è piuttosto flessibile, per cui nei periodi con minor flusso di spedizioni, elabora anche gli ordini con tempi di consegna meno stringenti e raggruppa le spedizioni con luoghi di destinazione simile. In questo modo i tempi di inattività sono ridotti al minimo e sono razionalizzati i carichi degli automezzi, con conseguente riduzione dei costi operativi e degli sprechi. Il sistema robotico Sequoia, che sfrutta tecnologie di computer vision e apprendimento automatico per analizzare i dati e

determinare il miglior posizionamento degli articoli all'interno dei pacchi, ha ridotto i tempi complessivi di elaborazione degli ordini nei centri di smistamento del 25% (Amazon, 2024).

Dunque, robotica e IA svolgono un ruolo di massima importanza nella gestione delle merci all'interno dei magazzini di Amazon. Identificano i percorsi più efficienti per il prelievo e l'organizzazione degli articoli, riducendo al minimo i tempi di elaborazione. Infine, consentono di ottimizzare i carichi dei veicoli per la consegna, massimizzando l'utilizzo dello spazio e riducendone il numero.

L'impiego delle tecnologie non ha contribuito esclusivamente allo sviluppo di Amazon, ma anche a quello dei suoi clienti. Due casi che testimoniano ciò sono:

- **Il caso Marketing Evolution**
- **Il caso Booking**

### **Il caso Marketing Evolution**

Marketing Evolution è un esempio notevole di come l'intelligenza artificiale generativa possa trasformare radicalmente il settore del marketing, pur considerando le complesse sfide legate alla privacy dei dati e all'attribuzione delle campagne pubblicitarie. Fondata nel 2000, l'azienda ha dovuto affrontare la necessità di fornire una soluzione di attribuzione multi-touch innovativa in grado di combinare dati da oltre 300 fonti e colmare le lacune nei percorsi dei clienti, un compito complicato dalle normative sulla privacy in continua evoluzione. Grazie alla partnership con Amazon Web Services (AWS), Marketing Evolution ha sviluppato una piattaforma scalabile e automatizzata che ha migliorato significativamente l'efficienza operativa e ridotto i costi (AWS case studies, 2023).

Prima che fossero adottate tecnologie avanzate, i processi di attribuzione erano molto laboriosi, richiedevano molto tempo e coinvolgevano molte risorse (sino a più di 20 persone) per sviluppare e addestrare modelli di apprendimento automatico. Preparare un singolo set di dati richiedeva giorni interi. Con l'adozione di AWS Glue, un servizio di integrazione dati "serverless", l'azienda ha automatizzato completamente questi processi, riducendo i tempi di preparazione dei dati a poche ore e consentendo così di fornire report settimanali ai clienti. Grazie a tale investimento, i costi di elaborazione e i costi della

manodopera sono diminuiti rispettivamente dell'85% e del 40% tra il 2020 e il 2022. Ciò ha permesso di aumentare il focus su attività strategiche come l'R&D (AWS case studies, 2023).

La soluzione, denominata Mevo, utilizza modelli predittivi basati sull'intelligenza artificiale generativa per ricostruire percorsi rappresentativi dei consumatori. Ad esempio, aggregando dati come le impressioni TV, suddivise per aree geografiche (usando i codici postali), Mevo può verificare l'influenza delle campagne pubblicitarie sulle decisioni d'acquisto dei consumatori. Tramite queste analisi, fornisce ai clienti informazioni dettagliate per ottimizzare i loro investimenti pubblicitari. Un cliente del settore sanitario ha sfruttato i dati di Mevo per spostare il proprio budget pubblicitario verso una piattaforma di social media più efficace, ottenendo un aumento del 20% delle conversioni e una riduzione del 15% del costo per acquisizione. AWS ha inoltre consentito a Marketing Evolution di lanciare una versione semplificata di Mevo, denominata Mevo Link, rivolta al segmento di mercato medio. Rilasciata su AWS Marketplace a luglio 2022, questa versione ha reso la soluzione accessibile a una gamma più ampia di clienti, aumentando ulteriormente il ritorno sull'investimento di Marketing Evolution e rafforzando la fidelizzazione dei clienti, che è salita dal 44% nel 2021 al 100% nel 2023. La transizione ad un modello serverless ha incrementato la scalabilità e l'efficienza della piattaforma. Nel 2023, Marketing Evolution ha iniziato la transizione ad Amazon EMR Serverless, eliminando la necessità di gestire infrastrutture fisiche o virtuali, riducendo ulteriormente i costi e aumentando la flessibilità operativa (AWS case studies, 2023).

Il caso Marketing Evolution ha quindi dimostrato che l'integrazione di tecnologie avanzate consente sia di superare le sfide operative sia di creare valore tangibile per i clienti. Sfruttando AWS Glue e altre soluzioni serverless, Marketing Evolution è riuscita nell'impresa di trasformare una complessa attività come l'attribuzione pubblicitaria in un vantaggio competitivo.

## **Il caso Booking**

Booking.com è una startup fondata nei Paesi Bassi nel 1996; grazie all'importanza attribuita all'innovazione è attualmente una delle principali agenzie viaggi digitali. L'azienda è stata capace di trasformare questo business, modificando radicalmente il modo in cui le persone organizzano i propri viaggi. Tutto ciò grazie a strumenti avanzati

come intelligenza artificiale generativa e altre tecnologie basate sul cloud offerte da Amazon Web Services. Booking ha creato un sistema che migliora sia l'efficienza sia la personalizzazione, diventando così uno dei leader del settore dei viaggi online (AWS case studies, 2023).

Booking si propone da sempre di “rendere il mondo alla portata di tutti”. L'azienda, tramite gli strumenti già citati, è capace di convertire le vaghe indicazioni dei clienti come: “vacanza invernale rilassante” o “città da esplorare in Europa” in consigli altamente personalizzati che spesso superano anche le aspettative degli utenti. La collaborazione con AWS è stata fondamentale per rendere possibile tutto questo. Ad esempio, tramite “Amazon SageMaker”, Booking può creare e addestrare modelli di machine learning che consentono di suggerire ai clienti mete di viaggio sempre più conformi alle loro richieste. Amazon Bedrock è un altro strumento di AWS utilizzato da Booking per garantire la sicurezza dei dati relativi ai clienti, in quanto consente di garantire che le informazioni sensibili rimangano all'interno del sistema protetto di Booking.com (AWS case studies, 2023).

Quindi, come sottolineato da Thomas Davey (vicepresidente della divisione Big Data e Machine Learning), l'utilizzo di soluzioni IA combinato con gli strumenti di AWS consente a Booking di elaborare esperienze su misura rendendo ogni viaggio unico e indimenticabile (AWS case studies, 2023).

### **4.3 Innovazione tecnologica e strategie di business**

L'integrazione dell'intelligenza artificiale nelle operazioni di Amazon e dei suoi clienti dimostra che le strategie di business classiche: leadership di costo, differenziazione, e focus danno risultati concreti soltanto attraverso l'innovazione tecnologica. Ciascuna delle strategie sopra menzionate comporta, attraverso l'implementazione dell'IA, un risparmio sui costi e la generazione di un valore maggiore. Il posizionamento competitivo di Amazon ne risulta quindi rafforzato.

Ad esempio, la leadership di costo è chiaramente visibile nell'uso di soluzioni automatizzate per ridurre i costi operativi. Amazon, tramite il Package Decision Engine, ha rivoluzionato la gestione degli imballaggi rendendola più sostenibile ed economica. Come visto in precedenza, grazie all'analisi di grandi dataset mediante algoritmi di intelligenza artificiale, Amazon è riuscita, a partire dal 2015, ad evitare l'uso di oltre due

milioni di tonnellate di materiale da imballaggio. L'azienda ha conseguito la leadership di costo anche nel campo del servizio clienti, facendo ricorso a chatbot e assistenti virtuali basati sull'IA generativa. Secondo stime della società ciò ha ridotto del 40% i costi legati alla gestione delle richieste frequenti. Grazie a questi strumenti Amazon e le aziende che usano i suoi servizi (come Booking e Marketing Evolution) sono in grado di gestire milioni di relazioni quotidiane con massima efficienza e precisione. Ciò, inoltre dà la possibilità ai dipendenti di dedicarsi ad attività più complesse con un conseguente aumento della produttività generale.

L'IA è inoltre fortemente utilizzata come strategia di differenziazione per fornire alla clientela soluzioni uniche con un elevato livello di personalizzazione. Per esempio Amazon, grazie agli algoritmi utilizzati fornisce ai clienti consigli e indicazioni mirati corrispondenti alle loro esigenze e tali da rendere più gratificante l'esperienza d'acquisto. Anche Booking, tramite l'IA generativa, trasforma le richieste generiche formulate dagli utenti in suggerimenti specifici e calzanti. Queste abilità collocano le due aziende citate in posizione dominante nei rispettivi settori, nei quali difficilmente gli altri operatori sono in grado di fornire le stesse prestazioni. Inoltre, l'integrazione con tecnologie come Amazon SageMaker e Amazon Bedrock garantisce la sicurezza e il costante aggiornamento dei sistemi di IA, rendendo più solido il vantaggio competitivo.

Nel campo dell'e-commerce, Amazon e eBay fanno un uso radicalmente diverso delle tecnologie e dell'intelligenza artificiale, che riflette i loro distinti modelli di business e priorità strategiche. Amazon utilizza l'IA in tutte le sue attività operative: ottimizzazione della logistica, personalizzazione dell'esperienza del cliente e sviluppo di nuovi servizi.

EBay vi ricorre, invece, per migliorare le funzionalità del suo marketplace, supportando venditori e acquirenti senza entrare nella gestione diretta delle operazioni logistiche. Di recente l'azienda sta utilizzando l'IA generativa per assistere i venditori nella creazione di inserzioni: la tecnologia automatizza la generazione di descrizioni, migliora la qualità delle immagini e rimuove gli sfondi, semplificando il processo di vendita. Tuttavia, l'approccio di eBay si limita a migliorare le interazioni all'interno della piattaforma. L'IA viene anche utilizzata per garantire la sicurezza delle transazioni e gestire le controversie tra acquirenti e venditori, ma l'implementazione resta meno complessa e pervasiva rispetto a quella di Amazon.

Per quanto riguarda la strategia di focus, mediante il Package Decision Engine, Amazon è in grado di realizzare imballaggi ecologici che rispettino le normative e le esigenze locali. Questa strategia di logistica mirata si integra con metodi di marketing localizzati, resi possibili da strumenti che analizzano i comportamenti degli individui a livello regionale. I dati reperiti consentono di personalizzare le campagne pubblicitarie: ad esempio, in un mercato urbano vengono promossi prodotti tecnologici, mentre in contesti rurali si dà risalto ad articoli di utilità domestica o agricola. Questa sinergia tra logistica e marketing crea un sistema che non solo soddisfa le esigenze specifiche dei mercati locali, ma rafforza anche la competitività di Amazon, ottimizzando operazioni e strategie su scala locale.

Nel complesso l'IA non solo costituisce un valido supporto alle strategie di business classiche ma ne aumenta sensibilmente l'efficacia. Ne sono un esempio Amazon e le altre aziende che vi fanno ricorso dimostrando che attraverso l'automazione, la personalizzazione e l'aumento dell'efficienza è possibile non solo affrontare le sfide moderne ma trasformarle in opportunità per accrescere il proprio vantaggio in un mercato sempre più competitivo.

## CAPITOLO 5: CONCLUSIONE

In questa tesi, è stato esaminato il panorama dinamico ed in continua evoluzione dell'intelligenza artificiale nell'ambito delle strategie aziendali. Analizzando le tre strategie di business, leadership di costo, differenziazione e focus nel contesto dell'IA, la ricerca ha evidenziato come le aziende siano in grado di sfruttare questa tecnologia trasformativa per conseguire un vantaggio competitivo sostenibile. Lo studio del settore è stato effettuato utilizzando il modello delle 5 forze di Porter e l'analisi PESTEL e ha mostrato un'ambiente competitivo caratterizzato da elevate barriere all'ingresso, un forte potere contrattuale dei fornitori di dati, una crescente domanda di soluzioni personalizzate e la minaccia di tecnologie disruptive. L'analisi PESTEL ha evidenziato come i fattori macroeconomici quali l'investimento governativo in R&D e le preoccupazioni etiche e sociali abbiano una notevole influenza e come sia importante un quadro normativo chiaro.

Lo studio dell'esperienza di Amazon ha mostrato come un'azienda leader possa utilizzare l'IA in ogni fase del proprio modello di business conseguendo notevoli vantaggi in termini di efficienza operativa, riduzione dei costi e creazione di valore per la clientela.

Si è dimostrato che tramite l'IA Amazon ha automatizzato i propri processi logistici, personalizzato le esperienze di acquisto e adeguato le proprie strategie di marketing alle esigenze locali.

La ricerca ha evidenziato che l'IA non è solo una tecnologia innovativa ma uno strumento per la trasformazione aziendale.

E' apparso evidente che le aziende che saranno capaci di inserire in modo strategico l'IA nei propri processi e modelli di business potranno:

- ottimizzare l'efficienza operativa: automatizzando compiti ripetitivi, migliorando la gestione delle risorse e riducendo gli sprechi;
- creare valore aggiunto: offrendo soluzioni personalizzate, anticipando le esigenze dei clienti e sviluppando nuovi prodotti e servizi;
- rafforzare la posizione competitiva: differenziandosi dalla concorrenza e conquistando nuovi mercati.

Tuttavia, l'utilizzo dell'IA non è privo di difficoltà implicando il confronto con questioni etiche e sociali e con l'esigenza di garantire la sicurezza dei dati. Richiede inoltre la capacità di gestire il cambiamento organizzativo nonché investimenti in formazione e sviluppo delle competenze.

Questo lavoro ha cercato di dare un contributo alla comprensione delle strategie di business applicabili al settore dell'IA e di offrire spunti utili alle aziende che intendono sfruttare il potenziale di questa tecnologia.

Di notevole interesse appare anche lo studio dell'impatto che essa potrebbe avere in settori come la sanità, la finanza e l'energia, così come l'analisi dell'evoluzione delle strategie di business in risposta all'emergere di nuove forme di IA.

## BIBLIOGRAFIA

**Alawamleh, M., Shammash, N., Alawamleh, K. and Bani Ismail, L. (2024)** ‘Examining the limitations of AI in business and the need for human insights using Interpretive Structural Modelling’, *Journal of Open Innovation: Technology, Market, and Complexity*, 10, p. 100338.

**Batra, G., Jacobson, Z., Madhav, S., Queirolo, A. and Santhanam, N. (2018)** *Artificial-intelligence hardware: New opportunities for semiconductor companies*. McKinsey & Company

**Caroli, M. (2021)** ‘Economia e gestione sostenibile delle imprese’, McGraw Hill

**Costa Climent, R., Haftor, D. M. and Staniewski, M. W. (2024)** ‘AI-enabled business models for competitive advantage’, *Journal of Innovation & Knowledge*, 9, p. 100532.

**Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L. and Zempel, R. (2023)** *The economic potential of generative AI: The next productivity frontier*. McKinsey & Company

**Clark, K.B. and Fujimoto, T. (1991)** *Product development performance: Strategy, organization, and management in the world auto industry*. Boston: Harvard Business School Press

**Jorzik, P., Klein, S. P., Kanbach, D. K. and Kraus, S. (2024)** ‘AI-driven business model innovation: A systematic review and research agenda’, *Journal of Business Research*, 182, p. 114764.

**Khan, M. S., Shoaib, A. and Arledge, E. (2024)** ‘How to promote AI in the US federal government: Insights from policy process frameworks’, *Government Information Quarterly*, 41, p. 101908.

**Liang, X., Luo, Y., Shao, X. and Shi, X. (2022)** ‘Managing complementors in innovation ecosystems: A typology for generic strategies’, *Industrial Management & Data Systems*, 122(9), pp. 2072–2090.

**Porter, M. (1985)** *Competitive advantage: Creating and sustaining superior performance*. New York: Free Press

**Unione Europea.** Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 relativo all'intelligenza artificiale. Gazzetta ufficiale dell'Unione Europea, vol. L1689, 12 luglio 2024, pp. 1-50.

**Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M. and Fuso Nerini, F.** (2020) 'The role of artificial intelligence in achieving the Sustainable Development Goals', *Nature Communications*, 11, p. 233.

## **SITOGRAFIA**

**AI Tech Park** (2023) 'The Future Potential for Blockchain, AI, and Quantum Computing', accessibile a: <https://ai-techpark.com/blockchain-ai-and-qi/>

**Angelo Alù** (2021) 'Network Digital 360', accessibile a: <https://www.ai4business.it/intelligenza-artificiale/amazon-e-lintelligenza-artificiale-un-rapporto-di-luci-e-ombre/>

**Amazon Team** (2023) 'Amazon's approach to providing customers low prices every day', accessibile a: <https://www.aboutamazon.com/news/retail/amazon-pricing>

**Amazon Team** (2024) 'Amazon ai raggi X: l'intelligenza artificiale aiuta ma i profitti arrivano dal cloud', accessibile a: <https://www.ilsole24ore.com/art/amazon-gioca-partita-dell-intelligenza-artificiale-e-rilancia-sfida-cloud-AF862ccC>

**Amazon Team** (2024) 'Amazon annuncia di aver investito oltre 700 milioni di euro nella robotica e nelle tecnologie basate sull'intelligenza artificiale in tutta Europa', accessibile a: <https://www.aboutamazon.it/notizie/innovazioni/amazon-annuncia-di-aver-investito-oltre-700-milioni-di-euro-nella-robotica-e-nelle-tecnologie-basate-sullintelligenza-artificiale-in-tutta-europa>

**Amazon Team** (2024) 'Amazon reports', accessibile a: <https://sustainability.aboutamazon.com/reports?sort=latest&offset=1&topic=supply-chain>

**Amazon Team** (2024) 'Come Amazon utilizza l'intelligenza artificiale per ridurre gli imballaggi', accessibile a: <https://www.aboutamazon.it/notizie/sostenibilita/come-amazon-utilizza-lintelligenza-artificiale-per-ridurre-gli-imballaggi>

**Anthropic** (2024) ‘Pricing’, accessibile a: <https://www.anthropic.com/pricing>

**AWS** (2023) ‘Aumentare il ritorno sull'investimento dei clienti utilizzando l'IA generativa su AWS con Marketing Evolution’, accessibile a: <https://aws.amazon.com/it/solutions/case-studies/marketing-evolution-case-study/>

**AWS** (2023) ‘Esploratore di casi studio IA’, accessibile a: [https://aws.amazon.com/it/ai/use-cases/?use-cases.sort=by=item.additionalFields.priority&use-cases.sort-order=asc&awsf.use-case-area=all&awsf.industry=all&awsf.business-function=all&awsf.business-outcome=all&awsf.type=\\*all](https://aws.amazon.com/it/ai/use-cases/?use-cases.sort=by=item.additionalFields.priority&use-cases.sort-order=asc&awsf.use-case-area=all&awsf.industry=all&awsf.business-function=all&awsf.business-outcome=all&awsf.type=*all)

**AWS** (2023) ‘Migliora l'esperienza dei clienti con chatbot e assistenti virtuali’, accessibile a: <https://aws.amazon.com/it/ai/generative-ai/use-cases/chatbots-and-virtual-assistants/>

**AWS** (2024) ‘Booking.com aiuta i clienti a scoprire un nuovo modo di viaggiare grazie ad AWS e all'IA generativa’, accessibile a: <https://aws.amazon.com/it/solutions/case-studies/booking-case-study/>

**AWS** (2024) ‘Sviluppa e scala su AWS la prossima ondata di innovazioni nel campo dell'IA’, accessibile a: <https://aws.amazon.com/it/ai/>

**Matt Banholzer, Dorian Gartner, Laura LaBerge** (2023) ‘AI unlock: From investing to innovation to reaping its rewards’, accessibile a: <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/next-in-growth/ai-unlock-from-investing-to-innovation-to-reaping-its-rewards>

**Elia Berteletti, Kimberly Borden** (2024) ‘Using AI to supercharge R&D: Takeaways from the R&D Leaders Forum’, accessibile a: <https://www.mckinsey.com/capabilities/operations/our-insights/operations-blog/using-ai-to-supercharge-r-and-d-takeaways-from-the-r-and-d-leaders-forum>

**Mike Black** (2023) ‘Amazon and Walmart Solidify Their Price Leadership Heading Into the Holidays’, accessibile a: <https://www.profitero.com/press-releases/amazon-and-walmart-solidify-their-price-leadership-heading-into-the-holidays>

**Blendee** (2023) ‘Product Recommendation: sulla scia di Amazon’, accessibile a: <https://blog.blendee.com/product-recommendation-sulla-scia-di-amazon/>

**BusinessInsider** (2024) ‘How Amazon s using AI for internet recommendations’, accessibile a: <https://www.businessinsider.com/how-amazon-is-using-ai-from-rufus-to-movie-recommendations-2024-7>

**Konstantinos Byron** (2024) ‘DeepL Translate: How It Beats Google Translate and other competitors’, accessibile a: <https://theaitrack.com/deepl-translate-review/>

**Rina Caballar, Cole Stryker** (2024) ‘Cos'è un motore di raccomandazione?’, accessibile: <https://www.ibm.com/it-it/think/topics/recommendation-engine>

**Anna Cataleta, Andrea Grillo** (2024) ‘Stati Uniti, tutte le leggi sull'intelligenza artificiale: il quadro’, accessibile a: <https://www.agendadigitale.eu/cultura-digitale/leggi-sullintelligenza-artificiale-ecco-la-complessa-roadmap-usa/>

**Glenn Chapman** (2024) ‘AI bubble or 'revolution'? OpenAI's big payday fuels debate’, accessibile a: <https://japantoday.com/category/tech/ai-bubble-or-%27revolution%27-openai%27s-big-payday-fuels-debate>

**DataDeep** ‘Cos'è la model compression nei modelli di deep learning?’, accessibile a: <https://datadeep.it/2023/03/22/cose-la-model-compression-nei-modelli-di-deep-learning/>

**DeepL** (2024) ‘Unbeatable translation quality for global business’, accessibile a: <https://www.deepl.com/en/quality>

**DeepL** (2024) ‘Why DeepL?’, accessibile a: <https://www.deepl.com/it/whydeepl>

**FasterCapital** (2024) ‘Guerre di prezzo i pericoli delle guerre di prezzo nell'ottimizzazione dei prezzi’, accessibile a: <https://fastercapital.com/it/contenuto/Guerre-di-prezzo--i-pericoli-delle-guerre-di-prezzo-nell-ottimizzazione-dei-prezzi.html>

**FasterCapital** (2024) ‘Fase di declino adattamento al cambiamento delle preferenze dei consumatori’, accessibile a: <https://fastercapital.com/it/contenuto/Fase-di-declino--adattamento-al-cambiamento-delle-preferenze-dei-consumatori.html>

**FasterCapital** (2024) ‘Strategia di differenziazione per sfruttare il vantaggio competitivo il potere della strategia di differenziazione’, accessibile a:

<https://fastercapital.com/it/contenuto/Strategia-di-differenziazione-per-sfruttare-il-vantaggio-competitivo--il-potere-della-strategia-di-differenziazione.html>

**Mesh Flinders, Ian Smalley** (2024) ‘What is a graphics processing unit (GPU)?’, accessibile a: <https://www.ibm.com/topics/gpu>

**Forbes Italia** (2023) ‘Amazon sale in Borsa grazie al suo supercomputer AI che sfida Nvidia’, accessibile a: <https://forbes.it/2024/12/04/amazon-sale-in-borsa-grazie-al-supercomputer-ai/>

**Gemini Team Google** (2024) ‘Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context’

**Ghemawat, P.** (2002) ‘Competition and Business Strategy in Historical Perspective’, *Business History Review*, 76(1), pp. 37–74. Accessibile a: <https://doi.org/10.2307/4127751>.

**Kate Gibson** (2024) ‘Fake and graphic images of Taylor Swift started with AI challenge’, accessibile a: <https://www.cbsnews.com/news/taylor-swift-artificial-intelligence-ai-4chan/>

**James Gong, Harry Qu, Hunter Dorwart** (2024) ‘AI Governance in China: Strategies, Initiatives, and Key Considerations’ accessibile a: <https://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations>

**Michael Goodwin** (2024) ‘What is an API?’, accessibile a: <https://www.ibm.com/topics/api>

**Google Deepmind** (2024) ‘Gemini Ultra’, accessibile a: <https://deepmind.google/technologies/gemini/ultra/>

**HDBlog** (2023) ‘Amazon Prime, 200 milioni di abbonati’, accessibile a: <https://www.hdblog.it/amazon/articoli/n536788/amazon-prime-200-milioni-abbonati/>

**IBM** ‘Cos’è l’etichettatura dei dati?’, accessibile a: <https://www.ibm.com/it-it/topics/data-labeling>

**IBM** (2023) ‘Understanding the different types of artificial intelligence’, accessibile a: <https://www.ibm.com/think/topics/artificial-intelligence-types>

**IBM** (2024) ‘Tokens and tokenization’, accessibile a: <https://www.ibm.com/docs/en/watsonx/saas?topic=solutions-tokens>

**IBM** (2024) ‘What is Strong AI?’, accessibile a: <https://www.ibm.com/topics/strong-ai>

**Il Fatto Quotidiano** (2023) ‘Low cost nelle Pmi: la guerra dei prezzi è più dannosa per chi la fa che per la concorrenza’, accessibile a: <https://www.ilfattoquotidiano.it/2023/02/04/low-cost-nelle-pmi-la-guerra-dei-prezzi-e-piu-dannosa-per-chi-la-fa-che-per-la-concorrenza/6981727/>

**Il Sole 24 Ore** (2023) ‘Amazon sbarca nell’intelligenza artificiale e sfida Microsoft e Google’, accessibile a: <https://www.ilsole24ore.com/art/amazon-sbarca-nell-intelligenza-artificiale-e-sfida-microsoft-e-google-AEsskKHD>

**Jasper AI** (2024) ‘Jasper is the AI built for business’, accessibile a: <https://www.jasper.ai/comparison/jasper-vs-chatgpt>

**Amba Kak, Sara Myers West, Meredith Whittaker** (2023) ‘Make no mistake—AI is owned by Big Tech’, accessibile a: <https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/>

**Sanjeev Katariya** (2019) ‘eBay’s Platform is Powered by AI and Fueled by Customer Input’, accessibile a: <https://innovation.ebayinc.com/tech/engineering/ebays-platform-is-powered-by-ai-and-fueled-by-customer-input/>

**Lucy Kelly** (2024) ‘12 Best Data Providers for Training Generative AI Models in 2024’, accessibile a: <https://datarade.ai/company/blog/best-data-providers-for-training-generative-ai-models>

**Will Knight** (2024) ‘A New Chip Cluster Will Make Massive AI Models Possible’, accessibile a: <https://www.wired.com/story/cerebras-chip-cluster-neural-networks-ai/>

**Kommunicate** (2024) ‘GPT-4 Turbo vs. Claude 3 Opus vs. Google Gemini 1.5 Pro – Best Generative AI LLMs Compared’, accessibile a: <https://www.kommunicate.io/blog/gpt4-vs-claude-3-vs-gemini/>

**La Repubblica** (2024) ‘OpenAI e CDP Venture Capital insieme per promuovere lo sviluppo dell'Intelligenza Artificiale in Italia’, accessibile a: [https://finanza.repubblica.it/News/2024/10/02/openai\\_e\\_cdp\\_venture\\_capital\\_insieme\\_per\\_promuovere\\_lo\\_sviluppo\\_dellintelligenza\\_artificiale\\_in\\_italia-73/](https://finanza.repubblica.it/News/2024/10/02/openai_e_cdp_venture_capital_insieme_per_promuovere_lo_sviluppo_dellintelligenza_artificiale_in_italia-73/)

**Kif Leswing** (2024) ‘Nvidia dominates the AI chip market, but there’s more competition than ever’, accessibile a: <https://www.cnn.com/2024/06/02/nvidia-dominates-the-ai-chip-market-but-theres-rising-competition.html>

**Ngor Luong, Margarita Konaev** (2023) ‘In&out of China: Financial support for AI development’, accessibile a: <https://cset.georgetown.edu/article/in-out-of-china-financial-support-for-ai-development/>

**Bertin Martens** (2024) ‘Why artificial intelligence is creating fundamental challenges for competition policy’, accessibile a: <https://www.bruegel.org/policy-brief/why-artificial-intelligence-creating-fundamental-challenges-competition-policy>

**Mozilla** (2024) ‘Creepy.exe: Mozilla Urges Public to Swipe Left on Romantic AI Chatbots Due to Major Privacy Red Flags’, accessibile a: <https://foundation.mozilla.org/en/blog/creepyexe-mozilla-urges-public-to-swipe-left-on-romantic-ai-chatbots-due-to-major-privacy-red-flags/>

**News Article** (2024) ‘Il nuovo programma Europa digitale investe oltre 176 milioni di EUR nelle capacità e nelle tecnologie digitali europee’, accessibile a: <https://digital-strategy.ec.europa.eu/it/news/new-digital-europe-programme-invests-over-eu176-million-european-digital-capacities-and-tech>

**OpenAI** (2024) ‘Pricing’, accessibile a: <https://openai.com/chatgpt/pricing/>

**Kala Philo** (2024) ‘The Role of Quantum Computing in Blockchain’s Future’, accessibile a: <https://zenledger.io/blog/the-role-of-quantum-computing-in-blockchains-future/>

**Precedence Research** (2024) ‘Semiconductor Market Size, Share, and Trends 2024 to 2034’, accessibile a: <https://www.precedenceresearch.com/semiconductor-market>

**Giovanni Rusconi** (2023) ‘Amazon porta l’intelligenza artificiale generativa dentro Alexa’, accessibile a: <https://www.ilsole24ore.com/art/amazon-porta-l-intelligenza-artificiale-generativa-dentro-alexa-AFP1Gew>

**Rytr** (2024) ‘Pricing’, accessibile a: <https://rytr.me/pricing/>

**Josh Schneider, Ian Smalley** (2024) ‘What is Quantum Computing?’, accessibile a: <https://www.ibm.com/topics/quantum-computing>

**Shopify** (2024) ‘In che modo le raccomandazioni di prodotto ecommerce migliorano le vendite’, accessibile a: <https://www.shopify.com/it/blog/raccomandazione-di-prodotto-e-commerce>

**Matthew Spaul** (2023) ‘Generative AI: Boon or bane for the planet’, accessibile a: <https://www.deloitte.com/uk/en/Industries/technology/blogs/generative-ai-and-foundational-models.html>

**Statista** (2024) ‘Vendite nette annuali di Amazon nei principali mercati selezionati dal 2014 al 2023’, accessibile a: <https://www.statista.com/statistics/672782/net-sales-of-amazon-leading-markets/>

**Statista** (2023) ‘Global data creation is about to explode’, accessibile a: <https://globalcio.com/articles/main/five-trends-in-enterprise-data-2023/>

**Statista** (2024) ‘Share of Amazon sellers and brands using artificial intelligence (AI) as of January 2024’, accessibile a: <https://www.statista.com/statistics/1454422/main-tasks-amazon-sellers-use-ai-for/>

**Rebecca Stropoli** (2023) ‘A.I. Is Going to Disrupt the Labor Market. It Doesn’t Have to Destroy It’, accessibile a: <https://www.chicagobooth.edu/review/ai-is-going-disrupt-labor-market-it-doesnt-have-destroy-it>

**Cole Stryker, Jim Holdsworth** (2024) ‘What is NLP?’, accessibile a: <https://www.ibm.com/topics/natural-language-processing>

**Cole Stryker, Eda Kavlakoglu** (2024) ‘what is AI?’, accessibile a: <https://www.ibm.com/topics/artificial-intelligence>

**Stephanie Susnjara, Ian Smalley** (2024) ‘Cos’è il cloud computing?’, accessibile a: <https://www.ibm.com/it-it/topics/cloud-computing>

**Systran** (2024) ‘Pricing’, accessibile a: <https://www.systransoft.com/pricing/translate-pro/>

**Petroc Taylor** (2023) ‘Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025’, accessibile a: <https://www.statista.com/statistics/871513/worldwide-data-created/>

**The Korea Times** (2024) ‘Korea to invest \$527 mil. to integrate AI into all sectors of society’, accessibile a: [https://www.koreatimes.co.kr/www/tech/2024/06/129\\_372092.html](https://www.koreatimes.co.kr/www/tech/2024/06/129_372092.html)

**TheVerge** (2023) ‘Amazon is stuffing generative AI into its shopping experience’, accessibile a: <https://www.theverge.com/2024/9/19/24249046/amazon-generative-ai-tools-personalized-shopping-recommendations>