The background of the slide is an abstract geometric pattern. It consists of numerous thin, colorful sticks (yellow, green, blue, and red) connected by small, yellow, three-pronged connectors. These sticks and connectors form a complex, interconnected network of triangles and other polygons, creating a sense of depth and structure. The overall effect is reminiscent of a molecular model or a complex architectural framework.

Dal vantaggio competitivo alla trasformazione strutturale: l'impatto dell'AI sui modelli aziendali

Convention FME 2025

Gianluca Dettori
Chairman, Partner
Primo Capital

gianluca.dettori@primo.vc

Piacere!

- Imprenditore seriale diventato VC
- 30 anni nel settore tecnologico (Olivetti, Vitaminic, Primo Capital)
- Kauffman Fellow
- Fondatore di Primo Capital (500m AUM), società di venture capital e private equity nel settore tecnologico
- Investiamo in software, digitale/AI, spazio, clima and healthcare, 30 people a Milano
- Ex batterista e surfista



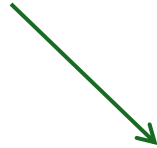


Cos'è l'AI?

L'artificial intelligence (AI) è un campo multidisciplinare tra scienza e ingegneria il cui obiettivo è creare macchine intelligenti

1956, Dartmouth Conference

John McCarthy Dartmouth Workshop Summer Project on Artificial Intelligence



Una tecnologia che è evoluta nel tempo

Reti neurali

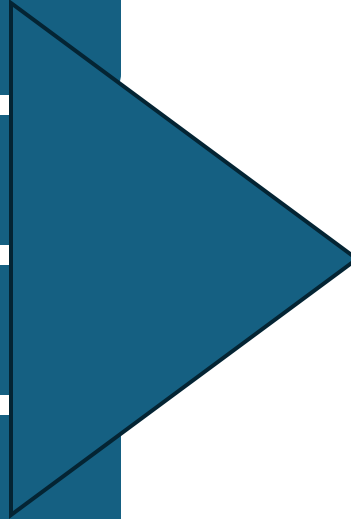
Sistemi esperti (narrow AI)

Machine learning

Deep learning

Tecnologia dei vettori

LLM (ChatGPT)



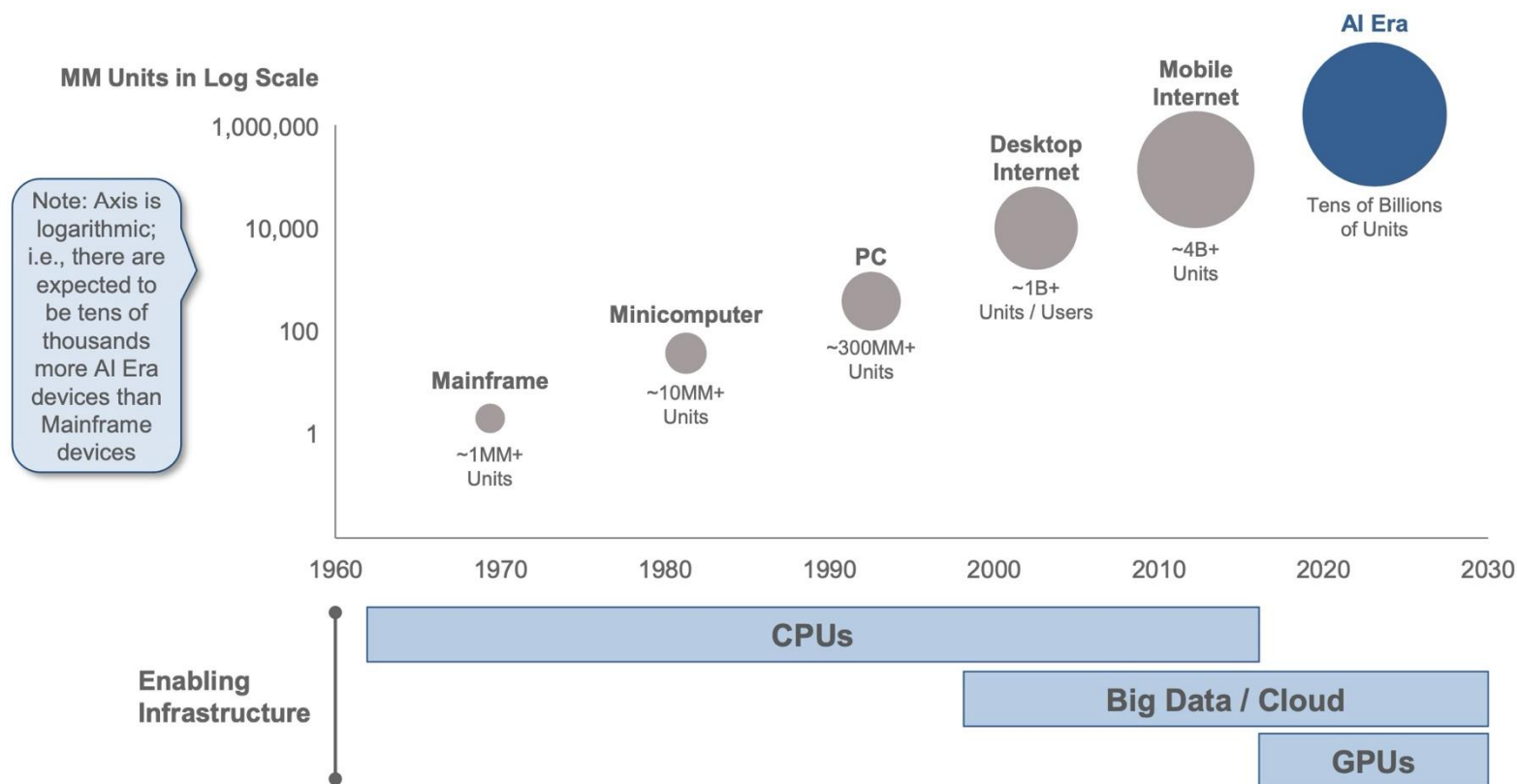
- Copilots
- Agenti intelligenti
- Physical AI
- Umanoidi
- ...

CC = Computing cycles:
input, processing,
output, and storage

Una nuova era
informatica

Mainframe>AI:
dal milione alle
decine di miliardi
di CC

Computing Cycles Over Time – 1960s-2020s, per Morgan Stanley

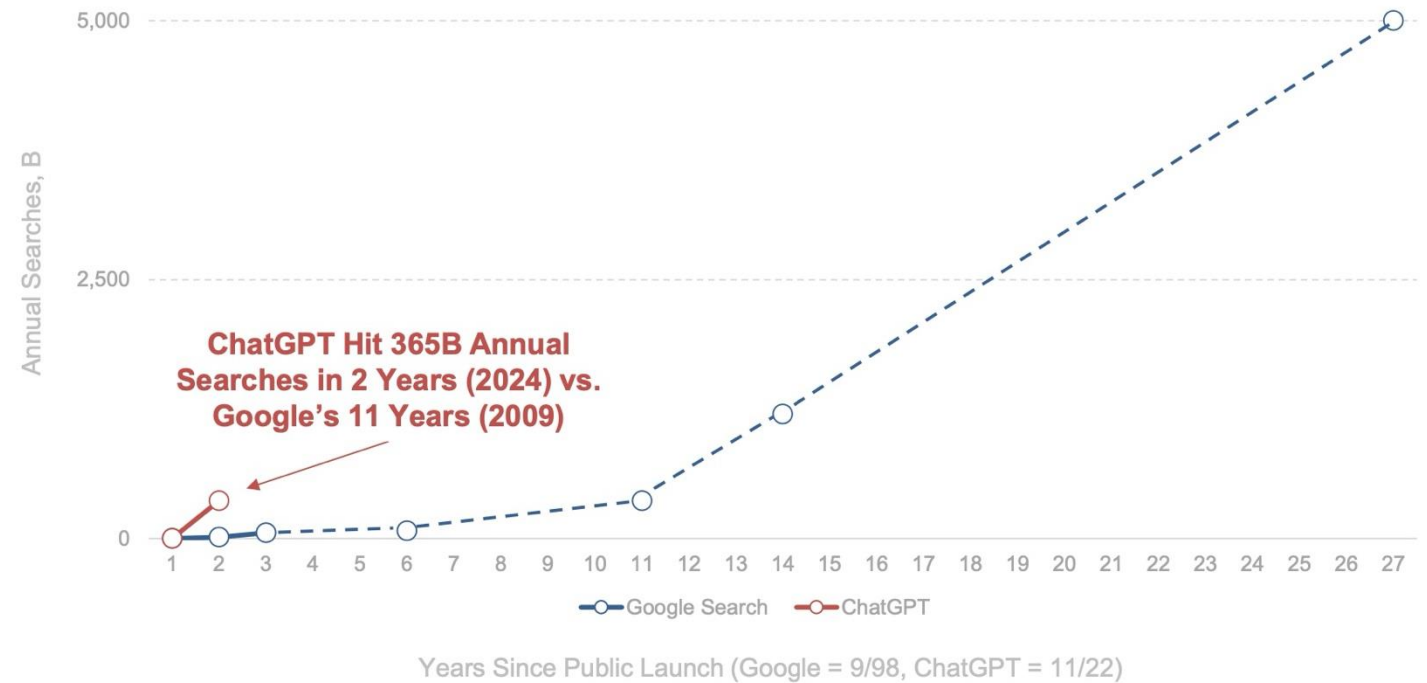




Le nuove fabbriche dei dati: Amazon in Indiana, 5 milioni di mq, 2.2 gigawatt di elettricità (=1.6 milioni di case)

ChatGPT
adoption =
5.5x di Google
Search

Annual Searches by Year (B) Since Public Launches of Google & ChatGPT – 1998-2025,
per Google & OpenAI

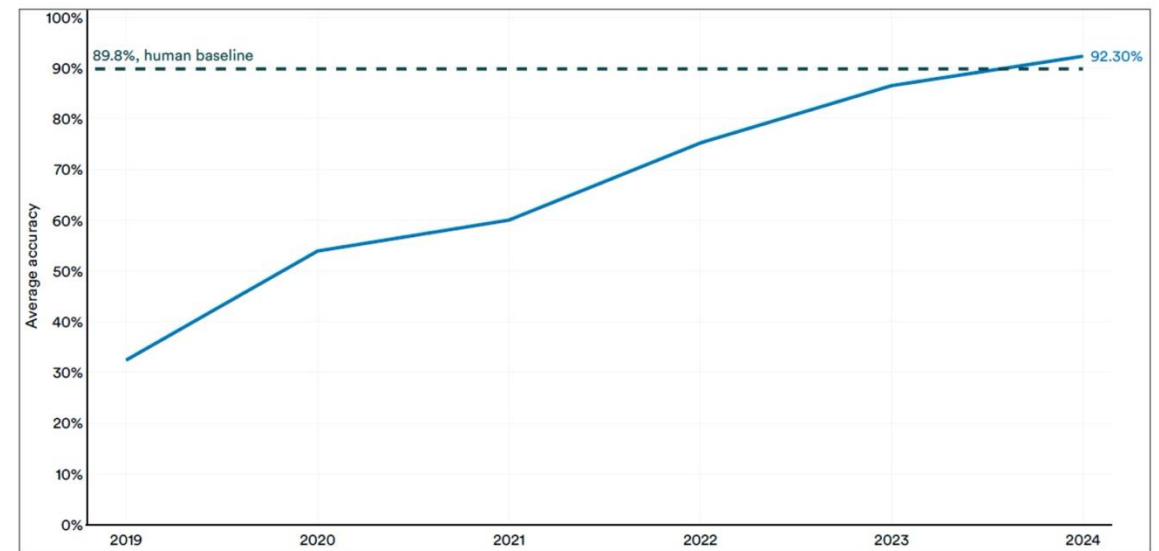


Note: Dashed-line bars are for years where Google did not disclose annual search volumes. Source: Google public disclosures, OpenAI (12/24). ChatGPT figures are estimates per company disclosures of ~1B daily queries

AI batte gli umani
>
verso la
superintelligenza
(AGI)

AI Performance = In 2024...
Surpassed Human Levels of Accuracy & Realism, per Stanford HAI

AI System Performance on MMLU Benchmark Test – 2019-2024, per Stanford HAI



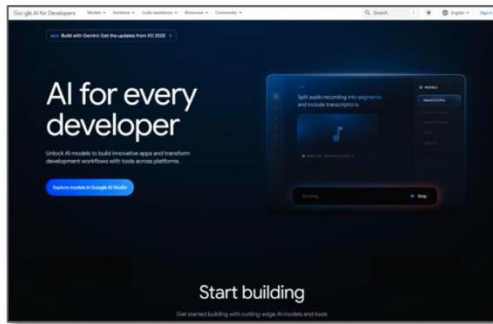
Note: The MMLU (Massive Multitask Language Understanding) benchmark evaluates a language model's performance across 57 academic and professional subjects, such as math, law, medicine, and history. It measures both factual recall and reasoning ability, making it a standard for assessing general knowledge and problem-solving in large language models. 89.8% is the generally-accepted benchmark for human performance. Stats above show average accuracy of top-performing AI models in each calendar year. Source: Papers With Code via Nestor Maslej et al., 'The AI Index 2025 Annual Report,' AI Index Steering Committee, Stanford HAI (4/25)

Forze potenti spingono l'AI

- La tecnologia sta evolvendo molto velocemente
- L'adozione degli utenti sta esplodendo
- I capex aumentano in modo gigantesco
- I ricavi crescono potentemente
- Il costo per token si riduce rapidamente

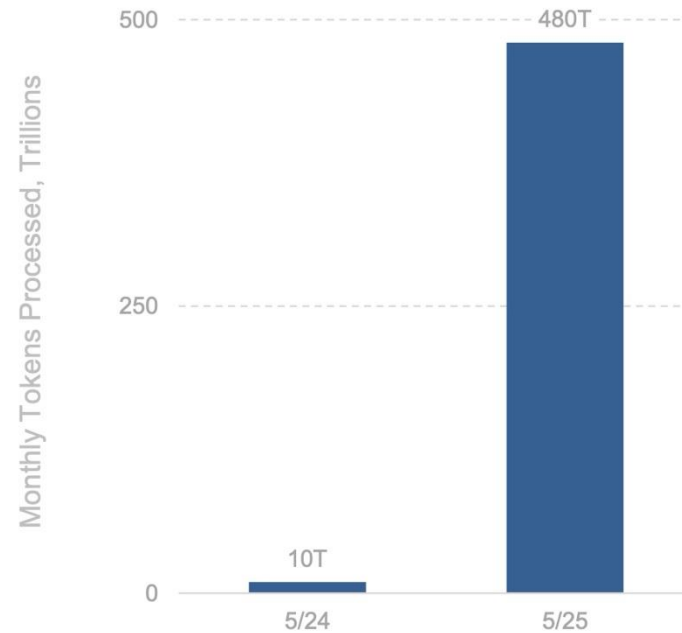


Google Monthly Tokens Processed (T) – 5/24-5/25, per Google



This time last year, we were processing 9.7 trillion tokens a month across our products and APIs. Now, we're processing over 480 trillion – that's 50 times more.

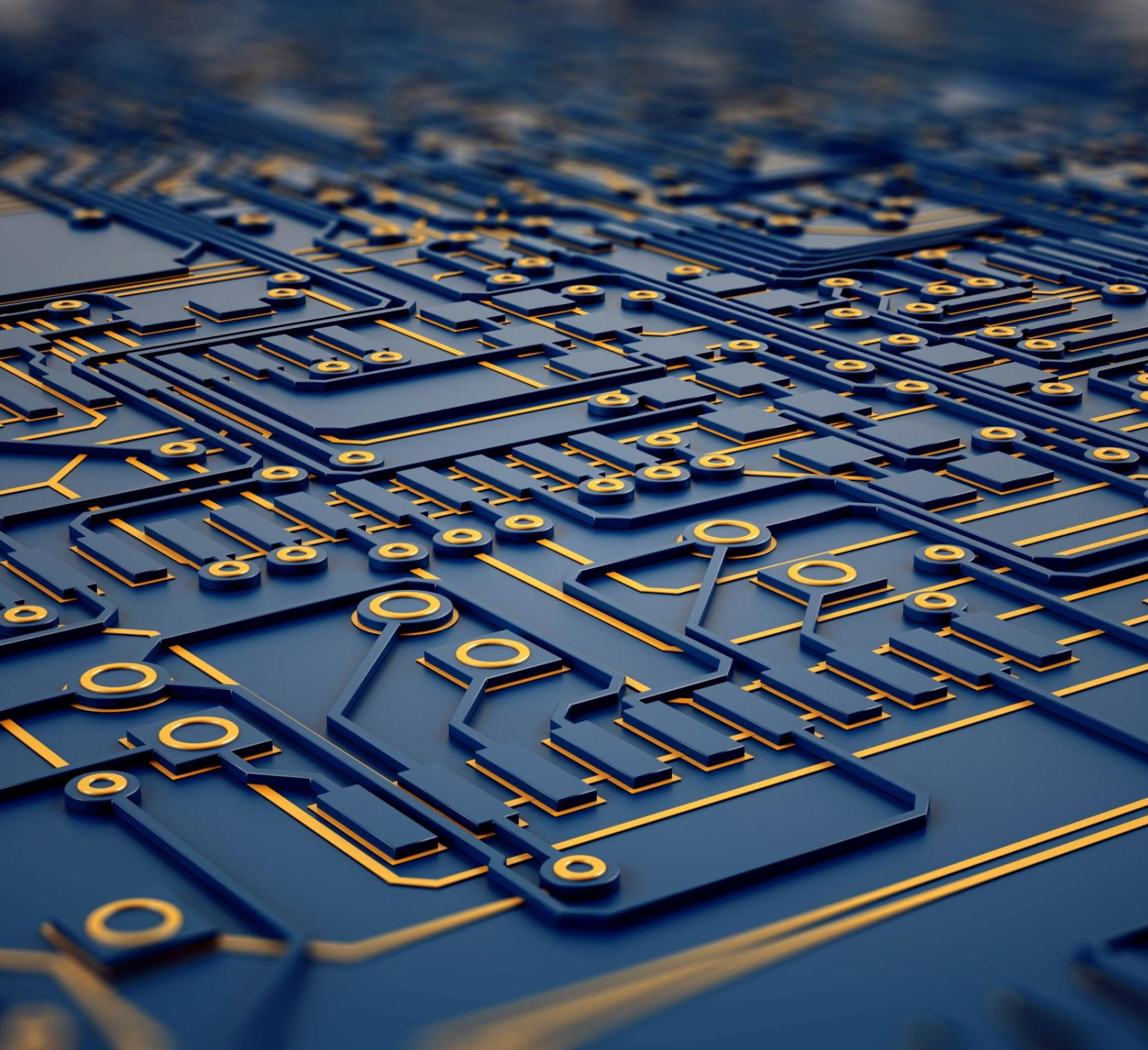
- Google I/O 2025 Press Release, 5/25



Computing e
inferenza sono
finanziati tramite
tokens

Cosa sono i tokens?

- Possiamo pensare ogni token come pezzi di una parola
- Un input viene spezzato in tokens (1 token circa 4 caratteri, 100 token circa 75 parole)
- 1-2 frasi = 30 token
- 1 paragrafo 100 token
- La dichiarazione di indipendenza Americano sono 1.695 tokens



Il costo per token sta crollando

- Tra il 2022 e il 2024 è crollato del 99,7%
- Da 20 \$ per milione di token a 0,07\$
- Il crollo è guidato sia da miglioramenti dell'hardware che dell'efficienza algoritmica
- A livello hardware i costi sono scesi del 30% e del 40% il costo energetico ogni anno dal 2022

La tecnologia evolve mooolto in fretta

**AI-Generated Image: 'Women's Necklace with a Sunflower Pendant' – 2/22-4/25,
per Midjourney / Gold Penguin**

Model v1 (2/22)



Model v7 (4/25)



Turing test

AI-Generated vs. Real Image – 2024

AI-Generated Image (2024)



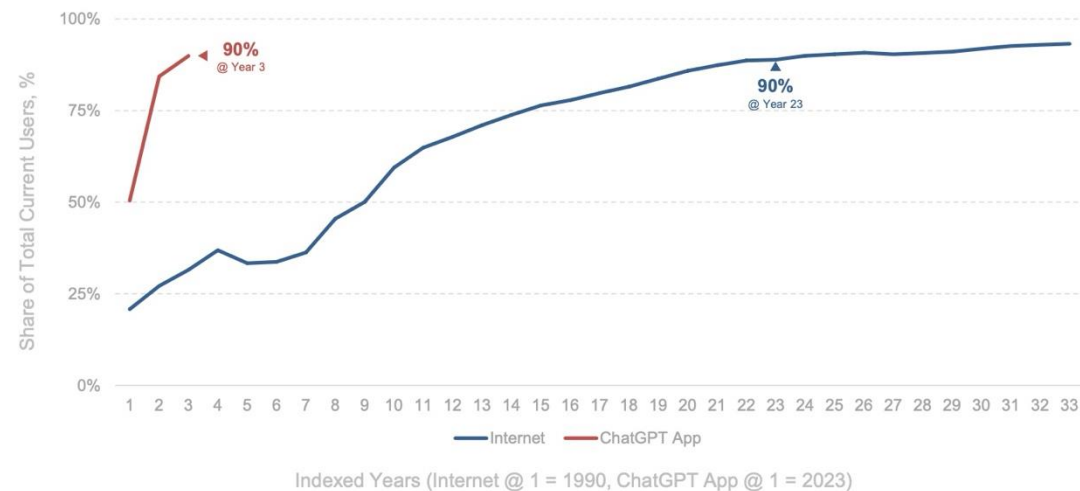
Real Image



Source: Left – StyleGAN2 via 'The New York Times,' 'Test Yourself: Which Faces Were Made by A.I.?' (1/24); Right – Creative Commons


Rampa di utilizzo mai vista prima

Internet vs. ChatGPT Users – Percent Outside North America (1990-2025),
Per ITU & Sensor Tower




Note: Year 1 for Internet = 1990; year 33 = 2022. Year 1 for ChatGPT app = 5/23; year 3 for ChatGPT app = 5/25. ChatGPT app monthly active users (MAUs) shown. Note that ChatGPT is not available in China, Russia and select other countries as of 5/25. China data may be subject to informational limitations due to government restrictions. Includes only Android, iPhone & iPad users. Figures may understate true ChatGPT user base (e.g., desktop or mobile webpage users). Regions per United Nations definitions. Figures show % of total current users in that year – note that as year 3 for ChatGPT has not yet finished, percentages could move in coming months. Data for standalone ChatGPT app only. Country-level data may be missing for select years, as per ITU. Source: United Nations / International Telecommunications Union (3/25), Sensor Tower (5/25)

- Scrivere, semplificare o editare istantaneamente qualunque testo (emails, contratti, poesie, codice software...)
- Insegnare qualunque soggetto (matematica, storia, lingue...)
- Automatizzare processi ripetitivi (generare report, slides, testi)
- Collegarsi a strumenti terzi (spreadsheets, calendar, APIs,...)
- Fare compagnia o offrire terapia (uno dei principali utilizzi oggi)
- Organizzare processi




Le cose
principali
che fa oggi
chatGPT

- Generare testo, codice e logica come un essere umano
- Creare film, musica e videogames in automatico
- Capire e parlare come un essere umano in tempo reale
- Gestire robot e umanoidi (operai, badanti, ...)
- Gestire autonomamente vendite e servizio clienti
- Gestire autonomamente un business
- Scoperte scientifiche (farmaci, materiali, modelli climatici)



Cose che
farà
ChatGPT
tra 5 anni

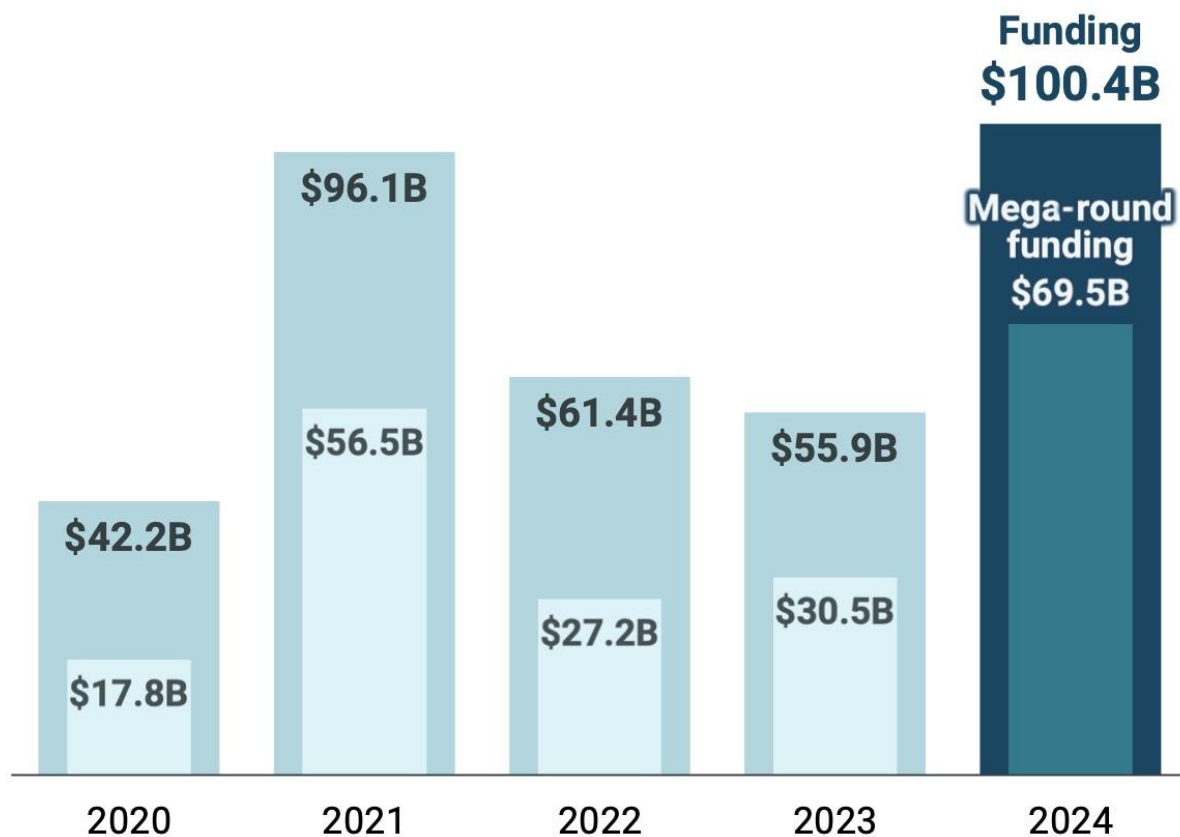
- Condurre ricerche scientifiche autonomamente
- Simulare persone (memoria, emozioni e comportamenti adattivi)
- Performare task fisici in totale autonomia adattandosi al mondo reale
- Gestire robot e umanoidi (operai, badanti, ...)
- Coordinare sistemi globali (logistica, energia, ...)
- Prendere decisioni critiche (mediche, legali e business)
- Gestire il dibattito pubblico e decisioni politiche (ie proporre leggi)







Cose che
farà
ChatGPT
tra 10 anni

Global AI funding hits a record \$100.4B in 2024

Mega-rounds (deals worth \$100M+) to AI infrastructure players drive funding boom



5 largest deals in 2024

- 1**  **databricks**
\$10.0B
Series J • 2024-12-17
- 2**  **OpenAI**
\$6.6B
Venture Capital • 2024-10-02
- 3**  **X1**
\$6.0B
Series B • 2024-05-26
- 4**  **X1**
\$6.0B
Series C • 2024-11-20
- 5** **ANTHROPIC**
\$4.0B
Corporate Minority • 2024-11-22

L'AI non è solo ChatGPT



Generative AI Tech Stack

Application Experience

Horizontal Apps

Vertical Apps

Data Platforms & Management

Enterprise / External Data

Ingestion

Cleaning

Data Lake

Vector Store

MLOps

Prompt Engineering

Deployment

Monitoring / Observability

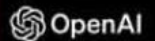
Fine Tuned Models

Domain Models

Model Training

Foundation Models

Closed Source



AI21labs

cohere

ANTHROPIC

Open Source

LLaMA

PaLM 2

cerebras

EleutherAI

BLISS

FLAN-T5

stability ai

Model Hubs

Hugging Face

Cloud Platforms

Microsoft Azure

Google Cloud

aws

CoreWeave

Compute

GPU, Memory, Networking

intel

nvidia

AMD

Arm

SambaNova

Lo stack AI

End-To-End

With Prompt

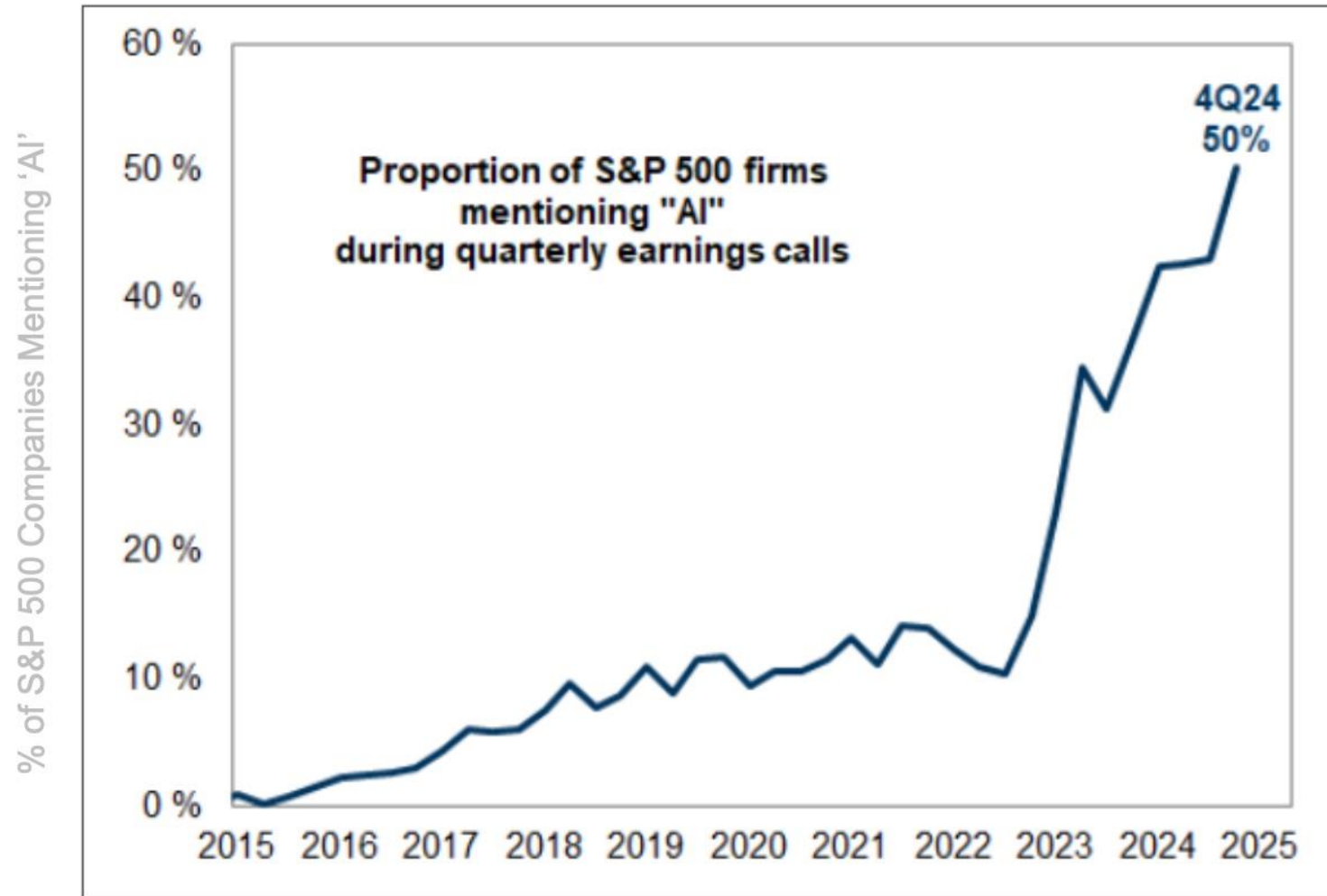
Model

- Applicazioni
- Data platforms & management
- MLOps
- Fine Tuned Models
- Foundation Models
- Cloud
- Compute

Il livello applicativo

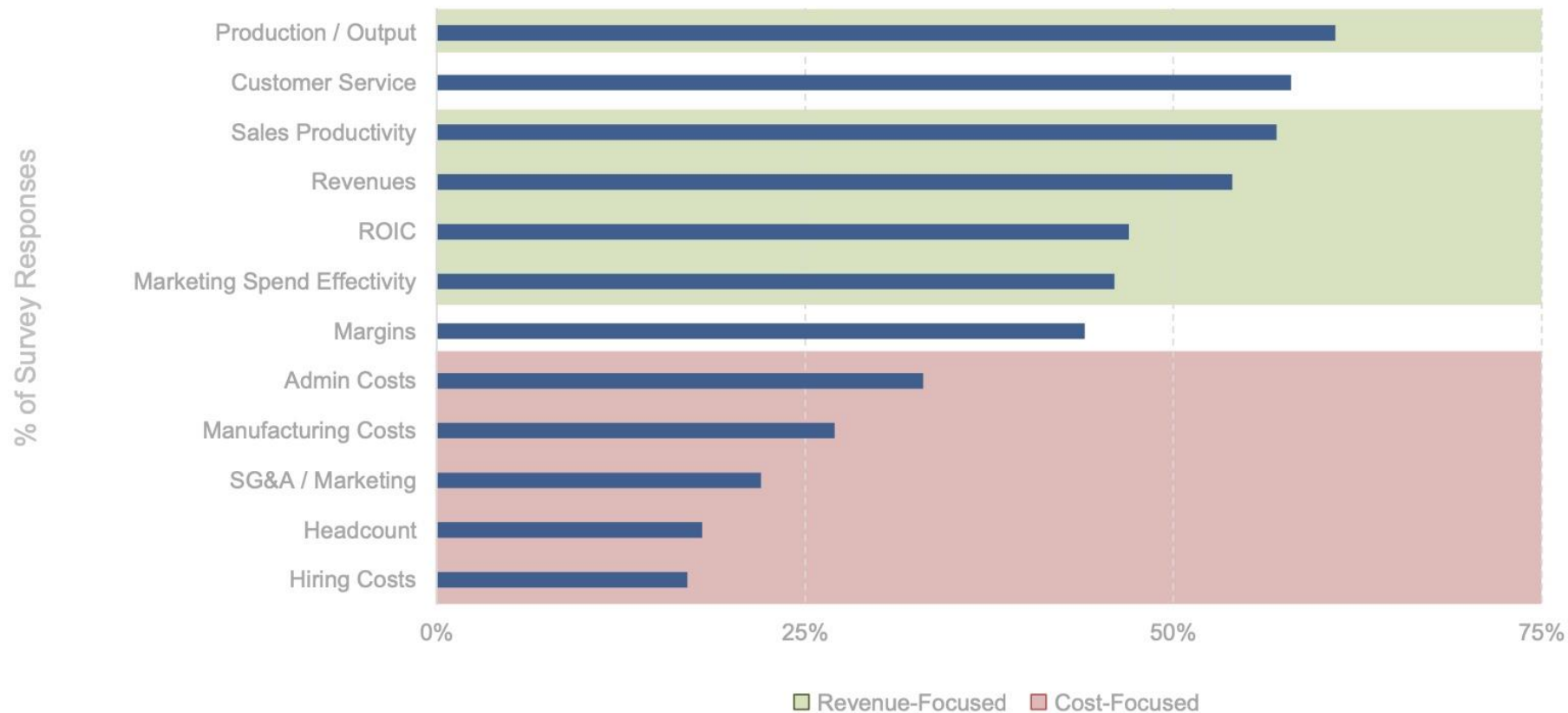
- **Piattaforma orizzontali** (ie Microsoft co-pilot, Adobe Firefly, Zoom AI Companion, Glean)
- **Piattaforme verticali** in qualunque ambito aziendale: legale, salute, vendita, marketing, logistica, acquisti, assunzioni di personale, servizi finanziari, etc.
- **Incumbents** contro **startups native AI**

Quarterly Earnings Call Mentions of 'AI' – S&P 500 Companies (2015-2025), per Goldman Sachs Research



L'impatto dell'AI sulle aziende

GenAI Improvements Targeted for Global Enterprises over Next 2 Years – 2024, per Morgan Stanley







THREATS

OPPORTUNITIES



7 idee concrete per grossisti e distributori (fonte ChatGPT)

- Migliore previsione della domanda → meno scorte e rotazione più veloce del magazzino con modelli che prevedono la domanda per SKU
- Ottimizzazione dell'inventario e riordino dinamico per ridurre stockout e overstock
- Spinta della domanda per componenti legati ai data center / AI infrastructure (rack, alimentazioni, cavi, raffreddamento)
- Manutenzione predittiva e servizio post-vendita a valore aggiunto
- Automazione del magazzino e logistica (robotica + routing intelligente)
- Pricing dinamico e ottimizzazione commerciale per aumentare i margini
- Gestione del rischio fornitore con modelli valutativi e suggerimento di fonti alternative di acquisto

Rischi e limiti da considerare(fonte ChatGPT)

- Qualità dati: senza dati puliti e integrazione i modelli falliscono
- Sovra-automazione: decisioni critiche (es. sourcing strategico) devono mantenere oversight umano.
- Dipendenza da pochi fornitori: l'IA non elimina il rischio se il fornitore è unico.
- Impatto occupazionale: cambiamenti nei ruoli (meno attività manuali, più ruoli analitici) richiedono riqualificazione.
- Cyber & compliance: più dati e modelli significano più superficie di attacco e obblighi GDPR/contrattuali.

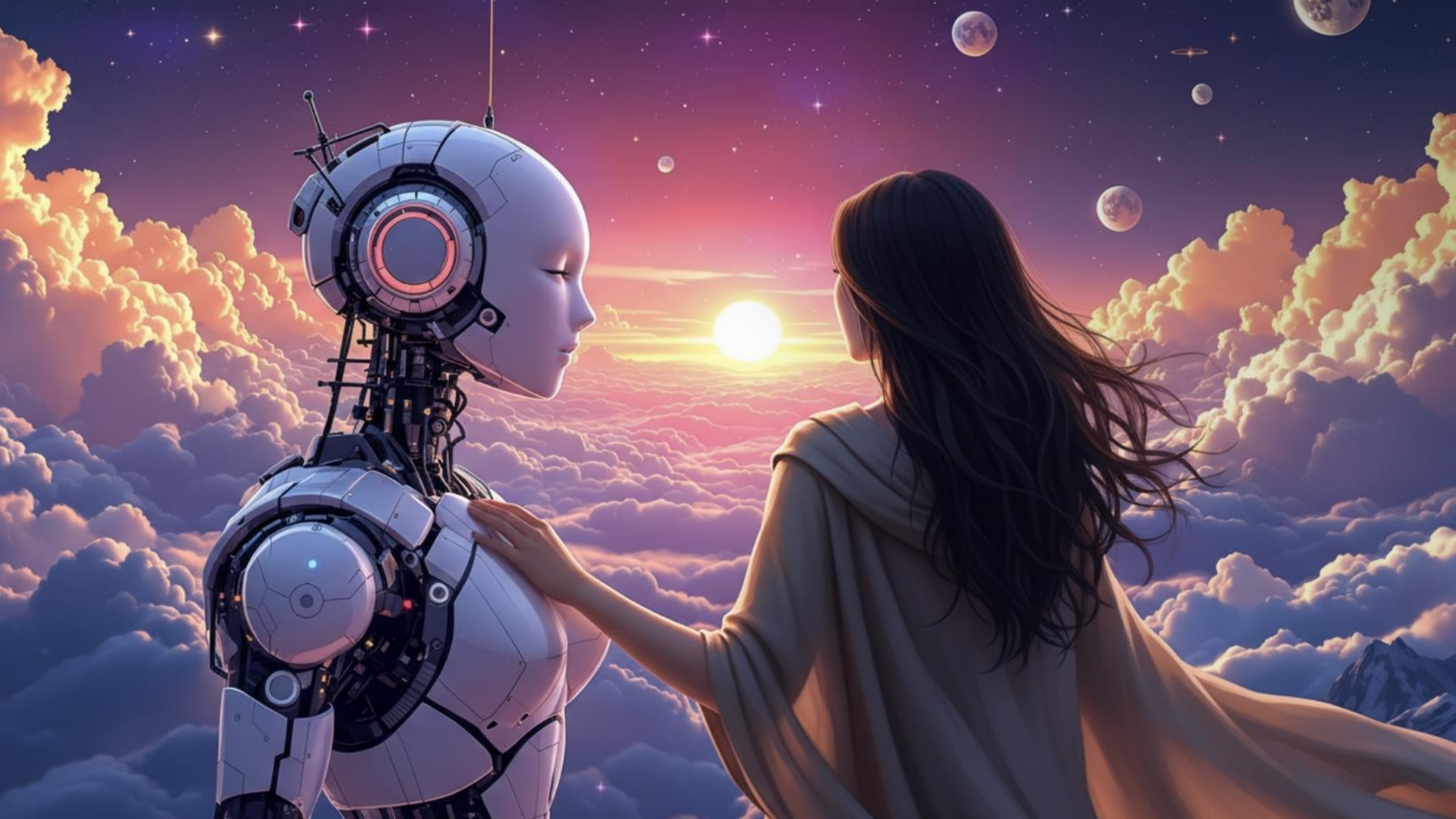
Roadmap operativa (Fonte ChatGPT)

- **Breve termine (0–12 mesi) — fundamenta**
 - Audit dati: mappa ERP, WMS, ordini, lead times, storico vendite
 - Pilot su forecasting SKU critici (top 20% del fatturato) con modelli ML off-the-shelf.
- **Medio termine (12–36 mesi) — scala e servizi**
 - Implementare riordino dinamico multi-magazzino.
 - Introdurre pricing dinamico per segmenti clienti e raccomandazioni di upsell.
 - Offrire servizi predittivi su componenti critici (contratti ricorrenti).
- **Lungo termine (>3 anni) — trasformazione**
 - Automazione avanzata magazzino (robotica + orchestrazione AI).
 - Ecosistema digitale: portale clienti con raccomandazioni, inventario in tempo reale e lead times personalizzati.
 - Nuovi modelli di business (pay-per-use, managed spares, “outcome selling”)



E dopo che
succede?





Grazie!

gianluca.dettori@primo.vc