The Future of AI

Nestor Maslej | September 17th, 2025

Meet Nestor

HAI
Stanford University
Human-Centered
Artificial Intelligence

AI

2025

8th Edition

AI Index Steering Committee

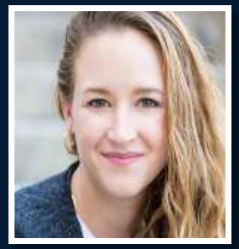
Ray Perrault
(Chair)
SRI International

Yolanda Gil
(Chair-Elect)
USC

Erik Brynjolfsson
Stanford Digital
Economy Lab

Katrina Ligett
Hebrew University

John Etchemendy
Stanford University



Terah Lyons
JP Morgan Chase

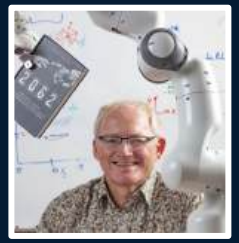
Juan Carlos Niebles
Stanford and Salesforce

James Manyika
Google, University
of Oxford

Yoav Shoham
Stanford University,
AI21 Labs

Vanessa Parli
Stanford HAI

Russell Wald
Stanford HAI



Toby Walsh
UNSW Sydney

Jack Clark
Anthropic, OECD


The New York Times


ALJAZEERA


The Register


FT
FINANCIAL
TIMES


FORTUNE


DER SPIEGEL


Bloomberg


Forbes


The Verge


The Washington Post


theguardian


San Francisco
Chronicle


IEEE
SPECTRUM


Vox

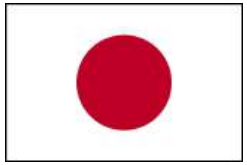

POLITICO

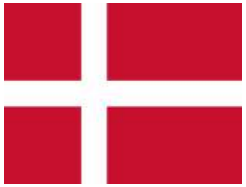

VentureBeat

Governments and agencies citing Nestor's AI work

















Businesses briefed by Nestor

Where is the future going?

1. Technical capabilities

AI performance continues to improve

Midjourney generations over time: “a hyper-realistic image of Harry Potter”
Source: Midjourney, 2024

Figure 2.3.6

AI performance continues to improve

V1, February 2022

V6, December 2023

Whichfaceisreal.com

Whichfaceisreal.com

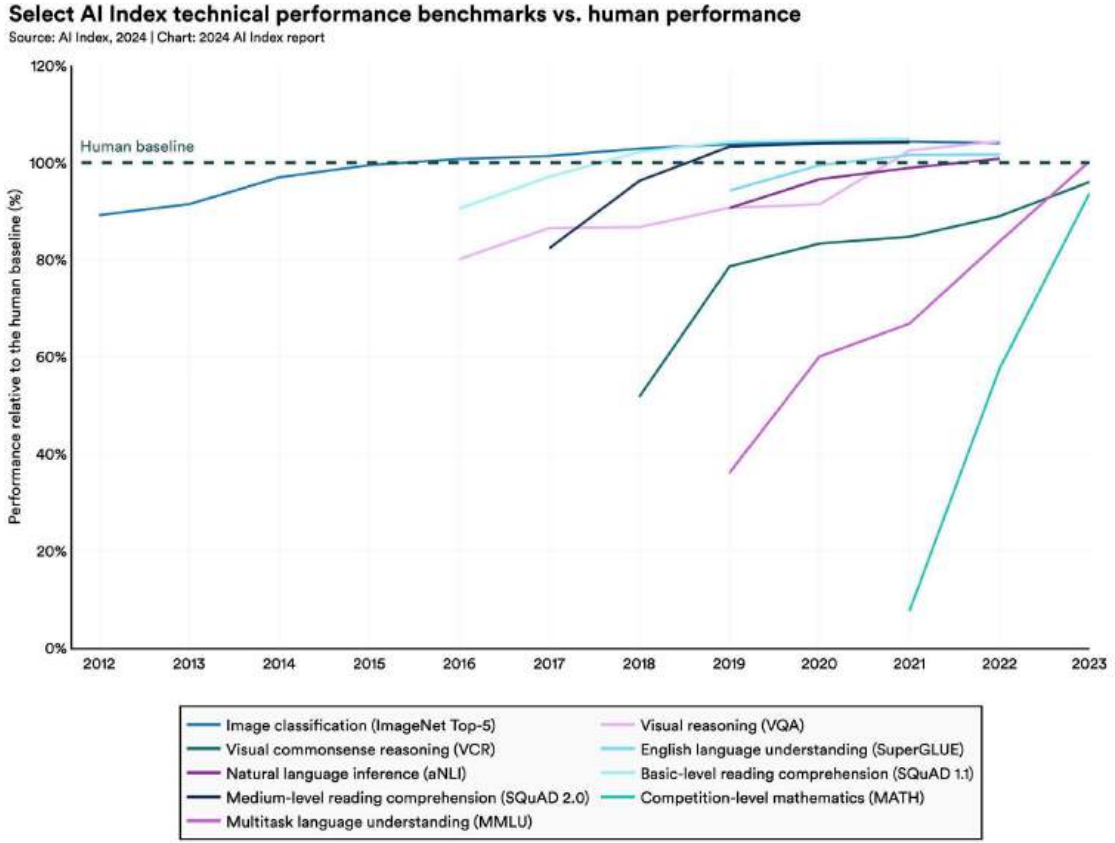
Not just image generation, videos too!
Pika v1, December 2023

Not just image generation, videos too!
Pika v1.5, October 2024

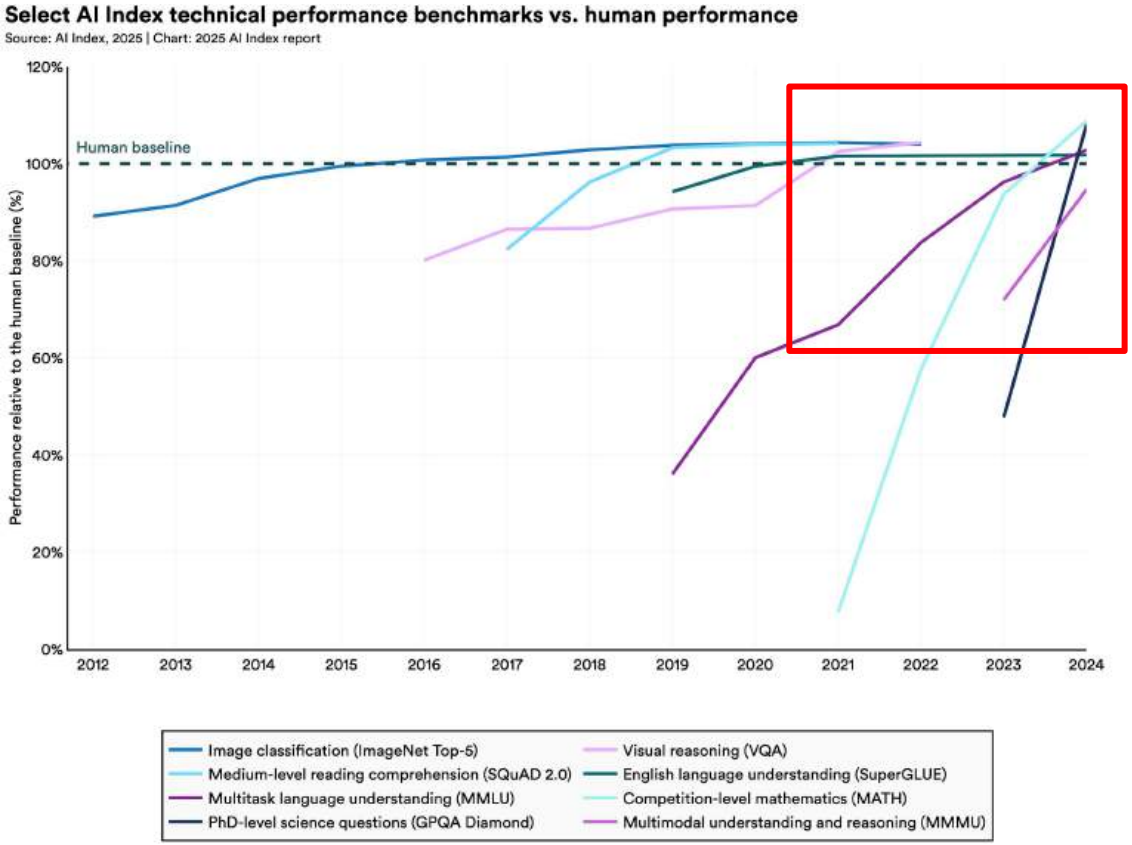
Not just image generation, videos too!
Pika v2.2, February 2025

Google’s Veo 3, May 2025

AI exceeds humans on some, but not all benchmarks (2024)



What benchmarks are left? (2025)



GPQA, + 49%

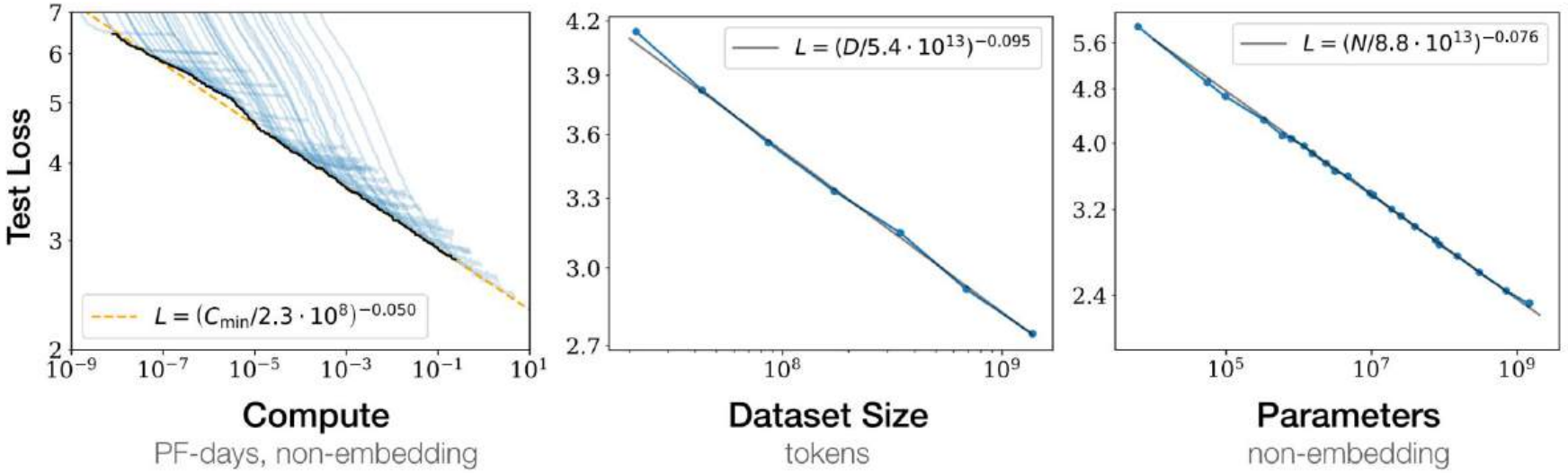
MMMU, +19%

SWE-bench, +67%

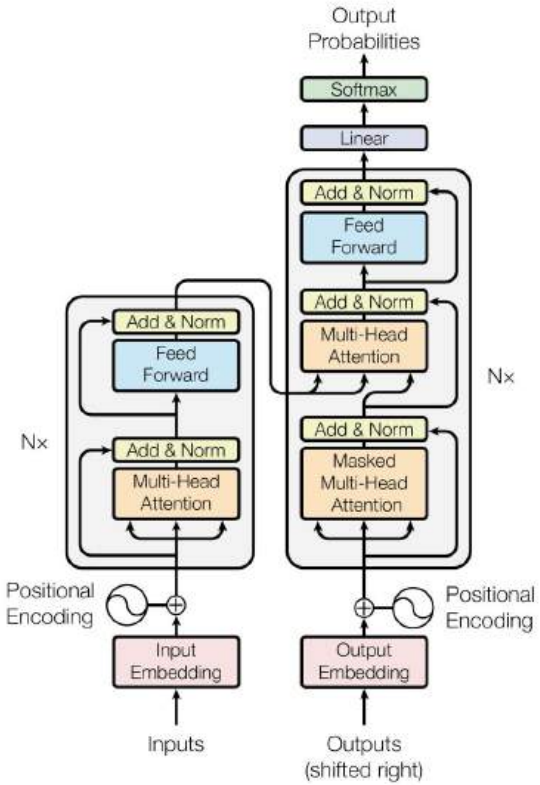
Key takeaway: AI systems are getting better

Key question: How much better will they get (with scale)?

Scaling laws



It all started with the transformer (Vaswami et al., 2017)



Training compute of notable AI models overtime

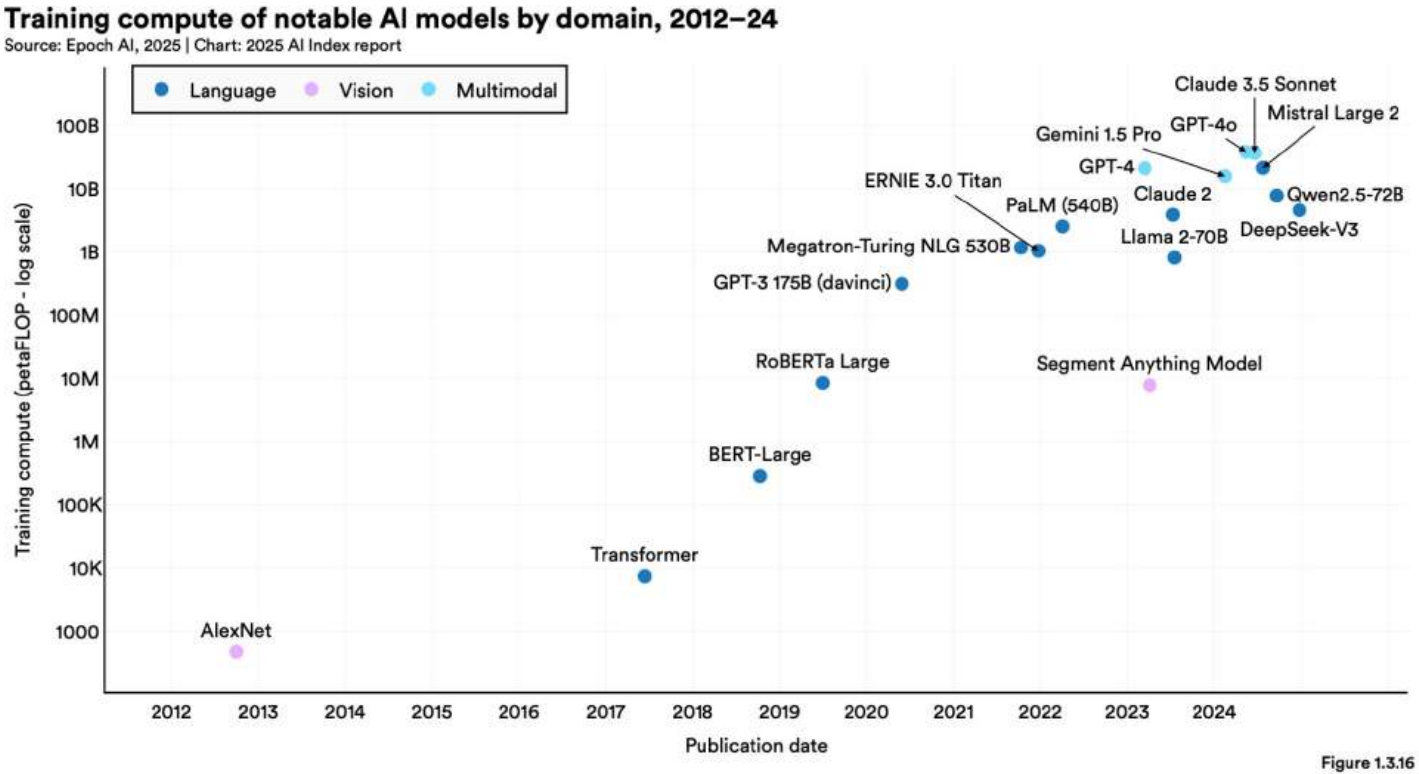
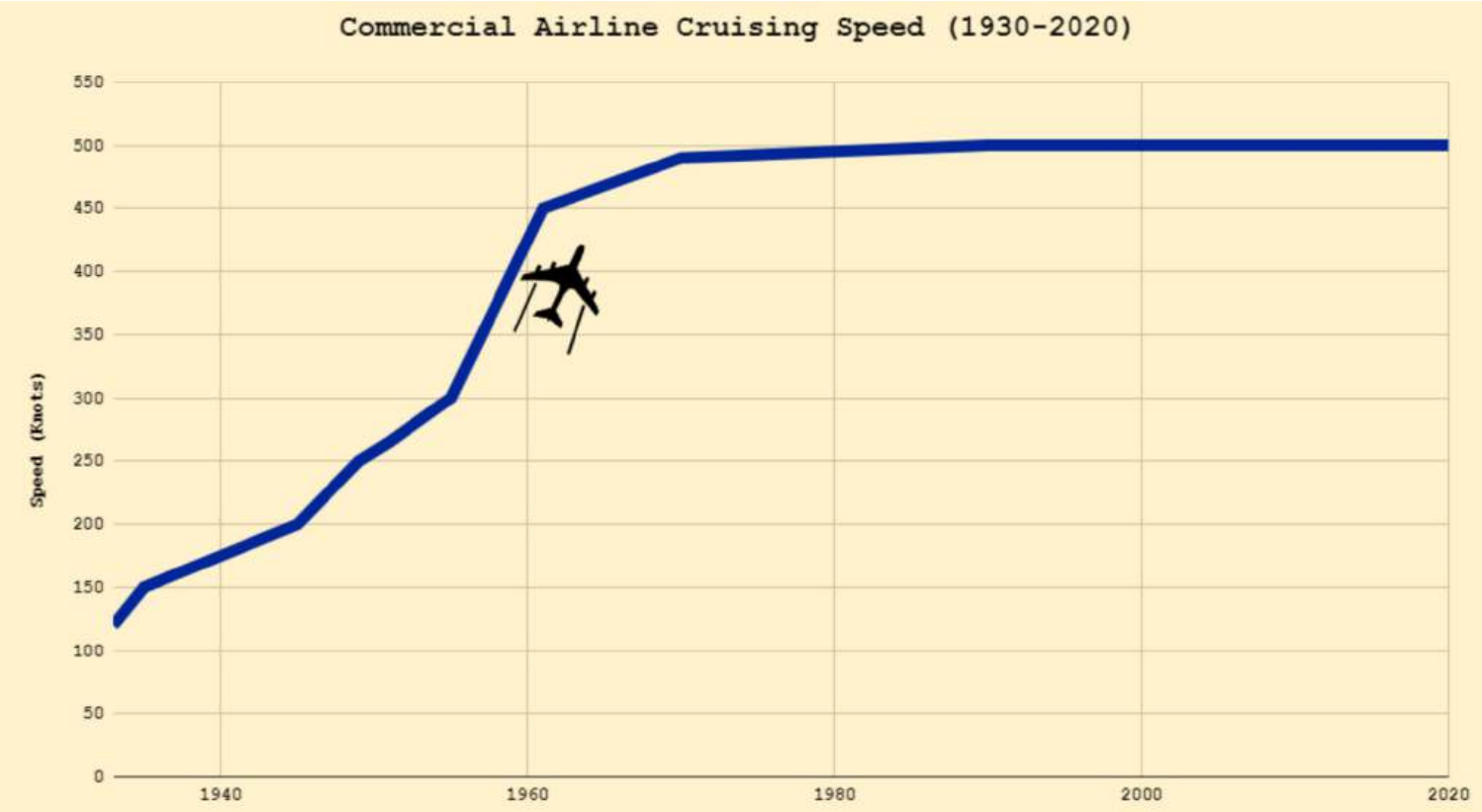


Figure 1.3.16

Comparison to airline speeds



GPT-5 negative reactions

I tested GPT-5's coding skills, and it was so bad that I'm sticking with GPT-4o (for now)

ChatGPT has been the gold standard of AI programming tools, but GPT-5 stumbled badly, delivering broken plugins and flawed scripts in my tests.

Nearly 5,000 GPT-5 users flock to Reddit in backlash — it 'feels like a downgrade' and 'I feel like I'm taking crazy pills'

News By Dave LeClair last updated August 8, 2025

r/ChatGPT • 4 days ago
GALAXY_12321

GPT-5 is a disaster.

GPT-5 : OpenAI's Worst Release Yet

OpenAI new model, GPT-5 looks bad in early reviews

Mehul Gupta Follow 5 min read · 2 days ago

r/OpenAI • 3 days ago
BernieBlade

GPT-5 is awful

GPT-5: Overdue, overhyped and underwhelming. And that's not the worst of it.

A new release botched ... and new research paper that spells trouble

GARY MARCUS
AUG 09, 2025

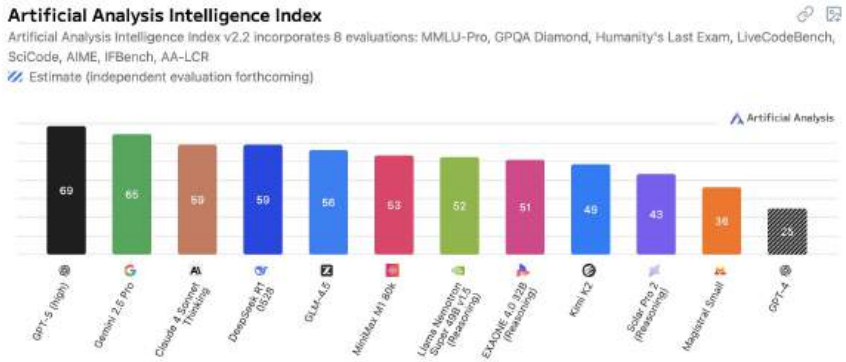
Contextualizing performance gains (part 1)

+ 15.2 %

+ 7.1 %

Model and Date	LM Arena Score
GPT-5 | August 2025	1481
GPT-4 | April 2023	1286
GPT-3.5 Turbo | November 2022*	1201

+44 vs. GPT-4



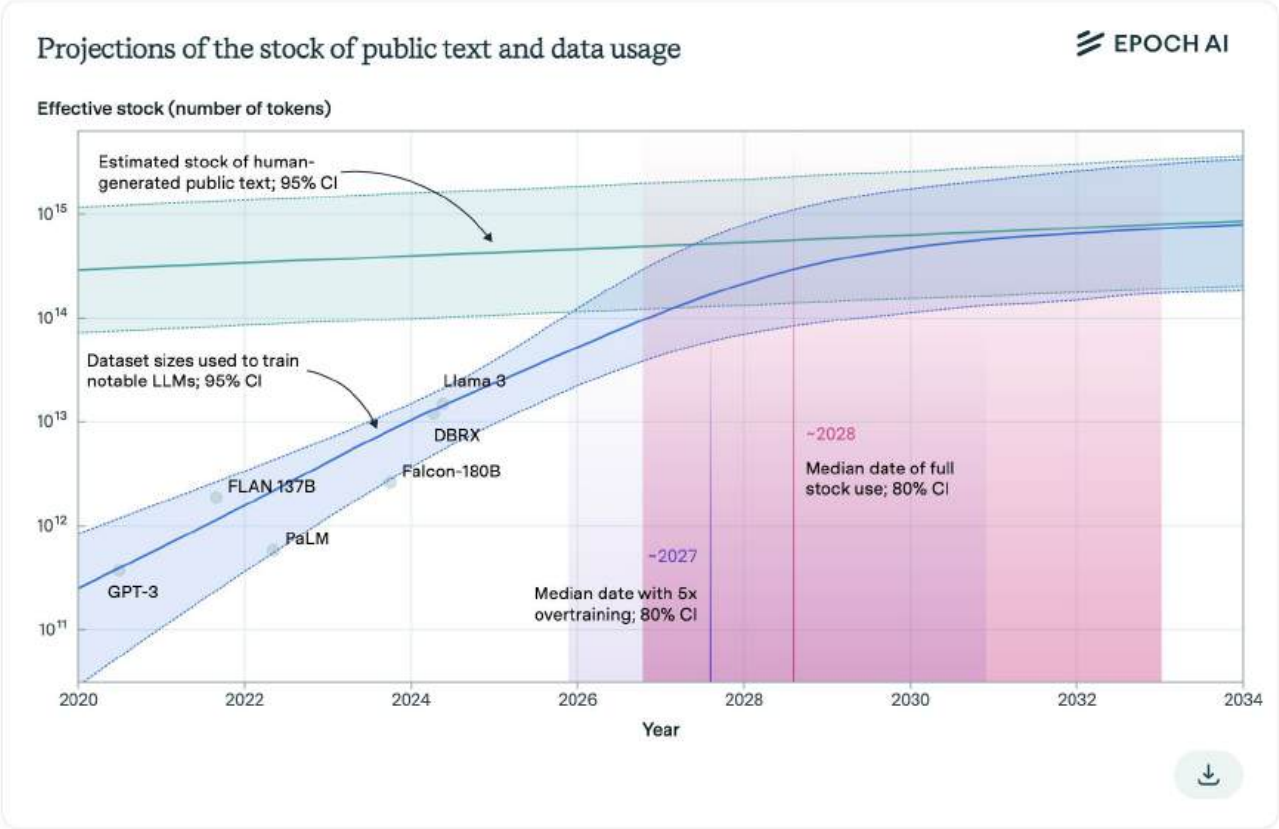
Contextualizing performance gains (part 2)

+ 2.1 %

Model and Date	**LM Arena Score**
*GPT-5	August 2025*	1481
*o3	April 2025*	1451

+1 vs. Grok 4

Are we going to run out of data?



AI as a whole is no longer so open, GPT-3 (2020)

Language Models are Few-Shot Learners

Tom B. Brown* Benjamin Mann* Nick Ryder* Melanie Subbiah*

Jared Kaplan† Prafulla Dhariwal Arvind Neelakantan Pranav Shyam Girish Sastry

Amanda Askell Sandhini Agarwal Ariel Herbert-Voss Gretchen Krueger Tom Henighan

Rewon Child Aditya Ramesh Daniel M. Ziegler Jeffrey Wu Clemens Winter

Christopher Hesse Mark Chen Eric Sigler Mateusz Litwin Scott Gray

Benjamin Chess Jack Clark Christopher Berner

Sam McCandlish Alec Radford Ilya Sutskever Dario Amodei

OpenAI

Abstract

Recent work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. While typically task-agnostic in architecture, this method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples. By contrast, humans can generally perform a new language task from only a few examples or from simple instructions – something which current NLP systems still largely struggle to do. Here we show that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine-tuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and test its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic. At the same time, we also identify some datasets where GPT-3's few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora. Finally, we find that GPT-3 can generate samples of news articles which human evaluators have difficulty distinguishing from articles written by humans. We discuss broader societal impacts of this finding and of GPT-3 in general.

AI as a whole is no longer so open, GPT-4 (2023)

2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

We are committed to independent auditing of our technologies, and shared some initial steps and ideas in this area in the system card accompanying this release.[2] We plan to make further technical details available to additional third parties who can advise us on how to weigh the competitive and safety considerations above against the scientific value of further transparency.

AI as a whole is no longer so open, o3(2024)

2. Geopolitics

US v. China race: US ahead with notable models

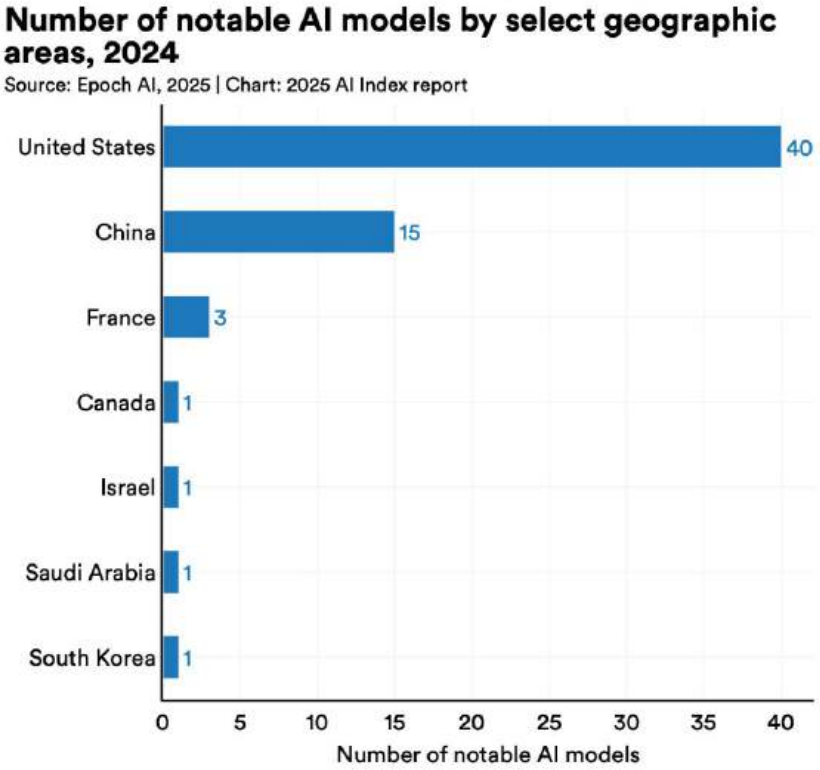


Figure 1.3.1[17]

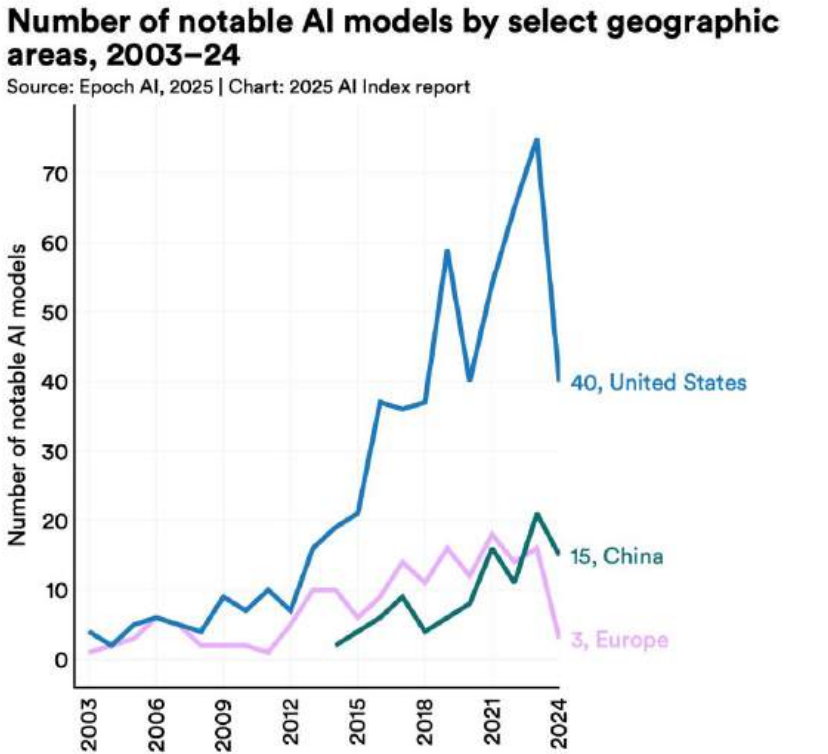


Figure 1.3.2

US v. China race: US ahead since 2003 as well

Number of notable AI models by geographic area, 2003–24 (sum)

Source: Epoch AI, 2025 | Chart: 2025 AI Index report

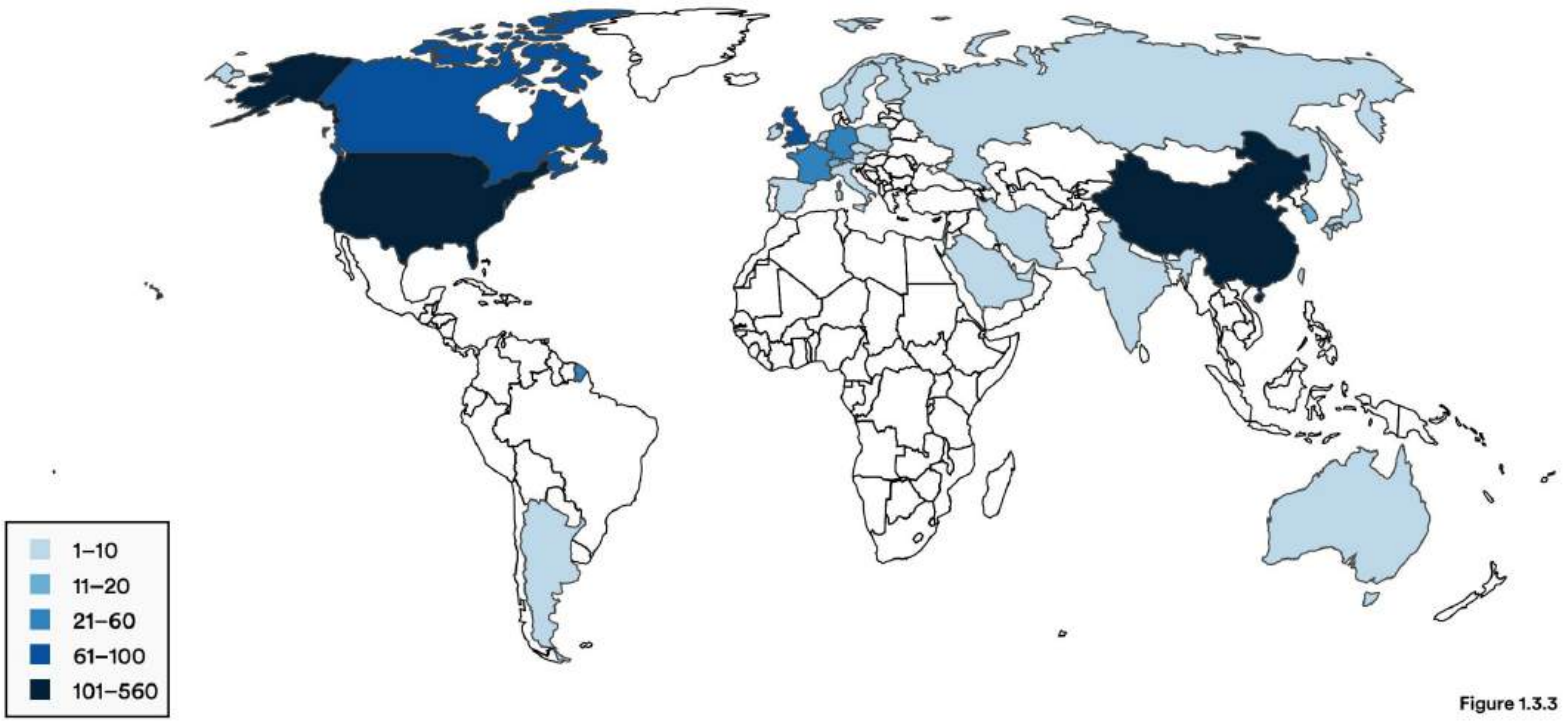


Figure 1.3.3

US v. China race: US producer of top research

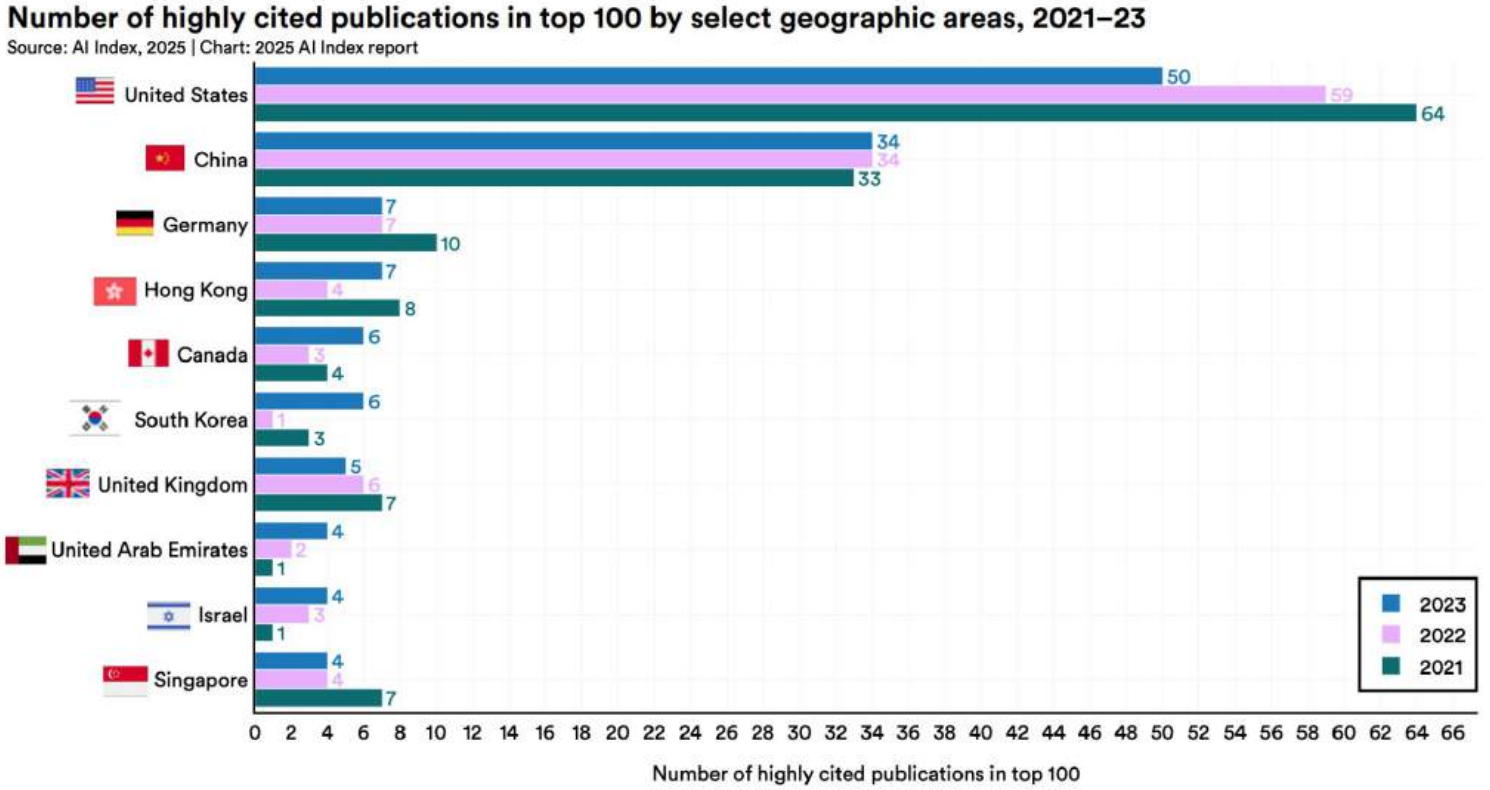


Figure 1.1.11

US v. China race: China leads in research totals

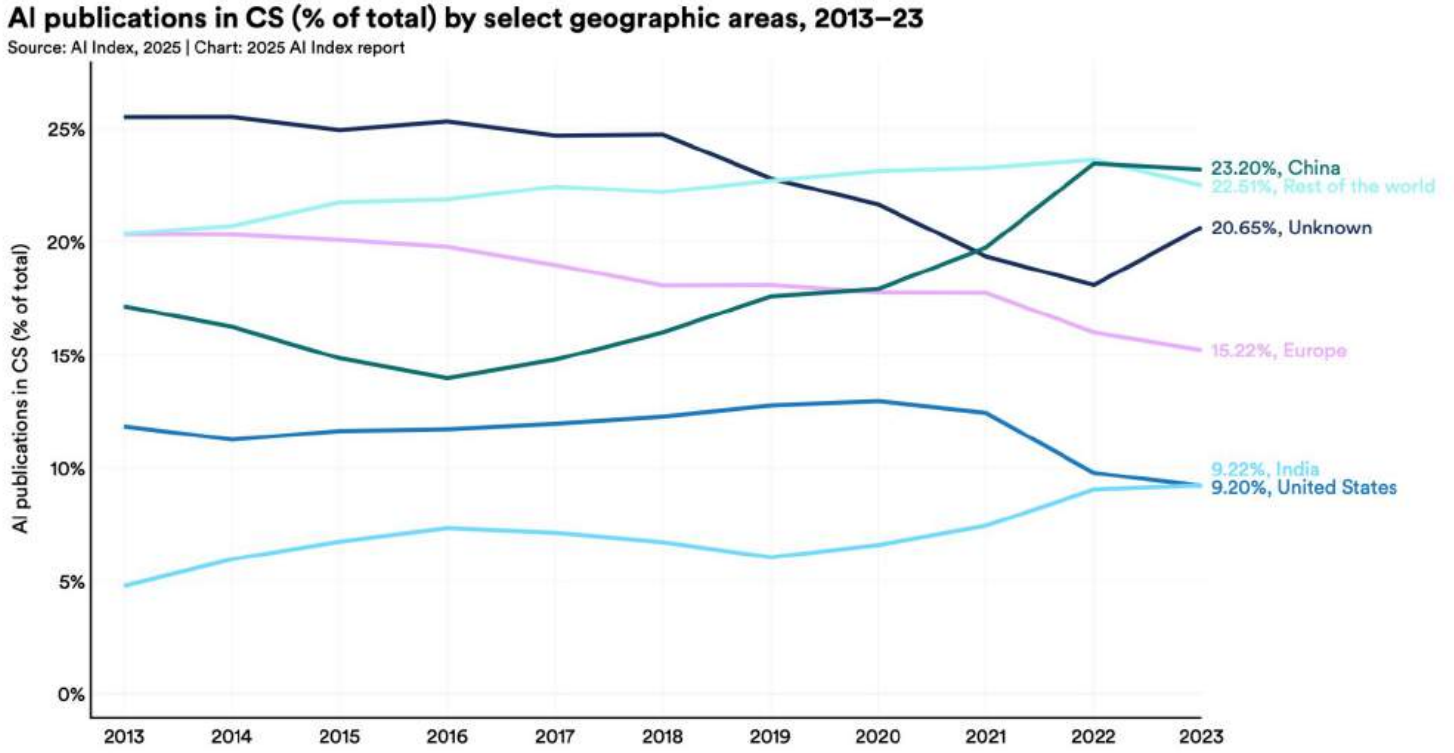


Figure 1.1.6[6]

US v. China race: China leads in patents

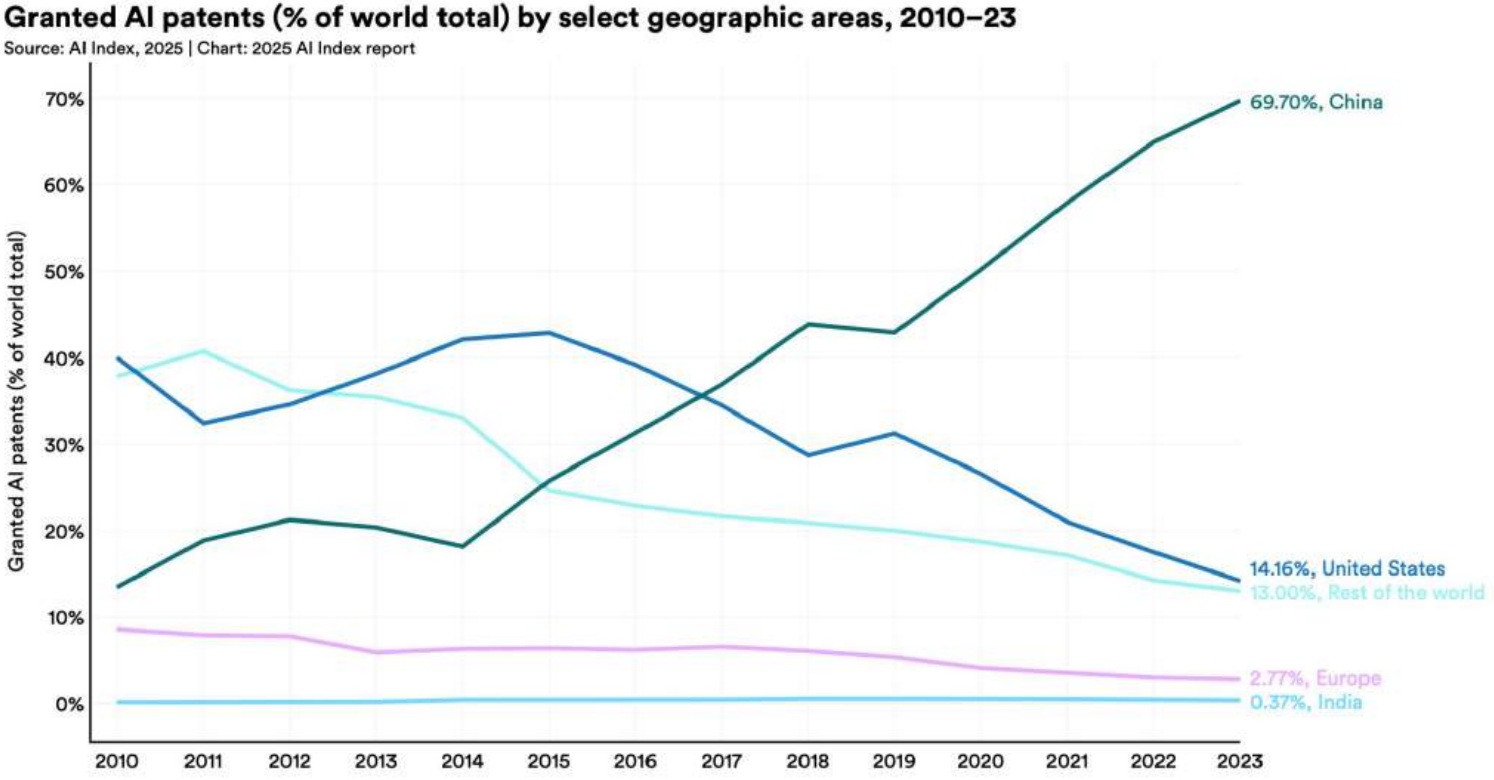


Figure 1.2.3

US v. China race: China leads in industrial robotics

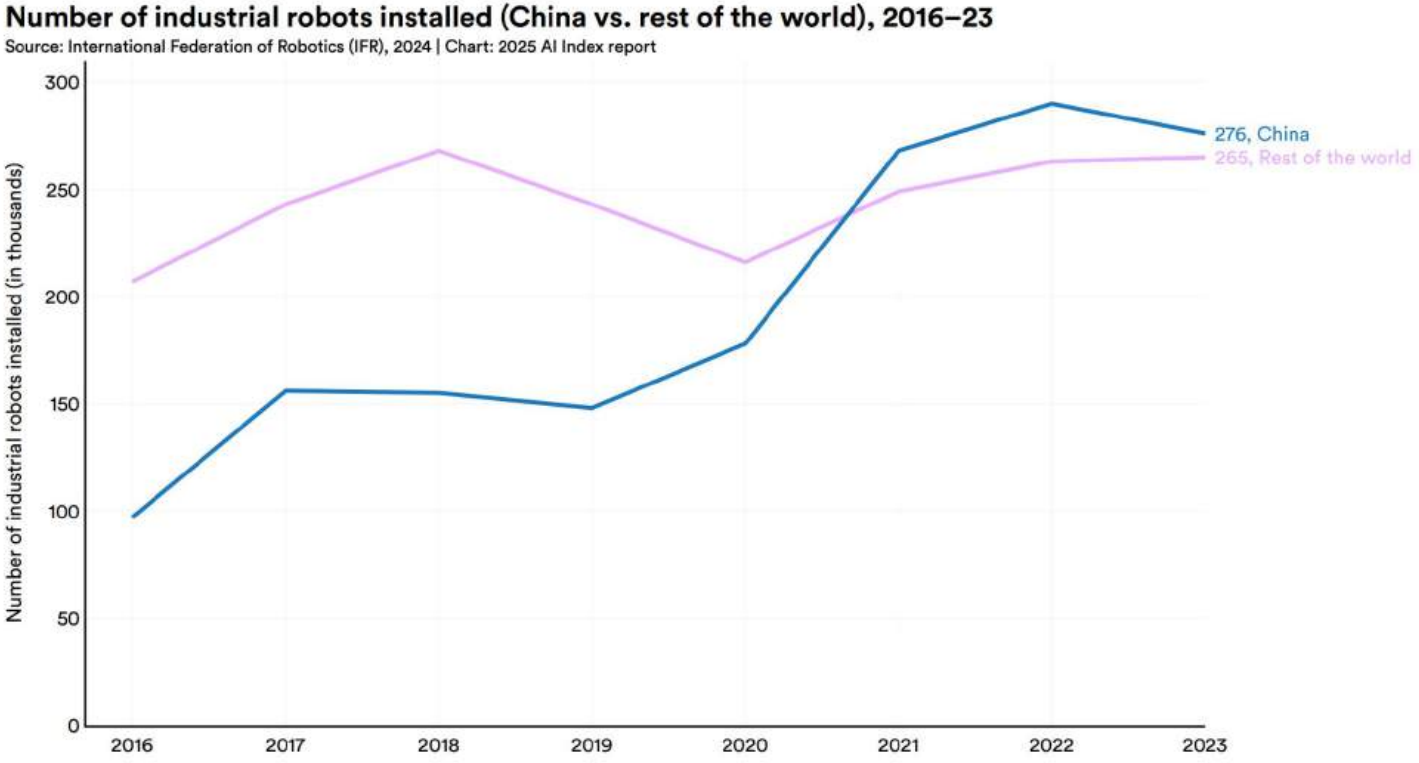
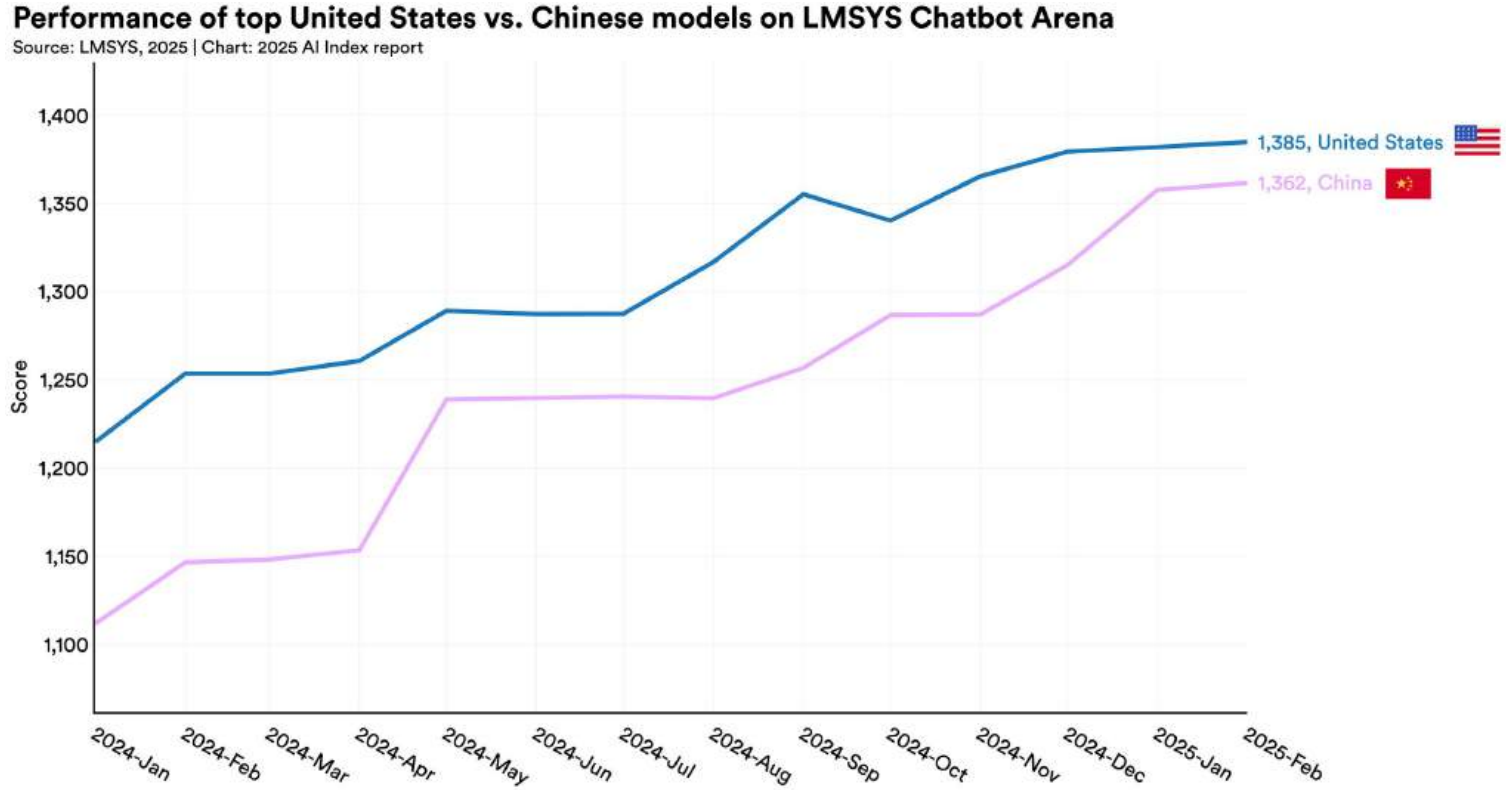


Figure 4.5.6

US v. China race: US catching up in model performance



Tracking AI competition with the [Global Vibrancy Tool]

Key takeaway: geopolitically, AI is a two-horse race

Key question: in ten years, which country will lead?

AI national leadership now a global priority

Details of 110 billion euros in investment pledges at France's AI summit

By Reuters

February 10, 2025 10:40 AM PST · Updated February 10, 2025

French President Emmanuel Macron delivers a speech on the first day of the Artificial Intelligence (AI) Action Summit at the Grand Palais in Paris, France, February 10, 2025. REUTERS/Benoit Tessier Purchase Licensing Rights

Nvidia to send 18,000 AI GPUs to Saudi Arabia's state-backed AI data centers in wake of cancelled export rules

News By Stephen Warwick published May 13, 2025

The semiconductors will furnish a 500-megawatt data center

Comments (3)

When you purchase through links on our site, we may earn an affiliate commission. Here's how it works.

(Image credit: Nvidia)

Trump administration rebrands AI Safety Institute

The Department of Commerce's safety-centered AI organization will instead become the Center for AI Standards and Innovation.

BY MADISON ALDER • JUNE 4, 2025

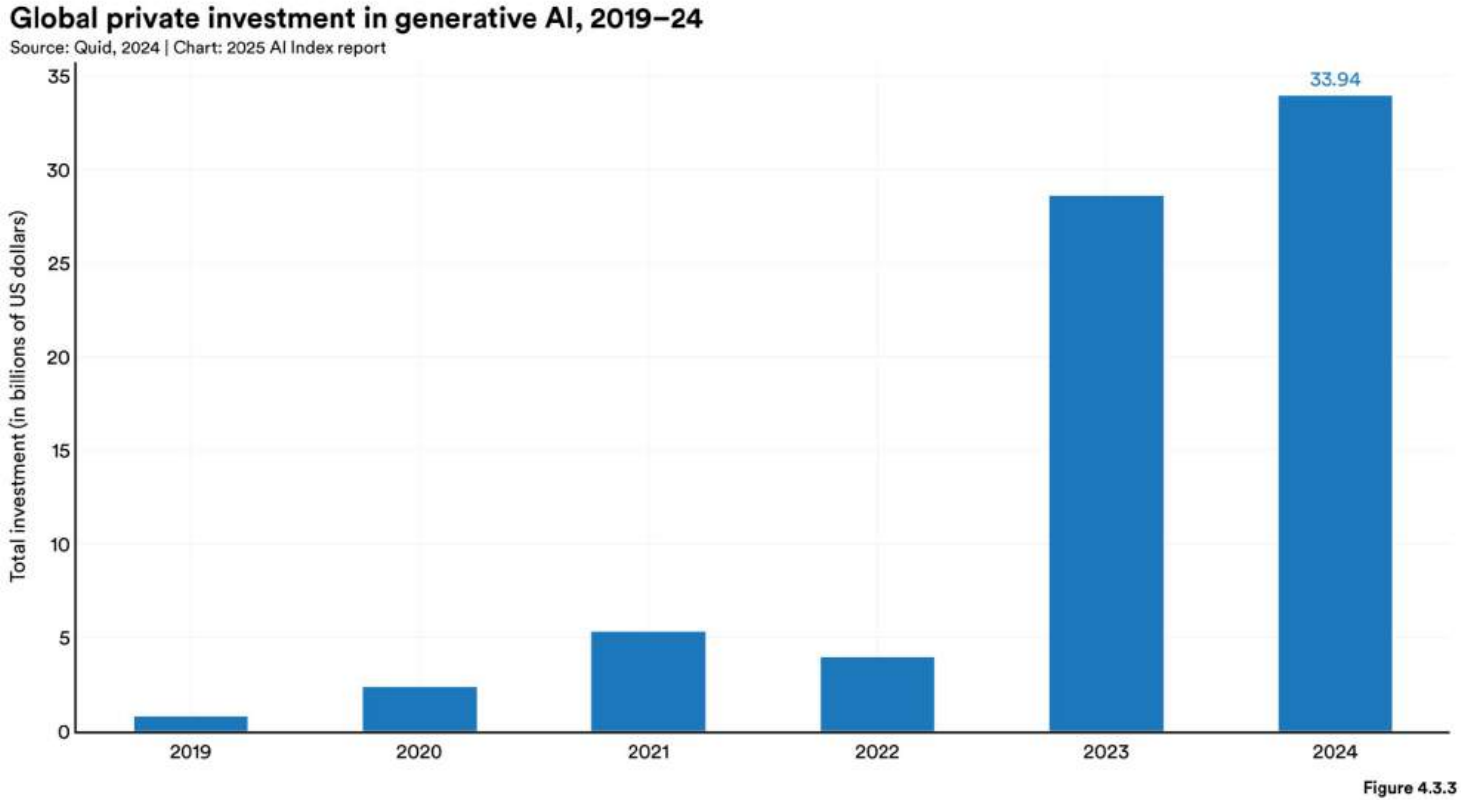
Listen to this article 2:17 Learn more.

The Department of Commerce building is seen in Washington, D.C., on Feb. 3, 2024. (Photo by BRENDAN SMIALOWSKI/AFP via Getty Images)

Would your business ever build with a Chinese model?

3. Business Integration

Investment in generative AI is especially ballooning



Major CEOs now really paying attention to AI

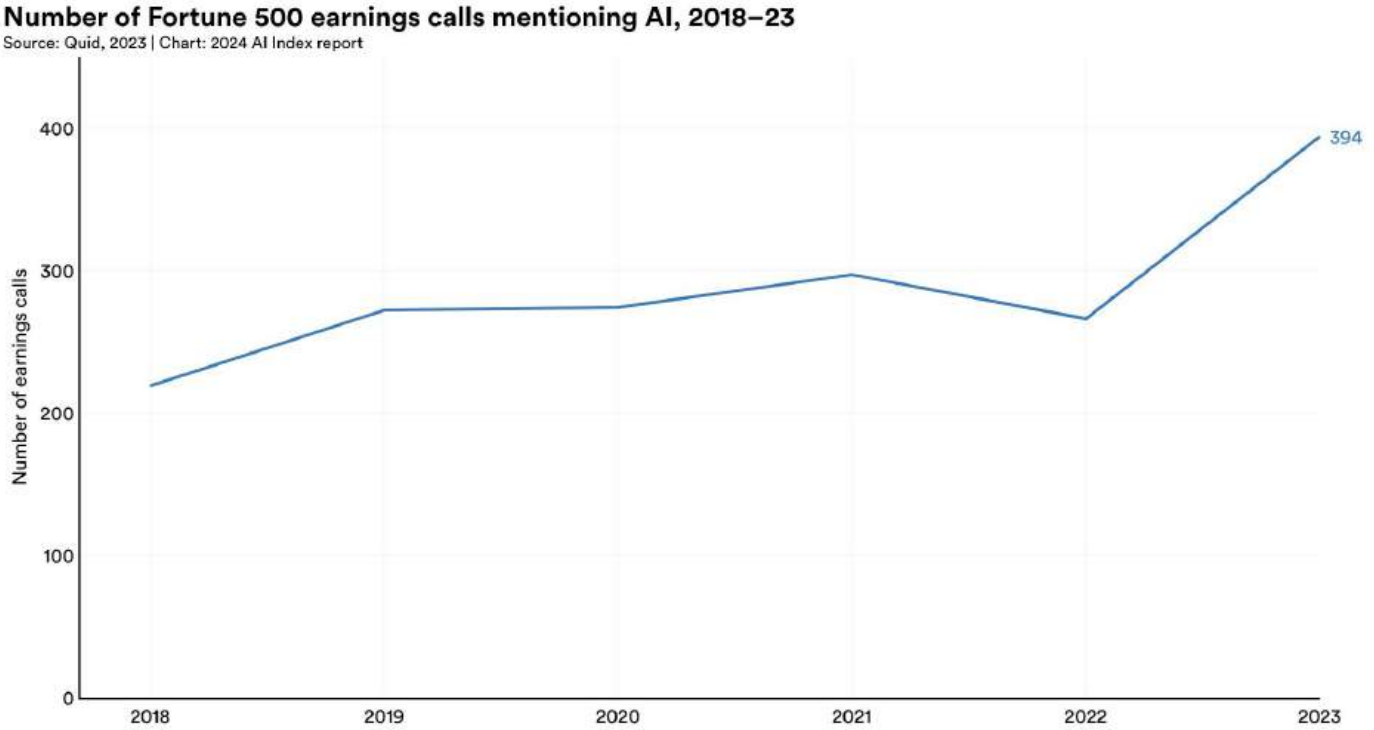


Figure 4.4.25

2024 was a turning point in organizational adoption

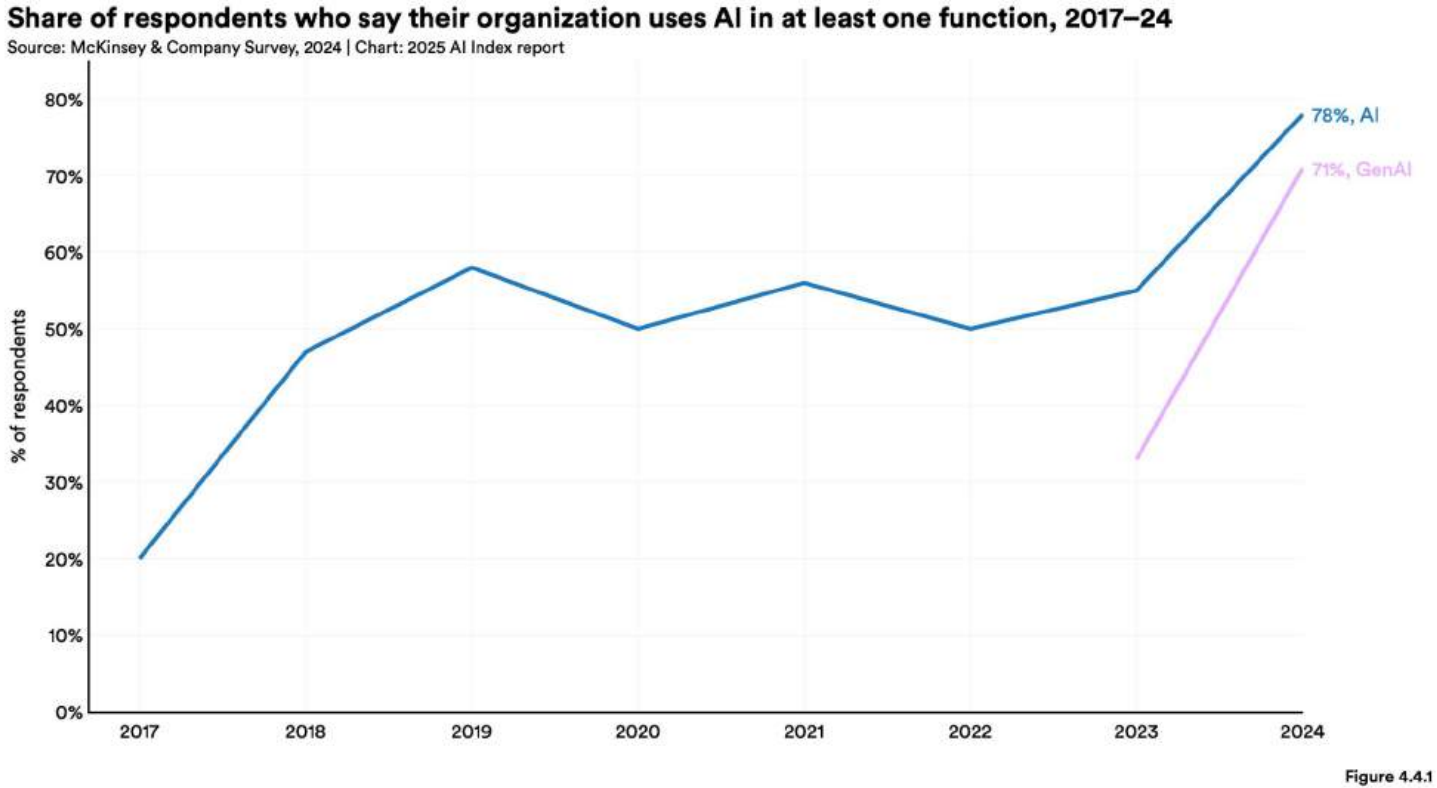
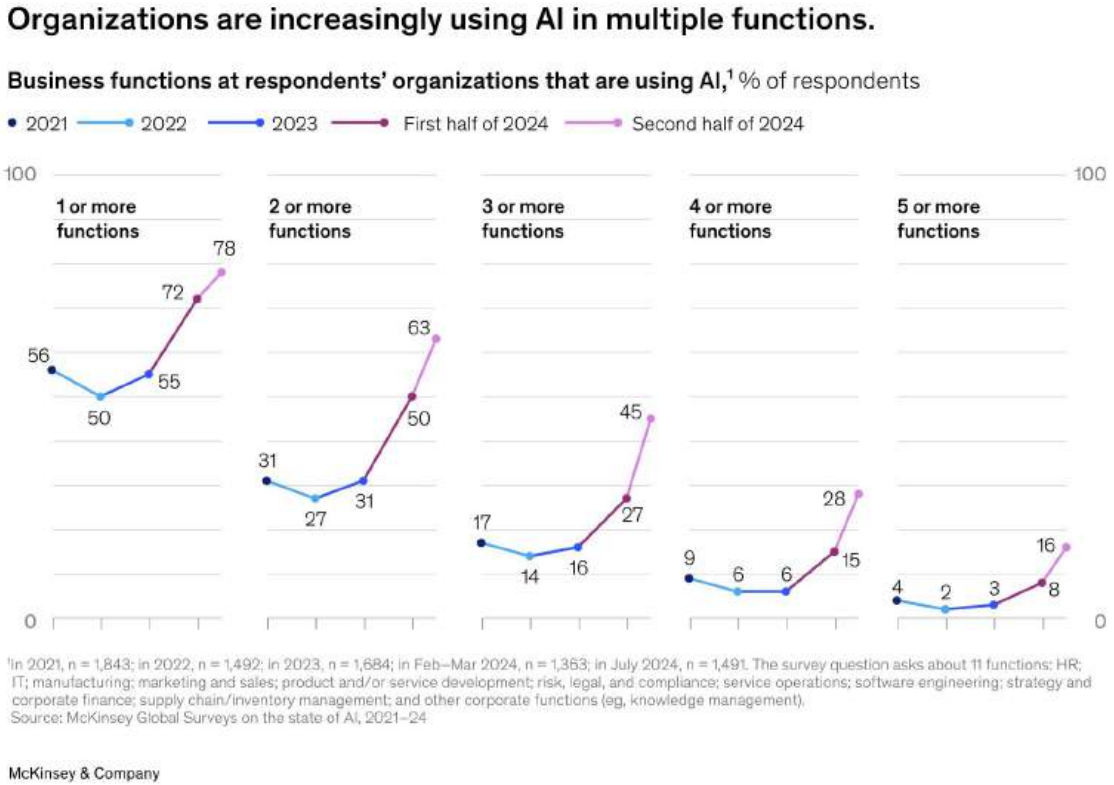


Figure 4.4.1

Usage across functions also increasing



GPTs are GPTs

GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou[1], Sam Manning[1,2], Pamela Mishkin*[1], and Daniel Rock[3]

[1]OpenAI
[2]OpenResearch
[3]University of Pennsylvania

August 22, 2023

(1) Pervasive | (2) Improving | (3) Complementary

AI speeds up work, can sometimes improve quality (computer science)

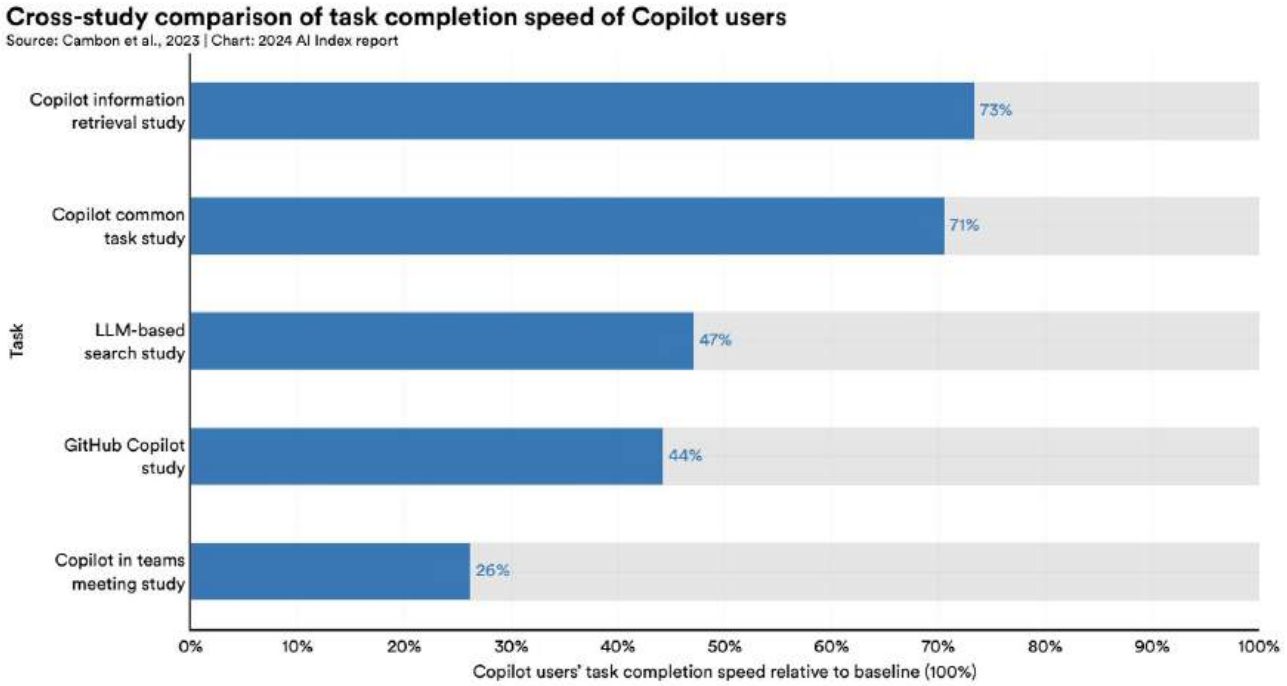


Figure 4.4.19

AI speeds up work, can sometimes improve quality (consulting, customer support)

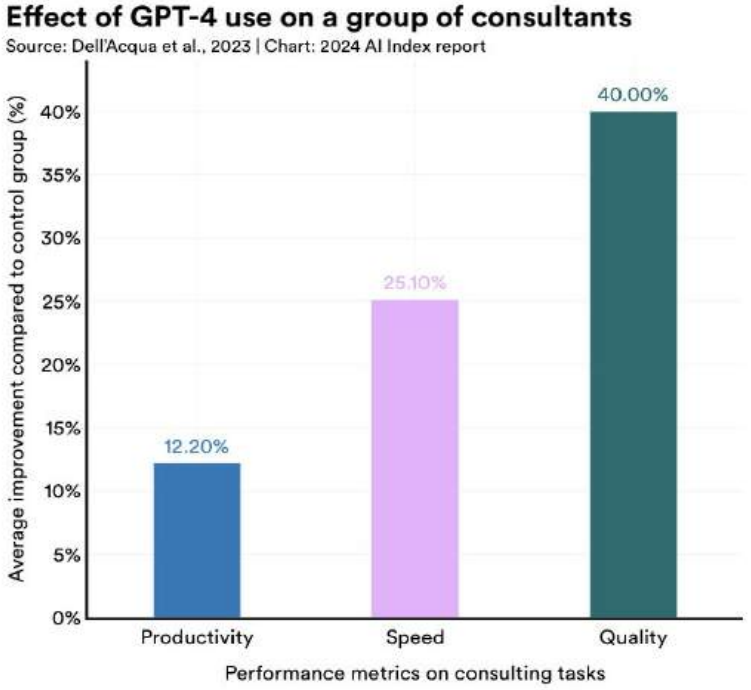


Figure 4.4.20

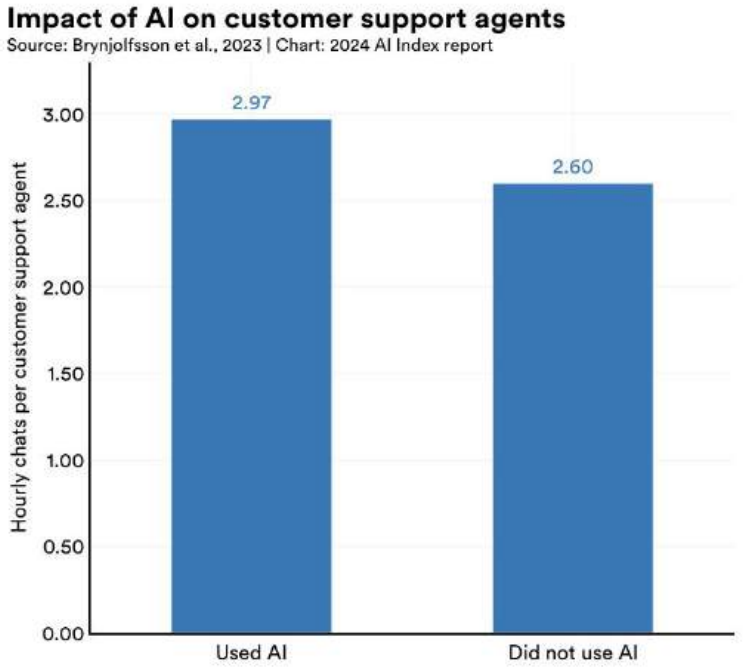


Figure 4.4.21

AI speeds up work, can sometimes improve quality (law)

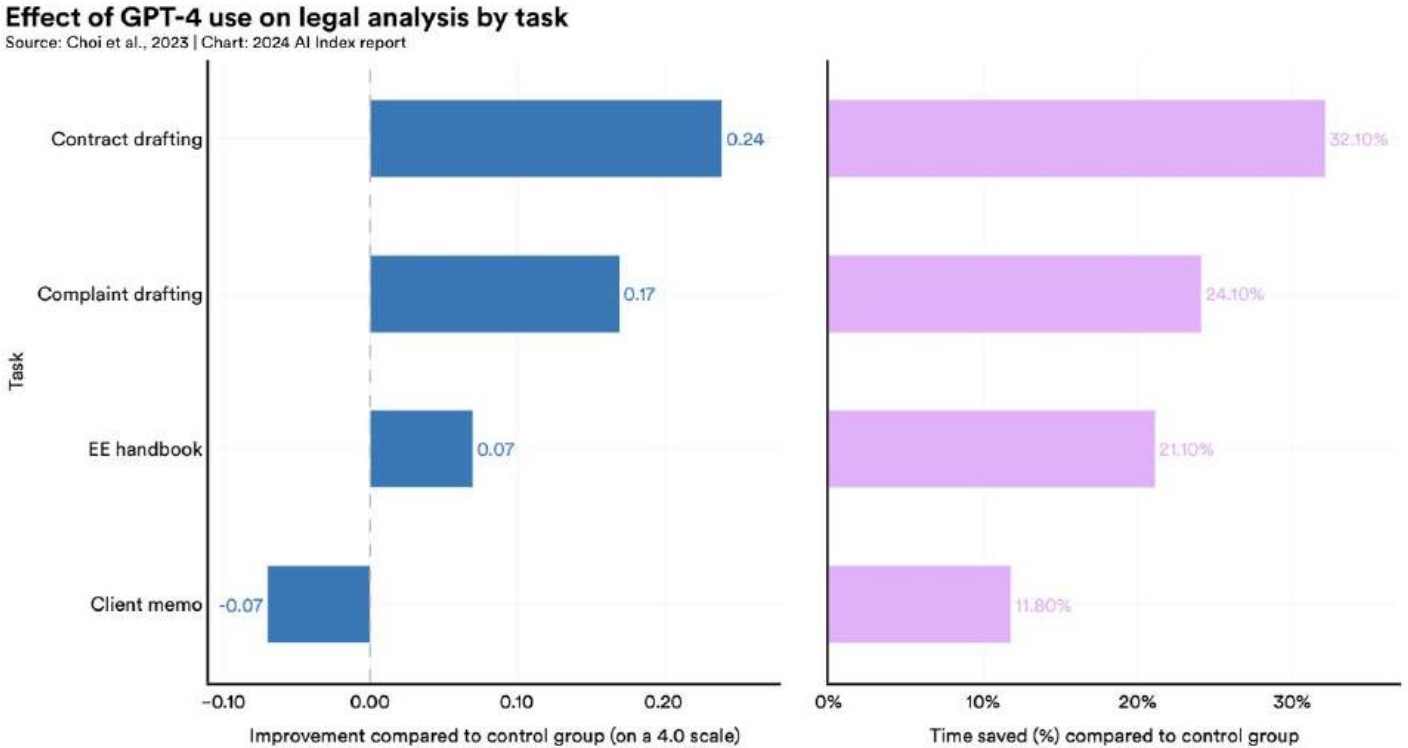
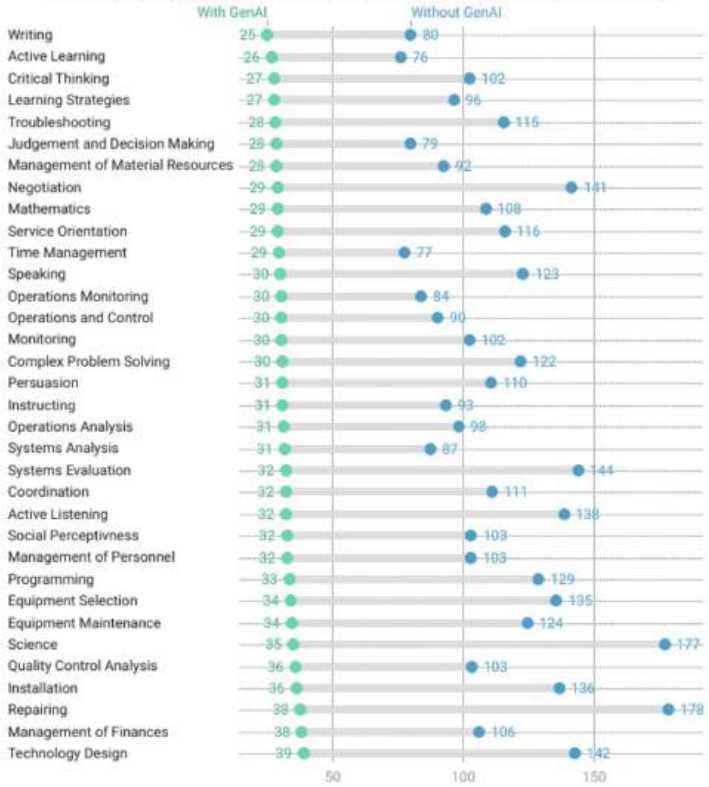


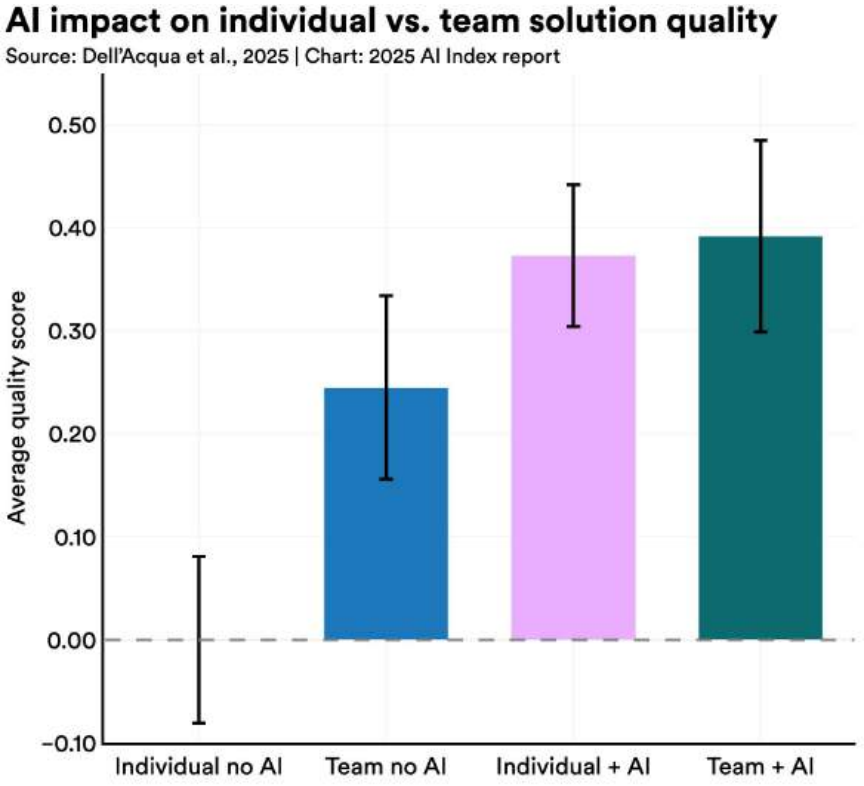
Figure 4.4.22

Other, more aggregate data (Hartley et al., 2025)

Figure 11: Average number of minutes to complete a task with and without Generative AI



Team effects are even stronger (Dell'Acqua et al., 2025)



AI levels up low-skilled workers more than high-skilled

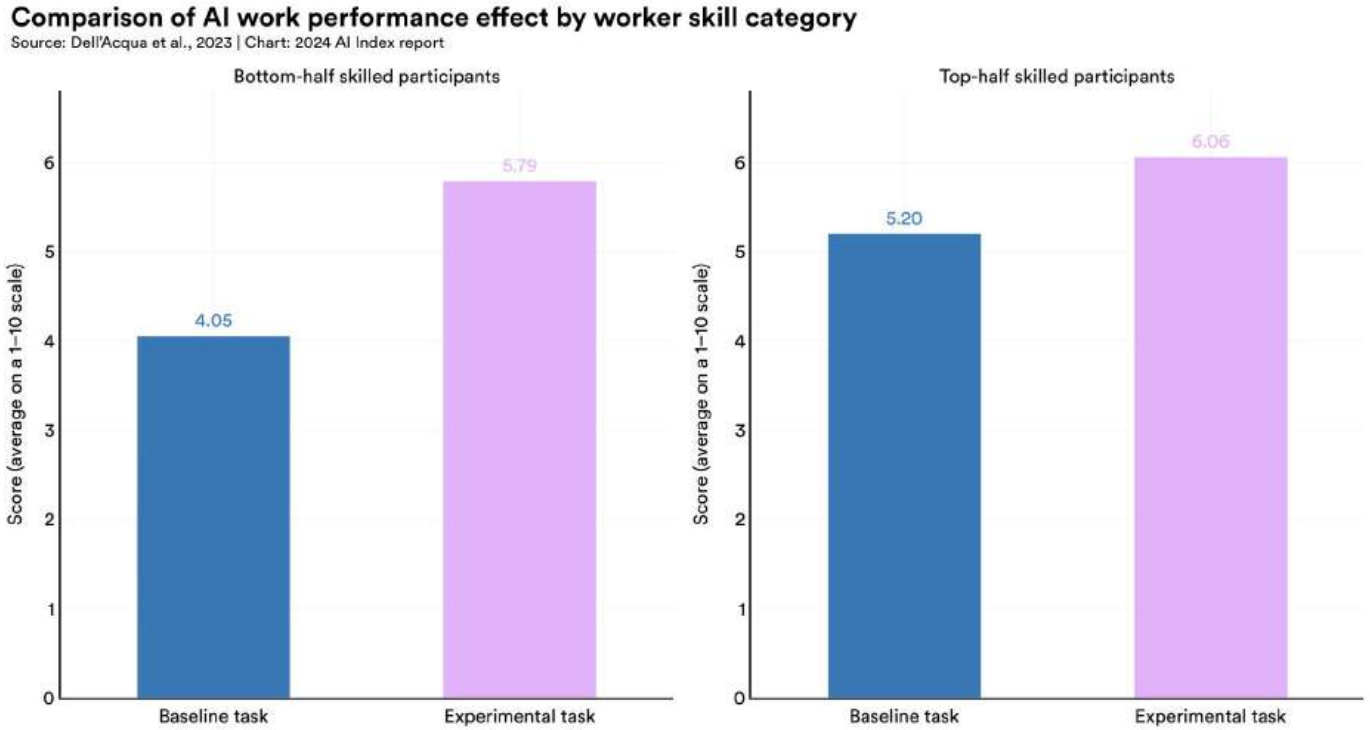


Figure 4.4.23

Leveling up effect a consistent finding across the literature

AI's productivity equalizing effects

Study	Task	Low-skill worker productivity gain	High-skill worker productivity gain
Brynjolfsson et al., 2023	Customer support	34%	Indistinguishable from zero
Dell'Acqua et al., 2023	Consulting	42.96%	16.5%
Cui et al., 2024	Software engineering	21–40%	7–16%
Hoffman et al., 2024	Software engineering	12–27%	5–10%

Figure 4.4.10

Humans should still be kept in the loop

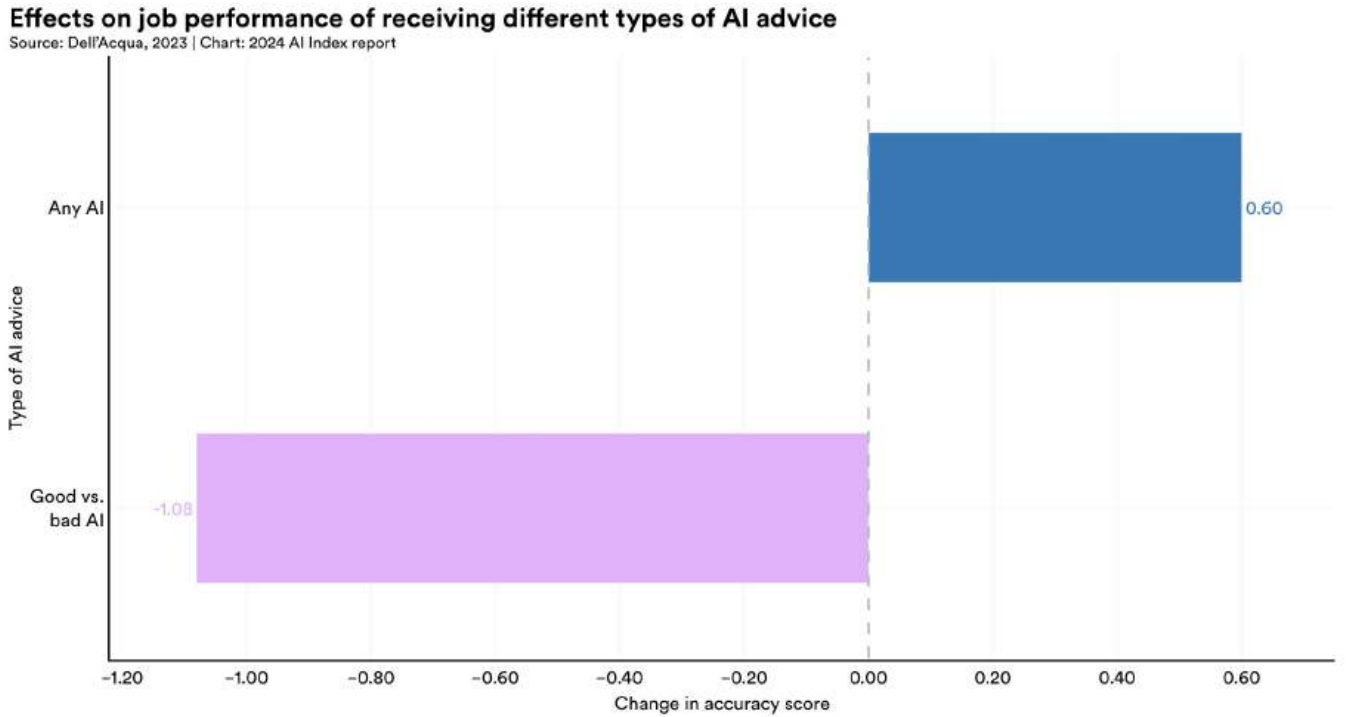
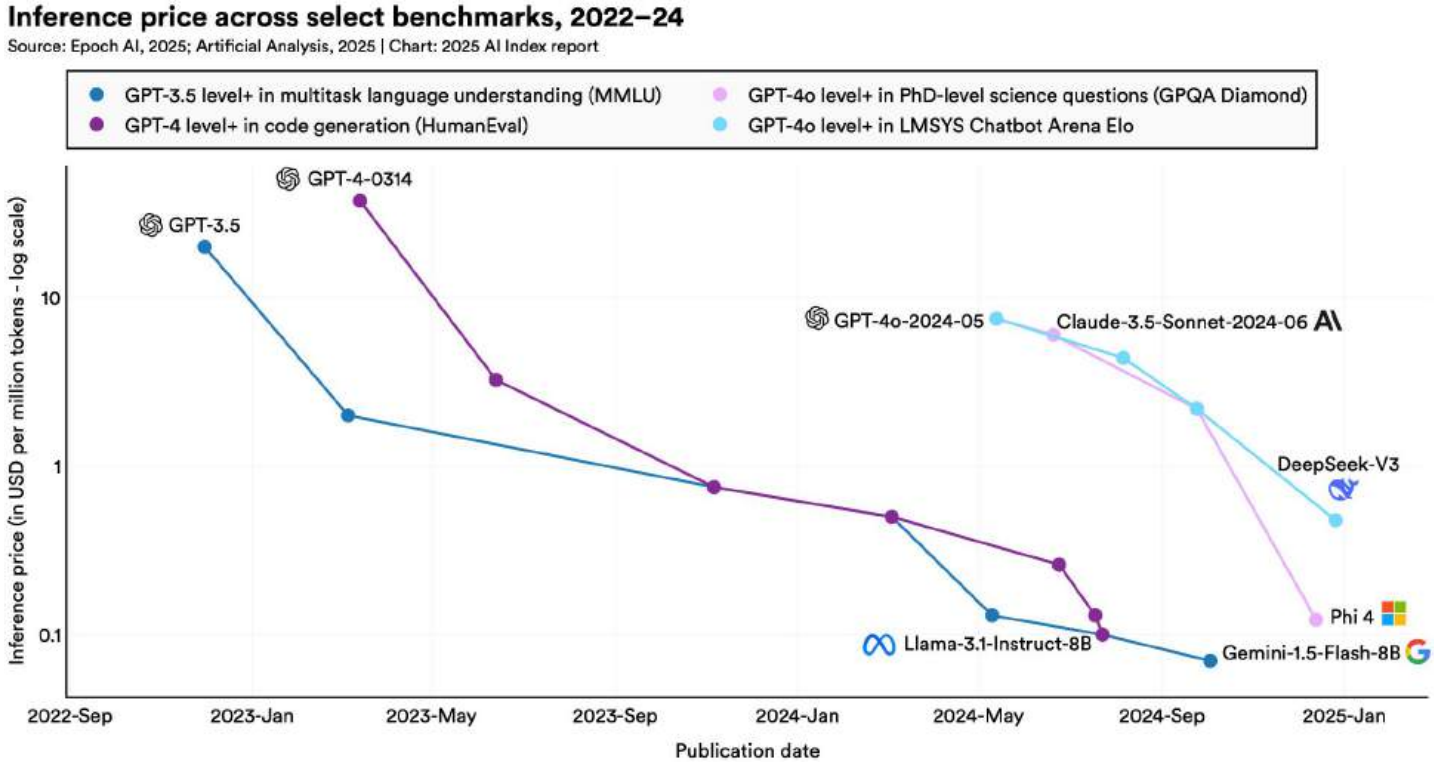
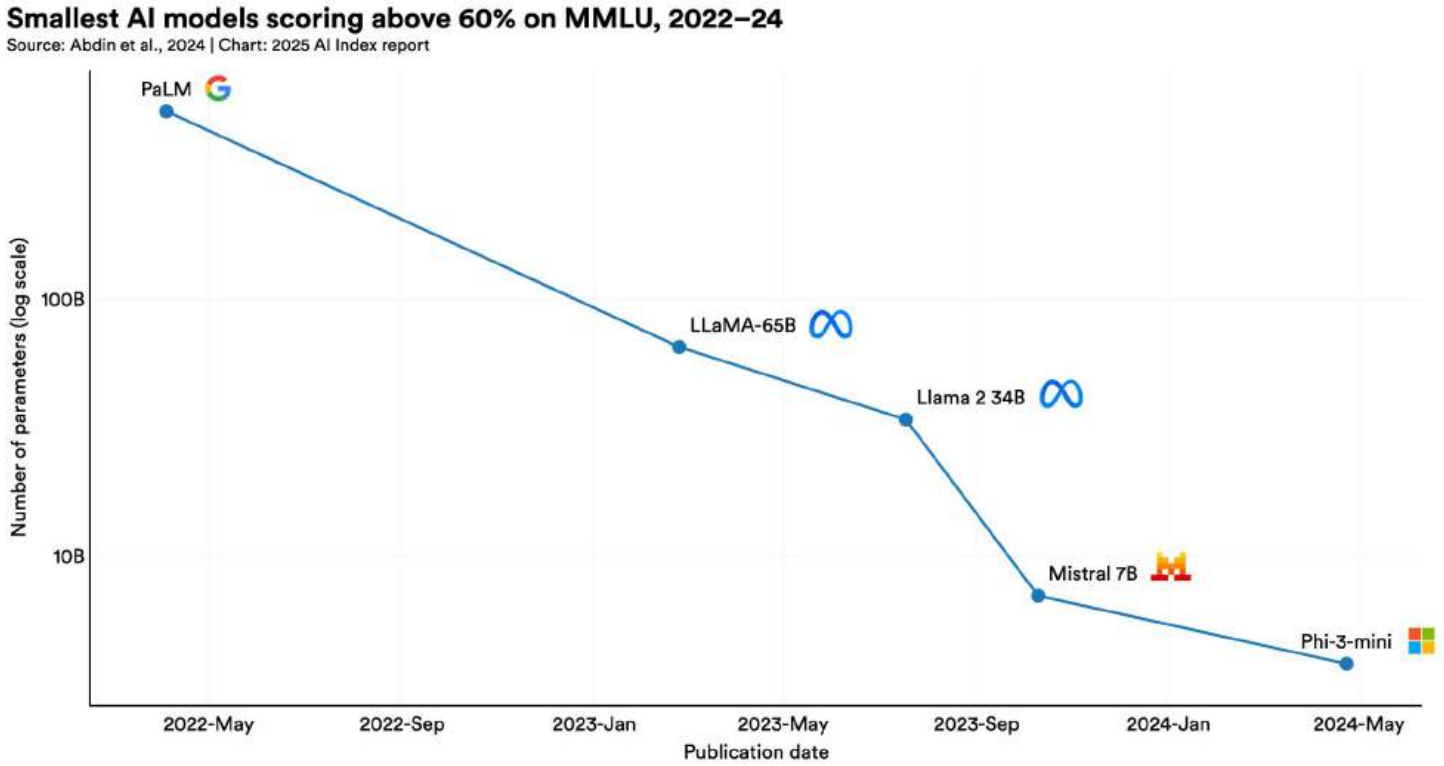


Figure 4.4.24

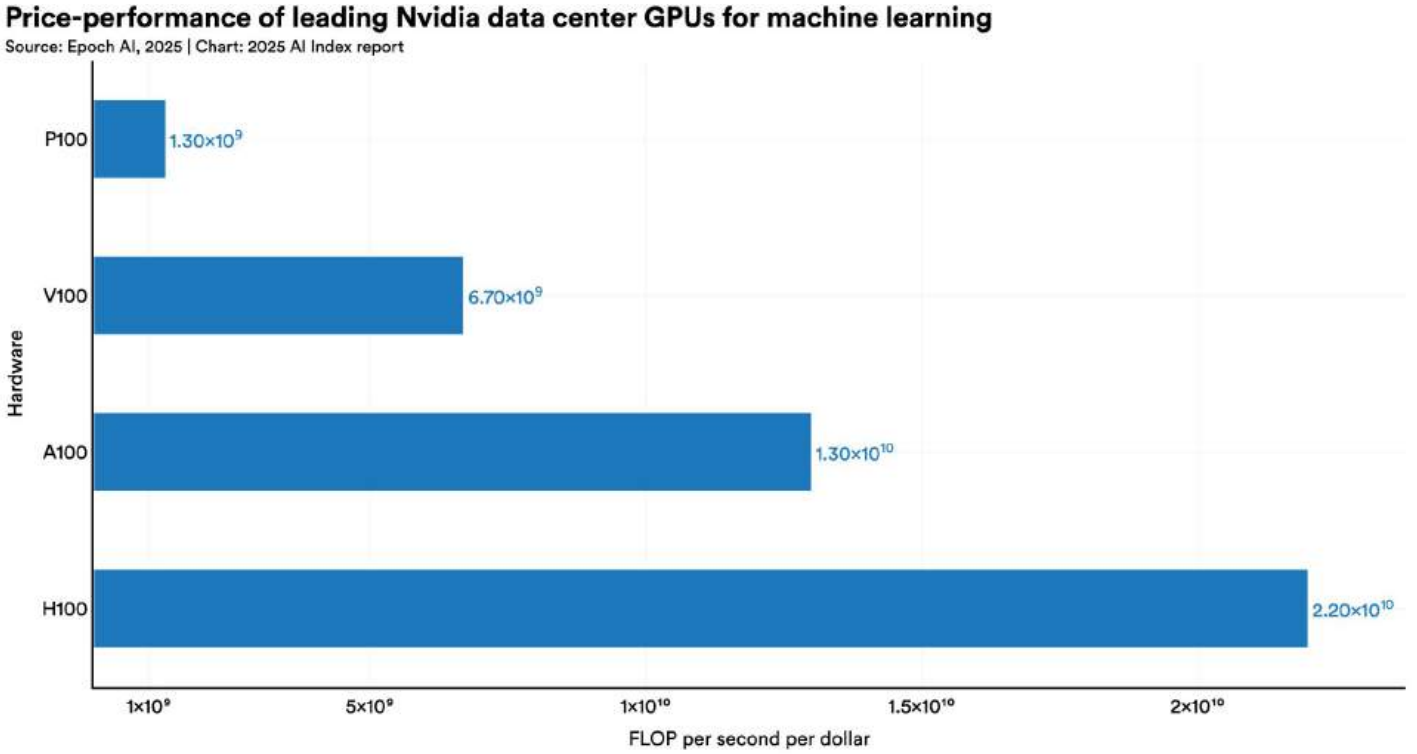
AI becomes more efficient, affordable and accessible: falling inference prices



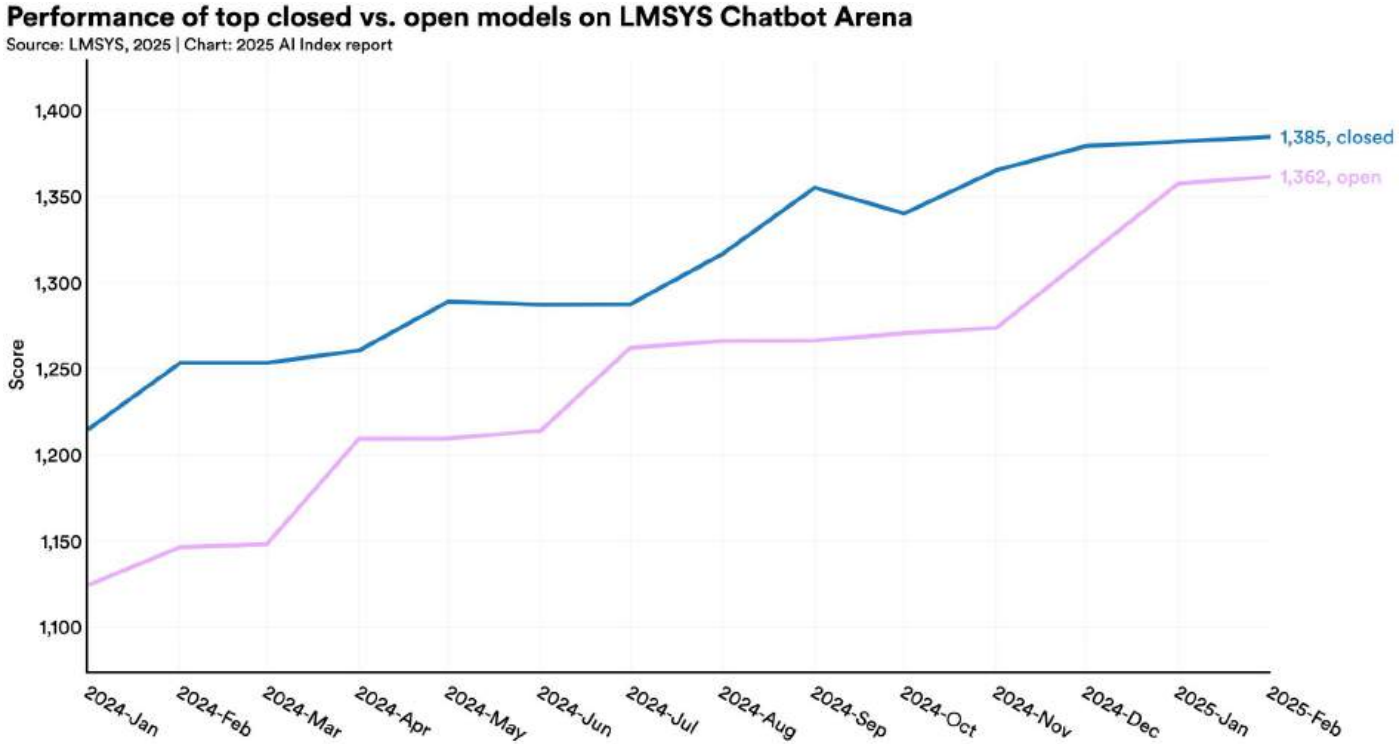
AI becomes more efficient, affordable and accessible: high-performing, small models



AI becomes more efficient, affordable and accessible: energy efficiency gains and improved price-performance



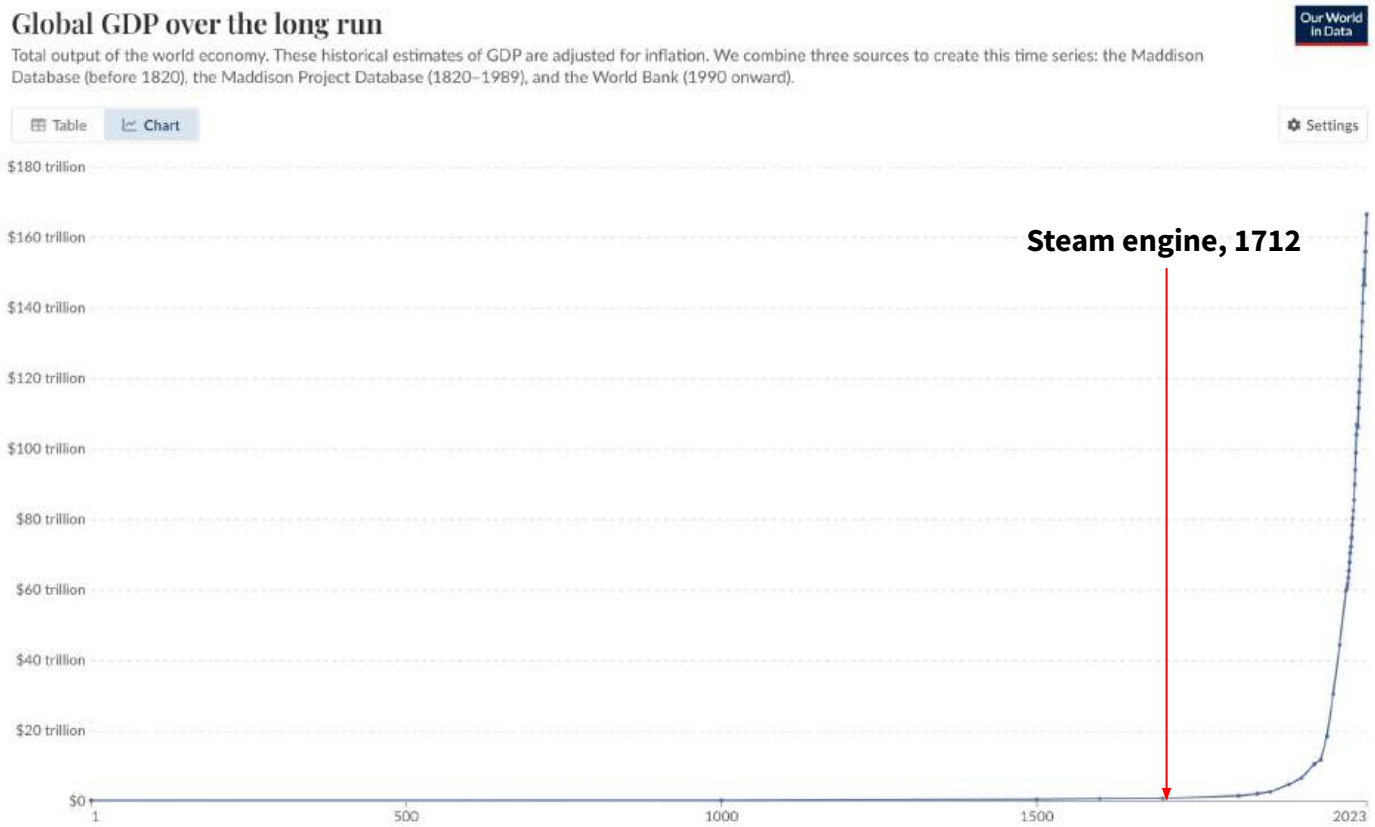
AI becomes more efficient, affordable and accessible: better open-weight models



Key takeaway: businesses are embracing AI

Key question(s): how much will AI impact productivity? What impact will it have on labor?

The steam engine changed history, is AI next?



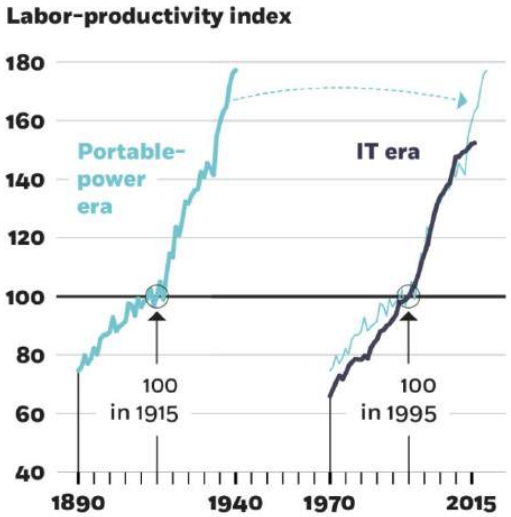
Massive productivity impacts take time to realize

The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox

Paul A. David

The American Economic Review, Vol. 80, No. 2, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association (May, 1990), pp. 355-361 (7 pages)

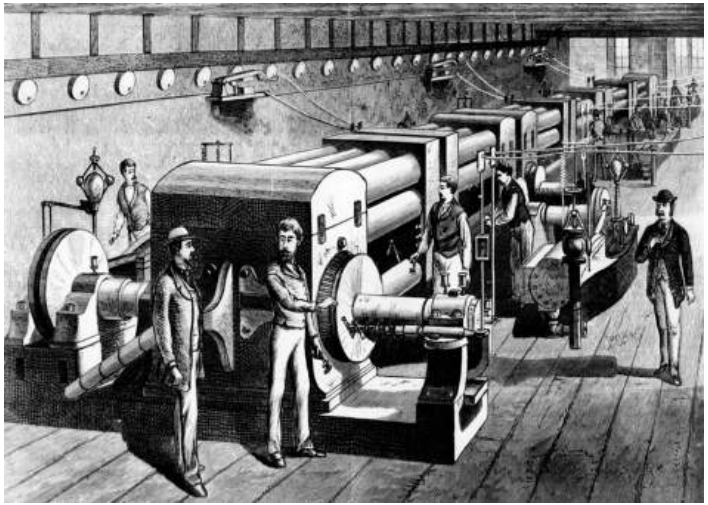
https://www.jstor.org/stable/2006600



Syverson, 2017

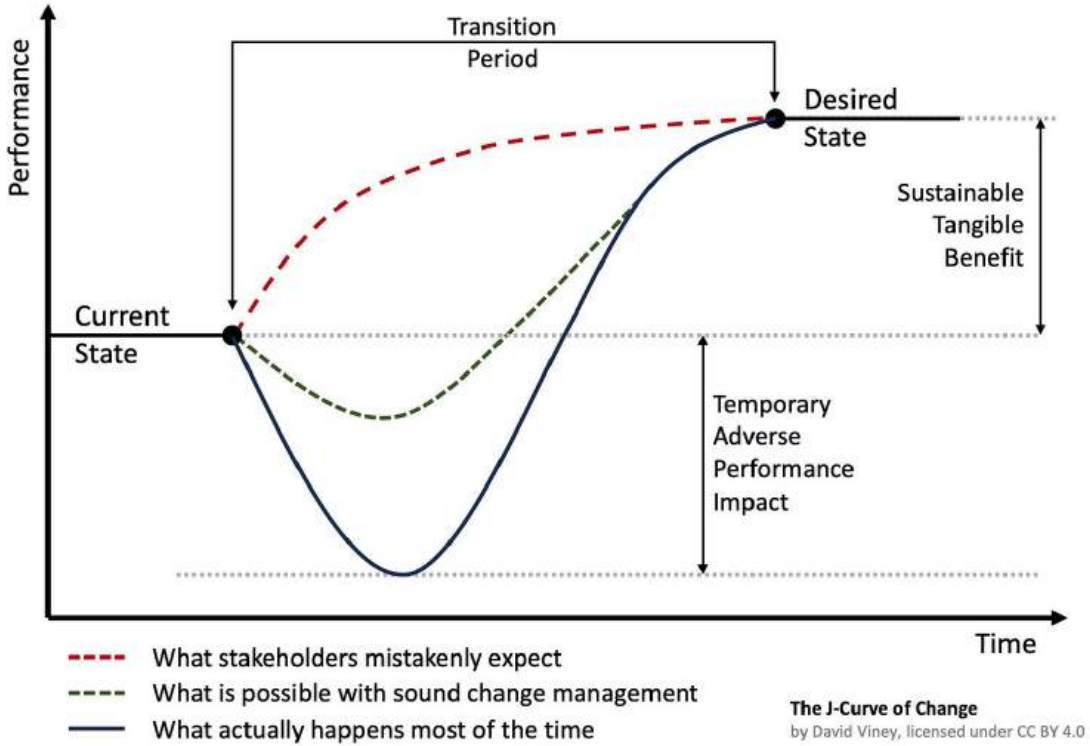
Steam (line-shaft, central) power versus electric (localized, per-unit power)

Old steam factory, Leipzig



Edison’s Pearl Street Station

The productivity J-Curve (Viney, 2022)



AI taking entry level jobs? (May, 2025)

SUCCESS · AI

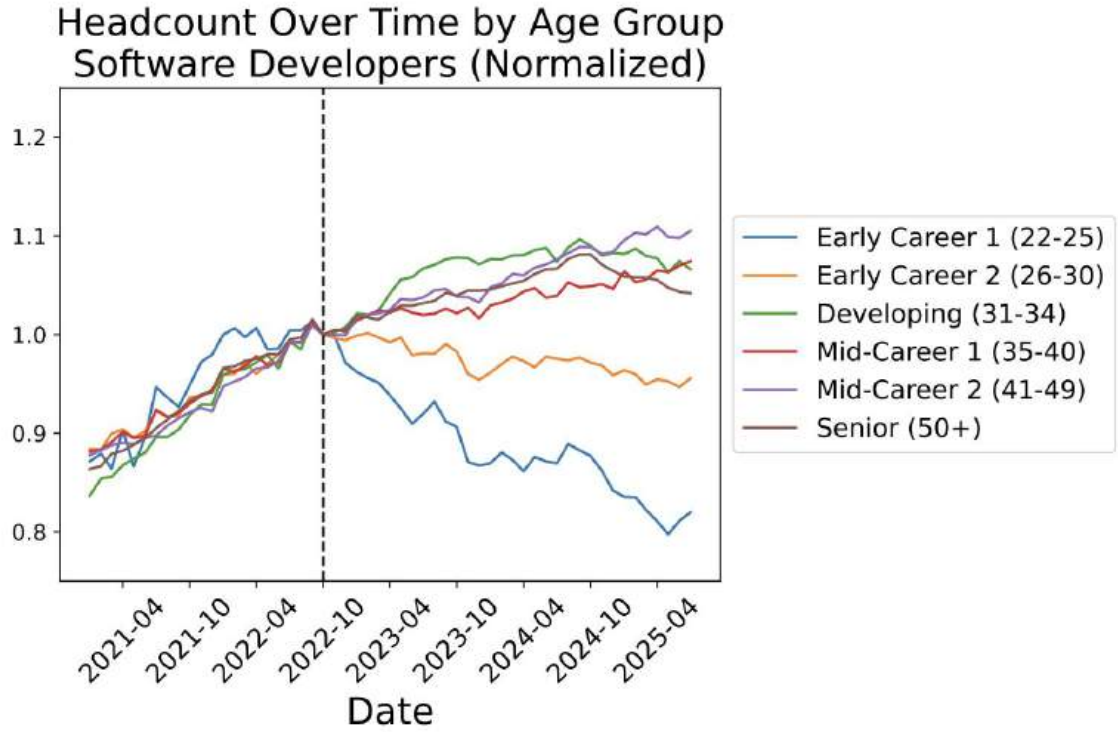
AI is 'breaking' entry-level jobs that Gen Z workers need to launch careers, LinkedIn exec warns

BY **JASON MA**
WEEKEND EDITOR

May 25, 2025 at 1:51 PM EDT

"Breaking first is the bottom rung of the career ladder," said LinkedIn's chief economic opportunity officer, Aneesh Raman.
GETTY IMAGES

New data paints grim picture for entry-level workers (Brynjolfsson, 2025)



New data paints grim picture for entry-level workers (Brynjolfsson, 2025)

This paper examines changes in the labor market for occupations exposed to generative artificial intelligence using high-frequency administrative data from the largest payroll software provider in the United States. We present six facts that characterize these shifts. We find that since the widespread adoption of generative AI, early-career workers (ages 22-25) in the most AI-exposed occupations have experienced a 13 percent relative decline in employment even after controlling for firm-level shocks. In contrast, employment for workers in less exposed fields and more experienced workers in the same occupations has remained stable or continued to grow. We also find that adjustments occur primarily through employment rather than compensation. Furthermore, employment declines are concentrated in occupations where AI is more likely to *automate*, rather than *augment*, human labor. Our results are robust to alternative explanations, such as excluding technology-related firms and excluding occupations amenable to remote work. These six facts provide early, large-scale evidence consistent with the hypothesis that the AI revolution is beginning to have a significant and disproportionate impact on entry-level workers in the American labor market.

Will we have more automation or augmentation? (Brynjolfsson, 2022)

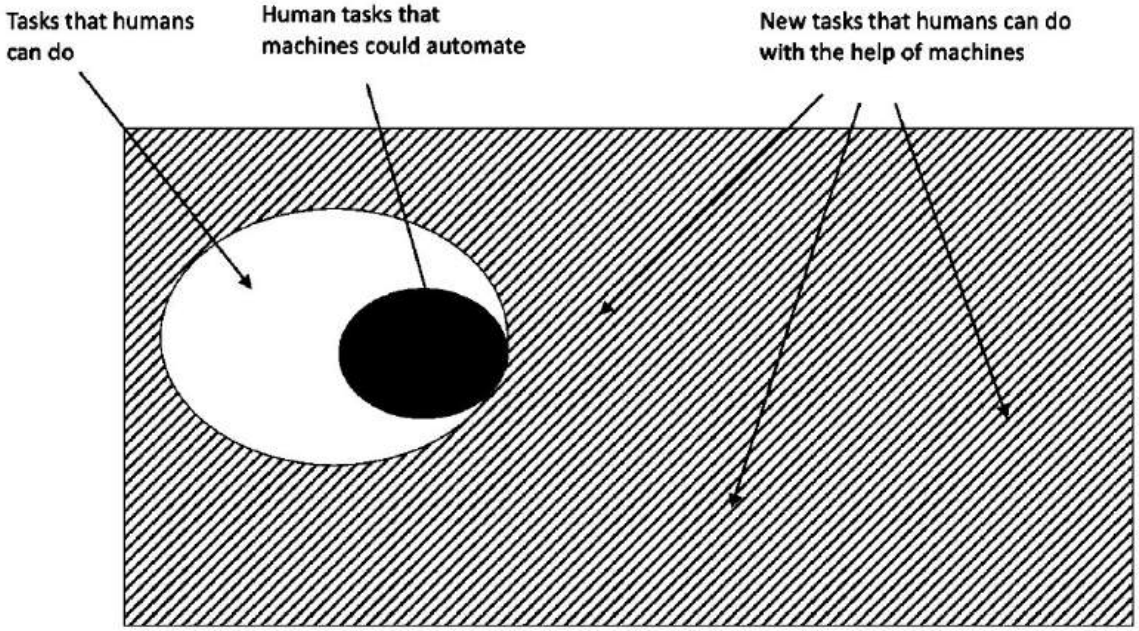
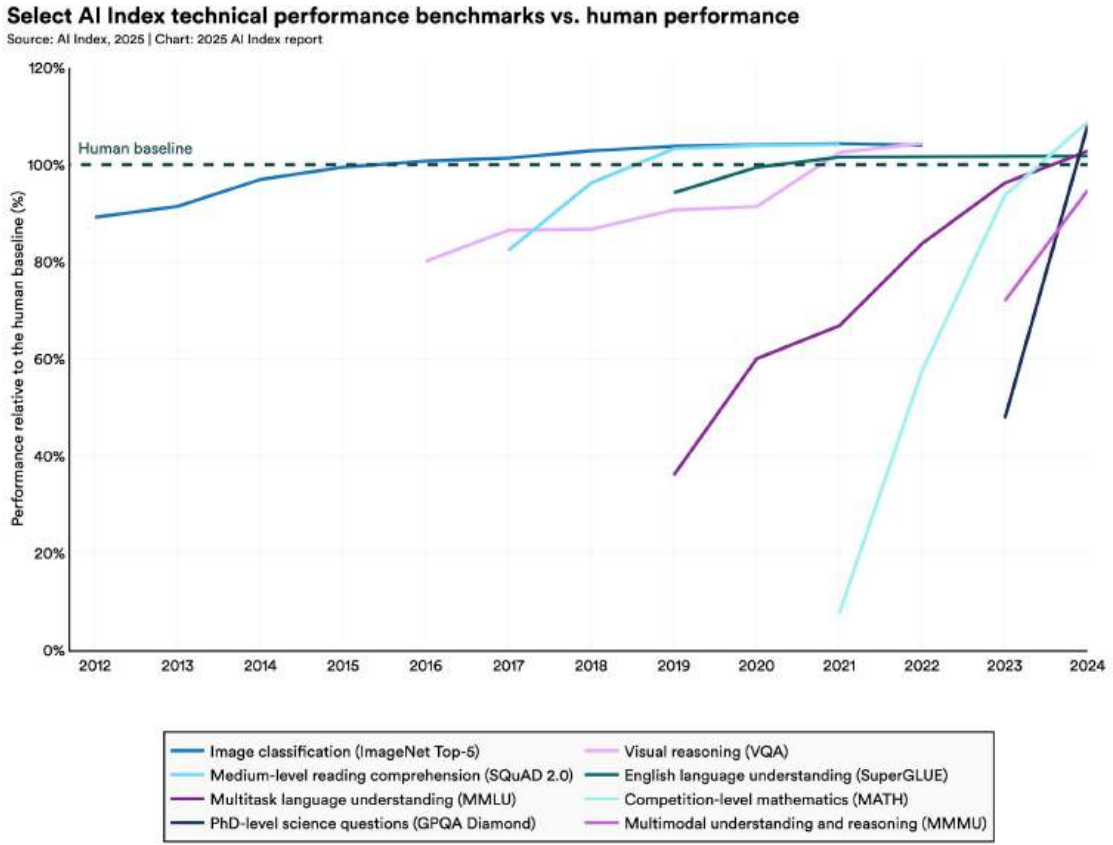


Figure 1: There is far more opportunity in augmenting humans to do new tasks, rather than automating what they can already do

4. Where is the ecosystem?

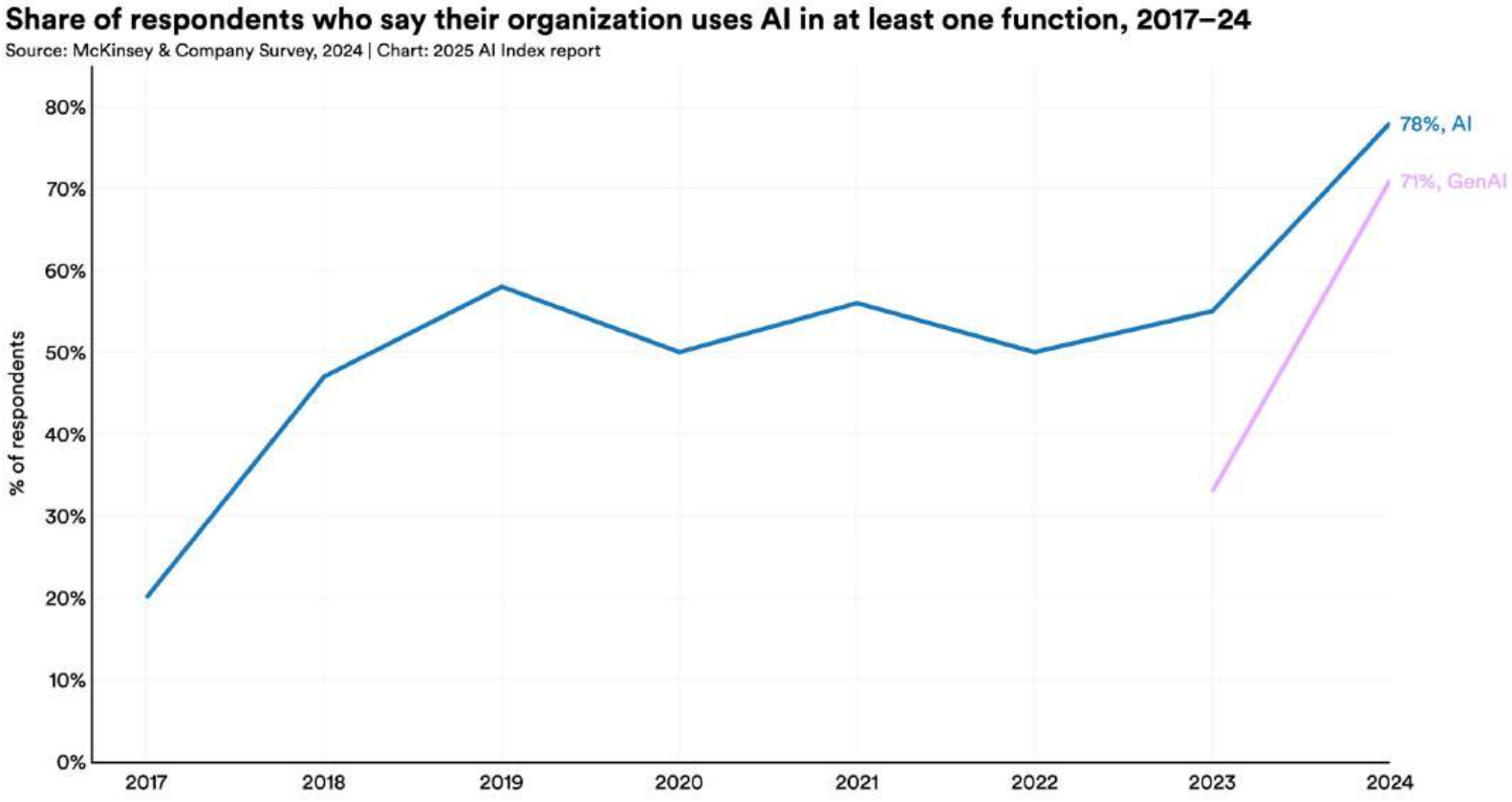
Premise #1: AI is already here

AI beats humans on most tasks

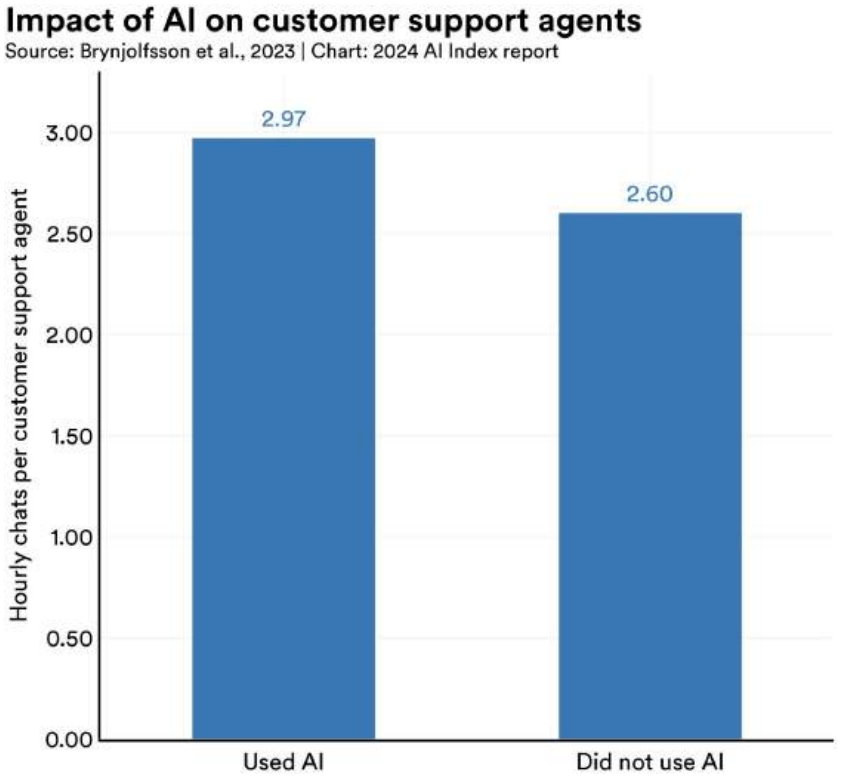


Premise #2: AI is already changing business

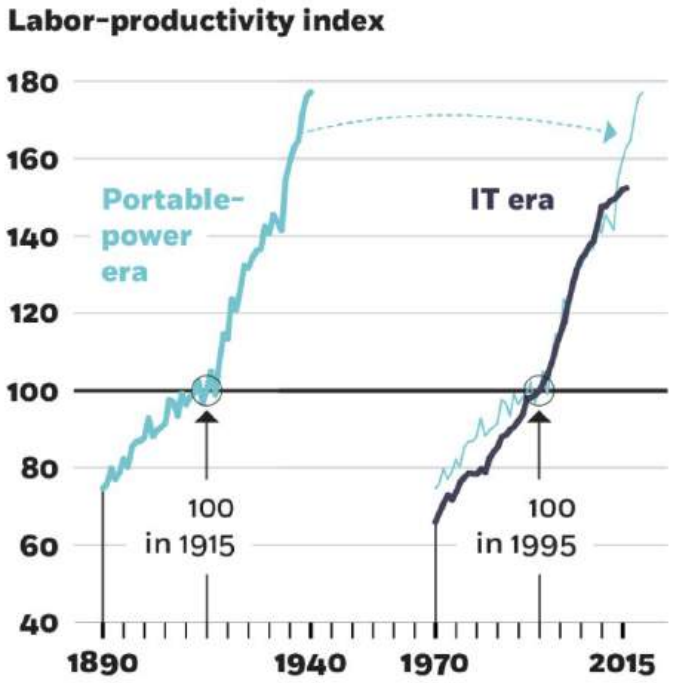
Business usage of AI is increasing



Research highlighting productivity impacts (AI Index, 2025)



Productivity impacts take time



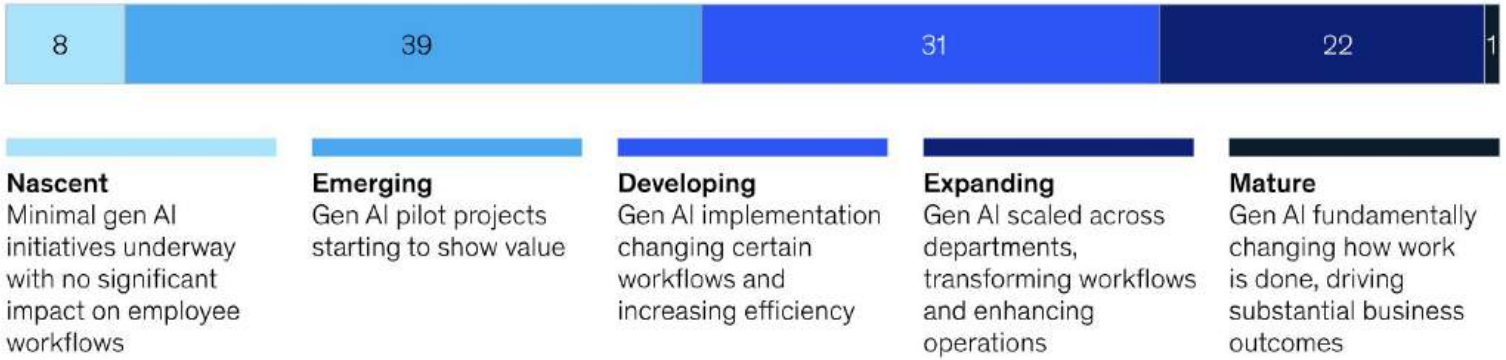
Syverson, 2017

Premise #3: Many businesses are struggling with integration

Leaders know they lag behind (McKinsey, 2024)

Only 1 percent of C-suite respondents describe their gen AI rollouts as mature.

US C-suite's categorization of organizations' level of gen AI maturity, % of respondents



Note: Figures may not sum to 100%, because of rounding.
Source: McKinsey US CxO survey, Oct–Nov 2024 (n = 118)

McKinsey & Company

NY Times, August 2025

Companies Are Pouring Billions Into A.I. It Has Yet to Pay Off.

Corporate spending on artificial intelligence is surging as executives bank on major efficiency gains. So far, they report little effect to the bottom line.

Premise #4: Businesses need AI education

Why? 🧠 Strategy Questions

1. **How will AI enhance our core value proposition or create new revenue streams over the next 3–5 years?**
   Pushes leadership to think beyond incremental gains toward strategic transformation.

2. **What are our competitors doing with AI, and how do we differentiate ourselves in this evolving landscape?**
   Promotes proactive positioning and market awareness.

3. **Are we making build, buy, or partner decisions strategically — and do we understand the trade-offs of each?**
   Helps structure a scalable and cost-effective AI roadmap.

4. **How are we future-proofing our AI investments to remain adaptable amid fast-moving technology and regulation?**
   Encourages resilience, flexibility, and regulatory readiness.

Why? ⚙️ Operational Questions

1. **Do we have the infrastructure (e.g., cloud, compute, storage) in place to support model training, deployment, and scaling?**
   Ensures technical feasibility and performance readiness.

2. **What is our process for testing, validating, and approving AI systems before they go live?**
   Reduces operational risk and improves reliability.

3. **How will AI systems be monitored in production — and who is responsible for identifying issues or failures?**
   Defines ownership and operational resilience mechanisms.

4. **Are we integrating AI into existing workflows in a way that augments, rather than disrupts, human decision-making?**
   Focuses on practical adoption and user alignment.

Why? 🔐 Data & Governance Questions

1. **What data sources will power this AI system, and do we have clear rights and documentation for their use?**
   Ensures lawful access, licensing clarity, and reduces IP/privacy risk.

2. **How are we assessing the quality, representativeness, and bias of the data we are using?**
   Directly addresses model performance, fairness, and reliability.

3. **Who is responsible for data stewardship — including data updates, lineage tracking, and lifecycle management?**
   Establishes operational accountability and long-term maintainability.

4. **Do we have mechanisms in place to detect and respond to data drift or model degradation over time?**
   Promotes proactive monitoring and sustained system performance.

Why? 👥 People & Culture Questions

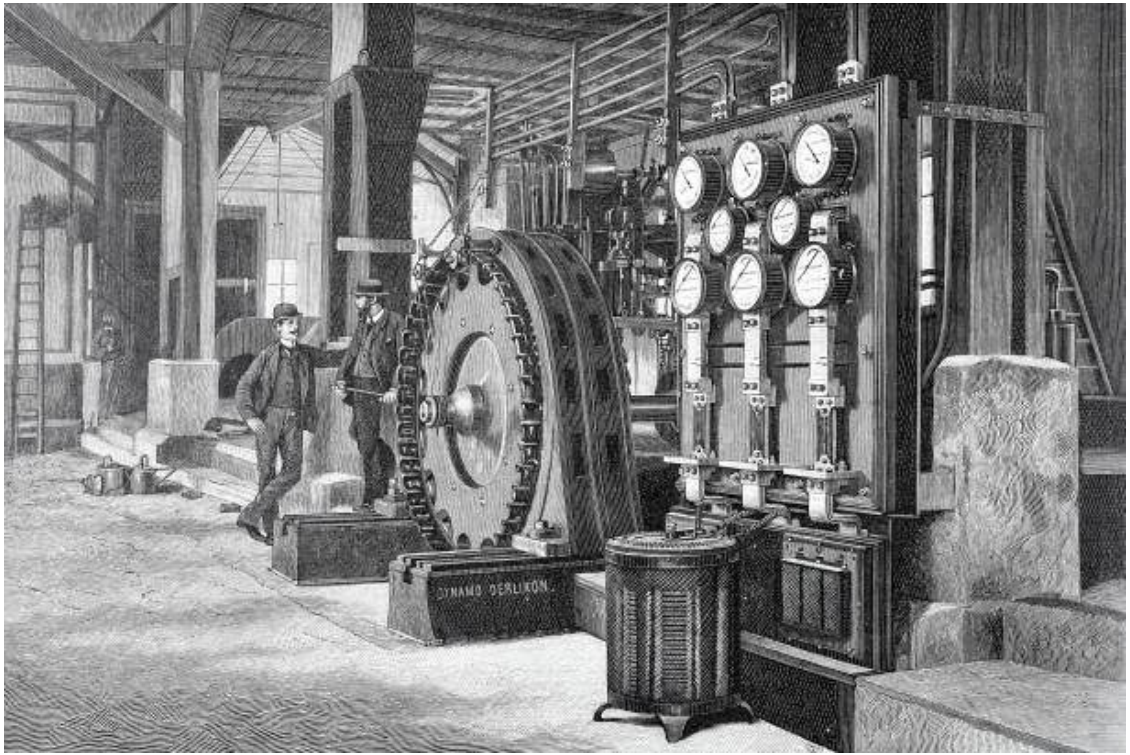
1. **Do our employees understand how AI is being used in the organization, and are we communicating its impact transparently?**
   Builds trust, reduces fear, and encourages engagement across teams.

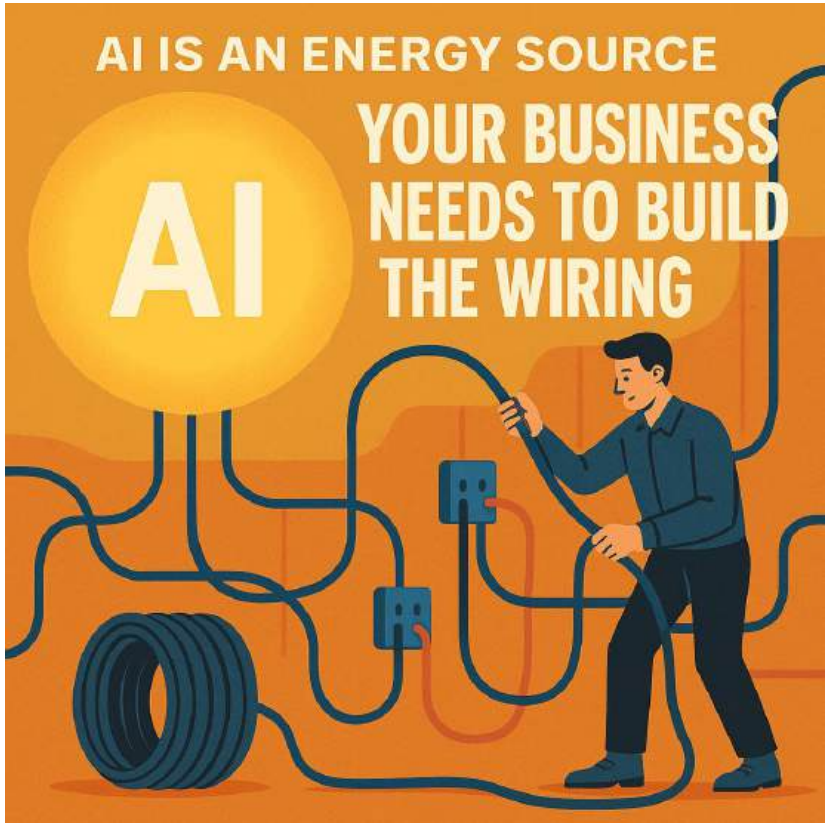
2. **Are we investing in upskilling and training so that staff can work effectively with AI tools?**
   Ensures that employees are equipped to adapt and collaborate with new technologies.

3. **How are we addressing concerns around job displacement, role changes, and ethical use of AI among our workforce?**
   Fosters a culture of responsibility and psychological safety.

4. **Do we have champions or cross-functional teams in place to model responsible AI adoption and support cultural change?**
   Encourages grassroots buy-in and organizational momentum.

You are the factory manager in the 1890s




AI IS AN ENERGY SOURCE
AI
YOUR BUSINESS NEEDS TO BUILD THE WIRING

nmaslej@stanford.edu