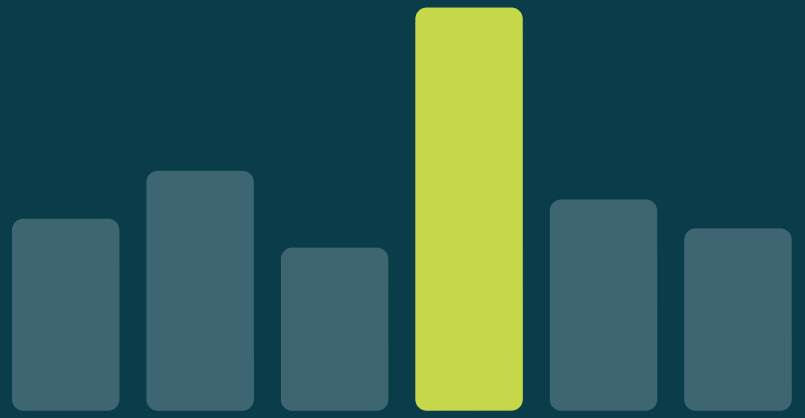
The gist of AI



What it is, why it isn’t paying off yet,
and what to do about it

Jaime Díaz-Beltrán · I build AI products and measure whether they pay off

Before: Mercado Libre, Nubank, Siemens Energy

AUGUST 1955 · DARTMOUTH COLLEGE

“We propose that a **2 month, 10 man study** of artificial intelligence be carried out during the summer of 1956…”

McCarthy, Minsky, Rochester and Shannon, in the proposal that coined the term “artificial intelligence”

It took seventy years.

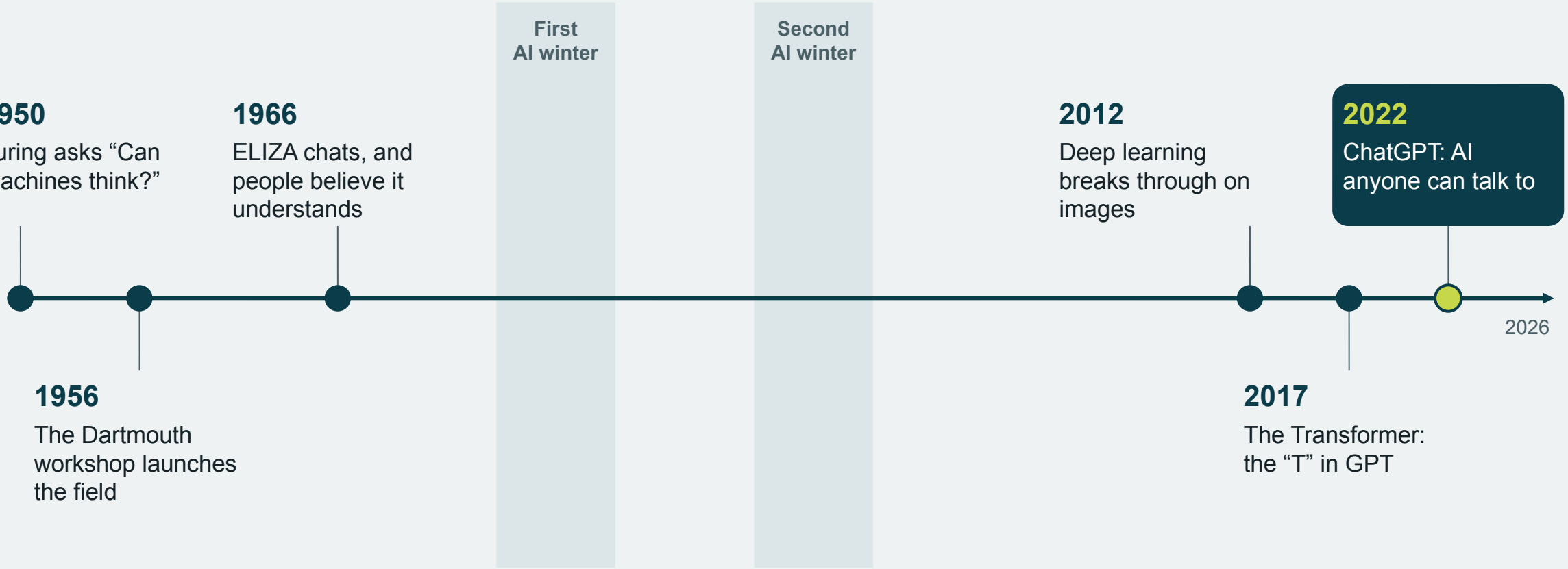
They thought the technology would be easy. Today, most firms think the redesign will be.

Sources: [1] McCarthy et al. 1955 · Photo: Daderot, CC0, via Wikimedia Commons



Seventy years, two winters, one breakout

Most of the ideas are decades old. The winters came when promises ran ahead of results.

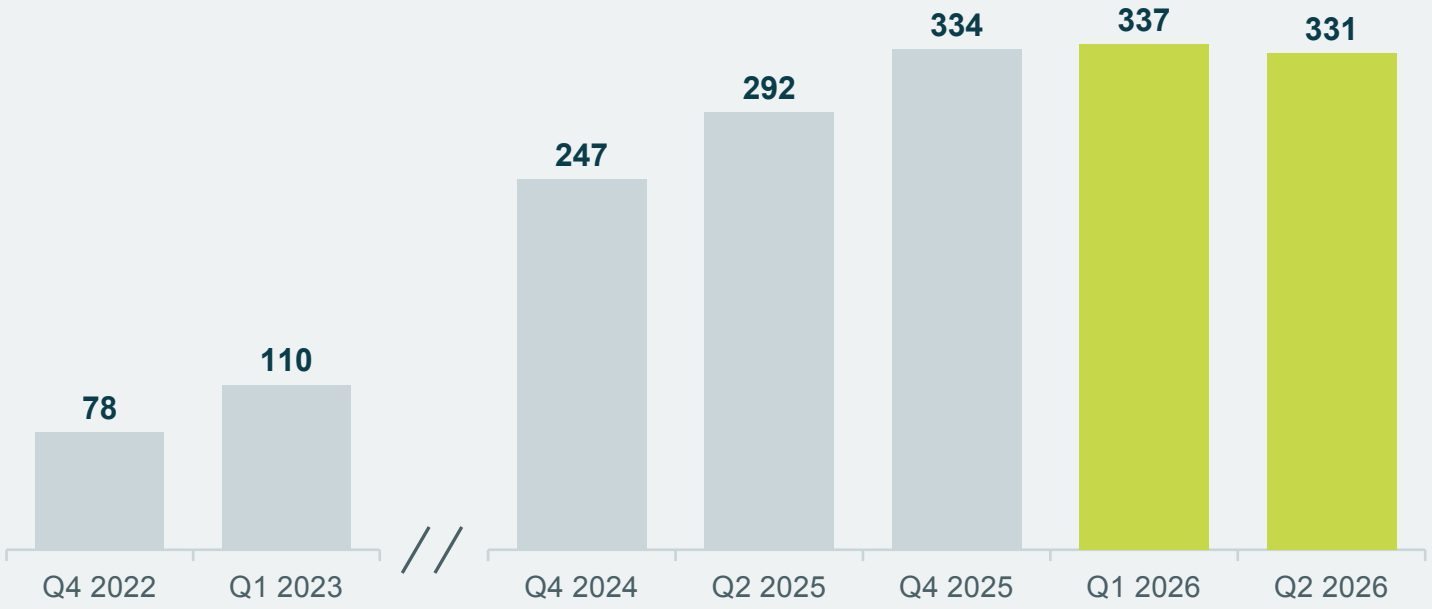


Sources: [2] [1] [3] [4] [5] [6] [7]

In 2022, AI learned to talk to everyone

And two in three S&P 500 earnings calls now mention it

S&P 500 earnings calls mentioning “AI”



Quarters between Q1 2023 and Q4 2024 not shown. The first two bars count companies, later bars count calls.

1.2B

weekly ChatGPT users by September 2026, as reported by OpenAI

Think of it as AI’s Windows moment: not more powerful, just usable by everyone.

Sources: [14] FactSet 2023–26 · [15] OpenAI 2026 · [16] Anand & Wu 2025

THE PUZZLE

Most firms say they use AI. Few see it in the numbers.

I call it the conversion gap: time people save isn't yet becoming firm results

78%

of firms say they use AI

~9,000 executives, 7 countries (NBER, 2026). No vendor sponsor.

but

+0.2%

a year in productivity so far

Same executives. They expect about 0.6% a year next.

Projects pay off. Firm numbers, not yet.

~3 in 4

leaders at large US firms report positive returns (Wharton)

7.5%

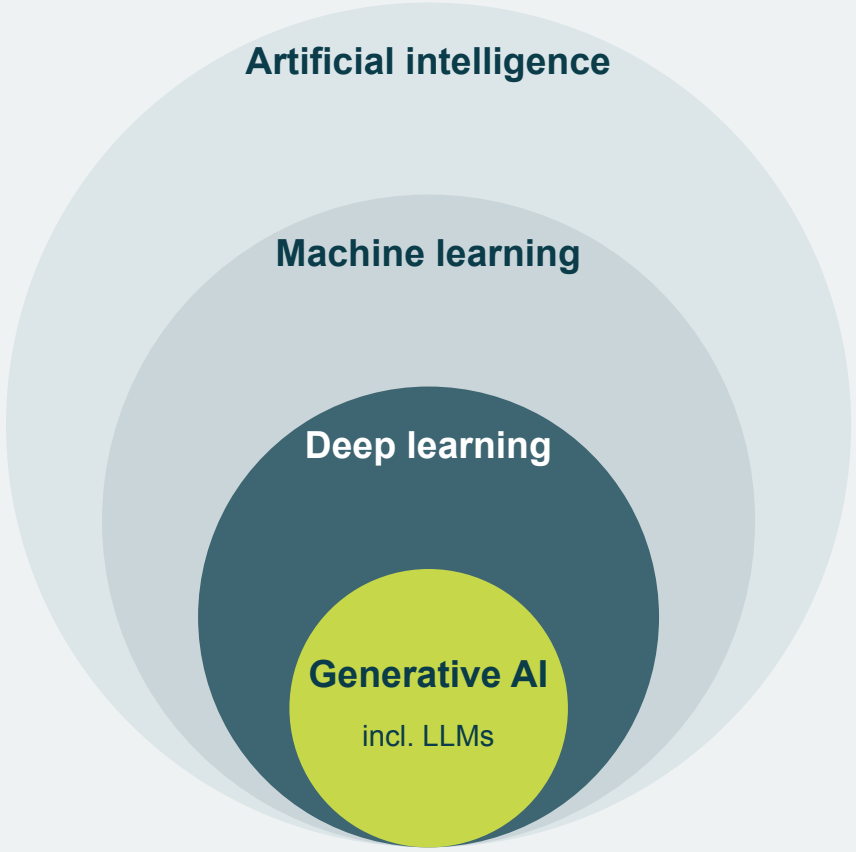
of companies get value at scale, up from 5% (BCG, which sells AI services)

Today's question: why isn't it in the numbers yet, and what do we do about it?

Sources: [17] NBER 2026 · [18] Wharton 2025 · [19] BCG 2025–26

“AI” is like “vehicle”: always ask which kind

You’ve had AI in your pricing, fraud and claims models for years. The new part is generative AI.



Artificial intelligence

Any machine doing something we’d call smart, including rules and search (Deep Blue)

Machine learning

Learns patterns from examples: pricing, fraud, claims and credit models

Deep learning

Large neural networks: image recognition, AlphaFold

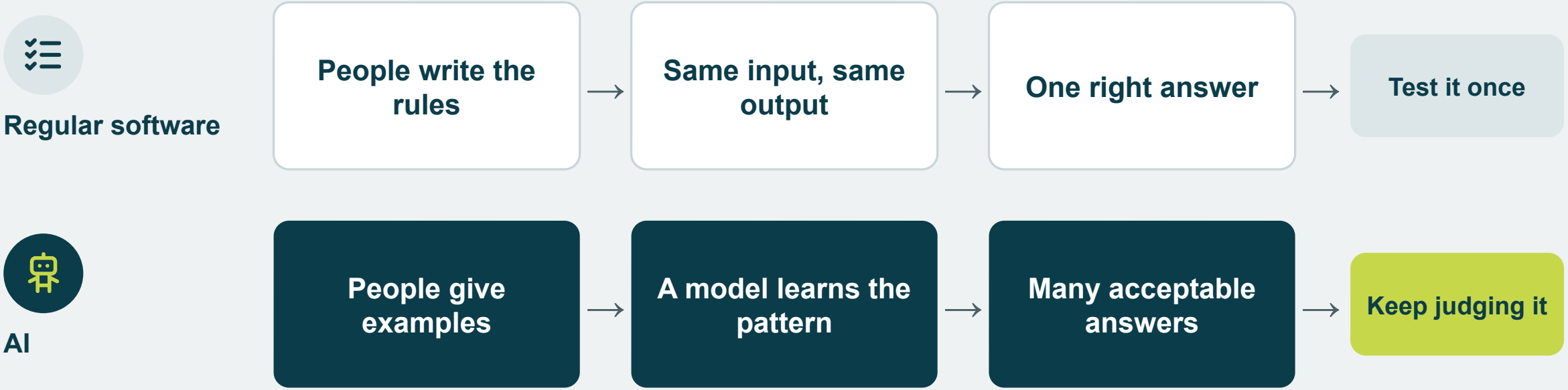
Generative AI

Creates text, images and code. Large language models (LLMs) like ChatGPT, Claude and Gemini

Sources: [20]

Regular software follows rules. AI learns from examples.

"Explain with examples, not instructions." (Cassie Kozyrkov, former Chief Decision Scientist, Google)

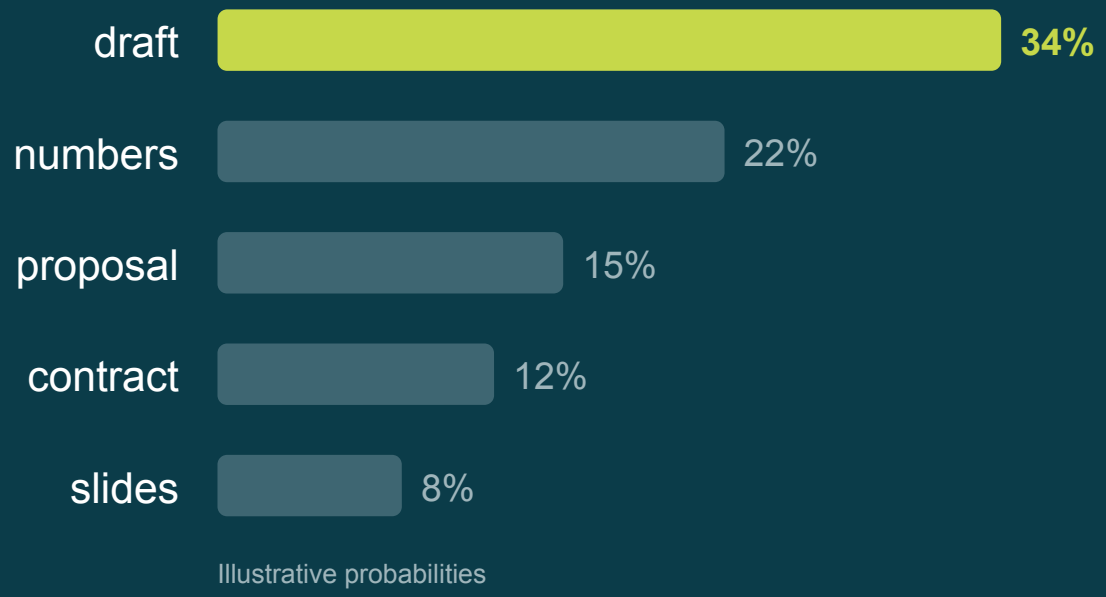


Think of it as the policy manual versus the junior who learned from a thousand past cases.

Sources: [21]

An LLM is the most well-read autocomplete ever built

The partner reviewed the _____



1. It learned to guess the next word from a vast amount of text.
2. To guess well, it had to absorb a lot about the world.
3. It produces what is plausible, not what is verified, so it can be confidently wrong.

Next-word prediction is how it's trained, not the ceiling of what it can do.

Think of it as a brilliant new hire who has read the whole library, has never met your client, and rarely says "I don't know."

Sources: [22]

Every AI term maps to something you already manage

Key terms, in one picture: the brilliant new hire



The brief

Prompt

The instructions and material you give it

The case file on the desk

Context window

All it can see right now. Nothing else counts

School vs. the job

Training vs inference

It learned before you met it. Using it doesn’t retrain it on the spot

The toolkit

Tools

Search, spreadsheets, email it can use

The firm’s library

Grounding (RAG)

Trusted documents it can look up

A confident bluff

Hallucination

A fluent answer that is simply invented

Sources: [23] [24] · Full glossary in the companion document

From answering to doing: you choose how far

All four exist today. Higher isn't always better: it depends on the task.

The autonomy slider: how much the AI does

Past this line it acts while you're not watching. Set controls: what it can touch, what it can send, who checks.

Since 2022

Ask

You do the work

Chat, then paste the answer into your work.

ChatGPT, Claude, Gemini

Mature

Since 2023

Assist

You steer each step

It drafts inside Word, Excel and email.

Copilot in Microsoft 365, Gemini in Workspace

Mature

Since 2025

Delegate

You review the result

Give it a goal. It researches and builds the file.

Deep research, ChatGPT agent

Works, with review

2026

Always-on

You set goals and check in

Agents with their own computer and memory keep working after you leave.

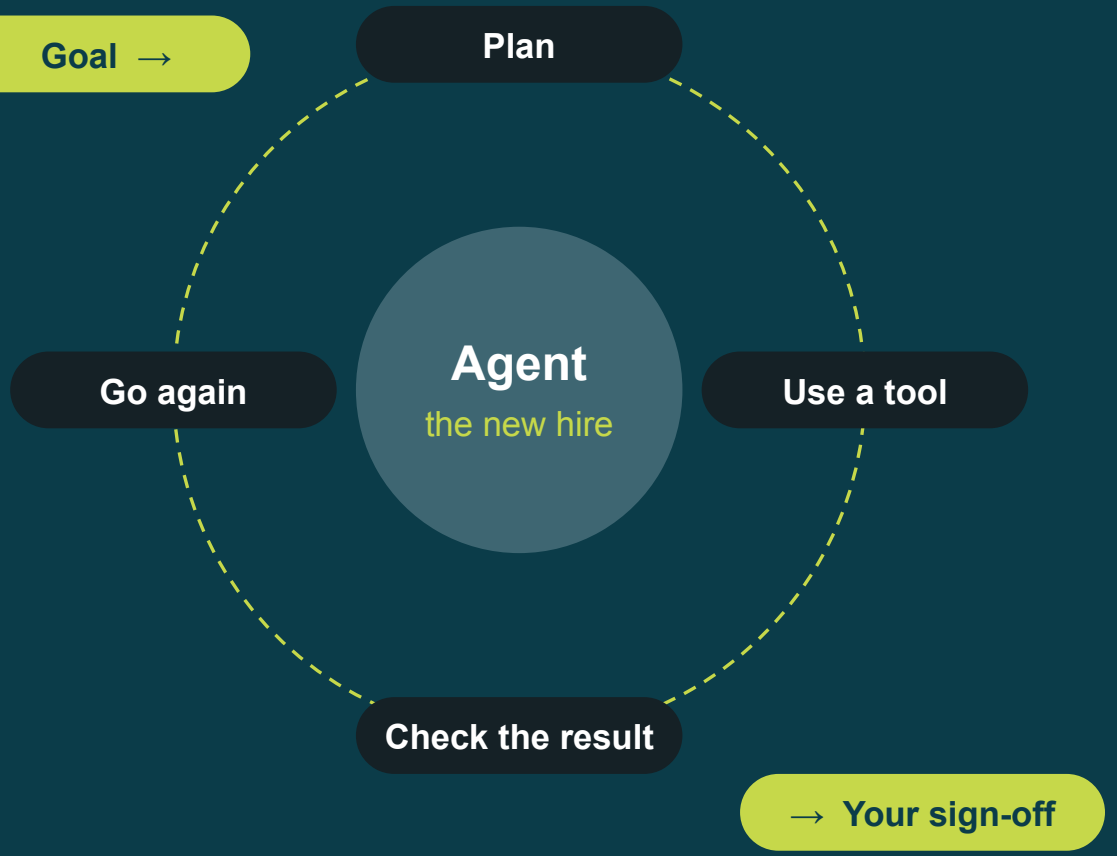
Claude Cowork, OpenAI Dots

Early: low-stakes first

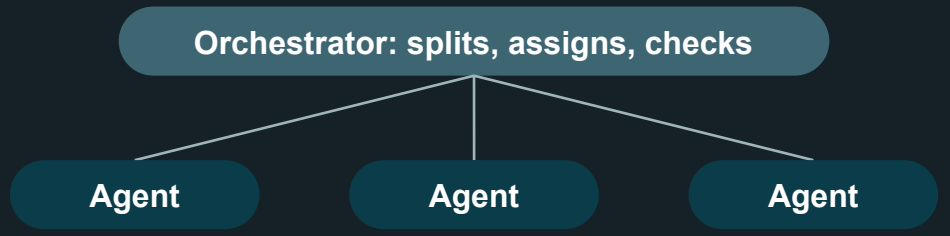
Sources: [25] [26]

An agent is the new hire working in a loop

Give it a goal, tools and permission to keep going. Orchestration manages many of them.



Orchestration: the engagement manager



One agent splits a big job, others do the pieces, the results come back assembled for your review.

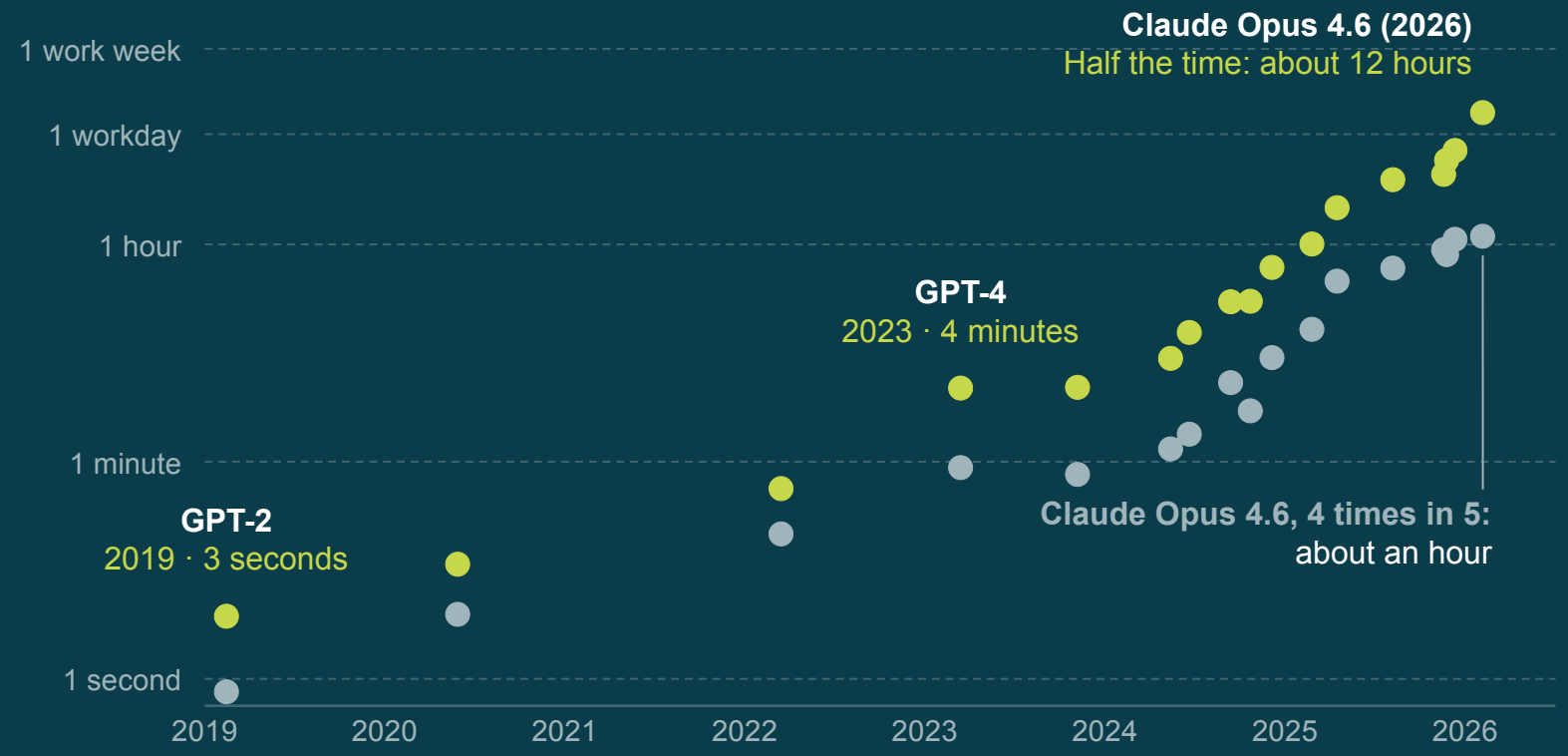
Where the analogy breaks

- It doesn’t learn from your red pen unless someone builds that.
- It can be copied: a hundred can run at once.
- It isn’t accountable or insured. You are.

Sources: [27]

AI can try a day's work. You'd leave it alone for an hour.

Software tasks, timed against a skilled professional. Each step up the axis is a far bigger jump.



~4 months

for that task length to double since 2023. About 6 months over the longer run.

Read it with care

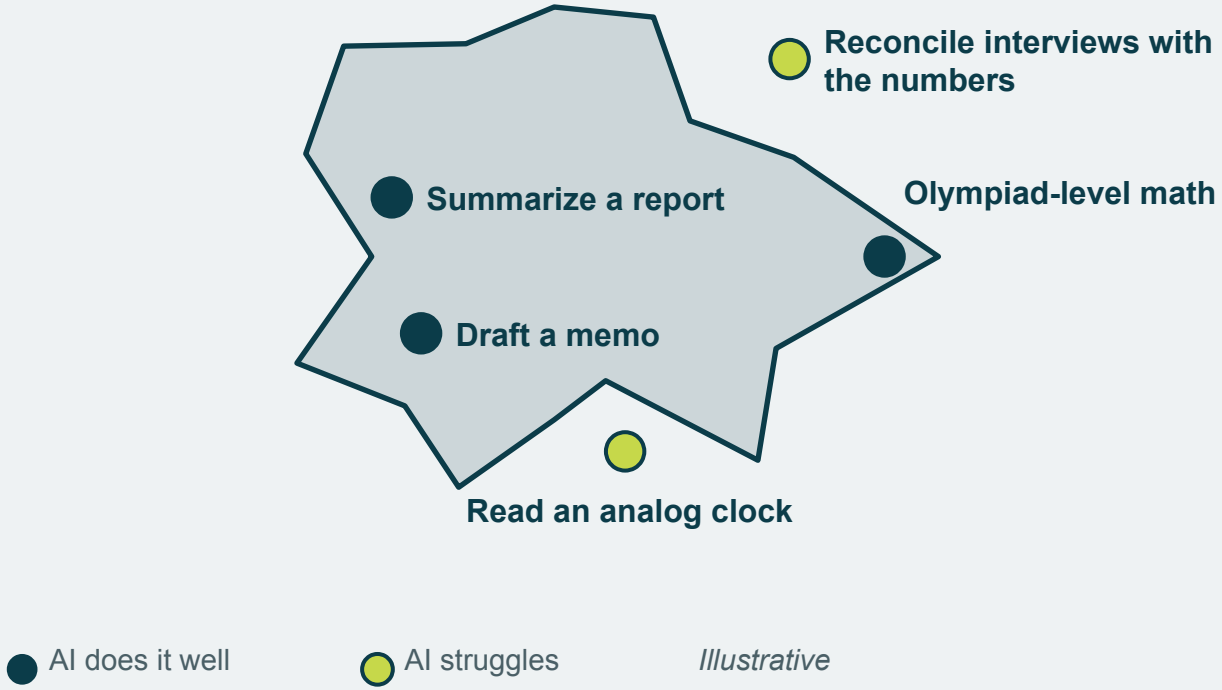
- Software tasks, not client advisory work.
- Half the time is a coin flip. Plan on the hour.
- The 12-hour point could be anywhere from 5 to 60.

Sources: [28] METR 2025–26 · Public models, software tasks. METR: results above 16 hours are unreliable.

AI’s skills are jagged, and you can’t see the edge

What AI does well doesn’t follow what’s hard for people

The frontier: what AI does well



Randomized trial, 758 BCG consultants (GPT-4, 2023)

Inside the frontier, with AI

+12% more tasks

25% faster

~⅓ higher quality

Just outside the frontier, with AI

19 points less often right

than consultants without AI, and the wrong answers looked more polished

Think of it as an alien new hire: Olympiad gold at math, a coin flip on a clock. Test it on your own work.

Sources: [29] Dell'Acqua et al. 2026 · [30] Stanford AI Index 2026

Four risks you can’t delegate

The new hire works fast. The accountability stays with you.



Client data

Paid business plans say in writing they don’t train on your data by default. Personal accounts may, and most people who use AI at work bring their own.



Invented content

Deloitte refunded part of its fee for a government report citing sources that don’t exist. Lawyers suing Avianca were fined for cases ChatGPT made up.



Regulation

EU: AI in hiring, managing staff, and life and health insurance pricing is high-risk from December 2027. Colombia: no AI law yet, but Ley 1581 and the SIC’s 2024 circular apply.



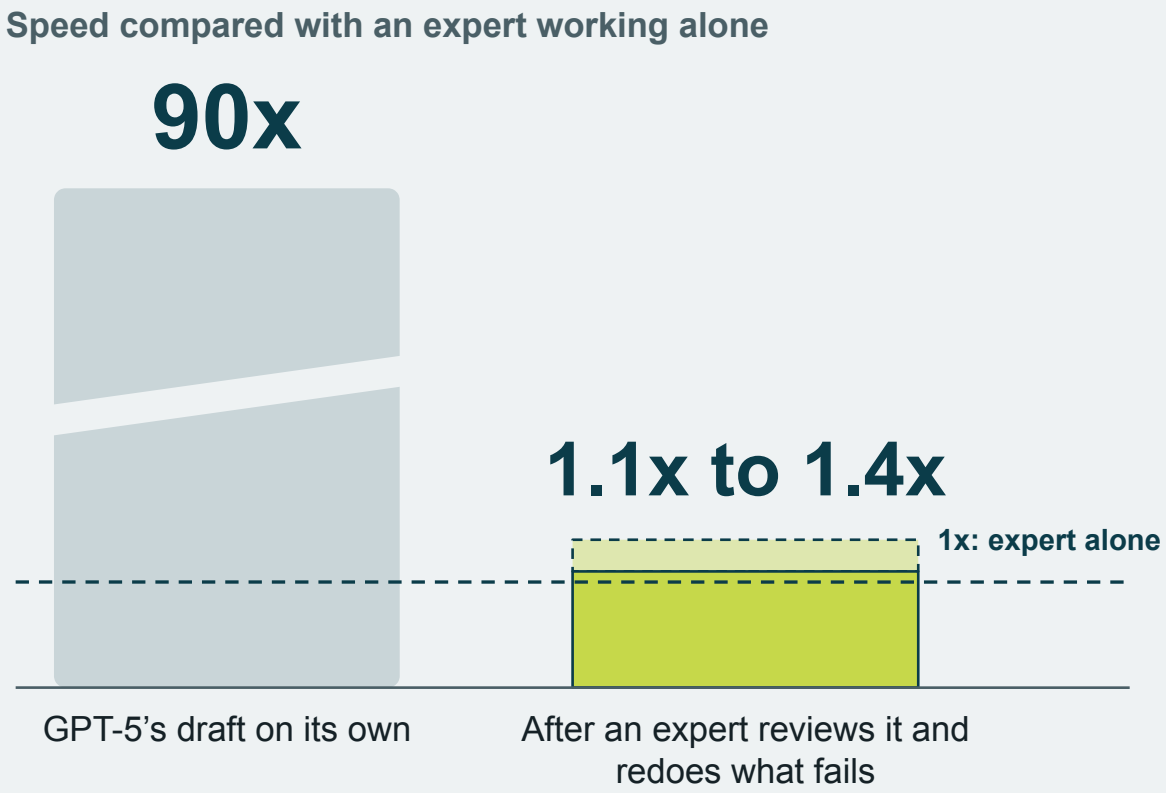
Over-trust

Polished is not the same as right. A named person signs off every client deliverable.

Sources: [31] Provider policies; Microsoft 2024 · [32] AP; Senate estimates 2025 · [33] Mata v. Avianca 2023 · [34] EU AI Act; Omnibus 2026 · [35] Ley 1581; SIC 2024

90 times faster to draft. 1.1 to 1.4 times faster to finish.

The bottleneck has moved from producing the work to checking it



OpenAI's estimates for GPT-5 (2025), modelled from time estimates, not timed

Checking is now the work.

In your firm, checking is senior time.

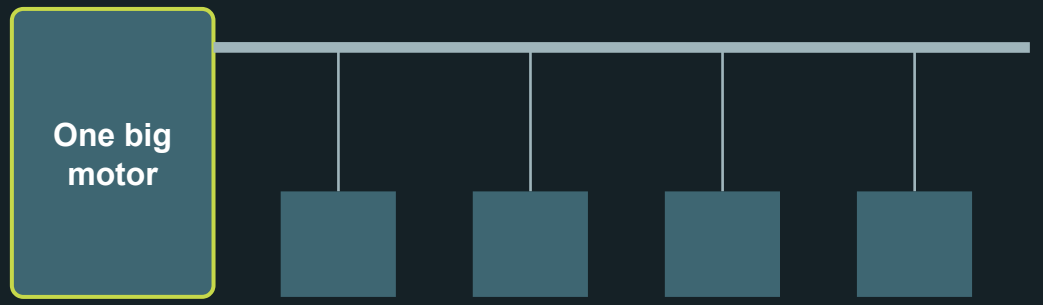
Better models mean less redo. The review still takes an expert.

Sources: [36] OpenAI 2025, GDPval

Electric motors paid off only on a redesigned floor

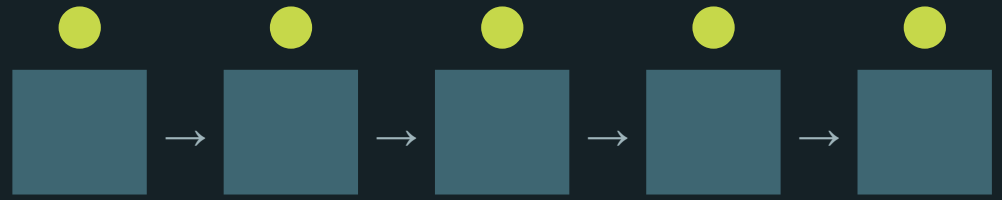
Most AI today is a new motor on an old factory floor

Around 1900: new motor, old floor



An electric motor replaces the steam engine. Same shafts, same belts, same layout. Little gain.

The 1920s: new floor



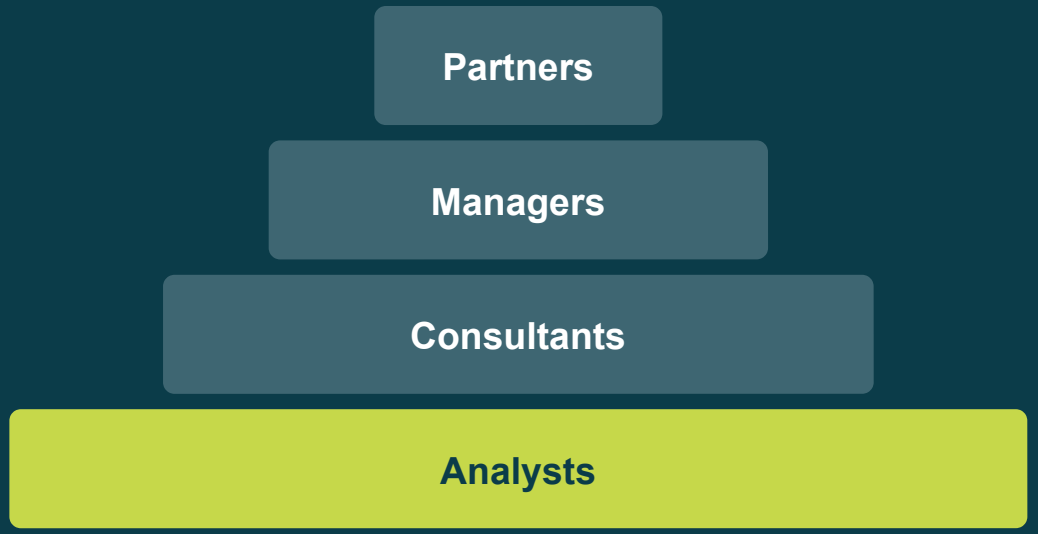
A small motor in every machine, and the floor rebuilt around the flow of work. Productivity jumps.

Think of it as rebuilding the floor. For factories it took decades. Today the floor is workflows, roles and pricing: one workflow can change in months, a whole firm takes years.

Sources: [37] [38] [39]

AI is thinning the base of the pyramid first

In payroll data it shows up in hiring more than firing



The base does research, first drafts, data and slides.

19% gap in US employment of 22 to 25 year olds in AI-exposed jobs, against peers in less-exposed jobs.
Mostly fewer hires, not layoffs.

Where AI did the work, young hiring fell. Where it helped people do the work, it held.

Same Stanford payroll study

Who learns?

Juniors learn on the tasks AI takes first.

Who pays?

If you bill hours, time saved is revenue lost. Clients will ask for the savings.

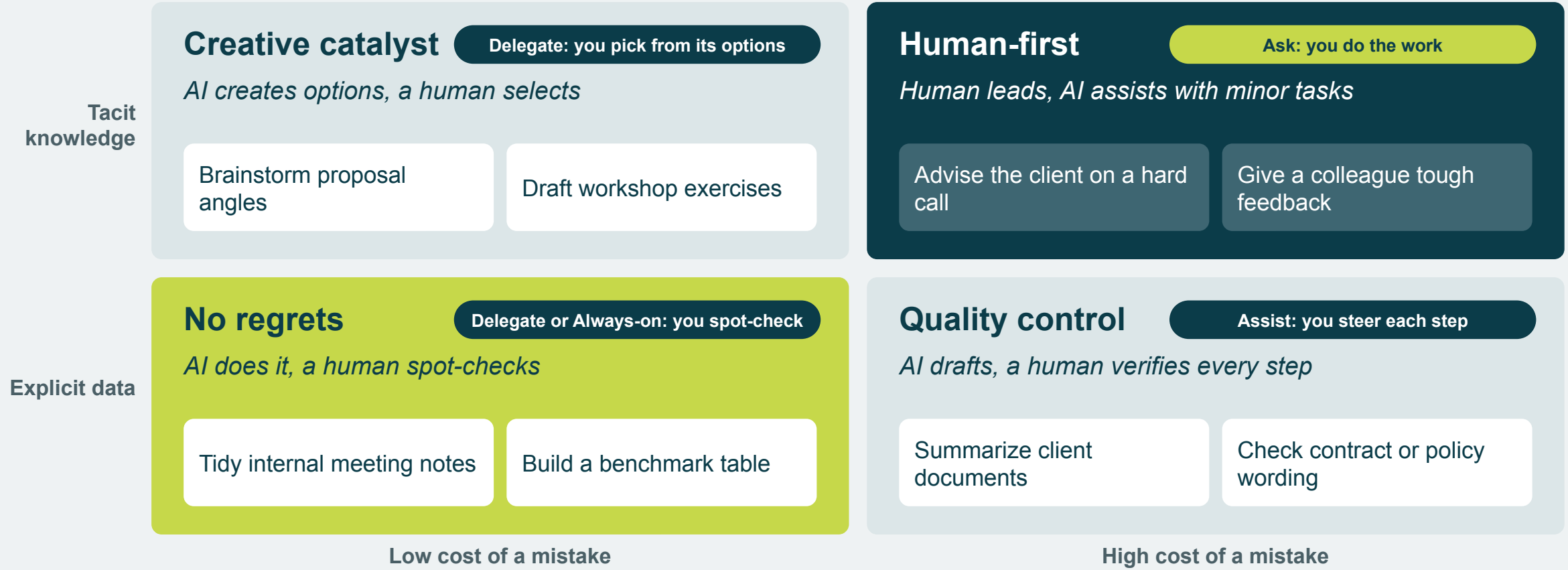
Who competes?

Leaner firms, and AI labs selling implementation.

Sources: [40] Brynjolfsson et al. 2026 · [41] Andreessen 2026 · [42] Sternfels 2026 · [43] Duncan et al. 2025 · [44] Anthropic 2026

Don’t ask which jobs AI takes. Ask which tasks.

The task map. The higher the cost of a mistake, the lower the slider.



Sources: [16] Anand & Wu 2025 (framework adapted; example tasks and slider settings are ours)

Two redesigns beat twenty pilots

Three moves, each with a number to watch

1 YOURSELF

Use it every week on real work

Start: one recurring deliverable

Watch: what you did with the hour

2 YOUR FIRM

Redesign two workflows end to end, in 90 days

Start: pick them with the task map

Watch: margin and rework per engagement

3 EVERY PROJECT

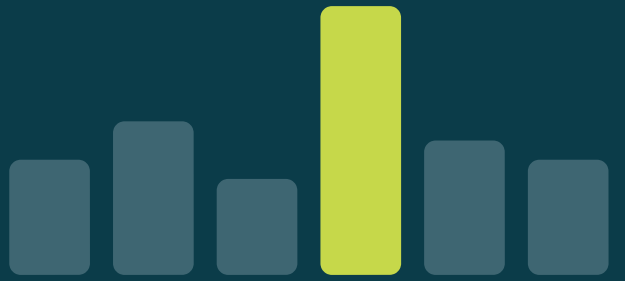
Prove value, not concept

Start: write the baseline and success metric on page one

Watch: the outcome, not the demo

One ask: pick two workflows, name the partner who owns each, and write down today's baseline.

Sources: [45] METR 2025–26 trial



1956: ONE SUMMER. 2026: THE REDESIGN.

You redesign work for your clients.
Now do it to your own.

We didn’t get electric light by improving the candle.

Jaime Díaz-Beltrán · I build AI products and measure whether they pay off

Sources

1. McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

2. Turing, A. M. (1950). Computing Machinery and Intelligence. Mind, 59(236), 433–460.

3. Weizenbaum, J. (1966). ELIZA. Communications of the ACM, 9(1), 36–45.

4. Lighthill, J. (1973). Artificial Intelligence: A General Survey. UK Science Research Council.

5. Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. NeurIPS.

6. Vaswani, A. et al. (2017). Attention Is All You Need. arxiv.org/abs/1706.03762

7. OpenAI (30 Nov 2022). Introducing ChatGPT. openai.com/index/chatgpt

8. IBM. Deep Blue; Watson, Jeopardy! champion. IBM Heritage. ibm.com/history/deep-blue; ibm.com/history/watson-jeopardy

9. Hassabis, D. (2016). What we learned in Seoul with AlphaGo. blog.google

10. Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589.

11. Royal Swedish Academy of Sciences (2024). The Nobel Prize in Physics 2024; The Nobel Prize in Chemistry 2024 (press releases). nobelprize.org

12. Luong, T. & Lockhart, E. (2025). Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the IMO. deepmind.google

13. Tesler, L. Adages and Coinages. nomodes.com

14. Butters, J. (2023–2026). S&P 500 companies citing “AI” on earnings calls. FactSet Insight and Earnings Insight (18 Sep 2026). insight.factset.com

15. OpenAI (29 Sep 2026). DevDay 2026 recap (self-reported weekly users). openai.com/index/devday-2026-recap

16. Anand, B. N. & Wu, A. (2025). The Gen AI Playbook for Organizations. Harvard Business Review, Nov–Dec 2025. hbr.org/2025/11/the-gen-ai-playbook-for-organizations

17. Yotzov, I., Barrero, J. M., Bloom, N., Davis, S. J. et al. (2026). Firm Data on AI. NBER Working Paper 34836, rev. Oct 2026. nber.org/papers/w34836

18. Korst, J., Puntoni, S., Tambe, P. et al. (2025). Accountable Acceleration: Gen AI Fast-Tracks Into the Enterprise. Wharton Human-AI Research & GBK Collective.

19. BCG (30 Sep 2026). AI Is Starting to Pay Off; BCG (30 Sep 2025). The Widening AI Value Gap. bcg.com

20. Narayanan, A. & Kapoor, S. (2024). AI Snake Oil. Princeton University Press.

21. Kozyrkov, C. (2018). The simplest explanation of machine learning you'll ever read. HackerNoon.

22. Sanderson, G. (2024). Large Language Models explained briefly. 3Blue1Brown. 3blue1brown.com/lessons/mini-llm

23. Google for Developers. Machine Learning Glossary. developers.google.com/machine-learning/glossary

Sources (continued)

24. NIST (2024). AI 600-1: Generative Artificial Intelligence Profile.

25. Microsoft (2023) Microsoft 365 Copilot GA; OpenAI (2025) deep research and ChatGPT agent; Nolan, B. (13 Jan 2026) on Claude Cowork, Fortune; Huamani, K. (29 Sep 2026) on OpenAI's always-on agent, Associated Press.

26. Karpathy, A. (17 Jun 2025). Software Is Changing (Again). Y Combinator AI Startup School.

27. Schluntz, E. & Zhang, B. (19 Dec 2024). Building effective agents. Anthropic.

28. METR (2026). Task-Completion Time Horizons of Frontier AI Models (Time Horizon 1.1), metr.org/time-horizons; Kwa, T., West, B. et al. (2025). Measuring AI Ability to Complete Long Software Tasks. arXiv:2503.14499.

29. Dell'Acqua, F., McFowland III, E., Mollick, E. et al. (2026). Navigating the Jagged Technological Frontier. Organization Science. doi:10.1287/orsc.2025.21838

30. Stanford HAI (2026). The 2026 AI Index Report. hai.stanford.edu/ai-index/2026-ai-index-report

31. OpenAI, Anthropic, Microsoft and Google: enterprise data-use and training policies (accessed Oct 2026); Microsoft & LinkedIn (8 May 2024). 2024 Work Trend Index: AI at Work Is Here.

32. Associated Press (7 Oct 2025). Deloitte to partially refund Australian government for report with apparent AI-generated errors; refund amount per Australian Senate estimates (Oct 2025).

33. Mata v. Avianca, Inc., No. 22-cv-1461 (S.D.N.Y. 22 Jun 2023) (Castel, J.), opinion and order on sanctions.

34. Regulation (EU) 2024/1689 (AI Act), Annex III; Regulation (EU) 2026/1744 (Digital Omnibus on AI), in force 27 Jul 2026.

35. Ley 1581 de 2012; Superintendencia de Industria y Comercio, Circular Externa 002 de 2024 (21 Aug 2024) on personal data in AI systems; CONPES 4144 (2025); Corte Constitucional, Sentencia T-323 de 2024.

36. Patwardhan, T. et al. (2025). GDPval: Evaluating AI Model Performance on Real-World Economically Valuable Tasks. OpenAI. arXiv:2510.04374

37. David, P. A. (1990). The Dynamo and the Computer. American Economic Review, 80(2), 355–361.

38. Hammer, M. (1990). Reengineering Work: Don't Automate, Obliterate. Harvard Business Review, Jul–Aug 1990.

39. Agrawal, A., Gans, J. S. & Goldfarb, A. (2017). What to Expect From Artificial Intelligence. MIT Sloan Management Review; Power and Prediction (HBR Press, 2022).

40. Brynjolfsson, E., Chandar, B. & Chen, R. (2026). Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of AI. Stanford Digital Economy Lab, rev. Aug 2026.

41. Rachitsky, L. (host) (29 Jan 2026). Marc Andreessen: The real AI boom hasn't even started yet. Lenny's Podcast, 39:28.

42. Sternfels, B. (6 Jan 2026). Where McKinsey and Consulting Go From Here. HBR IdeaCast, episode 1060. hbr.org/podcast

43. Duncan, D. S., Anderson, T. & Saviano, J. (10 Sep 2025). AI Is Changing the Structure of Consulting Firms. HBR.org.

44. Anthropic (4 May 2026). Building a new enterprise AI services company with Blackstone, Hellman & Friedman, and Goldman Sachs. anthropic.com/news

45. Becker, J., Rush, N., Barnes, E. & Rein, D. (2025). Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity. METR; and METR (24 Feb 2026) uplift update. metr.org/blog

Images: Dartmouth Hall by Daderot (CC0); Deep Blue by Erik Pitti (CC BY 2.0); Lee Sedol by LG Electronics (CC BY 2.0); AlphaFold figure from Jumper et al., Nature 2021 (CC BY 4.0); Geoffrey Hinton and Demis Hassabis by John Sears (CC BY-SA 4.0). All via Wikimedia Commons, cropped and toned. Charts redrawn from FactSet and METR data. Icons: Tabler Icons (MIT).