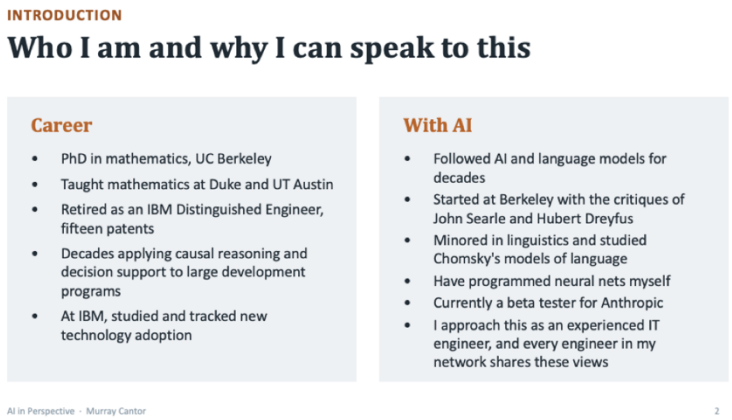
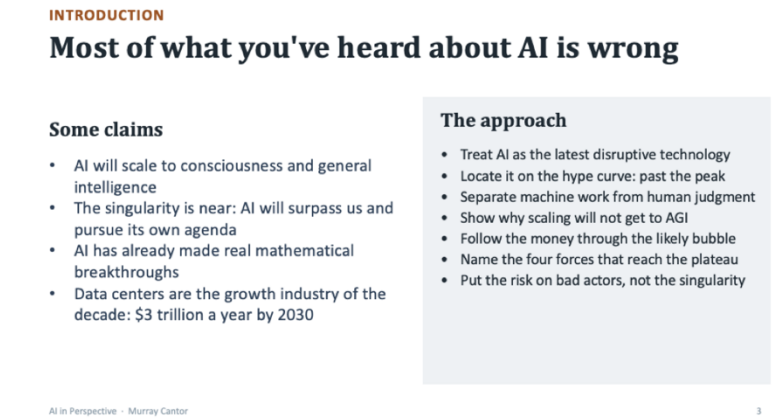
AI in Perspective: What Earlier Disruptive Technologies Tell Us

Slide-by-slide printout with commentary — Murray Cantor PhD

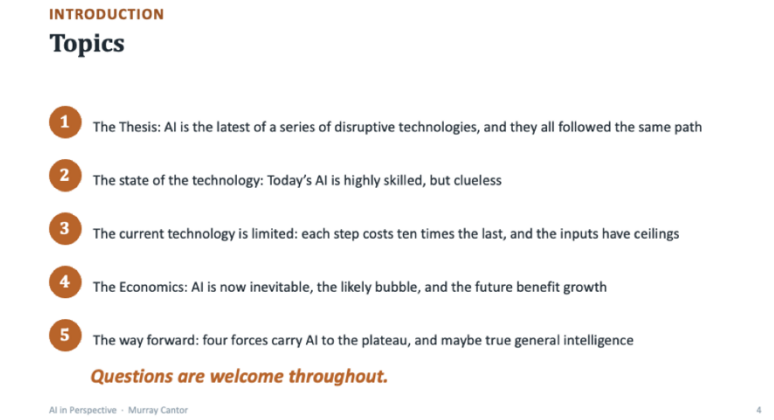
Slide	Commentary
 **Slide 1 — Title**	AI in Perspective: What Earlier Disruptive Technologies Tell Us, by Murray Cantor PhD. I start from the premise that AI is not the first disruptive technology and that the earlier ones left a record of how adoption played out, and that record tells us where AI is headed, what to believe, and what to doubt. The examples I use throughout are the steam engine, the power loom, business computing, internet fiber, and solar power.
 **Slide 2 — Who I am**	My credentials: a PhD in mathematics from UC Berkeley, teaching at Duke and UT Austin, retirement as an IBM Distinguished Engineer with fifteen patents, and decades applying causal reasoning and decision support to large development programs, including the study of technology adoption at IBM. On AI specifically, I have followed language models for decades, beginning with the critiques of John Searle and Hubert Dreyfus, minored in linguistics and studied Chomsky's models of language, programmed neural nets myself, and am a beta tester for Anthropic. I speak as an experienced IT engineer, skeptical of both the boosters and the critics.



Slide 3 — Most of what you've heard is wrong

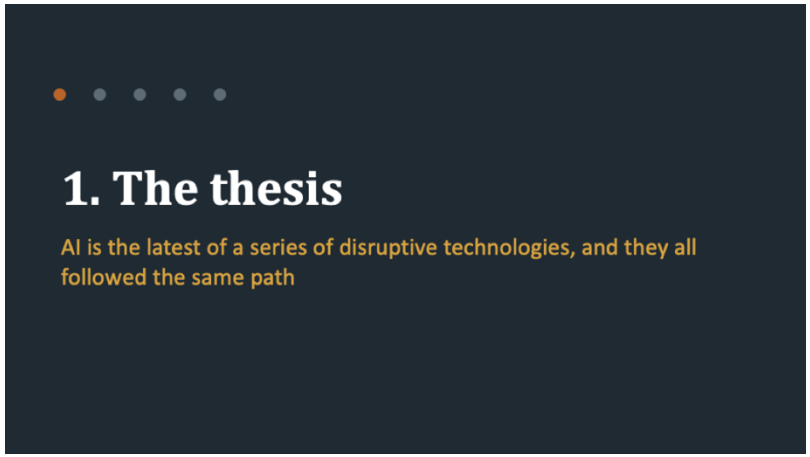
I set out four familiar claims to examine: that AI will scale to consciousness and general intelligence, that the singularity is near and AI will pursue its own agenda, that AI has already made real mathematical breakthroughs, and that data centers will be a $3 trillion a year industry by 2030.

My approach is to treat AI as the latest disruptive technology, locate it on the hype curve (past the peak), separate the machine's work from human judgment, show why scaling will not reach AGI, follow the money through the likely bubble, name the four forces that reach the plateau, and put the real risk on bad actors rather than the singularity.



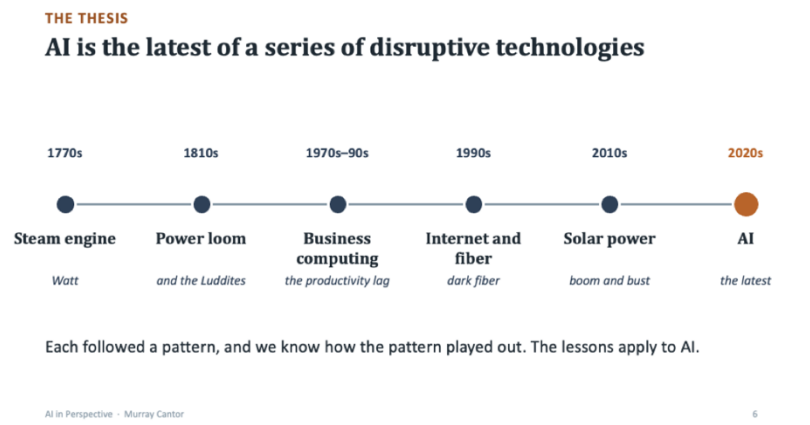
Slide 4 — Topics

The five sections of my talk, which follow the hype curve in order: the thesis that AI is one of a series of disruptive technologies that all followed the same path; the state of the technology, which is highly skilled but clueless; the limits of the current technology, where each step costs ten times the last and the inputs have ceilings; the economics, covering why adoption is now inevitable, the likely bubble, and the later growth in benefit; and the way forward, where four forces carry AI to the plateau and perhaps to true general intelligence.



Slide 5 — Section 1: The thesis

Section divider. The thesis is that AI is the latest of a series of disruptive technologies and that they all followed the same path. I first present the series and then the path, and the rest of the talk applies that pattern to AI.

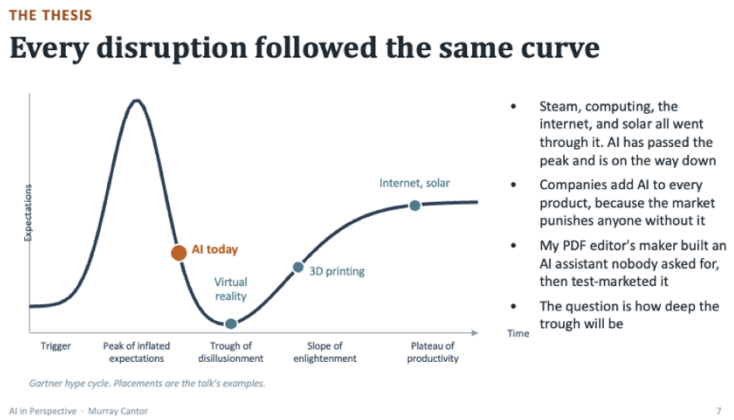


Slide 6 — The series

A 250-year timeline: the steam engine in the 1770s (Watt), the power loom in the 1810s (and the Luddites), business computing in the 1970s to 1990s (the productivity lag), the internet and fiber in the 1990s (dark fiber), solar power in the 2010s (boom and bust), and AI in the 2020s.

Each was hailed as a revolution, each turned out to be one, and each went through a painful stretch first: steam took decades to spread, the loom put skilled weavers out of work, computing cost a fortune before productivity moved, fiber was overbuilt, and solar makers went bankrupt even as solar won.

Five of the six endings are already known, and the lessons apply to AI.



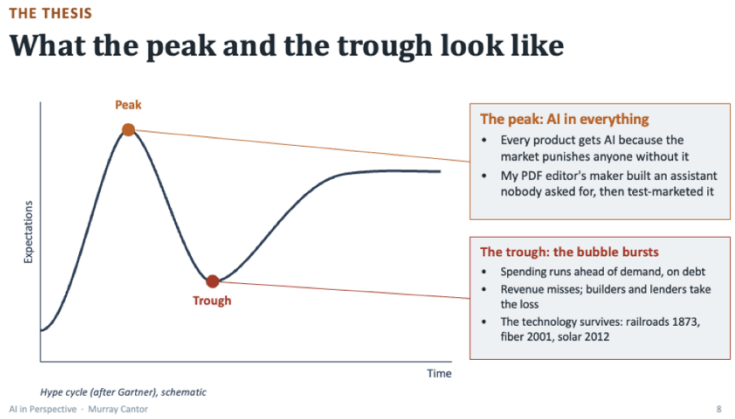
Slide 7 — The hype cycle

The Gartner hype cycle, on which the rest of the talk hangs: a trigger, a peak of inflated expectations, a trough of disillusionment, a slope of enlightenment, and a plateau of productivity.

I place AI past the peak and on the way down. The symptom is that companies add AI to every product because the market punishes anyone without it; my PDF editor's maker built an AI assistant nobody asked for and then test-marketed it.

Other placements show the range of outcomes: the internet and solar reached the plateau, 3D printing climbed slowly, and virtual reality is still in the trough.

The curve does not say a technology fails, and the open question for AI is how deep the trough will be.

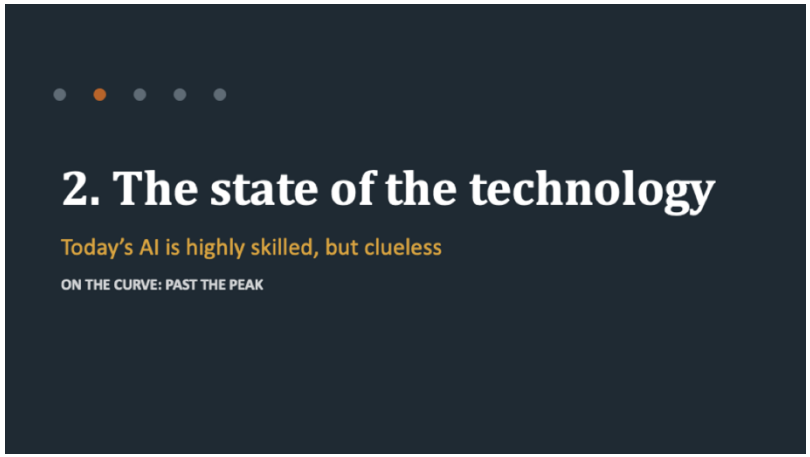


Slide 8 — The peak and the trough

What the two ends of the curve look like from inside, so you can recognize where we are.

The peak is overuse: AI goes into everything, and capability outruns use. The trough is the money running out: spending runs ahead of demand and is financed with debt, revenue misses the schedule, and the builders and lenders take the loss.

The technology survives each time, as the railroads did after 1873, fiber after 2001, and solar after 2012, and each trough cleared the way for the real gains.



Slide 9 — Section 2: The state of the technology

Section divider, placed on the curve past the peak.

The first lesson from the earlier technologies is that the machine does the work and people supply the judgment, and the test for AI is the same division of labor: what the machine does and what people still supply. The answer decides whether the singularity story is plausible.



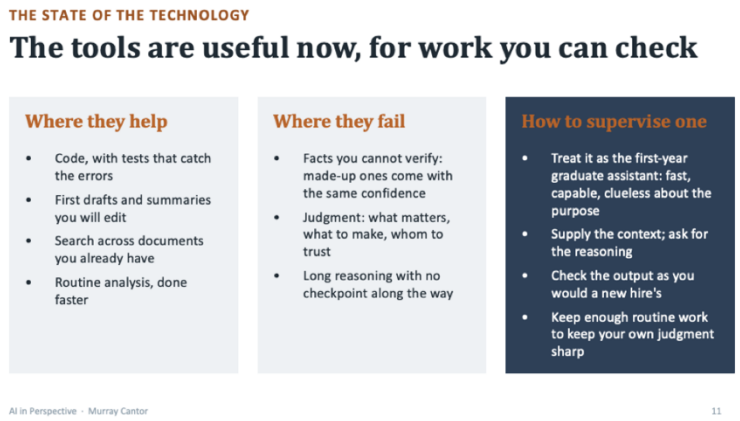
Slide 10 — Highly skilled, but clueless

The power loom wove far faster than a person but never decided what cloth to make.

AI is capable, and that should not be understated, but it lacks judgment: a model can pass an exam and still state made-up facts, because it predicts likely words rather than checked facts.

Neural nets are partial models of the brain rather than copies, keeping the computable part and leaving out the rest, so they have no consciousness and no intentionality, and there is no reason to expect that scaling them up will supply judgment.

The working image is a brilliant but clueless first-year graduate assistant.



Slide 11 — Useful now, for work you can check

Practical guidance.

The tools help with code that has tests to catch errors, first drafts and summaries that will be edited, search across documents you already have, and routine analysis done faster.

They fail on facts you cannot verify (made-up ones come with the same confidence), on judgment about what matters and whom to trust, and on long reasoning with no checkpoint.

To supervise one, treat it as the first-year graduate assistant: supply the context, ask for the reasoning, check the output as you would a new hire's, and keep enough routine work to keep your own judgment sharp.



Slide 12 — Tao: AI added brute force; the ideas were already there

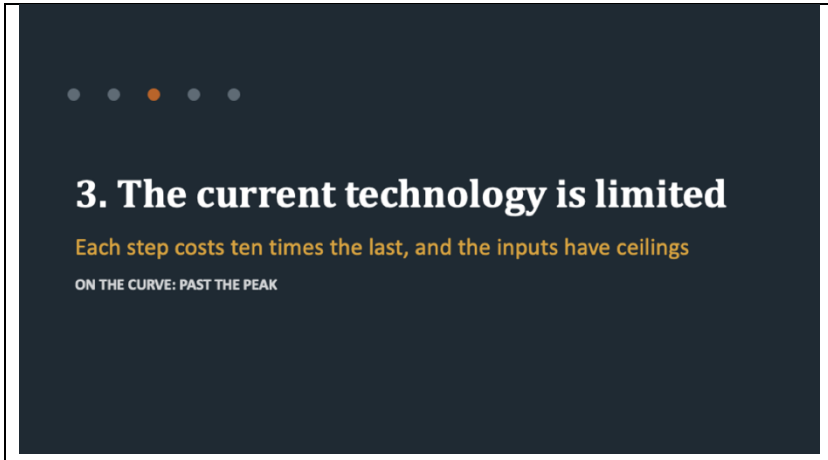
A close look at the 2026 episode taken as proof that AI is surpassing us.

OpenAI claimed a Navier-Stokes blow-up: 165 pages from about 10,000 agents over 88 hours, then 722 results at once out of some 4,000 problems attempted.

The method followed a human program already in motion (Buckmaster and Alpöge had just blown up Euler the same way), it works only with a contrived external force, three mathematicians then proved the method can never reach the real unforced problem, three manuscripts were withdrawn for a sign error, and the Clay Institute still lists the problem as open.

Terence Tao calls this artificial general cleverness rather than intelligence: a brute-force search over known techniques, fallible, needing Lean or a human to check it, and strip-mining problems for answers while losing the methods and questions that make a field grow.

The best mathematician alive uses AI daily and still says the ideas and judgment are his.

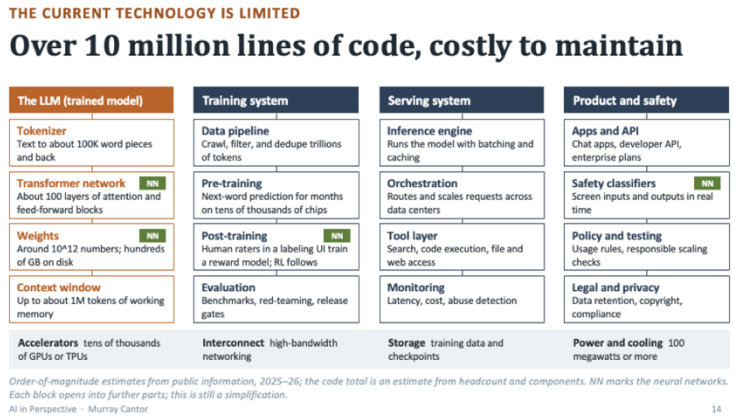


Slide 13 — Section 3: The current technology is limited

Section divider, still past the peak.

The singularity claim rests on the idea that more of the same eventually becomes something different, and the numbers do not support it.

The obvious reply to the previous section is that more scale will fix the lack of judgment, so this section shows what is actually being scaled and then the numbers.



Slide 14 — Over 10 million lines of code

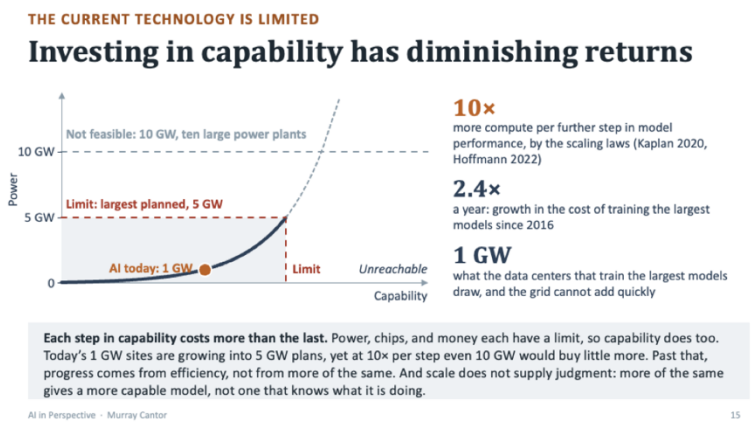
A block diagram of the whole system behind a frontier model.

At the core are the tokenizer (about 100K word pieces), the transformer network (about 100 layers), the weights (around a trillion numbers, hundreds of gigabytes), and the context window (up to about a million tokens).

Around it sit the training system (data pipeline, pre-training, post-training with human raters, evaluation), the serving system (inference engine, orchestration, tool layer, monitoring), and product and safety (apps and API, safety classifiers, policy, legal and privacy), all on tens of thousands of GPUs or TPUs with high-bandwidth networking and 100 megawatts or more of power.

The code is engineering that can be read, tested, and fixed, and it is expensive to host and maintain; the weights are different, a trillion numbers nobody wrote, and no one can fully explain why they work so well.

Figures are order-of-magnitude estimates from public information.



Slide 15 — Diminishing returns

Three numbers: about 10× more compute for each further step in model performance by the scaling laws (Kaplan 2020, Hoffmann 2022); a 2.4× a year growth in the cost of training the largest models since 2016; and about 1 GW drawn by the data centers that train the largest models, which the grid cannot add quickly.

The chart plots capability against power: today's 1 GW sites are growing into 5 GW plans, and at 10× per step even a 10 GW site, ten large power plants, would buy little more.

Power, chips, and money each have a limit, so capability does too, and past that limit progress has to come from efficiency.

Scale also does not supply judgment: more of the same gives a more capable model, not one that knows what it is doing. The brain runs on about 20 watts, which says today's methods are wasteful and there is room to improve.



Slide 16 — The architecture is near its limit

The case against scaling in three steps: each step costs about 10× the compute of the last; the inputs (power, chips, money, usable text) have ceilings, so the argument is which binds first, not whether; and scale does not supply judgment, so even without the first two limits the result is a bigger version of the same system with the same gap.

The 2026 math results do not contradict this: they were like thousands of first-year graduate assistants spending weeks on a brute-force search, added no real insight, and are what Tao calls artificial general cleverness.

We are close to the largest model this architecture can afford, and further gains need new science rather than more chips; this is a claim about this architecture, not about what is possible in principle.

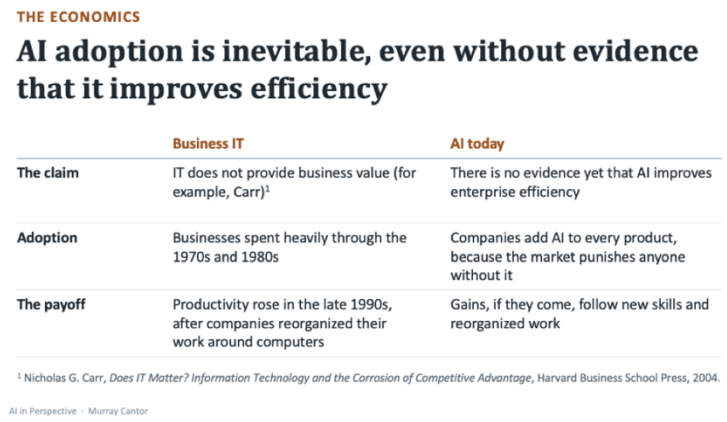


Slide 17 — Section 4: The economics

Section divider, placed on the curve going into the trough.

Three parts follow: why adoption is now inevitable, the likely bubble, and the later growth in benefit.

The question is not whether AI is useful but whether anyone earns back the spend, and history is most useful here because booms have been financed this way before.



Slide 18 — Adoption is inevitable

A side-by-side of business IT and AI today. The claim then was that IT does not provide business value (Nicholas Carr's Does IT Matter?); the claim now is that there is no evidence yet that AI improves enterprise efficiency.

Businesses spent heavily on IT through the 1970s and 1980s anyway, because competitors did, and companies now add AI to every product because the market punishes anyone without it.

The IT payoff came in the late 1990s after companies reorganized their work around computers, and if the precedent holds, AI gains will follow new skills and reorganized work and arrive later than the spending.

Adoption can be inevitable while the builders' economics stay uncertain, as fiber showed.



Slide 19 — Nobody knows yet whether AI will be profitable

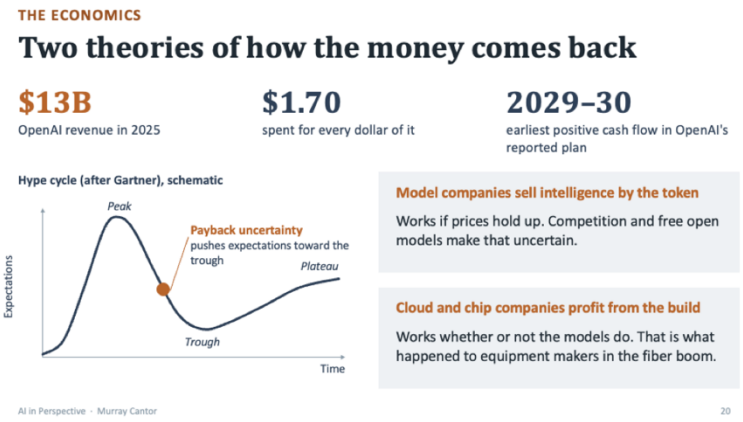
Three layers of cost.

Building the models: a frontier training run costs hundreds of millions of dollars and is repeated every generation, like a new factory every year or two.

Operating them: every answer runs on data-center chips, and for OpenAI inference was about half of 2025 revenue, which is the surprise, because software normally gets cheaper to serve at scale.

Everything around them: research staff, sales, support, safety, and legal work.

Underneath all of it, the four largest cloud companies plan about $700 billion of data center spending in 2026, up from about $410 billion in 2025, roughly all of their cloud revenue, so the payoff has to arrive on schedule.



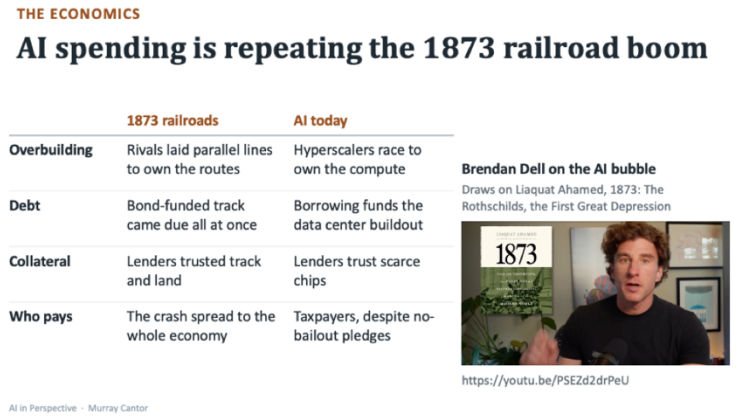
Slide 20 — Two theories of how the money comes back

The revenue picture: OpenAI had about $13 billion of revenue in 2025, spent about $1.70 for every dollar of it, and its own plan shows no positive cash flow before 2029–30; Anthropic had its first profitable quarter in 2026, a start rather than a track record.

Theory one is that model companies sell intelligence by the token, which works if prices hold up, but competition and free open models push token prices toward serving cost.

Theory two is that cloud and chip companies profit from the build whether or not the models do, which is what happened to the equipment makers in the fiber boom.

Which theory wins decides the depth of the trough, and the chart shows payback uncertainty pushing expectations toward it.



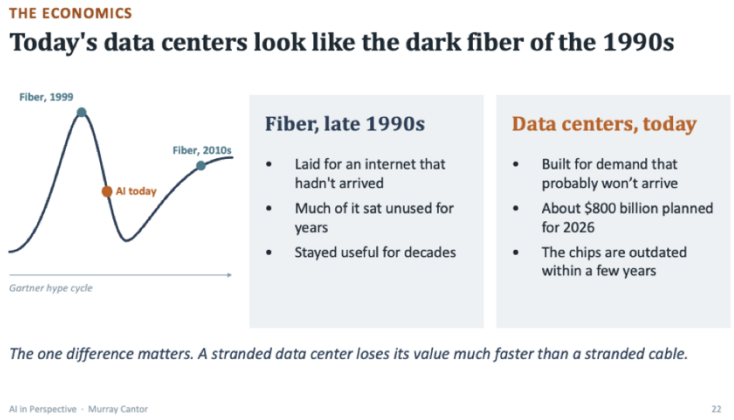
Slide 21 — The 1873 railroad boom

A four-row parallel between the railroads of 1873 and AI today.

Overbuilding: rivals laid parallel lines to own the routes, and hyperscalers race to own the compute. Debt: bond-funded track came due all at once, and borrowing funds the data center buildout. Collateral: lenders trusted track and land, and now they trust scarce chips. Who pays: the 1873 crash spread to the whole economy, and this time it is likely to be taxpayers despite no-bailout pledges.

The railroads were not a bad idea; the crash was about financing, not the technology, and the same distinction applies to AI.

The slide links a Brendan Dell video (https://youtu.be/PSEZd2drPeU) drawing on Liaquat Ahamed's 1873: The Rothschilds, the First Great Depression.



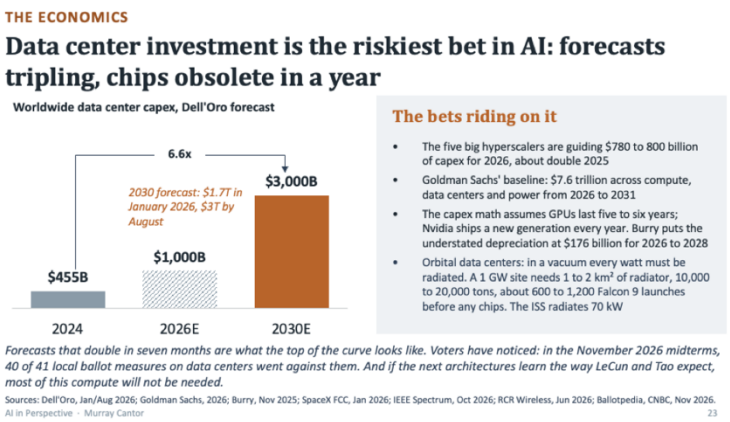
Slide 22 — Data centers and dark fiber

The closest precedent. In the late 1990s telecoms laid fiber for an internet that had not arrived, much of it sat unused for years, and the builders suffered even though the internet came and fiber became essential.

Today's data centers are built for demand that probably will not arrive on schedule, with about $800 billion planned for 2026.

The one difference matters: fiber stayed useful for decades, while chips are outdated within a few years, so a stranded data center loses its value much faster than a stranded cable.

The chart places fiber in 1999, fiber in the 2010s, and AI today on the same curve.



Slide 23 — Data center investment is the riskiest bet in AI

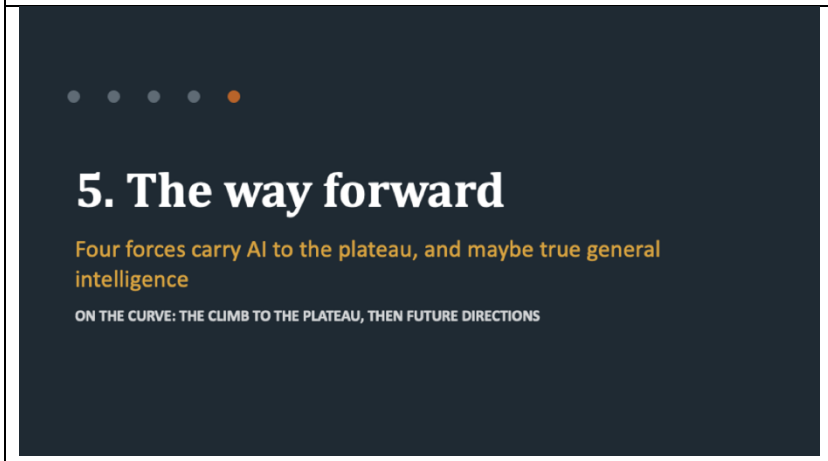
The size of the bet: the five big hyperscalers are guiding $780 to 800 billion of capex for 2026, about double 2025, and Goldman Sachs' baseline is $7.6 trillion across compute, data centers, and power from 2026 to 2031.

The Dell'Oro forecast for 2030 went from $1.7 trillion in January 2026 to $3 trillion by August, and a forecast that doubles in seven months is what the top of the curve looks like.

The capex math assumes GPUs last five to six years while Nvidia ships a new generation every year, and Burry puts the understated depreciation at $176 billion for 2026 to 2028.

Orbital data centers do not work the numbers: in a vacuum every watt must be radiated, so a 1 GW site needs 1 to 2 km² of radiator weighing 10,000 to 20,000 tons, about 600 to 1,200 Falcon 9 launches before any chips, while the ISS radiates 70 kW.

Voters have noticed, with 40 of 41 local ballot measures on data centers going against them in the November 2026 midterms, and if the next architectures learn the way LeCun and Tao expect, most of this compute will not be needed.

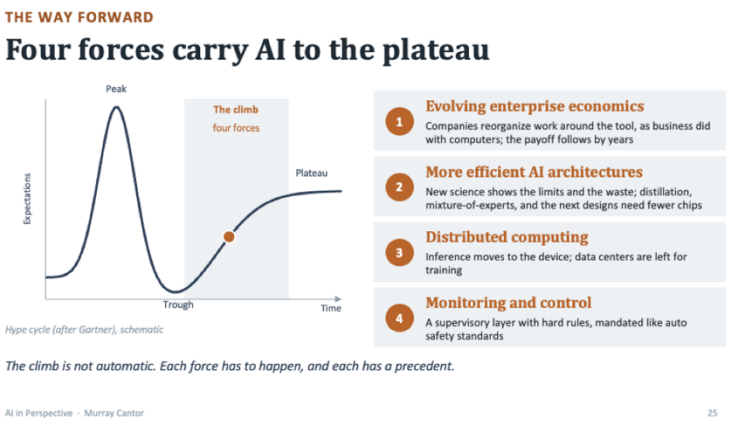


Slide 24 — Section 5: The way forward

Section divider, placed on the climb to the plateau.

Four forces carry AI there: evolving enterprise economics, more efficient AI architectures, distributed computing, and a monitoring and control layer. Each gets its own slide, and the section ends with whether true general intelligence is within reach.

This is the constructive section: what reaches the plateau, who benefits, who pays, and what a real path to human-level AI involves.



Slide 25 — Four forces carry AI to the plateau

The climb is not automatic; each force has to happen, and each has a precedent.

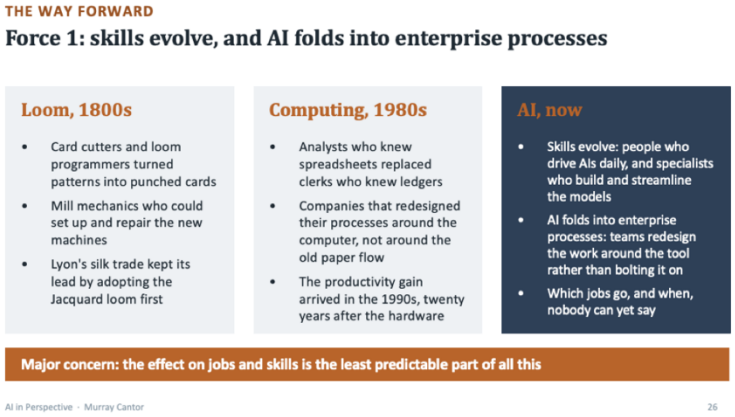
Force 1, evolving enterprise economics: companies reorganize work around the tool, as business did with computers, and the payoff follows by years.

Force 2, more efficient architectures: new science shows the limits and the waste, and distillation, mixture-of-experts, and the next designs need fewer chips.

Force 3, distributed computing: inference moves to the device and data centers are left for training.

Force 4, monitoring and control: a supervisory layer with hard rules above the model, mandated the way auto safety standards were, without which one large failure resets expectations into the trough.

Today's AI builders work with a lot of ad-hoc skill, like the weavers who built the trade before the theory existed.



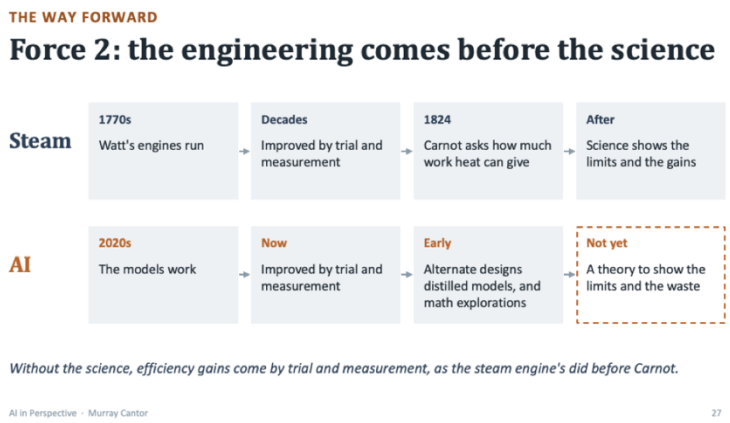
Slide 26 — Force 1: skills evolve, and AI folds into enterprise processes

Each trough rewarded the early learners. With the loom in the 1800s, card cutters and loom programmers turned patterns into punched cards, mill mechanics set up and repaired the machines, and Lyon kept its lead in silk by adopting the Jacquard loom first.

With computing in the 1980s, analysts who knew spreadsheets replaced clerks who knew ledgers, the winners redesigned their processes around the computer rather than the old paper flow, and the productivity gain arrived in the 1990s, twenty years after the hardware.

With AI now, skills evolve toward people who drive AIs daily and specialists who build and streamline the models, and teams redesign the work around the tool rather than bolting it on.

The cost is real: the Luddites were skilled workers whose trade was lost while society gained, the losers are not automatically the winners, and because judgment is learned by doing routine work under supervision, the jobs AI removes first, the effect on jobs and skills is the least predictable part of all this.



Slide 27 — Force 2: the engineering comes before the science

The fourth lesson, the reverse of what we assume.

Watt's engines ran from the 1770s and were improved for decades by trial and measurement before Carnot asked in 1824 how much work heat can give; only then did science show the limits and where to look for gains.

AI is at the same stage: the models work, nobody can fully say why, the knowledge is spread across billions of learned numbers, and improvement comes by trial and measurement.

Early studies exist (geometric deep learning, representation geometry, mechanistic interpretability, singular learning theory, muP), but there is not yet a theory to show the limits and the waste.

Efficiency is already arriving by trial through distillation, mixture-of-experts, and quantized open models, and the money consequence is fewer chips for the same work, so data centers built for today's methods could sit partly idle, like 1990s fiber.

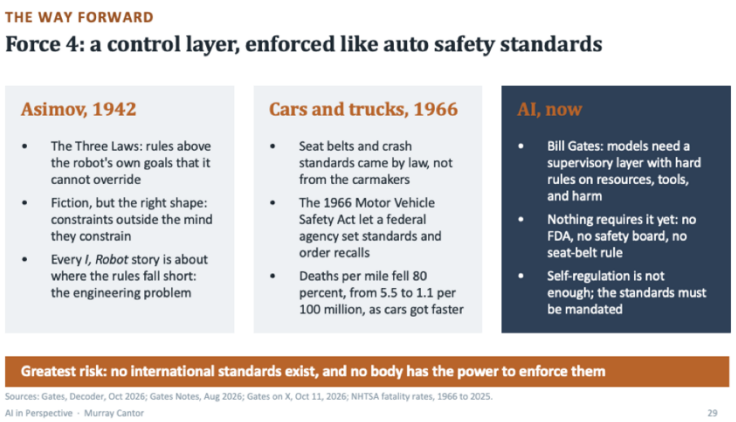


Slide 28 — Force 3: IT cycles between local and cloud

A table of where computing ran: mainframes and time-sharing in the 1960s–70s (central); the personal computer in the 1980s (local, on the desk); PCs in the 1990s, when Oracle's diskless Network Computer failed to sell and work stayed local; the cloud led by Amazon Web Services in the 2000s–10s (hybrid, split between devices and data centers); and AI in the 2020s, trained centrally with inference moving to the device.

Apple is hedging its bets: its 2026 Mac mini takes 64GB of unified memory and macOS lets four minis pool 256GB over Thunderbolt, enough to run large open-weight models on the desk, yet for its largest model Apple rents NVIDIA GPUs in Google Cloud.

Each swing to local stranded the central build, and if inference goes local most of the planned data center capacity is for training alone, a third reason the capex forecasts are at risk.



Slide 29 — Force 4: a control layer, enforced like auto safety standards

Three columns.

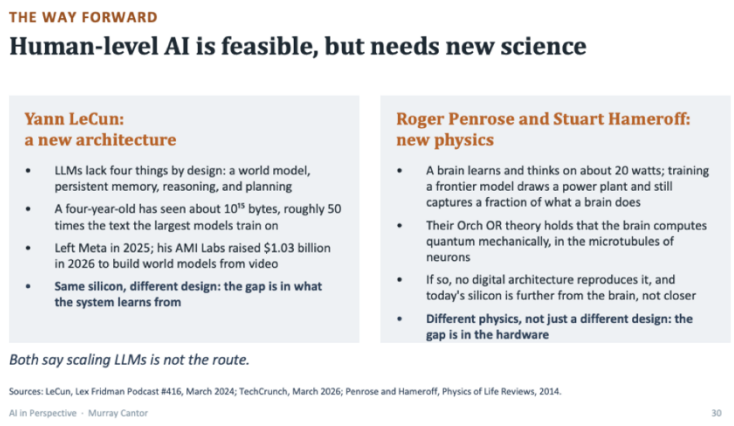
Asimov's Three Laws of 1942 are fiction but the right shape: constraints outside the mind they constrain, with every story about where the rules fall short, which is the engineering problem.

> First Law: a robot may not injure a human being or, through inaction, allow a human being to come to harm.
>
> Second Law: a robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
>
> Third Law: a robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Seat belts and crash standards for cars and trucks came by law in 1966, not from the carmakers; the Motor Vehicle Safety Act let a federal agency set standards and order recalls, and deaths per mile fell about 80 percent, from 5.5 to 1.1 per 100 million miles, as cars got faster.

For AI now, Bill Gates says models need a supervisory layer with hard rules on resources, tools, and harm, calls it the most plausible near-term fix, and says it is not built; nothing requires it yet, there is no FDA or safety board for AI, and self-regulation is not enough.

The greatest risk is that no international standards exist and nobody has the power to enforce them.



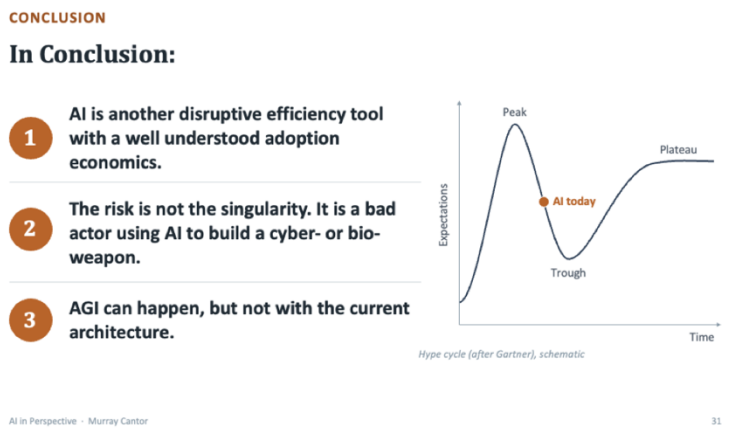
Slide 30 — Human-level AI is feasible, but needs new science

Two serious people say scaling LLMs is not the route, for different reasons.

Yann LeCun argues that LLMs lack four things by design (a world model, persistent memory, reasoning, and planning), that a four-year-old has seen about 10^{15} bytes, roughly 50 times the text the largest models train on, and that the gap is in what the system learns from; he left Meta in 2025 and his AMI Labs raised $1.03 billion in 2026 to build world models from video.

Roger Penrose and Stuart Hameroff argue that the gap is in the hardware: a brain learns and thinks on about 20 watts while training a frontier model draws a power plant, and their Orch OR theory holds that the brain computes quantum mechanically in the microtubules of neurons, so that no digital architecture reproduces it.

This is a minority view but a serious proposal with serious authors, and both camps agree the way forward is new science.



Slide 31 — In conclusion

Three conclusions.

First, AI is another disruptive efficiency tool with well-understood adoption economics: it is past the peak and heading into the trough, and we should expect heavy spending ahead of demand, a painful shakeout, and then real but ordinary gains for those who learned the tool.

Second, the risk is not the singularity but a bad actor using AI to build a cyber- or bio-weapon; the capability exists now, the control layer does not, and no body can enforce standards across borders.

Third, AGI can happen, but not with the current architecture, because of the ceilings on power, chips, money, and text and because scale does not supply judgment.

The takeaways are to use the tools now on work you can check, be skeptical of claims that need endless scaling, watch the money, and push for safety standards before the first large failure.