Beyond the Babble

Why the Real AI Transition Runs Through Power Plants, Not Philosophy

by Dr. Gregory S. Carmichael • CryptoSoWhat.com • July 2026

The Singularity as we are taught to see it: intelligence rendered as pure, luminous abstraction — a museum exhibit observed from a respectful distance. The argument of this piece is that the more instructive view looks quite different.

A FRIEND sent me a thoughtful summary of "the Singularity" and asked what I thought. He had spent a Saturday working through it — reading, watching documentaries, sketching out the canonical argument as he understood it. He wrote it up and circulated it to friends and colleagues, which is what thoughtful people do when they encounter a big idea.

His piece was better than most. It hit the three classical pillars — recursive self-improvement, artificial superintelligence, the merging of biology and silicon — and acknowledged the "complexity brake" counter-argument, which most popular treatments skip entirely. It named the alignment problem. It gave honest weight to both the utopian and the existential frames.

And yet, reading it, I had the same reaction I have to most writing in this genre: the argument is structurally incomplete in a way that matters for anyone actually trying to plan, invest, or build during the next ten years.

This is my longer answer to his question. It is not a refutation of the Singularity idea — there is a legitimate intellectual lineage there worth respecting, from I. J. Good in 1965 through Vernor Vinge in 1993 to Ray Kurzweil's 2005 synthesis. It is an argument that the dominant framing has become a poor planning document, and that the practical version of what we are actually living through looks quite different from the philosopher-futurist version.

The Inherited Frame

The Singularity argument, as it has crystallized in popular form, runs roughly like this. Once we build a machine capable of redesigning itself, it will trigger a recursive feedback loop. Each improved version designs a better successor. The curve bends vertical. Within a short window — some theorists say days, some say years — machine intelligence surpasses all of humanity combined. Combined with brain-computer interfaces and nanotechnology, this transition either delivers paradise (disease cured, scarcity ended, consciousness uploaded) or extinction (a superintelligence with goals indifferent or hostile to human survival). Kurzweil's 2045 date has become a kind of cultural anchor. The event horizon metaphor — that we cannot see past the moment of transition — has done a lot of rhetorical work.

I want to stipulate up front that the people who developed this argument are serious. Good, Vinge, Kurzweil, Bostrom, Yudkowsky — these are not dilettantes. Their work has shaped a generation of thinking about technology's trajectory, and parts of it have held up well.

But it is worth being precise about which parts, because this field has a habit of grading itself generously.

Vinge's 1993 claim was specific. Within thirty years, he wrote, we would possess the technological means to create superhuman intelligence, and shortly thereafter the human era would end. That window closed in 2023. We did not build superhuman general intelligence, and the human era did not end. What arrived instead was a class of systems with extraordinary — in narrow domains, genuinely superhuman — capability, showing up faster than most skeptics expected and staying more bounded than most believers predicted.

That distinction is not a quibble. It is this entire essay in miniature. The capability was real. The scaling story attached to it was wrong. And a literature that cannot mark its own misses is not a literature you should be planning against.

Arguments age. The Singularity argument in its 2005 form was built for a world that no longer exists. The most serious thinkers in this space have already moved on. The popular version has not.

Here is what I think it gets wrong.

The Frame Is Dated

The conversation shifted materially in late 2022. Before then, “will recursive self-improvement happen?” was a reasonable question to debate. After the emergence of large language models, the question changed. It became: “are we already in a soft takeoff, and if so, what does governance look like during the transition rather than at some theoretical event horizon?”

That is a different question with different answers. LLMs did not deliver AGI — they are not AGI, and the serious researchers pushing the frontier are candid about that. What they did was make capability growth legible. Kurzweil’s extrapolations assumed smooth exponential curves. What we actually got was step-function jumps followed by plateaus, punctuated by emergent capabilities nobody predicted from the underlying architecture. GPT-3 to GPT-4 to the current generation was not a smooth curve. It was a series of discrete shocks, each followed by a period of absorption.

This matters because the shape of the change determines the shape of the response. Smooth exponential curves let you plan ahead. Step-function shocks do not. You cannot extrapolate from the last plateau to know when the next jump comes, or how big it will be. That is a different policy environment than the one Kurzweil was writing for.

Intelligence Does Not Scale in a Vacuum

This is where I think the standard framing misses the most, and it is the part that anyone thinking about AI through a capital or operations lens needs to internalize.

The Singularity narrative treats intelligence as software. A mind is an algorithm, and algorithms compound cheaply. If you have the right architecture, you just need enough compute, and compute grows exponentially per Moore’s Law, so intelligence grows exponentially too, and eventually the curve bends vertical.

This is wrong at the level of physics.

Intelligence in its current form runs on GPUs. GPUs run on power. Power runs on grid buildout. Grid buildout runs on transformers (the electrical kind), substation capacity, interconnection queues, high-voltage cables, permitting, water rights, rare earths, and grain-oriented electrical steel. None of these things scale on Moore’s Law curves. They scale on the much slower curves of steel, concrete, permitting, and federal procurement. And as of this writing, every one of them is constrained in ways that actively limit how fast AI capability can grow.

> *Intelligence in its current form runs on GPUs. GPUs run on power. Power runs on grid buildout. None of these things scale on Moore’s Law curves.*

Consider what is actually happening right now.

The generation pipeline is shrinking

Lawrence Berkeley National Laboratory tracks every power plant seeking a connection to the American transmission grid. As of the end of 2025, roughly 8,200 projects were active in those queues, representing about 1,312 gigawatts of generation and another 749 gigawatts of storage — a little over two terawatts, against an installed U.S. fleet of roughly 1,250 gigawatts.

Two features of that number matter more than its size.

First, it is shrinking. The queue peaked near 2.6 terawatts at the end of 2023, fell about 12 percent through 2024, and has continued down since. That decline is not evidence of a loosening constraint. It is the sound of developers giving up — a historic wave of withdrawals driven by interconnection upgrade costs and study delays that make the economics impossible. Of all the capacity that entered these queues between 2000 and 2018, only about 14 percent was ever built. Median time from interconnection request to commercial operation has stretched from under two years in 2008 to roughly five years now.

Second, it measures the wrong side of the meter. LBNL's data covers generator interconnection only. Data centers do not connect as generation. They connect as *load*, through an entirely separate and considerably less mature process. The generation queue tells you about the supply environment a data center is buying into. It does not tell you what the data center itself is waiting for.

For that, go to Texas.

The load side, in the one place we can see it clearly

ERCOT runs the only major American grid not synchronously tied to its neighbors, which makes Texas something close to a controlled experiment. Its large-load interconnection queue — the line data centers actually stand in — grew from 63 gigawatts at the end of 2024 to roughly 226 gigawatts by November 2025. That is nearly a quadrupling in a single year. Roughly three-quarters of it is data centers aiming to connect by 2030.

Against that, ERCOT added about 23 gigawatts of new generation capacity between 2024 and 2025, with another 9 gigawatts slated for early 2026.

Do the division carefully, because the honest version is damning enough without help. The queue grew by 163 gigawatts in one year. New generation arrived at 23 gigawatts over two. That is roughly a seven-to-one mismatch between what wants to connect and what can be served.

And that ratio is generous to the grid, because the numerator is inflated. ERCOT is being flooded with speculative and duplicate requests — the same project filed at several sites to see which one clears first — and the operator has said plainly that it has outgrown the process built to review them. Texas passed SB 6 to force disclosure of duplicate filings, and ERCOT has retained outside consultants to rebuild the large-load interconnection process entirely. So discount the queue. Cut it in half and the mismatch is still three-and-a-half to one. Cut it by three-quarters and Texas still cannot serve what remains.

The tell is not the ratio. The tell is that Texas — the jurisdiction least inclined in the union to tell a business what it may build — is now seriously debating whether large data centers must bring their own generation as a condition of connection.

What the AI transition actually looks like at ground level: transmission towers marching toward a hazy metropolitan skyline, rooted in dust and copper rather than code and philosophy.

The transformer problem, and the steel underneath it

Transformers — the unglamorous metal boxes that step voltage up and down between transmission and distribution — are the second chokepoint, and the harder one.

Before the pandemic, a large power transformer could be sourced in roughly twelve to fourteen months. As of Wood Mackenzie's second-quarter 2025 supply-chain survey, the average had stretched to **128 weeks**, with generator step-up units — the ones that tie a power plant to the grid — at 144 weeks. Prices are up roughly 77 percent since 2019.

Those averages now understate the problem at the top end, because the market has split in two. Distribution-class units under 10 MVA have eased back toward eight to sixteen weeks for standard configurations. Meanwhile the Tier 1 manufacturers of large high-voltage units — ABB, Siemens Energy, Hitachi Energy, GE Vernova — are quoting 48 to 60 months, with some lines booking into 2030 and 2031. The small equipment is recovering. The equipment that actually gates a gigawatt-scale campus is getting worse.

The United States imports roughly 82 percent of the large power transformers it consumes. But the deeper exposure sits one layer upstream, in the steel. Every power transformer requires a core of grain-oriented electrical steel, and Chinese producers accounted for something like 56 percent of global GOES output in 2023. ThyssenKrupp Electrical Steel — the only sizable European producer — announced in December 2025 that its Isbergues plant in France would drop to half capacity and then close entirely from June through September 2026, citing Chinese imports that had tripled since 2022. Europe committed to a historic grid expansion and simultaneously lost its last domestic source of the material that expansion is built from.

The United States has one producer. Cleveland-Cliffs, Butler Works, Pennsylvania.

That is the entire American supply base for the single material without which no transformer gets built.

The capital, and what it cannot buy

Now put the money against the constraint.

Alphabet, Amazon, Meta, and Microsoft have collectively guided to roughly $725 billion in 2026 capital expenditure — up about 77 percent from some $410 billion in 2025. Two details inside that number are more informative than the headline.

Meta raised its guidance and named the reason: higher component and data center costs. That is physical scarcity arriving directly in the capital line. Not more capacity being bought — the same capacity costing more.

And Microsoft is carrying roughly $80 billion in orders it cannot fulfill, constrained by power availability.

Read that second one again. It is not projected demand or an analyst's forecast. It is signed business, at a company that can buy essentially anything, which cannot be delivered because the electricity is not there.

On the delivery side, the 2026 picture is genuinely contested and worth handling carefully. Bloomberg reported on 1 April 2026, drawing on data from Sightline Climate, that roughly 12 gigawatts of U.S. data center capacity was slated to come online this year with only about a third under active construction, and that close to half of planned builds would be delayed or cancelled. Sightline's own framing was narrower — a 30 to 50 percent *delay* range — and the headline figure has since drawn credible criticism from analysts who track the pipeline site by site, on the grounds that nearly all the flagged capacity sat in the speculative, pre-construction "announced" bucket and was never going to land on a 2026 timeline under any rigorous analysis.

That criticism is fair. It also does not rescue the schedule. Strip the forecasting out entirely and the arithmetic still holds: roughly 11 gigawatts of 2026 capacity remained in the announced stage with no construction underway, against typical build times of twelve to eighteen months. Whether or not anyone formally cancels it, capacity that has not broken ground cannot energize this year.

You do not need a prediction. You need a calendar.

An AI data center deployment cycle runs twelve to eighteen months. A large transformer procurement cycle now runs two and a half to five years. Those numbers are not mathematically incompatible — you can order the transformer three years early. Almost nobody did, because in 2021 nobody needed to. The industry is now discovering that the binding constraint on a 2026 project was a purchase order that should have been signed in 2022.

What this adds up to

This is what the Singularity actually looks like at ground level. It is not a mind bootstrapping itself to omniscience. It is a capital-intensive, supply-chain-constrained, regulation-mediated industrial buildout that is straining the physical capacity of the American power system and the global manufacturing base for high-voltage equipment. The chokepoints are not algorithmic. They are made of copper, steel, electrical steel, and signed permits.

If you want to know when and how the AI transition arrives in your life, you will learn more from watching PJM capacity auction prices, DOE Title 17 loan guarantees, transformer order books, and state utility dockets than from reading arXiv papers. Virginia's SB 253, Georgia's SB 34, and Texas's SB 6 are all doing the same thing from different directions — shifting grid-upgrade costs onto the large loads that trigger them, with estimates of 10 to 20 percent added to electricity costs in affected jurisdictions by 2027. That is the actual governance of the AI transition. It is happening in state legislatures, not in white papers about superintelligence.

"Intelligence Explosion" Is Asserted, Not Interrogated

The classical Singularity argument hinges on recursive self-improvement — the idea that once an AI reaches a certain capability threshold, it can redesign itself into a more capable version, which then redesigns itself again, and so on, with the cycle accelerating until the curve bends vertical.

This is a theoretical claim about what is architecturally possible, not an empirical observation about what current systems do.

There is a growing camp of serious AI researchers — François Chollet, Yann LeCun, Subbarao Kambhampati, and others — who argue that current transformer-based systems self-distill and self-refine within a capability envelope, but do not recursively expand it in the way the classical argument requires. Kambhampati's work is particularly sharp on this. He makes a compelling case that what looks like reasoning in modern LLMs is often sophisticated retrieval and interpolation — extraordinary in scope, but categorically different from the kind of open-ended problem-solving that genuine recursive self-improvement would require.

This is a live technical debate, not a settled question. I am not qualified to adjudicate it, and neither, probably, is anyone writing about the Singularity for a general audience. But it matters enormously for planning, because if recursive self-improvement is architecturally gated rather than imminent, the governance problem shifts fundamentally.

The question "how do we align a god-like intelligence that will emerge in the 2030s" is a theological question. It is important, but it is not actionable at the level of policy, investment, or engineering. The question "how do we manage an extended period of rapid but bounded capability growth, distributed across a small number of labs and nation-states, over the next 10 to 20 years" is a practical question. It has concrete answers involving export controls, compute governance, antitrust, industrial policy, and institutional design. We are already bad at it, and the cost of getting it wrong is enormous, but at least it is a problem that can be worked.

The Singularity frame, by collapsing everything into a single future event, makes the practical question harder to see.

The Real Risk Landscape Is the Messy Middle

The Singularity discourse tends to present two scenarios: utopia (disease cured, poverty ended, consciousness transcended) or extinction (a misaligned superintelligence eliminating humanity as a side-effect of optimizing something else). This binary has done enormous damage to clear thinking about what is actually happening.

The more textured risk landscape, which has largely displaced the Bostrom-era framing in serious AI governance circles, looks like this.

Concentration of capability. A small number of labs and a smaller number of governments are building systems that, even in their current bounded form, confer meaningful advantages in scientific research, cybersecurity, content generation, and strategic analysis. If those advantages compound, we get unprecedented asymmetries of power between those who have frontier models and those who do not. This is not a theoretical risk. It is the current trajectory.

Epistemic collapse. Synthetic media at scale, deployed into a political environment that cannot process it, is already degrading the shared factual basis of democratic discourse. This is happening now, at the current level of capability, and will intensify as generative systems improve.

Economic displacement outpacing institutional adaptation. Labor markets, tax systems, credentialing systems, and social safety nets were not designed for the pace of change that capable AI enables in knowledge work. The problem is not that humans become economically useless overnight; it is that the transition happens faster than institutions can adapt, creating political pressures that exceed what existing political frameworks can channel constructively.

Misuse by humans rather than misalignment by the AI. The dominant threat model in serious AI governance work is no longer "the AI decides to harm us." It is "someone uses a capable system to do something terrible" — biothreats, cyber attacks at unprecedented scale, mass manipulation, or sovereign-scale economic coercion. The alignment problem in this frame is about aligning the humans deploying the systems, not the systems themselves.

Gradual disempowerment. Humans remain nominally in control of systems they no longer fully understand, and slowly lose the ability to meaningfully audit, reverse, or shape their outputs. This is the framing developed by Kulveit, Douglas, Ammann, Turan, Krueger, and Duvenaud, adjacent to Paul Christiano's earlier work on what failure actually looks like and to Dan Hendrycks on competitive and evolutionary pressure. It is considerably more worrying to serious researchers than the Skynet scenario, because it does not require any discrete moment of takeoff. It just requires the slow atrophy of distributed human capability.

None of these require AGI. All of them are happening, at small scale, right now. All of them scale with capability. And none of them are well addressed by the utopia-versus-extinction binary that dominates popular discourse.

The existential framing, paradoxically, can be a distraction from the risks that are actually actionable.

What To Do About It

Here is where I part company most clearly with the philosopher-futurist tradition, and it is the thesis I spent the last year working through in book form. The core argument of *The Invisible Hand Meets AI* (Quantum Reserve Press, 2026) is that Adam Smith's framework does not break under AI — it gets sharper.

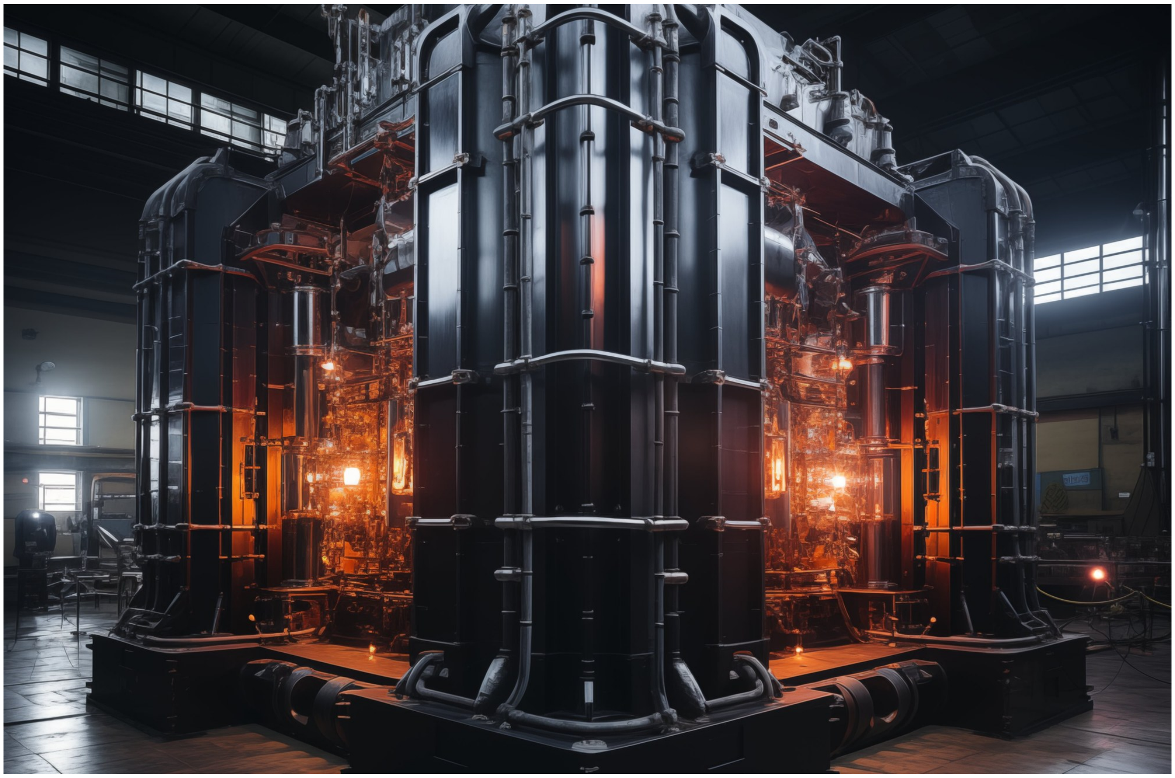
Markets, property rights, and distributed decision-making become more important, not less, when cognitive capability concentrates. The failure mode is not "the AI takes over." It is "a small number of humans and institutions capture the productivity gains, and the rest of the system atrophies — along with the distributed resilience that makes a society hard to coerce, hard to manipulate, and hard to capture."

Trust, atoms, institutions. The three things AI cannot manufacture for itself.

The defense against that is close to what Smith was arguing for in 1776: distributed ownership, credible institutions, rules that apply to everyone, and property rights that individuals can actually enforce. To that list I would add one item Smith could not have anticipated but would have recognized the logic of immediately — monetary architecture that does not require trust in any single actor. These are not nostalgic virtues. They are active load-bearing elements of any system that has to hold weight during a technological discontinuity.

This thesis runs through what I am building operationally, and I mention that only because it is the honest context for the argument.

At Quantum Reserve Capital, it means building financial infrastructure that complies with the GENIUS Act's stablecoin framework — reserves held in short-term Treasuries and insured deposits, monthly disclosures, the kind of boring transparency that lets institutions commit across decades. The bet is that trustworthy monetary rails will be scarce during a transition where traditional financial plumbing is simultaneously being automated and challenged by on-chain alternatives.



The chokepoint that the philosopher-futurist class does not see: a large power transformer, photographed with the reverence usually reserved for sculpture. The machines that decide whether compute arrives on time.

At Advanced Nano-Materials Manufacturing, it means advanced materials and physical upcycling — nano-engineered coatings, waste-to-energy conversion, graphene composites. Atoms. Things that cannot be abstracted into software no matter how capable the software becomes. The bet is that during a cognitive-capability boom, the relative scarcity of physical things — competent manufacturing, strategic materials, energy infrastructure, real-world logistics — increases rather than decreases.

Trust, atoms, institutions. The three things AI cannot manufacture for itself.

The 100-Year Test

I test decisions against a frame I find unusually clarifying: will someone 100 or 1,000 years from now look at this and say it was the right thing to do?

Applied to the Singularity discussion, this frame is particularly useful because most utopian and existential scenarios collapse the time dimension. They imagine a fast transition — years, not decades — to either paradise or extinction. The more likely reality is a messy multi-decade transition in which the institutional, legal, monetary, and physical-infrastructure decisions made in the next five to ten years determine whether the outcome is broadly distributed or narrowly captured.

The decisions that will look right in 100 years are probably not the decisions that optimize for any particular prediction about 2045. They are the decisions that build durable institutions, distribute ownership and capability widely, and preserve the capacity of human communities to govern themselves — regardless of what the technology ends up looking like.

That frame generates very different priorities than “prepare for the Singularity.”

So What

The Singularity as popularized is a useful myth but a poor planning document. It compresses timelines, abstracts away physical constraints, frames risk as a binary between paradise and extinction, and locates the decisive moment at some future event horizon that makes present action feel futile.

The practical version is already here, and it arrives through channels the philosophical discourse barely notices. It arrives through power purchase agreements. Through transformer order books. Through fabrication capacity at TSMC and Samsung. Through the GENIUS Act and its stablecoin custody requirements — which are genuinely consequential and radically underappreciated. Through export controls on advanced GPUs. Through CHIPS Act allocations and DOE Title 17 loan guarantees. Through interconnection queue reform at FERC. Through antitrust enforcement on frontier labs. Through the slow, grinding, unglamorous work of building institutions that can hold weight during a discontinuity.

The people who will shape the actual outcome are not the ones predicting 2045. They are the ones building the grid, the fabs, the custody rails, the regulatory frameworks, and the distributed ownership structures right now. They are utilities, manufacturers, legislators, regulators, engineers, soldiers, financiers, and founders who ship things. They do not talk in event-horizon metaphors. They talk in permits, megawatts, basis points, and delivery dates.

The Singularity discourse has been dominated by philosophers and futurists for thirty years. It needs more engineers, more operators, more builders — and probably more soldiers and more accountants — for the next ten.

That is the conversation I think is worth having, and it is one I am going to keep pushing at in this newsletter. The philosopher-futurist class has had its turn. The infrastructure class is what will actually decide how this goes.

A Note on the Numbers

This piece argues that you should watch order books rather than forecasts. It would be poor form not to show my own. Every quantitative claim above, with its source and the date it was measured:

Claim	Source	Measured as of
~8,200 projects; 1,312 GW generation + 749 GW storage in U.S. interconnection queues	LBNL, *Queued Up: 2026 Edition*	End of 2025 (pub. June 2026)
Queue peak ~2.6 TW; −12% in 2024	LBNL, *Queued Up* 2024 and 2025 editions	End of 2023; end of 2024
~14% of queued capacity ever built	LBNL, queue cohort 2000–2018	Cohort through 2018
Median interconnection-to-operation ~5 years	LBNL	2023 cohort
ERCOT large-load queue 63 GW → 226 GW; ~3/4 data centers	ERCOT system planning update to board	November 2025
ERCOT +23 GW generation; +9 GW slated	ERCOT	2024–2025; early 2026
LPT lead time 128 weeks; GSU 144 weeks	Wood Mackenzie T&D Equipment Supply Chain Survey	Q2 2025
Pre-pandemic LPT lead time 12–14 months	Industry supplier reporting	Pre-2020
Transformer prices +77%	Industry price indices	Since 2019
Tier 1 HV quotes 48–60 months	Manufacturer quote tracking	June 2026
U.S. imports ~82% of its LPTs	Trade data analysis	2026
China ~56% of global GOES output	Global electrical-steel capacity data	2023
ThyssenKrupp Isbergues half capacity, closed Jun–Sep 2026	Company announcement	December 2025
Big Four 2026 capex ~$725B, +77% YoY	Company guidance, Q4 2025 and Q1 2026 calls	April 2026
Microsoft ~$80B unfulfilled backlog, power-constrained	Company disclosure	Fiscal Q3 2026
12 GW 2026 capacity, ~1/3 under construction; 30–50% delay range	Sightline Climate, via Bloomberg, 1 Apr 2026	Q1 2026

Figures move. Several of these will be wrong within two quarters, and the direction of travel matters more than any single value. Where a number here disagrees with a number you hold, check the measurement date before checking the arithmetic — in this market, most disagreements turn out to be vintage rather than error.

Dr. Gregory S. Carmichael is CEO of Quantum Reserve Capital and founder of Advanced Nano-Materials Manufacturing. A retired USAF Lieutenant Colonel with 26 years of service, he holds a DBA in International

Finance, an MA in International Relations, and a BA in Physics and Chemistry. He writes at CryptoSoWhat.com and is the author of The Invisible Hand Meets AI (Quantum Reserve Press, 2026).