

The Map Nobody Has Drawn

Why the AI Stack Isn't What You Think and What Actually Decides Where Value Goes

CORE INSIGHTS

- **Technology constraints are shifting from capability to infrastructure.** Your AI strategy is obsolete if it doesn't account for power, cooling, and grid interconnection.
- **The bottom of the stack is consolidating power.** For the first time in 40 years, a memory vendor has pricing leverage. Who profits now? not the layer you think.
- Governance is the new supply chain problem. You can deploy AI anywhere. **Regulators decide where you can defend deploying it.**
- Vertical integration works differently this cycle. IBM lost to distributed computing. **AI's capital intensity may reward integration instead.** This breaks historical pattern.
- **Value capture is inverted today.** It won't stay that way. Bottom layers capture profit today. Middle layers burn capital. When margin migrates upstack that's the defining event of this era.
- Your layer determines which constraint reaches you first. Power hits infrastructure vendors now. Governance hits SaaS layers later. That timing difference decides who survives.
- **Dominant constraints are moving, not disappearing.** Power gates buildout today. In 25 months, it becomes a cost. Which constraint moves into the power slot next? That determines the winner.
- **The physical world is back in the value chain.** For the first time since mainframes, copper, water, transformers, and geography matter as much as algorithms. Software-centric thinking is now a liability.

The Constraint Nobody Named

Microsoft has an \$80 billion problem.

Not a problem with chips. Not a problem with AI capability. Not a problem with customer demand.



In January 2026, CFO Amy Hood disclosed a backlog: \$80 billion in committed Azure cloud orders that Microsoft cannot fulfill. Disclosed in April 2026, the Azure Remaining Performance Obligation had grown to \$627 billion. By June 2026, it reached \$678 billion. The reason? Power. Transformers. Grid interconnection. The physical infrastructure that moves electrons from a utility company to a data center runs out before chips do.

The bottleneck isn't innovation or demand. It's power.

Transformers. Grid capacity. The physical infrastructure that delivers electricity to data centers now moves slower than the chips that sit inside them. ***"This is the hidden layer. Not the technology stack you see in diagrams. The constraint stack underneath it."***

At every layer, ask three simple questions:

1. Where is value created?
2. Where is value captured?
3. What single constraint decides whether anyone gets to use it?

"If you want to understand where AI goes next, stop reading the stack as a technology hierarchy. Read it as a value chain."

Why This Map Exists

It happens everywhere.

At a board meeting where I'm serving as an advisor. Over dinner. On a call with a former colleague looking for career advice. In the middle of a project focused on sales growth. On a podcast. Even standing on a softball field, waiting for my turn at bat.

Someone asks me the same question:

"Joe, what do you think of AI?"

It sounds like a simple question. It isn't.

It is one of the biggest questions of our time, yet we keep asking it as if there is a single answer.

For years, I gave the best answer I could. And almost every time, I walked away knowing it wasn't good enough. Not because I didn't have an opinion, but because the question itself was too broad. Without a reference model, any answer was inevitably shallow. I was answering a question that couldn't be answered from a single perspective.

So, I started building a map. Thirteen layers that we will explore in a series of articles.

The **AI Value Map** is my attempt to make sense of where value is being created, where it is moving, who controls it, what constraints stand in the way, and what happens next.

But the map isn't the answer.

The map is the tool.

And importantly, it is a **public map** one meant to be shared, challenged, and improved. More than 25 reviewers across numerous disciplines, industries, and perspectives helped shape the AI Value Map. You can find their names in the acknowledgments section of the article.

The map is designed to uncover the hidden changes, technology, markets shifts, companies, where value created and destroyed, we learn more, to question, to identify second order effects. I don't believe anyone has the complete map today. I certainly don't.

What I do believe is that we need one.

Because the same map can help different people ask fundamentally different questions.

For a board:

Where do we sit? What constraint stops us? Who controls the layer above us? What are the second-order effects?

For a colleague navigating a career transition:

Where is value being created? What skills are becoming scarce? Where is talent concentrating—or disappearing?

For an executive:

How do we reorganize? Where is our constraint? What should we build? And what does the next generation of our business look like?

Same map. Different lenses. Different questions. Different decisions.

Not because I have all the answers.

Because I was tired of pretending that one answer was enough. Enjoy the Map.

THE HIDDEN LAYER: A MAP OF THE AI STACK

13 layers | 13 dominant constraints | 1 question: Which one stops you? Layer -1: Natural Resources

Layer 0: Energy & Physical Infrastructure

Layer 1: Design Tools & IP

Layer 2: Accelerator Design

Layer 3A: Semiconductor Manufacturing

Layer 3B: Advanced Packaging & HBM

Layer 4: Systems & Networking

Layer 5: Cloud AI Infrastructure

Layer 6: Foundation Models

Layer 7: Data & Model Infrastructure

Layer 8: AI Governance & Security

Layer 9: Intelligence Delivery

Layer 10: Enterprise Transformation

Why Thirteen Layers

A layer earns its place when three things are true:

- **Its own economics**
- **Its own owners**
- **Its own way of failing**

Could you defend eleven layers? Fifteen? Absolutely. The exact number matters less than the principle: when a constraint is structural, it deserves its own layer.

This map improves in public. AI's stack moves too fast for stone tablets. Argue with it.

Five Forces That Run Through Every Layer

These constraints cut vertically across the entire stack:

1. Capital Structure

Nvidia guarantees \$250B of OpenAI's data center debt. That guarantee creates systemic risk.

2. Talent

Every layer starves for engineers. The best cluster at winning companies. Everyone else falls behind.

3. Data Ownership

You cannot govern what you don't own. Data ownership cascades from Layer 7 to Layer 10 accountability.

4. Geopolitics

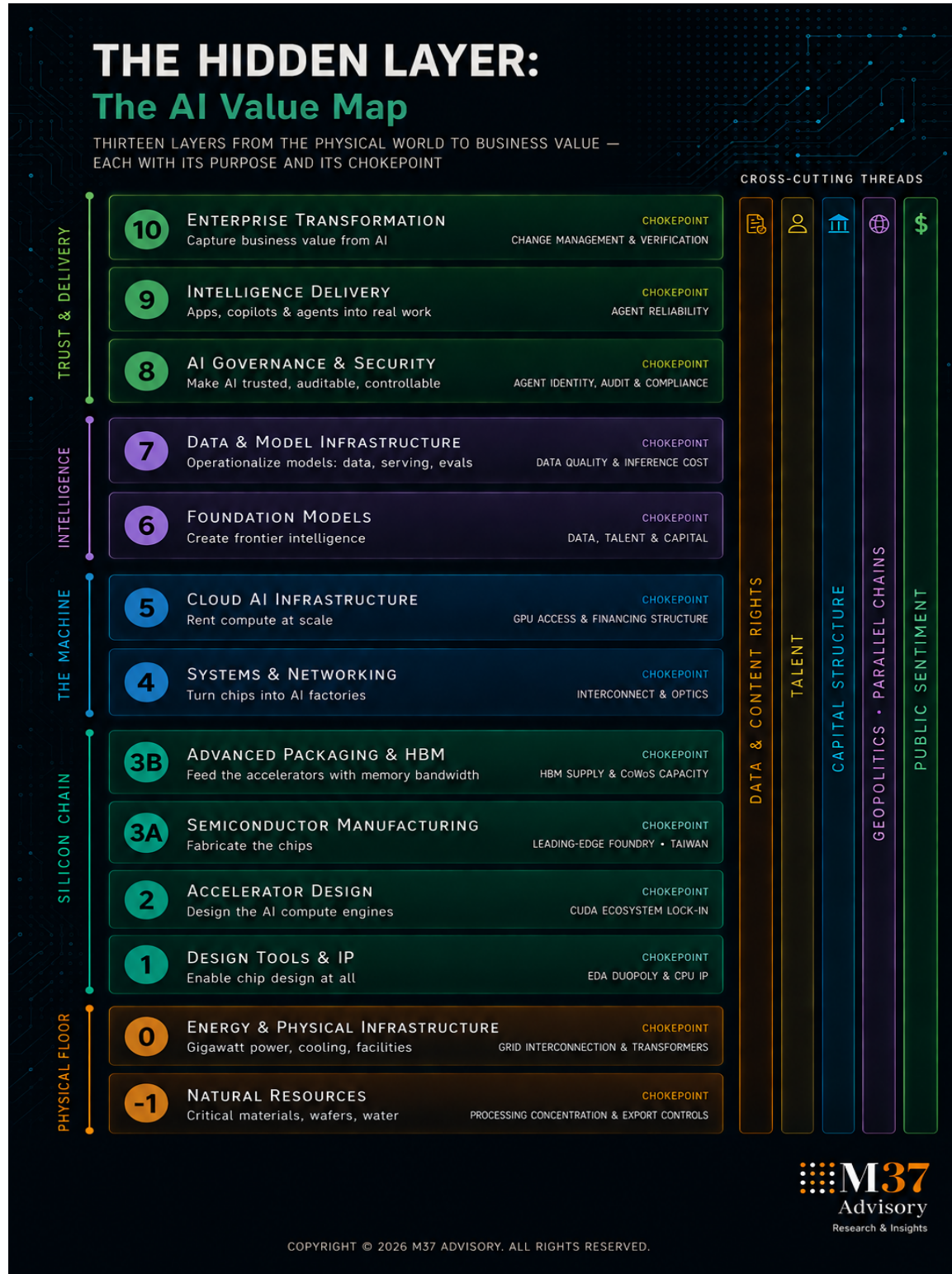
Taiwan makes chips. Netherlands makes the equipment. China wants both. Export controls split the stack in two.

5. Public Sentiment

AI fatigue, displacement concerns, and data center resistance can halt buildout. Regulators listen.

To subscribe to the series, complete the form [LINK](#)

Here Is the Map



To subscribe to the series, complete the form [LINK](#)

Three Examples That Show This Works

① Power Is the Gate

Most people think semiconductors gate AI deployment. **They don't. Electricity does.** You can have the best chips in the world, but if you can't plug them into data center capacity fast enough, they sit in inventory.

Microsoft proved this:

- ✓ They have the chips
- ✓ They have the customers
- ✓ They have the capital
- ✗ They cannot find the power

**A transformer is now more valuable than a GPU.
That's the inversion nobody predicted.**

② Memory Became the Hidden Hand

For forty years, the memory industry operated like a commodity:

Supply rises → Prices fall. Always.

2023-2024:

- Single vendor held >50% HBM (High Bandwidth Memory) market share
- Raised prices 20% while customers begged for allocation
- First time in 40 years this happened that I could remember

Why? Because every accelerator depends on HBM bandwidth. If HBM is slow or unavailable, no amount of GPU capability matters. Layer 3B the afterthought became the hidden hand.

③ Trust Became the Rate Limiter

Technology companies ship anything that works. **Regulated institutions won't.**

"Can I explain and defend this outcome to a regulator in court?"
Governance is the gate. Not capability. Not cost.

The constraint isn't whether an AI system can produce an answer. The constraint is whether a board can defend that answer to regulators. Governance, not capability, is the gate. Ask any CIO at a financial institution: regulators don't ask, "Is your model accurate?" They ask, "Can you explain and defend this outcome in court?" That question reshapes Layers 8, 9, and 10 entirely.

Why This Pattern Keeps Repeating

History doesn't repeat, but it rhymes.

IBM in 1975 owned the entire stack and Digital Equipment Corporation (DEC) the same in the 1980's the chips (inside their systems), the systems themselves, the software that ran on top, the services that supported customers, and the relationship with every customer. IBM was the most valuable company on earth. Vertical integration meant total control. Then someone said: "What if we focus on just one thing?" Intel focused on chips. Microsoft focused on operating systems. Neither company worked with IBM's permission. Both became worth more than IBM. By the 1990s, IBM had posted the largest loss in American corporate history.

Today's giants (Nvidia, OpenAI, Google, Microsoft) are trying to rebuild that integrated stack. History suggests focused horizontal competitors will eventually break this apart, just like Intel and Microsoft broke IBM. But this time, the stack has capital requirements, cloud economics, export controls, and energy constraints that didn't exist in 1975. Will the pattern hold?

Three Questions You Should Asking

1. Where does our company sit in this stack?

(Are we Layer 2? Layer 6? Layer 9? Do we know?)

2. What constraint can stop us?

(Power? Capital? Talent? Governance? Whose budget controls it?)

3. Who controls the layer above us?

(And what happens if they change strategy?)

If you can't answer all three, the map isn't finished yet.

The Prediction

I missed the HBM story in 2023. I should have seen it. Here's what I see now:

" The quarter HBM contract prices flatten while new capacity ships, the inversion begins to unwind. I expect that signal within twenty -five months"

If it doesn't appear, this cycle has broken from forty years of history.

Hold me to it.

To subscribe to the series, complete the form [LINK](#)

What Comes Next

This is the map. The next articles go deeper layer by layer, constraint by constraint. We'll explore what makes Layer 0 different from Layer 5. What happens when Layer 3B pricing breaks. How you compete when the layer above you is controlled by a single company.

But first: Which layer would you redraw? Which constraint would you challenge?

The map improves in public.

Article 2 in the Series

**"The Accidental Kingmaker:
Why Layer 3B Memory Is the Deal All Others Hinge On"**

To subscribe to the series, complete the form [LINK](#)

Acknowledgments

This map was pressure-tested by many practitioners across a number of industries standing on different layers of the stack: a data architect a CPO Chief Product Officer, a CIO at a financial institution, legal counsel, board members, CTO Chief Technology Officers, architects, strategists, operators and major vendors along the stack who have lived through multiple technology cycles. They pushed back hard. The map is better because of it. Special thanks to some significant contributors.

- Don Demcsak, Global Technology Strategist, DELL Technologies
- Ben Schreiner, Head of AI and Modern Data Strategy Business Development, AWS
- John Rebecchi PhD, New York Institute of Technology
- Peter Moser, Senior. Distinguished Technologist, Hewlett Packard Enterprise
- Rich McGrath, Senior System Engineer, Everpure
- Marc Lauritsen, President Capstone Practice Systems, Harvard Law
- Chris Coomey, Sr. Vice President & CIO, Saint Mary's Credit Union
- Santiago Aldana, Chief Product Solution Officer, Saint Mary's Credit Union
- Kevin Coulter, Former Field CTO, AI and Data Science at Dell Technologies, Global Partner Presales at Cloudera
- Ray Richardson, Enterprise AI Architect, Founder, Praesidium AIOS

About the Author

Joe Batista advises boards, C-suite executives, and go-to-market teams on technology value creation, growth strategies, and the opportunities hidden beneath major technology shifts. He has spent more than 40 years in the technology industry, guiding companies on growth and technology decisions, spanning the minicomputer era Digital Equipment Corporation (DEC) through client-server, cloud, virtualization, and now AI, with executive experience at Dell Technologies and Hewlett Packard Enterprise.

Through M37 Advisory, Joe helps leadership teams recognize recurring patterns, uncover second-order effects, find the opportunities others overlook, and identify the strategy worth betting on. His work is grounded in a simple belief: **What if the biggest obstacle to your growth isn't the competition, but the way you think?**

"Most executives have been bombarded with 'theories' about what might work. Based on our research, Joe's extensive playbook is rooted in creating alignment of interest and business outcomes that leaders should take notice of."

— Jim Dickie, Research Fellow & Sales Mastery Founder, CSO Insights

Email: Chief.Creatologist@M37-Advisory.com

Website: <https://www.m37-advisory.com>

Idea Vault: <https://www.m37-advisory.com/idea-vault>

COMPLETE SOURCES & CITATIONS

The Map Nobody Has Drawn | Full Source Documentation

[1] Microsoft Azure Backlog & Power Constraints

\$80 billion backlog disclosed: January 29, 2026 (CFO Amy Hood, earnings call)

Azure RPO \$627 billion: April 29, 2026 (Q2 2026 earnings report)

Azure RPO \$678 billion: June 2026 (Q3 2026 earnings report)

Access: <https://www.microsoft.com/en-us/investor/earnings/>

[2] Nvidia's OpenAI Data Center Guarantee

Amount: \$250 billion guarantee on OpenAI's Picketon, Ohio data center construction and lease costs

Additional: Separate from \$350 billion in accelerator capex not covered by guarantee

Source: AMO Research, "Every Tech Bubble Since 1900 Did This," Medium, August 2026

Access: <https://medium.com/the-investors-handbook/every-tech-bubble-since-1900-did-this-6de50ed6a078>

COMPLETE SOURCES & CITATIONS

The Map Nobody Has Drawn | Full Source Documentation

[3] Kansas City Federal Reserve Warning

Speaker: Jeff Schmid, President, Federal Reserve Bank of Kansas City

Date: August 4, 2026

Quote: "We have to correlate what's happening in AI...to create a systemic problem. Is this industry becoming another too big to fail?"

Source: AMO Research (as cited above)

Official: Federal Reserve Bank of Kansas City communications: <https://www.kansascityfed.org/>

[4] Lucent Technologies Precedent (Telecom Bubble)

Period: Telecom Bubble buildout (1999-2000), collapse (2000-2002)

Mechanism: Vendor financing of customer capex; customers failed; vendors held worthless receivables

Impact: Customer bankruptcies (WorldCom, Global Crossing, Level 3 Communications) left Lucent holding billions in uncollectible debt

Reference: Telecom Industry Collapse case studies; Business history archives

Source: SEC filings, Lucent Technologies historical reports: <https://www.sec.gov/>

[5] Micron Memory Oversupply & AI Recovery (2023-2024)

Event: Pandemic-driven oversupply of DRAM and NAND

Response: Micron slashed capex, cut wafer starts by ~20%, reported heavy financial losses in 2023

Recovery: AI-driven demand inflection 2024, return to profitability

Access: <https://investor.micron.com/earnings/earnings-releases>

[6] Semiconductor Fabrication Concentration (TSMC)

Market Share: ~72% of top 10 foundry revenue (Q1 2026)

Second Place: Samsung Foundry, with TSMC more than 3.5x larger

Capex Allocation: 70-80% of 2026 capex (\$60-64 billion) to advanced process technologies

Source: TrendForce, TSMC Q2 2026 earnings conference (July 16, 2026)

Access: <https://investor.tsmc.com/>

[7] ASML Lithography Equipment (EUV Monopoly)

Company: ASML Holding N.V., Veldhoven, Netherlands

Founded: 1984 (joint venture between Philips and ASM International)

Market Position: Sole global supplier of extreme ultraviolet (EUV) lithography systems

EUV Market Share: 100% monopoly—no other supplier has commercialized EUV technology

DUV Market Share: ~90% of global lithography tools market

Source: ASML corporate: <https://www.asml.com/>

Reference: Semiconductor Industry Association industry reports

[8] HBM (High Bandwidth Memory) Market Dynamics

Concentration: Single vendor holds >50% of high-bandwidth memory market

Pricing Anomaly: Mature product generation saw ~20% price increase in 2023-2024

Historical Context: First time in 40 years a memory vendor had pricing power

Constraint: HBM bandwidth now the bottleneck in accelerator performance stacking

Source: Semiconductor industry analyst reports (2024-2026)

COMPLETE SOURCES & CITATIONS

The Map Nobody Has Drawn | Full Source Documentation

[9] Semiconductors as Share of AI Data Center Capex

Semiconductor Share: >50% of total AI data center capex (SIA 2026 report)
Server Rack Value: >95% of AI server rack value is semiconductors
AI Chip Market 2026: ~\$500 billion (Deloitte 2026 outlook)
AI Chip Market Projection: \$1.2 trillion annually (2028-2029)
Global Investment: >\$4 trillion in government + industry AI data center infrastructure (2024-2028)
Semiconductor Allocation: Up to \$2.8 trillion dedicated to semiconductors
Source: SIA 2026 State of the Industry Report: <https://www.sia-online.org/>
Reference: Deloitte 2026 Semiconductor Industry Outlook: <https://www.deloitte.com/>

[10] Melissa Reeve, Hyperadaptive: Rewiring the Enterprise to Become AI-Native

Author: Melissa M. Reeve
Title: Hyperadaptive: Rewiring the Enterprise to Become AI-Native
Publisher: IT Revolution Press
Publication Date: May 12, 2026
ISBN: 9781966280262
Pages: 304
Framework: Five-stage organizational progression from AI foundation to organizational orchestration
Focus: How enterprises evolve from running AI projects to becoming Hyperadaptive Organizations
Access: <https://itrevolution.com/> or major book retailers

[11] Theory of Constraints (Eliyahu M. Goldratt)

Original Work: The Goal (1984) and Theory of Constraints (TOC) methodology
Core Principle: Every system has one constraint that limits throughput; managing that constraint is the key to optimization
Application: Foundation for "dominant constraint per layer" thesis in the Hidden Layer model
Reference: Goldratt Institute: <https://www.goldratt.com/>

[12] Porter's Value Chain (Michael E. Porter, 1985)

Original Work: Competitive Advantage: Creating and Sustaining Superior Performance (1985)
Framework: Value chain activity mapping + five-force competitive analysis
Application: Adapted for technology stack analysis and AI infrastructure economics
Publisher: Free Press
Reference: Harvard Business School case studies and frameworks

[13] MCP (Model Context Protocol) Specification — July 28, 2026, Update

Specification: Model Context Protocol (MCP)
Critical Change: July 28, 2026, update removed stateless trust model
New Requirement: Permission re-check on every agentic call (not once per session)
Implication: Operationalizes "scoped, verifiable agency" in Layer 8 governance
Industry Significance: Forcing function that makes authorization architecture visible and measurable
Access: <https://modelcontextprotocol.io/> | GitHub: <https://github.com/anthropics/mcp>

COMPLETE SOURCES & CITATIONS

The Map Nobody Has Drawn | Full Source Documentation

[14] NIST AI Risk Management Framework (NIST AI RMF)

Organization: National Institute of Standards and Technology (U.S. government)
Framework: NIST AI Risk Management Framework
Coverage: Risk categories, governance approaches, compliance mechanisms for AI systems
Relevance: Cross-reference for Layer 8 (Governance & Security) governance and compliance requirements
Access: <https://www.nist.gov/aip/ai-risk-management-framework>

[15] U.S. Export Controls on AI Semiconductors

Authority: U.S. Department of Commerce, Bureau of Industry and Security
Status: Ongoing AI chip export restrictions (current as of August 2026)
Impact: Bifurcates semiconductor supply chains (U.S./Western alliance vs. China/Russia coalition)
Affected Layers: Layers 1-3 (design, fabrication, packaging) most directly affected
Policy Mechanism: Executive order and multilateral coordination through the Quadrilateral Security Dialogue (QUAD)
Access: <https://www.bis.doc.gov/>

How to Access These Sources

Most sources are publicly available through official websites, SEC filings, earnings reports, or academic/industry publications. URLs provided above link directly to source organizations. For subscription-based industry reports (Deloitte, TrendForce), check your organization's research access, or contact the publication directly.

To subscribe to the series, complete the form [LINK](#)