The AI Infrastructure Revenue Gap: Implications for Market Structure and Competition

Abstract

This paper examines competitive dynamics in artificial intelligence during a period of unprecedented infrastructure investment that may not be economically sustainable. A September 2025 Bain & Company analysis estimates AI firms face an $800 billion annual revenue gap by 2030 to fund capital expenditures. We calculate this gap expands to over $1.5 trillion annually when accounting for the accelerated replacement cycles of AI infrastructure: chips have 1-3 year useful lives due to technological obsolescence and physical failure, yet firms currently depreciate these assets over 5-6 years. This discrepancy between accounting assumptions and operational reality creates what we term a “capital subsidy” that enables incumbent coalitions to pursue aggressive application-layer strategies during the critical market formation window.

The timing of this revenue challenge matters for competition. Market structure at the application layer—where enterprises choose AI models to integrate into their operations—is being determined now, while the long-run economics of the infrastructure buildout remain uncertain. Hyperscaler-model developer coalitions (Microsoft-OpenAI, Amazon-Anthropic, Google-Gemini) are establishing deep enterprise integrations through pricing and capacity expansion enabled by the capital subsidy. If these investments prove unsustainable once assets require replacement at actual operational lifecycles, the question becomes whether early-mover advantages and customer lock-in will persist or whether market structure remains contestable.

We analyze three competitive scenarios. First, if model capabilities continue scaling meaningfully with increased compute, training capacity creates lasting advantages through superior model quality—though this scenario appears increasingly implausible as capabilities plateau. Second, if model capabilities plateau and application-layer switching costs prove durable, incumbent coalitions may retain dominance even when better alternatives emerge, analogous to Google's search position but with substantially higher technical integration barriers. Third, if capabilities plateau but switching costs remain manageable through standardization and interoperability, the capital subsidy creates temporary distortions without permanent competitive foreclosure. We document how depreciation conventions, coalition financing arrangements, and enterprise integration patterns suggest the second scenario is most likely.

The analysis reveals how circular vendor financing structures amplify these dynamics. Unlike the 2001 telecom bust where long-lived assets became available to new entrants at fire-sale prices, AI chips' short useful lives mean excess capacity cannot enable competitive entry if current buildouts prove excessive—potentially cementing the market structure formed during the subsidy window regardless of whether underlying economics are sustainable.

Drawing on industry financial data, capital expenditure disclosures, and coalition partnership structures, we find that accounting practices may be obscuring the true cost of maintaining AI infrastructure leadership during the critical period when application-layer customer relationships are being established. We propose enhanced disclosure requirements for infrastructure replacement cycles and interoperability standards to preserve competitive entry opportunities while market structure remains fluid.

Keywords: artificial intelligence, competition policy, accounting standards, infrastructure investment, vendor financing, market structure, switching costs, application layer competition

1. Introduction

1.1 The $1 Trillion Question

In 2025, eight major technology companies are projected to invest over $300 billion in AI infrastructure (Goldman Sachs 2025). This represents one of the largest industrial capital buildouts in United States history, comparable in scale to the railroad expansions of the nineteenth century or the interstate highway system of the twentieth century. Yet this massive deployment rests on economic foundations that appear fundamentally unstable.

A September 2025 analysis by Bain & Company estimates that AI firms will face an annual revenue shortfall of approximately $800 billion in 2030 to fund their ongoing capital expenses (Bain & Company 2025). This revenue gap suggests the industry faces a sustainability crisis even before accounting for any systematic mispricing of asset replacement costs. Our analysis indicates that when realistic useful life assumptions replace current accounting conventions, this revenue gap expands to over $1.5 trillion annually—nearly doubling the already substantial sustainability challenge identified by Bain.

The core accounting puzzle is straightforward: the specialized chips powering AI workloads have a useful economic lifespan of one to three years due to rapid technological obsolescence and physical degradation under high-utilization conditions, but companies depreciate these assets over five to six years. This mismatch creates what we term a "capital subsidy"—an apparent profitability cushion that enables incumbent coalitions to pursue strategies during the critical market formation window that would appear unsustainable under more realistic assumptions.

1.2 Where Competition Happens: The Application Layer

A critical insight for competition analysis is that we believe market power in AI is not primarily determined at the model development layer. As model capabilities plateau and open-weight alternatives approach proprietary model quality, model development is commoditizing. We argue that competition is instead concentrating at the application layer—where enterprises integrate AI capabilities into their operations and where customer relationships, switching costs, and integration depth determine competitive outcomes.

The generative AI stack can be conceptualized in three layers:

Infrastructure Layer: Hyperscalers providing compute, storage, and networking (Microsoft Azure, Amazon AWS, Google Cloud)

Model Layer: Developers creating and serving AI models (OpenAI, Anthropic, Google DeepMind, Meta)

Application Layer: Enterprises integrating AI capabilities into business operations, and specialized AI application builders creating vertical solutions

The market structure is coalescing around vertically-integrated coalitions spanning infrastructure and model layers: Microsoft-OpenAI, Amazon-Anthropic, and Google's integrated structure. These coalitions compete primarily at the application layer—not for which model is technically superior, but for which coalition's integrated stack becomes embedded in enterprise operations.

The capital subsidy matters because it enables incumbent coalitions to establish application-layer dominance during the critical window when customer relationships are being formed. By pricing aggressively while reporting sustainable economics, coalitions can lock in enterprise customers whose switching costs will prevent competitive correction even when: (1) true replacement costs become visible, (2) model capabilities commoditize, and (3) better alternatives emerge.

Thus, this is fundamentally a story about application-layer foreclosure, not model-layer competition.

1.3 Research Questions and Contribution

This paper addresses three central questions:

First: What is the magnitude of the capital subsidy created by the mismatch between accounting depreciation and actual useful life of AI infrastructure, and how does this subsidy flow through coalition structures to enable application-layer competitive strategies during the market formation window?

Second: Under what technological and market scenarios might the competitive advantages established during this formation window persist or dissipate? Specifically:

- If model capabilities continue scaling, does training capacity create durable advantages through superior quality?
- If capabilities plateau and application-layer switching costs prove high, can incumbent coalitions maintain dominance even when better alternatives emerge?
- If capabilities plateau but switching costs remain manageable, does the capital subsidy create only temporary distortions without permanent foreclosure?

Third: How do circular vendor financing structures amplify the competitive implications of the capital subsidy, and why does the short useful life of AI infrastructure create different dynamics than previous technology booms when financial pressure emerges?

Our analysis contributes to several literatures. First, we add to research on accounting policy's role in capital markets by examining how depreciation assumptions may affect competitive dynamics in capital-intensive technology markets during their formation period. Second, we contribute to the literature on vertical relationships and competition in digital platforms by analyzing coalition structures where infrastructure providers partner with application-layer competitors. Third, we build on emerging work on AI market structure by examining how competitive dynamics increasingly concentrate at the application layer as model development commoditizes, and exploring the conditions under which early advantages in this layer might prove durable. Fourth, we extend the literature on vendor financing by analyzing how circular capital flows interact with short asset lifecycles to change the competitive implications of financial instability. Unlike telecommunications, where long-lived assets could be redeployed by new entrants after market corrections, short useful life in AI infrastructure means excess capacity cannot enable competitive entry—fundamentally changing the policy calculus.

1.4 The Strategic Logic of Temporary Subsidies

The capital subsidy creates a distinctive strategic opportunity. Incumbent coalitions need not rely on the subsidy indefinitely. Instead, the strategic logic is temporal:

Years 1-3 (Market Formation Window): Use the capital subsidy to price aggressively, expand capacity rapidly, and establish deep enterprise integrations at the application layer. Report profitability that attracts capital at favorable terms. Build customer relationships during the critical period when market structure is being determined.

Years 4-5 (Crystallization Period): When accounting might catch up to operational reality and revenue gaps become visible, customer relationships are already locked in through switching costs, integration depth, organizational inertia, and multi-year contracts. The market structure formed during the subsidy window persists regardless of whether underlying economics are sustainable.

The question is not whether this economics works indefinitely. The question is whether the market structure crystallized during the formation window proves durable enough that competitive correction becomes impossible even when better technology and more realistic economics emerge.

The parallel to Google's dominance in search is instructive. Google did not need to maintain its advantages through superior technology indefinitely. Once default positions and integration created lock-in, competitive alternatives struggled to gain traction even when offering comparable or superior quality. The Department of Justice's antitrust case demonstrated that lock-in proved decisive despite apparently low switching costs.

Application-layer integration in enterprise AI creates switching costs orders of magnitude higher than consumer search engine selection. If lock-in proved durable in search, application-layer lock-in in AI may prove even more persistent.

1.5 Structure

The paper proceeds as follows. Section 2 reviews relevant literature on competition in technology markets, accounting policy effects, and the economics of application-layer competition. Section 3 presents our analytical framework and documents the asset life mismatch. Section 4 examines the revenue sustainability gap and its interaction with the capital subsidy. Section 5 analyzes how the subsidy flows through coalition structures to enable application-layer strategies. Section 6 examines three scenarios for competitive dynamics under different technological trajectories. Section 7 examines how vendor financing amplifies the competitive harm in the most concerning scenario. Section 8 proposes policy responses. Section 9 concludes.

2. Background and Literature

2.1 Competition and Market Structure in Technology Markets

The economics literature on competition in technology markets emphasizes several dynamics relevant to understanding application-layer competition in AI. Network effects and switching costs can create durable competitive advantages even when alternative products are technically superior (Farrell and Klemperer 2007; Shapiro and Varian 1998). Timing matters profoundly in markets characterized by learning curves and scale economies—early movers can establish positions that later entrants struggle to overcome regardless of technological superiority (Arthur 1989; Lieberman and Montgomery 1988).

Recent antitrust cases have demonstrated how integration depth and default positions create lock-in even when switching costs appear minimal. The Google search case revealed that default search engine placements created persistent competitive advantages despite the apparent ease of consumer switching—users could change search engines with "one click" yet rarely did. Default positions combined with organizational and interface integration proved decisive.

These insights apply with greater force to application-layer competition in enterprise AI. Enterprise integration requires months of engineering work, multi-year contracts create financial switching costs, security certifications and compliance frameworks are built around specific technology stacks, and organizational workflows and training are developed around particular providers. If lock-in proved durable in consumer search with trivial switching costs, application-layer integration with substantial switching costs may create even more persistent competitive advantages.

2.2 Vertical Integration and Foreclosure in Digital Markets

The literature on vertical foreclosure examines conditions under which vertically integrated firms can disadvantage non-integrated rivals (Rey and Tirole 2007; Salop and Scheffman 1983). Traditional theories focus on access to essential inputs or distribution channels. Our analysis extends this framework by examining how accounting treatment of shared infrastructure investments can create effective subsidies that flow through vertical partnerships to affect downstream competition at the application layer.

Recent work on digital platform competition emphasizes the role of ecosystem orchestration and complementary investments (Jacobides, Cennamo, and Gawer 2018). In "platform ecosystems," value creation depends on coordinated investments by multiple parties, and the distribution of rents depends critically on bargaining power and contract structures (Gans 2022). Vertical integration in digital markets often serves to secure access to complementary capabilities and to coordinate investments rather than to foreclose rivals from essential inputs (Crémer, de Montjoye, and Schweitzer 2019).

Our contribution is to show how accounting conventions affect the apparent economics of these partnerships during critical market formation periods, enabling strategies at the application layer that would appear unsustainable under realistic cost accounting. The foreclosure concern is not that rivals are denied access to infrastructure, but that application-layer competition occurs on a tilted playing field where coalition members" costs are systematically understated.

2.3 Application Layer Competition and Customer Lock-In

A growing literature examines competition dynamics in enterprise software markets where switching costs are substantial (Shapiro and Varian 1998; Greenstein 1993). Enterprise software exhibits several characteristics that create customer lock-in:

- Integration costs: Enterprise systems require months of implementation and integration work
- Training and workflow: Organizations build processes around specific platforms
- Data lock-in: Accumulated data becomes difficult to migrate
- Certification and compliance: Security and regulatory frameworks are built around specific solutions
- Contract structure: Multi-year agreements with penalties for early termination

AI application-layer competition exhibits all these characteristics with additional switching costs unique to AI systems: model-specific prompt engineering, evaluation frameworks tuned to particular model behaviors, fine-tuning datasets optimized for specific models, and continuous integration where AI capabilities are embedded throughout organizational workflows rather than isolated in specific applications.

The enterprise software literature demonstrates that even substantial quality or price advantages may not induce switching when integration costs are high (Greenstein 1993). Our analysis builds on these insights by examining how temporary cost advantages during the formation period can establish lock-in that persists even when the cost advantages disappear.

2.4 Accounting Policy and Real Effects on Competition

A substantial literature documents "real effects" of accounting policy—ways that financial reporting conventions affect actual business decisions rather than merely their representation (Kanodia and Sapra 2016; Leuz and Wysocki 2016). Depreciation policy affects investment decisions by changing reported profitability and thus capital costs (Bushman and Smith 2001). Firms facing more favorable accounting treatment can raise capital at lower cost and appear more attractive to investors, affecting competitive dynamics (Ball, Robin, and Wu 2003).

However, existing literature has not examined how accounting policy affects competitive dynamics in nascent markets where capital intensity is extreme, technological change is rapid, and competition concentrates at the application layer rather than at the production layer. Previous work on accounting policy and competition has focused primarily on how reporting affects investment decisions or capital costs within established market structures (Graham, Harvey, and Rajgopal 2005).

Our analysis demonstrates that depreciation conventions can shape market structure itself by affecting which firms can pursue aggressive application-layer strategies during critical formation periods while maintaining the appearance of financial viability. The capital subsidy matters not because it affects long-run cost structures, but because it enables strategic positioning during the window when application-layer customer relationships are being established.

2.5 Vendor Financing and Market Stability

Vendor financing—when suppliers provide funding to enable customers to purchase their products—has received attention primarily in contexts of market instability. The telecommunications equipment market in the late 1990s demonstrated risks when vendors financed their own sales at scale (Partnoy 2003). Equipment makers extended billions in financing to telecommunications companies, enabling purchases that would not have occurred at arm's length. When customers failed, vendor financing became bad debt and equipment sales that had appeared to represent genuine market demand instead reflected artificial demand created by circular financing.

Petersen and Rajan (1997) examine vendor financing more generally, identifying conditions under which such arrangements serve legitimate business purposes versus when they signal financial distress or artificial demand creation. The key insight is that vendor financing becomes problematic when it creates circular flows where supplier investments return as equipment purchases, making it difficult to distinguish genuine market demand from artificially sustained activity.

We build on this literature by analyzing how vendor financing interacts with short asset life to create different competitive dynamics than in previous technology booms. When assets have three-year useful lives rather than decades, excess capacity from failed investments does not provide a competitive foundation for new entrants. This changes the policy implications of circular financing in important ways.

3. The Asset Life Mismatch: Evidence and Magnitude

3.1 Technical Evidence on Useful Life

Multiple independent sources converge on a one-to-three-year useful lifespan for AI infrastructure chips. Technical evidence comes from three sources: engineering assessments, observed replacement patterns, and technological obsolescence rates.

Physical Degradation: A senior Google architect, speaking on condition of anonymity to industry press, assessed that graphics processing units (GPUs) running at 60-70% utilization—standard for AI training and inference workloads—survive one to two years of operation, with three years representing a maximum useful life before physical failure (SemiAnalysis 2024). The limiting factors are thermal stress from continuous high-power operation and electrical degradation of components under sustained load. AI workloads stress hardware more intensively than traditional computing applications, running GPUs near maximum capacity continuously rather than in burst patterns typical of other workloads.

Technological Obsolescence: Physical failure represents only one dimension of asset life. Technological advancement drives replacement cycles independently of physical degradation. Nvidia's GB200 ("Blackwell") architecture, introduced in 2024, provides 4-5x faster inference performance than the H100 architecture introduced just two years earlier (Nvidia 2024). When competitors deploy hardware with substantially superior performance characteristics, older chips become economically obsolete even if physically functional.

The economics of technological obsolescence are straightforward. Consider inference workloads where cost per token determines competitiveness. Running inference on three-year-old H100 chips costs approximately 5x more per token than on current-generation GB200 chips due to differences in computational efficiency, power consumption, and throughput. For price-sensitive applications, older hardware becomes economically uncompetitive regardless of physical condition.

Improvement Trajectory: The pace of improvement shows no signs of slowing. Each generation of AI-specific chips has delivered 2-4x improvements in performance per watt, with similar gains in performance per dollar (Khan et al. 2024). This improvement trajectory makes multi-year depreciation assumptions economically questionable. A chip depreciated over six years reaches only one-third of its nominal life before becoming economically obsolete due to technological progress.

3.2 Accounting Treatment in Practice

Current accounting practice for AI infrastructure diverges sharply from technical evidence. Major technology companies report depreciation periods of five to six years for computing equipment in their 10-K filings (Microsoft 2024; Amazon 2024; Meta 2024). These periods reflect general IT asset depreciation conventions developed for traditional enterprise computing rather than the specific characteristics of AI workloads and specialized hardware.

Representative Examples from 10-K Filings:

- **Microsoft:** Computer equipment depreciated over 4-6 years
- **Amazon:** Computer and network equipment depreciated over 5 years
- **Meta:** Network equipment and servers depreciated over 5 years
- **Google:** Computer equipment depreciated over 4 years

These depreciation periods were established for general-purpose servers and networking equipment with different operational characteristics than AI-specific hardware. Traditional servers operate at moderate utilization rates with mixed workloads. AI chips run at maximum capacity continuously on thermally intensive workloads. The physical and economic characteristics differ fundamentally, yet accounting treatment remains unchanged.

3.3 Magnitude of the Capital Subsidy

We can estimate the annual magnitude of the capital subsidy through straightforward calculation. McKinsey analysis indicates that approximately 60% of AI infrastructure spending addresses computing hardware (chips, servers, memory), 25% covers power and cooling infrastructure, and 15% funds physical construction (McKinsey & Company 2024). For analytical clarity, assume 50% of total infrastructure spending addresses computing hardware with true three-year useful life, while remaining spending covers longer-lived assets.

Calculation for Representative Coalition Member:

Consider a firm investing $100 billion annually in AI infrastructure:

- Computing hardware spending: $50 billion
- True economic depreciation (3-year life): $50B ÷ 3 = $16.7 billion per year
- Accounting depreciation (6-year life): $50B ÷ 6 = $8.3 billion per year
- **Annual subsidy: $8.3 billion**

This represents the difference between what the firm reports as depreciation expense and what it actually faces in replacement costs to maintain equivalent computational capacity. The $8.3 billion annual subsidy accumulates in reported earnings, making the business appear substantially more profitable than underlying economics warrant.

Aggregate Industry Impact:

Scaled across major AI infrastructure investors spending approximately $300 billion in 2025:

- Annual aggregate subsidy: $25 billion
- Three-year cumulative subsidy: $75 billion

Over the critical market formation period (years 1-3), the cumulative capital subsidy approaches $75 billion across the industry.

3.4 Financial Analyst Recognition

Investment analysts have begun recognizing this mismatch. Barclays equity research published revised earnings forecasts for major AI infrastructure investors, cutting 2025-2027 estimates by up to 10% to account for more realistic depreciation assumptions (Barclays Capital 2024). The adjustments reflect concern that current profitability metrics overstate sustainable economics.

Other analysts have published similar adjustments. Goldman Sachs equity research noted in July 2024 that "AI infrastructure depreciation assumptions may prove optimistic given the pace of technological change" and adjusted discounted cash flow models accordingly (Goldman Sachs 2024). These analytical adjustments suggest sophisticated market participants recognize the asset life mismatch, even though companies continue using extended depreciation periods in actual reporting.

However, these analytical adjustments have not yet translated into changes in actual accounting practice, regulatory scrutiny, or public discussion of competitive implications. The capital subsidy continues to enable strategies that appear sustainable in reported financials but may not be sustainable at true replacement costs.

4. The Revenue Sustainability Gap

4.1 The Bain Analysis: An $800 Billion Annual Shortfall

In September 2025, Bain & Company published an analysis estimating that AI firms will face an annual revenue shortfall of approximately $800 billion in 2030 to fund their capital expenses (Bain & Company 2025). What Bain calls an "AI revenue gap" emerges from comparing projected capital expenditures to the revenue that current and projected AI applications can generate.

Bain's methodology examines:

- Projected AI infrastructure capital expenditures in 2030: ~$250-300 billion annually
- Required return on invested capital: typical tech industry returns of 15-20%
- Current AI application revenue projections: insufficient to fund capital costs at required returns

The gap of approximately $800 billion annually represents the difference between what AI applications would need to generate to justify current investment levels and what Bain projects they will actually generate based on current adoption trajectories and pricing.

This analysis suggests the AI infrastructure buildout faces a fundamental economic sustainability problem even before considering whether current accounting practices accurately reflect replacement costs.

4.2 Incorporating Realistic Depreciation: A $1.5 Trillion Gap

Bain's analysis uses companies' reported capital expenditures and depreciation as inputs. If depreciation periods systematically understate true replacement costs, then the revenue gap is actually larger than Bain estimates.

Adjusting the Bain Calculation:

If we recalculate the revenue requirements incorporating realistic three-year useful life for computing hardware rather than reported five-to-six-year depreciation:

- Bain's estimated annual CapEx in 2030: $250-300 billion
- Adjustment for realistic depreciation on computing hardware (50% of spending):
    - Additional replacement cost: ~$40-50 billion annually
- True annual capital requirements: $290-350 billion
- Required revenue at 15-20% ROIC: $1,740-2,333 billion annually
- Projected AI application revenue (Bain estimate): ~$200-300 billion
- Adjusted revenue gap: $1.4-2.1 trillion annually

Using the midpoint, the revenue gap expands from approximately $800 billion (Bain's estimate) to over $1.5 trillion when realistic asset life is incorporated.

4.3 Implications for Market Structure Formation

The expanded revenue gap has critical implications for understanding competitive dynamics:

First**,** the industry faces a sustainability challenge independent of the capital subsidy. Even with favorable accounting treatment, AI infrastructure economics appear problematic. This suggests the current buildout may not be sustainable at any depreciation assumption.

Second**,** the capital subsidy makes the sustainability problem worse by understating true costs. Companies and investors making decisions based on reported profitability are systematically underestimating the capital requirements for sustaining operations.

Third**,** and most importantly for competition analysis, the combination of unsustainable economics and accounting subsidy creates a distinctive strategic dynamic. Incumbent coalitions can establish application-layer dominance during the market formation window even though the underlying economics don't work. By the time economic reality becomes visible, customer relationships are locked in.

Consider the strategic calculation for an incumbent coalition:

Option A: Acknowledge that economics don't work at realistic depreciation, scale back investment, lose application-layer market formation race.

Option B: Deploy capital aggressively using capital subsidy to appear viable, establish application-layer dominance, bet that lock-in will create value even if infrastructure economics prove unsustainable.

Option B dominates strategically because application-layer lock-in has value independent of infrastructure profitability. Microsoft benefits from Azure ecosystem expansion even if OpenAI's infrastructure economics ultimately don't work. Amazon benefits from AWS integration even if Anthropic's costs exceed revenues. The application-layer value persists even if model-layer or infrastructure-layer economics fail.

4.4 The "Extend and Pretend" Dynamic

The combination of capital subsidy and revenue gap creates what we term an "extend and pretend" dynamic:

Extend: Use favorable depreciation to extend the apparent viability of infrastructure investment beyond what realistic accounting would support.

Pretend: Maintain the appearance of sustainable profitability while building application-layer lock-in, betting that customer relationships will prove valuable even when infrastructure economics are revealed to be unsustainable.

This is not necessarily intentional deception. Companies may genuinely believe that revenue will eventually catch up to capital costs, or that technological improvements will resolve the economics. But the effect is the same: market structure forms based on economics that may prove unsustainable once true replacement costs become visible.

The competition policy concern is that by the time economic reality becomes undeniable, application-layer lock-in will make competitive correction impossible.

5. Coalition Structure and Subsidy Flow to Application Layer

5.1 Market Organization Around Vertically-Integrated Coalitions

The AI market has organized around vertically-integrated coalitions rather than a competitive marketplace of independent firms. This structure reflects the complementary investments required across infrastructure and model development layers, but it also creates pathways for subsidies to flow from infrastructure accounting to application-layer competition.

The Major Coalitions:

Microsoft-OpenAI Coalition:

- Microsoft invested ~$13 billion in OpenAI while providing exclusive Azure infrastructure access
- Microsoft sells “Azure OpenAI Service” with deep enterprise integration
- Microsoft benefits economically from OpenAI's success through Azure consumption and ecosystem expansion
- OpenAI's application-layer API sales drive Azure adoption

Amazon-Anthropic Coalition:

- Amazon invested ~$4 billion in Anthropic with AWS Bedrock integration
- Anthropic runs primarily on AWS infrastructure

- AWS benefits from Anthropic's customer acquisitions through infrastructure consumption
- Application-layer API sales through Bedrock drive AWS ecosystem growth

Google Integrated Coalition:

- Google owns both infrastructure (Google Cloud) and model development (DeepMind, Gemini)
- Vertical integration allows direct coordination of infrastructure and model strategies
- Application-layer integration across Google Workspace and Cloud services
- More integrated structure than partnerships but same economic logic

Meta:

- Owns infrastructure and develops models (Llama series) but focuses on internal applications
- Uses AI to enhance core advertising and social networking businesses
- Releases open-weight models to shape ecosystem without competing directly for enterprise API revenue
- Application-layer value captured through improved engagement and ad targeting rather than API sales

These coalitions reflect deeper integration than traditional supplier-customer relationships. Infrastructure economics directly enables application-layer competitive positioning through favorable pricing, capacity prioritization, joint product development, and coordinated go-to-market strategies.

5.2 How the Subsidy Flows: From Infrastructure Accounting to Application-Layer Pricing

The capital subsidy enables a multi-step flow of competitive advantage from infrastructure accounting to application-layer market positioning:

Step 1: Infrastructure Layer Subsidy Creation

Hyperscalers (Microsoft, Amazon, Google) report computing equipment depreciation over 5-6 years while facing 3-year replacement cycles. This creates apparent profitability that:

- Improves reported earnings and attracts capital at favorable terms

- Makes infrastructure investments appear sustainable
- Creates “room” for aggressive pricing while showing profits

Step 2: Favorable Transfer Pricing to Model Developers

The infrastructure subsidy enables hyperscalers to offer model developer partners (OpenAI, Anthropic) infrastructure access on terms more favorable than standalone cloud providers could match:

- Below-market infrastructure pricing for coalition partners
- Capacity prioritization during shortage periods
- Joint development of optimized infrastructure-model integration
- Revenue sharing arrangements that benefit from ecosystem expansion

Step 3: Aggressive Application-Layer Pricing

Model developers with access to subsidized infrastructure can price application-layer APIs more aggressively than would be sustainable at true market rates:

- Lower per-token pricing to win enterprise customers
- Aggressive customer acquisition spending
- Capacity expansion that appears financially viable
- Multi-year contracts at prices that lock in customers

Step 4: Application-Layer Lock-In

Enterprise customers integrating OpenAI or Anthropic models become embedded in the coalition's technology stack:

- API integration requires months of engineering work
- Prompt engineering and evaluation frameworks are model-specific
- Compliance and security certifications are built around specific stacks
- Multi-year contracts create financial switching costs
- Organizational workflows and training are developed around particular providers

Step 5: Ecosystem Value Capture

Once application-layer customers are locked in, hyperscalers capture value through:

- Increased Azure/AWS consumption as customers expand AI usage

- Adoption of adjacent cloud services (storage, databases, security)
- Platform fees and integration services
- Data lock-in as customer information accumulates in coalition systems

The subsidy thus flows from infrastructure accounting through favorable coalition pricing to aggressive application-layer competition, ultimately establishing customer relationships that create value independent of whether infrastructure economics are sustainable.

5.3 The Microsoft-OpenAI Example in Detail

Consider Microsoft's strategic calculation. Microsoft's Azure AI infrastructure investments, when depreciated over six years rather than realistic three years, create apparent profitability of approximately $6-7 billion annually (using our earlier calculation scaled to Microsoft's investment level).

This apparent profitability enables several strategic moves:

Favorable OpenAI Infrastructure Pricing: Microsoft can offer OpenAI infrastructure access at below-market rates while maintaining the appearance of profitable Azure operations. The difference between true replacement costs ($13.3B annually) and reported depreciation ($6.7B annually) creates room for aggressive pricing to OpenAI without showing losses in Azure's reported results.

OpenA''s Application-Layer Pricing Flexibility: With access to below-market infrastructure, OpenAI can price APIs more aggressively than competitors paying market rates for infrastructure. OpenAI's pricing appears sustainable in OpenAI's economics (infrastructure is artificially cheap from OpenAI's perspective) even though it wouldn't be sustainable if infrastructure were priced at true replacement cost.

Enterprise Customer Acquisition: OpenAI's aggressive API pricing wins enterprise customers who integrate OpenAI models deeply into their operations. The initial pricing advantage (enabled by the subsidy) creates customer relationships that persist even after pricing adjusts to market levels, because switching costs prevent customers from migrating once integrated.

Azure Ecosystem Expansion: Every enterprise customer integrating OpenAI's models through Azure OpenAI Service becomes more deeply embedded in the Azure ecosystem. These customers expand their use of Azure storage, Azure databases, Azure security services, Azure networking, and other Azure offerings. Microsoft captures this expansion value independent of whether OpenAI's infrastructure economics are sustainable.

The Strategic Payoff: Microsoft's investment creates value through application-layer lock-in even if OpenAI's infrastructure economics ultimately don't work at true replacement costs. The application-layer customer relationships, ecosystem expansion, and platform value persist independent of infrastructure-layer profitability.

This is not a story about whether Microsoft or OpenAI are individually profitable at realistic depreciation. It's a story about how the capital subsidy enables strategies to establish application-layer dominance during the critical formation window, with payoffs that persist even if infrastructure economics prove unsustainable.

5.4 Why Application-Layer Competition Matters Most

A critical insight is that application-layer competition determines the market structure that matters for enterprises and for innovation. As model development commoditizes, competitive differentiation increasingly comes from:

- Ease of integration and developer experience
- Enterprise compliance and security capabilities
- Reliability and service level agreements
- Ecosystem breadth (adjacent services and integrations)
- Organizational trust and brand reputation

These are application-layer characteristics, not model-layer capabilities. Two models with comparable accuracy become differentiated by which coalition's application-layer integration is more attractive to enterprises.

The capital subsidy affects application-layer competition because it enables coalitions to establish customer relationships at prices that wouldn't be sustainable at true infrastructure costs. Once those relationships are established through deep integration, switching costs prevent customers from migrating even when:

- True infrastructure costs become visible
- Better models become available
- More efficient infrastructure emerges
- The revenue gap becomes undeniable

The lock-in is at the application layer, where enterprises have made complementary investments in integration, training, compliance, and workflow design. This lock-in persists independent of model-layer or infrastructure-layer economics.

6. Competitive Scenarios and Application-Layer Lock-In

The durability of competitive advantages created by the capital subsidy depends on technological trajectories and application-layer switching costs. We analyze three scenarios that span the range of possible outcomes, focusing on implications for application-layer competition.

6.1 Scenario A: Training Keeps Scaling (Model Differentiation Persists)

Technological Assumption: Frontier model capabilities continue to improve meaningfully with increased training compute. Scaling laws remain favorable. Model quality continues to differentiate providers at the application layer.

Market Dynamics: In this scenario, training capacity matters more than inference efficiency. Model quality differentiates providers. Enterprise customers pay premium prices for frontier capabilities. Investment in massive training clusters creates lasting competitive advantages.

Application-Layer Implications: Incumbent coalitions that built large-scale training capacity early possess significant advantages. Enterprise customers at the application layer choose providers based primarily on model quality. Integration depth reinforces advantages—customers become embedded with the coalition offering the best models.

Subsidy Effects: When accounting catches up to operational reality (years 4-5), incumbent coalitions face profitability pressure but retain competitive advantages through superior model quality. New entrants with better unit economics struggle because the competitive game centers on model quality rather than cost. Capital requirements are enormous and ongoing. Whoever built scale first maintains advantages through continued investment in training capacity. Application-layer customers stay with their chosen coalition because switching would mean moving to inferior models.

Assessment: This scenario appears increasingly implausible. Current evidence suggests model capabilities are plateauing or hitting diminishing returns to pure scale (Marcus and Davis 2024). The gap between frontier models and open-weight alternatives has narrowed substantially. Model quality differentiation appears to be converging rather than expanding. Industry discussion has shifted from training scale to inference efficiency. Major model releases show incremental rather than transformative improvements.

If this scenario somehow materializes, competition concerns focus on whether private capital can sustain the required investment indefinitely, or whether the industry requires policy support once true costs become visible. However, the application-layer competition question becomes less important because model quality determines outcomes regardless of integration depth. We proceed by analyzing the two more plausible scenarios where model capabilities plateau.

6.2 Scenario B: Models Plateau, Application-Layer Lock-In Holds (Foreclosure)

Technological Assumption: Model capabilities hit diminishing returns. Most models become "good enough" for enterprise applications. Inference efficiency and cost become important but not determinative. Open-source models match or approach proprietary model quality. Application-layer integration depth and switching costs determine competitive outcomes.

Market Dynamics: Competition shifts entirely to the application layer. Model development commoditizes—multiple models offer comparable capabilities. Competition centers on integration quality, developer experience, enterprise security and compliance, ecosystem breadth, and reliability.

However, incumbent coalitions that established customer relationships during the formation window retain advantages through mechanisms unrelated to model quality:

- Deep enterprise integrations built during the buildup period when coalitions had pricing power
- Multi-year contracts signed when model quality appeared differentiated
- Compliance certifications and security frameworks built around specific coalition technology stacks
- Organizational inertia: employees trained on specific platforms, workflows designed around particular APIs
- Brand trust and reliability reputation built during market formation
- Data accumulation: customer data stored in coalition ecosystems, creating additional switching costs

- Adjacent service adoption: enterprises use multiple services from coalition members, increasing switching costs further

Application-Layer Lock-In Dynamics: This scenario creates the most concerning competitive dynamics. The capital subsidy enabled incumbent coalitions to establish application-layer market positions during years 1-3. By years 4-5, when true costs become visible and model capabilities have commoditized, switching costs prevent customers from moving even when:

- New entrants offer better technology at lower cost
- The revenue gap makes incumbent economics unsustainable
- Better alternatives clearly exist

The Google Search Parallel: The analogy to Google search illustrates the competitive concerns in this scenario. Google maintained search dominance through default placements and integration depth. The DOJ's case demonstrated these mechanisms created durable advantages even when:

- Switching appeared trivially easy ("one click away")
- Alternative search engines offered comparable quality
- Users expressed no strong preference for Google in blind tests

Default positions combined with interface integration and organizational habit proved sufficient to maintain dominance despite the apparent ease of switching.

Application-Layer Switching Costs Are Higher: Switching costs in AI application markets are substantially higher than in consumer search:

Consumer Search	Enterprise AI Application Layer
One-click switching	Months of re-integration work
No contracts	Multi-year agreements with penalties

No organizational workflow	Workflows built around specific APIs
No compliance requirements	Security certifications around specific stacks
No training investment	Employee training on specific platforms
No data lock-in	Accumulated data in provider systems
No adjacent services	Multiple coalition services create compound lock-in

If Google could maintain dominance with low switching costs, application-layer lock-in in AI may prove far more durable when enterprises are deeply integrated into coalition technology stacks.

Why Better Technology Doesn't Enable Entry: In Scenario B, new entrants may have:

- More efficient infrastructure (newer chips, better designs)
- Lower unit costs
- Comparable or superior model quality
- Better pricing

Yet they cannot compete effectively because application-layer switching costs exceed the benefits of migration. An enterprise considering switching must weigh:

Benefits: Lower per-token costs, possibly better performance, escape from coalition dependency

Costs: Months of re-integration engineering, disruption to organizational workflows, multi-year contract penalties, re-certification for compliance, retraining employees, migrating accumulated data, losing integration with adjacent services

For most enterprises, the costs exceed benefits even when the alternative is clearly superior on technical or economic merits.

Technology Deflation Reinforces Rather Than Undermines Lock-In

An intuitive counterargument holds that rapid technological improvement should undermine incumbent advantages. Each generation of chips delivers better performance at lower cost, which should enable new entrants to compete on superior economics.

However, technology deflation affects all market participants. The question is whether it helps entrants more than incumbents. When application-layer lock-in exists, technology improvement paradoxically reinforces incumbent advantages:

Incumbents with locked-in customers:

- Upgrade infrastructure through normal replacement cycles
- Existing customers automatically benefit from improved capabilities and lower costs
- No customer acquisition required
- Lock-in strengthens as customers accumulate more data and deeper integration over time

New entrants with better technology:

- Have lower unit costs but must still acquire customers
- Face application-layer switching costs that exceed economic benefits of migration
- Must overcome integration depth even when offering clearly superior technology
- Technology advantage is neutralized by switching costs

The critical insight from the Google search case: technology deflation that should enable new entry actually reinforces incumbent advantages when lock-in is strong. Google's competitors could access the same technological improvements (better algorithms, faster servers, cheaper bandwidth), but default positions and integration depth meant technology improvements reinforced Google's position rather than enabling entry.

Assessment: Scenario B represents the most concerning competitive outcome. The capital subsidy enables incumbent coalitions to establish application-layer dominance during the critical formation window. When model capabilities commoditize and true costs become visible, switching costs prevent competitive correction.

Current evidence suggests this scenario is plausible:

- Model capabilities appear to be plateauing
- Open-weight models (Meta's Llama, Mistral, etc.) are approaching proprietary model quality
- Competition is shifting to inference efficiency and integration quality
- Enterprise integration depth is substantial and growing

If Scenario B materializes, the capital subsidy will have facilitated application-layer foreclosure—incumbent coalitions establish dominance during the formation window using subsidized economics, and lock-in prevents correction even when better alternatives emerge.

6.3 Scenario C: Models Plateau, Low Switching Costs (Competitive Entry Remains Possible)

Technological Assumption: Model capabilities plateau and commoditize. Importantly, application-layer switching costs prove lower than expected or are reduced through standardization and interoperability.

Market Dynamics: APIs standardize across providers. Models reach comparable capability levels. Enterprises successfully demand interoperability. Integration depth matters less than anticipated because portability improves. Price and service quality become primary differentiators.

How Switching Costs Could Be Lower Than Expected: Several developments could reduce application-layer switching costs:

API Standardization: Industry converges on compatible API standards, allowing code portability. This is not implausible—the "OpenAI-compatible API" has become a de facto standard that many providers support.

Abstraction Layers: Tools emerge that abstract away provider-specific details, allowing enterprises to switch backends without changing application code. Companies like LangChain and LlamaIndex provide some of this functionality.

Contractual Flexibility: As competition intensifies, providers may offer more flexible contract terms with lower switching penalties to attract customers.

Enterprise Bargaining Power: Large enterprises with substantial purchasing power may demand interoperability and low switching costs as conditions for adoption.

Regulatory Intervention: Competition authorities or sectoral regulators could mandate interoperability standards, directly reducing switching costs.

Subsidy Effects: In this scenario, the capital subsidy enables wasteful overinvestment that temporarily distorts competition but does not create permanent application-layer foreclosure. Early investments by incumbents look wasteful in retrospect as enterprises can easily switch to better alternatives.

New entrants can compete effectively on price and service quality, leveraging newer, more efficient infrastructure. The application-layer market remains competitive despite the formation-period subsidy because switching costs are manageable.

Assessment: This is the optimistic scenario from a competition perspective. The capital subsidy creates inefficiency and shapes market structure temporarily, but competitive harm is not permanent because application-layer switching remains feasible.

However, current evidence suggests this scenario is less likely than Scenario B:

- Enterprise AI integration is deep and growing deeper
- Contract structures include substantial penalties for switching
- No strong industry movement toward standardization
- Major providers have incentives to maximize lock-in rather than embrace portability
- Regulatory intervention on interoperability has not materialized

The burden of proof should be on those claiming switching costs will remain low. Historical evidence from enterprise software markets suggests deep integration creates persistent lock-in (Greenstein 1993). The Google search case demonstrates that even apparently trivial switching costs can prove decisive.

6.4 Which Scenario Is Materializing?

Current evidence increasingly suggests we are heading toward Scenario B rather than Scenario A or C:

Evidence Against Scenario A (Continued Training Scaling):

- Model capability improvements have slowed
- Diminishing returns to pure scale are visible
- Open-weight models are closing the gap with proprietary models
- Industry discussion has shifted from training scale to inference efficiency

Evidence For Scenario B (Commoditization with Lock-In):

- Model capabilities appear to be plateauing
- Competition is shifting to application-layer differentiation
- Enterprise integration depth is substantial and growing
- Multi-year contracts with substantial penalties are common
- No meaningful interoperability standards are emerging
- Switching costs in enterprise AI are demonstrably high

Evidence Against Scenario C (Low Switching Costs):

- Deep enterprise integrations require months of engineering work
- Compliance and security certifications are provider-specific
- Multi-year contracts with penalties are standard
- No strong industry movement toward standardization
- Major providers resist portability
- Regulatory intervention has not occurred

The most likely trajectory is that model development commoditizes while application-layer lock-in proves durable, creating persistent competitive advantages for coalitions that established customer relationships during the formation window—precisely the outcome the capital subsidy enabled.

7. Vendor Financing Amplifies Scenario B's Competitive Harm

The competitive dynamics of Scenario B—where model capabilities plateau but application-layer lock-in proves durable—become even more concerning when we incorporate circular vendor financing into the analysis. Vendor financing amplifies the capital subsidy's effects during the formation window, but more importantly, the short useful life of AI infrastructure means that even market corrections cannot enable competitive entry.

7.1 The Circular Financing Structure

Major chip manufacturers, particularly Nvidia, have adopted strategies of making equity investments in companies that purchase their products, creating circular capital flows that amplify the market dynamics enabled by the capital subsidy.

The CoreWeave-OpenAI-Nvidia Circle: The structure is most visible in relationships among Nvidia, CoreWeave, and OpenAI:

- Nvidia owns approximately 7% of CoreWeave, a position worth ~$3 billion as of June 2025 (Bloomberg 2025)
- CoreWeave has purchased at least 250,000 Nvidia GPUs, primarily H100s at ~$30,000 each, totaling ~$7.5 billion in hardware purchases (The Information 2025)
- CoreWeave signed $22.4 billion in infrastructure contracts with OpenAI (Reuters 2025)
- Nvidia participated in OpenAI's $6.6 billion funding round in October 2024 (Wall Street Journal 2024)
- Nvidia announced a $100 billion investment commitment in September 2025 (Financial Times 2025)

The capital flows in a loop: Nvidia invests in OpenAI → OpenAI commits to CoreWeave contracts → CoreWeave purchases Nvidia GPUs → Nvidia holds equity in CoreWeave. Each transaction appears as legitimate revenue, investment, or contract commitment depending on which company's financial statements are examined.

Similar structures appear to underlie AMD’s recently announced multibillion-dollar deal with OpenAI, though details have not been fully disclosed (Reuters 2025).

7.2 Learning from Telecom: The 2001 Vendor Financing Bust

Circular vendor financing is not new. The telecommunications equipment market in the late 1990s demonstrated both the attraction and the risks of this structure (Partnoy 2003).

The Telecom Pattern: During the late 1990s telecom boom, equipment makers provided substantial financing to enable customers to purchase their products:

- Lucent committed $8.1 billion in vendor financing—approximately 24% of its annual revenue (SEC Filings, Lucent Technologies 2001)

- Nortel extended $3.1 billion in vendor financing (SEC Filings, Nortel Networks 2001)
- Cisco promised $2.4 billion in vendor financing (SEC Filings, Cisco Systems 2001)

The strategy worked until it didn't:

- Equipment vendors lent money to cash-strapped telecom companies
- Telecom companies used borrowed funds to purchase equipment
- Vendors booked revenue from equipment sales
- Stock prices rose based on revenue growth
- Cycle repeated until customers couldn't sustain operations

When the bubble burst:

- 47 competitive local exchange carriers went bankrupt between 2000 and 2003 (FCC Data 2004)
- Vendor financing became bad debt
- Lucent wrote off $3.5 billion in customer loans (SEC Filings, Lucent Technologies 2002)
- Equipment sales that appeared to represent genuine demand actually reflected artificial demand created by circular financing

Today's Numbers Are Larger: Nvidia's disclosed investments and financing commitments total approximately $110 billion against $165 billion in trailing-twelve-month revenue—representing 67% of revenue compared to Lucent's 24% at the telecom peak (Company Financial Statements 2025).

The scale of circular financing in AI infrastructure substantially exceeds the telecom boom in both absolute and relative terms.

7.3 Critical Difference: Asset Life Eliminates the "Second Chance" for Competition

Two critical differences distinguish the AI infrastructure buildout from the telecom boom, with fundamentally different implications for competition policy.

First: There are no allegations or findings of fraud in current AI vendor financing, whereas several telecom cases involved fraudulent accounting. The circular flows in AI infrastructure appear to represent legitimate commercial relationships rather than attempts to manufacture artificial revenue. Nvidia's investments in customers may be genuine strategic bets rather than disguised subsidy mechanisms.

Second and More Important: Asset lifespan fundamentally changes the competitive implications when buildouts prove excessive. This is the crucial insight for understanding why Scenario B is even more concerning than the Google search parallel initially suggests.

In Telecommunications (1990s-2000s):

- Fiber optic cables had useful lives measured in decades
- Switching equipment remained functional for 10-20 years
- When overbuilt telecom companies went bankrupt, new entrants could acquire infrastructure at fire-sale prices
- Long asset life meant excess capacity created during the boom could be redeployed by new entrants during the correction
- Competitive entry was actually enabled by the bust—new entrants got infrastructure cheaply and competed effectively

In AI Infrastructure (2020s):

- AI chips have three-year useful lives before technological obsolescence
- By the time financial pressure emerges (years 3-5), technology has progressed 2-3 generations
- Three-year-old chips purchased at fire-sale prices cannot compete when incumbents deploy current-generation hardware with 4-5x better performance
- Excess capacity from failed investments does not provide a competitive foundation for new entrants
- Short asset life means the bust does not enable competitive entry

Why This Matters for Application-Layer Competition: In telecom, the vendor financing bust actually improved competitive conditions. New entrants acquired infrastructure cheaply and competed effectively with incumbents. The bust was economically wasteful but competitively beneficial.

In AI, even if circular financing unwinds and creates financial distress, this will not enable application-layer competitive entry for two reasons:

First: Failed infrastructure has no competitive value due to short useful life. Three-year-old chips are economically obsolete when incumbents run current-generation hardware. A new entrant acquiring stranded H100 chips in 2027 cannot compete effectively when incumbent coalitions are running GB200 or next-generation hardware with 4-5x better performance characteristics.

Second: Application-layer lock-in persists regardless of infrastructure-layer distress. Incumbent coalitions with locked-in customers will upgrade to current hardware through normal replacement cycles. Their application-layer advantages continue even if some infrastructure investments fail. Enterprise customers remain integrated with Microsoft-OpenAI or Amazon-Anthropic regardless of whether the underlying infrastructure investments prove economically sustainable.

The short asset life means vendor financing unwinding does not create the "second chance" for competitive entry that occurred in telecommunications. This fundamentally changes the policy calculus and makes Scenario B's competitive harm more severe and more permanent than historical analogies suggest.

7.4 How Circular Financing Amplifies the Capital Subsidy

Circular vendor financing interacts with the capital subsidy to amplify competitive distortions during the formation window:

First: Obscuring true sustainability. Circular financing makes it difficult to assess the true financial sustainability of the buildout. When the same dollars flow through multiple entities as investment, revenue, and contracts, traditional financial analysis becomes challenging. This obscures the revenue gap and makes it harder for markets or regulators to identify problems early.

Second: Reinforcing coalition structures. The circular structure reinforces coalition dynamics and application-layer competition. Companies with equity stakes in multiple layers of the infrastructure stack have aligned incentives to maintain the coalition structure rather than compete independently. This entrenchment makes application-layer competition less likely to emerge even if underlying economics deteriorate.

Third: Fragility without competitive benefit. If financial pressure emerges, interconnectedness means distress could cascade through the circular structure quickly—but without creating competitive opportunities for application-layer entry due to short asset life. The system is fragile to shocks but failure does not improve competitive conditions.

Fourth: Accelerating the formation-window buildout. Circular financing enables larger-scale buildout than arm's-length transactions would support. Nvidia's investments in customers effectively reduce the capital customers need to raise independently. This amplifies the formation-period advantage—incumbent coalitions can deploy capital faster and establish application-layer positions more quickly than if financing were entirely independent.

The combination of vendor financing and capital subsidy thus creates a distinctive dynamic: rapid deployment of capital that establishes application-layer lock-in, based on circular financing that may prove unsustainable, with short asset life that prevents competitive correction even if the financing unwinds. Thus, Scenario B's competitive foreclosure may prove permanent even if the financial structure collapses.

8. Policy Proposals

8.1 Current State of Competition Review

Competition authorities in the United States and Europe are examining market concentration in AI, focusing on traditional antitrust metrics: market share, pricing power, exclusionary conduct, and merger effects. In 2024, the Federal Trade Commission launched inquiries into partnerships between hyperscalers and AI developers. The European Commission is examining potential foreclosure concerns under the Digital Markets Act (European Commission 2024). However, the competitive implications of accounting policy and vendor financing have not yet received extensive attention in competition proceedings.

8.2 Enhanced Disclosure Requirements for Infrastructure Projects

Our first policy proposal addresses information asymmetry. Any AI infrastructure project receiving government support—whether through direct funding, tax incentives, power grid access, or regulatory approval—should be required to disclose:

1. Realistic Useful Life Assumptions:

- Expected useful economic life for computing hardware based on technological obsolescence and physical degradation
- Sensitivity analysis showing how capital requirements change under different useful life assumptions

- Comparison of accounting depreciation to expected replacement schedules

2. Expected Replacement Costs:

- Forward-looking estimates of annual replacement capital requirements
- Analysis of how these costs change under different technology improvement trajectories
- Total capital requirements to maintain operations over 5-10 year horizons

3. Revenue Sustainability Analysis:

- Projected revenue from AI applications needed to fund capital requirements
- Analysis of revenue gap between current projections and capital costs
- Scenario analysis under different adoption and pricing assumptions

4. Partnership Economics:

- Disclosure of how infrastructure costs flow from chip manufacturers to hyperscalers to model developers
- Transfer pricing methodologies for coalition infrastructure access
- Capacity allocation mechanisms and contractual revenue sharing arrangements
- Terms of partnership agreements including exclusivity provisions

5. Vendor Financing Relationships:

- Equity positions held by equipment vendors in customers or partners
- Lending arrangements and financing commitments
- Contractual linkages between vendor investments and equipment procurement
- Total magnitude of circular financing as percentage of revenue

6. Application-Layer Integration and Lock-In:

- Analysis of customer switching costs
- Contract structures including length, penalties, and renewal terms
- Description of integration depth and technical dependencies
- Assessment of competitive implications at application layer

Enhanced disclosure serves multiple functions. It enables regulators to assess competitive effects at the application layer more accurately. It allows investors to make more informed capital allocation decisions based on realistic economics. It creates market pressure for more realistic accounting treatment through transparency. Most importantly, it makes visible the mechanisms by which application-layer competition is being shaped during the formation window.

8.3 Interoperability Standards to Reduce Application-Layer Switching Costs

Our second policy proposal directly addresses application-layer switching costs. Competition authorities should actively promote interoperability standards for:

1. Model API Standardization:

- Standardized interfaces allowing enterprises to switch between model providers without rewriting integration code
- Technical challenges are manageable—generative AI APIs have relatively simple surface areas amenable to standardization
- The “OpenAI-compatible API” has already emerged as a de facto standard; perhaps formalize and extend this
- Ensure standards cover not just inference but also fine-tuning, evaluation, and monitoring

2. Data Portability Requirements:

- Clear requirements for exporting training data, fine-tuning datasets, and application data
- Standardized formats to enable migration between providers
- Reduction of data lock-in as a switching cost
- Requirements for providers to offer export functionality

3. Cloud Integration Framework Interoperability:

- Standardized approaches to security, compliance, data residency, and operational procedures
- Reduce the cost of migrating security certifications between providers
- Enable compliance frameworks that work across coalition stacks
- Allow enterprises to maintain security posture while switching providers

The Google Search Remedy Analogy: The remedy in the Google search case focused substantially on reducing switching costs by limiting default placements and exclusive arrangements. The Court recognized that even apparently low switching costs (“one click”) could create durable competitive advantages. The remedy sought to reduce lock-in mechanisms that prevented users from accessing alternatives.

AI application-layer switching costs are substantially higher than consumer search—months of engineering work versus one click. This makes interoperability standards even more important in AI than default-placement restrictions were in search.

Addressing Industry Resistance: Industry participants will likely resist standardization, arguing that:

- It reduces innovation incentives by commoditizing interfaces
- Technical differences make standardization impractical
- Competitive differentiation requires proprietary approaches
- Standardization is premature in a rapidly evolving market

However, interoperability has proven achievable in previous technology markets (telecommunications, payment systems, internet protocols, email) without limiting innovation. The key is to standardize interfaces rather than implementations, preserving competitive differentiation while reducing switching costs.

The argument that standardization is premature is weakest precisely because application-layer lock-in is forming now. Waiting for the market to mature means waiting until lock-in is already established—at which point intervention becomes much more difficult, as the ongoing Google search remedy demonstrates.

9. Conclusion

This paper has documented a significant mismatch between the economic useful life of AI infrastructure chips (1-3 years) and their accounting depreciation periods (5-6 years), creating a "capital subsidy" worth tens of billions of dollars annually. This accounting treatment matters because it enables incumbent coalitions to pursue application-layer strategies during the critical market formation window that would appear unsustainable under realistic depreciation.

The competitive implications concentrate at the application layer rather than model development. As model capabilities plateau and model development commoditizes, competition shifts to application-layer integration, customer relationships, and switching costs. The capital subsidy enables hyperscaler-model developer coalitions to establish application-layer dominance through aggressive pricing and rapid capacity expansion that appears financially viable in reported earnings but may not be sustainable at true replacement costs.

The industry faces a fundamental economic challenge even before considering accounting distortions. Bain & Company's analysis estimates an $800 billion annual revenue shortfall in 2030 to fund capital expenses. Incorporating realistic depreciation expands this gap to over $1.5 trillion—nearly double the sustainability challenge. The capital subsidy enables incumbent coalitions to pursue strategies during the formation window despite economics that may prove unsustainable, betting that application-layer lock-in will create value regardless of infrastructure-layer profitability.

Three scenarios span possible competitive outcomes. Scenario A (continued training scaling) appears increasingly implausible as model capabilities plateau—we dismiss this scenario and focus on the two plausible futures where model development commoditizes. Scenario B (commoditization with durable application-layer lock-in) represents the most concerning outcome, where incumbent coalitions establish customer relationships during years 1-3 using subsidized economics, and by years 4-5, when true costs become visible, application-layer switching costs prevent competitive correction even when better alternatives emerge. Scenario C (commoditization with low switching costs) represents the optimistic outcome where switching costs prove manageable and competition remains possible.

Current evidence—deep enterprise integrations, multi-year contracts with penalties, absence of interoperability standards, and lessons from the Google search case—suggests Scenario B is more plausible than Scenario C. If Scenario B materializes, the capital subsidy will have enabled application-layer foreclosure that proves durable as in Google search, but with switching costs orders of magnitude higher than changing search engines.

Circular vendor financing amplifies Scenario B’s competitive harm in a distinctive way. Unlike telecommunications, where long-lived assets could be redeployed by new entrants after failures, three-year useful life means excess capacity will not enable competitive entry even if financial distress emerges. The combination of vendor financing, capital subsidy, and short replacement cycles creates a system that is fragile to shocks but where failure does not improve competitive conditions. Market corrections that might have enabled competitive entry in previous technology booms cannot do so in AI infrastructure—making Scenario B's foreclosure potentially permanent even if the financial structure collapses.

We propose two policy interventions addressable through available regulatory authority. Enhanced disclosure requirements would create transparency about true economics, partnership structures, and application-layer lock-in mechanisms for any project receiving government support. Interoperability standards would reduce switching costs and enable competition based on merit rather than integration depth at the application layer. These interventions could be implemented through existing regulatory frameworks without requiring new legislation.

The critical insight for competition policy is timing. Market structure at the application layer is being determined now during the formation window. The decisions being made—based on accounting assumptions that may overstate financial sustainability by a factor of two—are crystallizing the competitive landscape for artificial intelligence services. Once customer relationships are established and integration depth is achieved, competitive correction becomes substantially more difficult.

The Google search remedy—now in its implementation phase after years of litigation—demonstrates how difficult unwinding lock-in becomes after market structure crystallizes. Application-layer integration in enterprise AI creates switching costs far exceeding consumer search engine selection. If intervention waits until application-layer lock-in is complete, remedies become much more challenging and less likely to succeed.

Competition authorities face a window for intervention measured in years, not decades. The capital subsidy creates a distortion during market formation that may shape application-layer competitive outcomes permanently, independent of whether the underlying infrastructure investment economics prove sustainable at true replacement costs. The revenue gap suggests they will not. Circular vendor financing amplifies the formation-window advantage while eliminating the possibility of competitive correction through market failure. By the time economic reality becomes undeniable, the application-layer market structure may already be foreclosed.

References

Amazon. (2024). Form 10-K: Annual Report. SEC Filing.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. The Economic Journal, 99(394), 116-131.

Bain & Company. (2025). The AI Revenue Gap: Analyzing the Economics of Artificial Intelligence Infrastructure. Industry Report.

Ball, R., Robin, A., & Wu, J. S. (2003). Incentives versus standards: Properties of accounting income in four East Asian countries. Journal of Accounting and Economics, 36(1-3), 235-270.

Barclays Capital. (2024). Technology Hardware: AI Infrastructure Economics and Depreciation Analysis. Equity Research Report.

Bloomberg. (2025). Nvidia's CoreWeave Stake Valued at $3 Billion. News Report, June 2025.

Bushman, R. M., & Smith, A. J. (2001). Financial accounting information and corporate governance. Journal of Accounting and Economics, 32(1-3), 237-333.

Cisco Systems. (2001). Form 10-K: Annual Report. SEC Filing.

Crémer, J., de Montjoye, Y. A., & Schweitzer, H. (2019). Competition Policy for the Digital Era. European Commission Report.

European Commission. (2024). Digital Markets Act: Preliminary Findings on AI Partnerships. Press Release.

Farrell, J., & Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. Handbook of Industrial Organization, 3, 1967-2072.

FCC. (2004). Trends in Telephone Service. Federal Communications Commission Report.

Financial Times. (2025). Nvidia Announces $100bn AI Infrastructure Commitment. September 2025.

FTC. (2024). FTC Launches Inquiry into AI Partnerships Between Cloud Providers and AI Developers. Press Release.

Gans, J. (2022). The Fine Print in Smart Contracts. NBER Working Paper No. 29701.

Goldman Sachs. (2024). AI Infrastructure Investment: Risks and Opportunities. Equity Research Report, July 2024.

Goldman Sachs. (2025). Global Technology Hardware Spending Projections. Industry Analysis.

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. Journal of Accounting and Economics, 40(1-3), 3-73.

Greenstein, S. M. (1993). Did installed base give an incumbent any (measurable) advantages in federal computer procurement? RAND Journal of Economics, 24(3), 19-39.

Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. Strategic Management Journal, 39(8), 2255-2276.

Kanodia, C., & Sapra, H. (2016). A real effects perspective to accounting measurement and disclosure: Implications and insights for future research. Journal of Accounting Research, 54(2), 623-676.

Khan, A., et al. (2024). Efficiency Trends in AI Hardware. Nature Electronics, 7, 889-901.

Leuz, C., & Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. Journal of Accounting Research, 54(2), 525-622.

Lieberman, M. B., & Montgomery, D. B. (1988). First-mover advantages. Strategic Management Journal, 9(S1), 41-58.

Lucent Technologies. (2001). Form 10-K: Annual Report. SEC Filing.

Lucent Technologies. (2002). Form 10-K: Annual Report. SEC Filing (including write-off disclosures).

Marcus, G., & Davis, E. (2024). Has AI Hit a Wall? Substack Essay, November 2024.

McKinsey & Company. (2024). The State of AI Infrastructure Investment. Technology Sector Report.

Meta. (2024). Form 10-K: Annual Report. SEC Filing.

Microsoft. (2024). Form 10-K: Annual Report. SEC Filing.

Nortel Networks. (2001). Annual Information Form. Canadian SEC Filing.

Nvidia. (2024). GB200 Grace Blackwell Superchip Technical Specifications. Product Documentation.

Partnoy, F. (2003). Infectious Greed: How Deceit and Risk Corrupted the Financial Markets. New York: Times Books.

Petersen, M. A., & Rajan, R. G. (1997). Trade credit: Theories and evidence. The Review of Financial Studies, 10(3), 661-691.

Reuters. (2025). CoreWeave Signs $22.4 Billion Infrastructure Deal with OpenAI. News Report, March 2025.

Reuters. (2025). AMD Announces Multi-Billion Dollar Partnership with OpenAI. News Report, September 2025.

Rey, P., & Tirole, J. (2007). A primer on foreclosure. Handbook of Industrial Organization, 3, 2145-2220.

Salop, S. C., & Scheffman, D. T. (1983). Raising rivals' costs. The American Economic Review, 73(2), 267-271.

SemiAnalysis. (2024). GPU Lifespan in AI Workloads: Technical Analysis. Industry Report.

Shapiro, C., & Varian, H. R. (1998). Information Rules: A Strategic Guide to the Network Economy. Boston: Harvard Business School Press.

The Information. (2025). Inside CoreWeave's Massive GPU Purchases. News Report, May 2025.

United States v. Google LLC, No. 1:20-cv-03010 (D.D.C. 2024). Court Opinion and Findings of Fact.

Wall Street Journal. (2024). Nvidia Joins OpenAI's $6.6 Billion Funding Round. News Report, October 2024.