STATE OF THE AI ECONOMY

Investment Drivers, Structural Risks & Positioning Implications

DATE	SOURCE	FRAMEWORK	CLASSIFICATION
June 27, 2026	Exponential View	STORM Methodology	Internal Use Only

STORM RESEARCH BRIEF | KEY QUESTIONS FOR THIS REPORT	
SITUATION	What does the Exponential View data reveal about the current state of realized AI demand, CapEx, and value-stack dynamics as of mid-2026?
THESIS	Which market participants and structural forces are positioned to generate sustained returns from the AI buildout — and on what time horizon?
OPPOSITION	Where do the bear cases, overbuild risks, commoditization threats, and measurement gaps undermine the consensus bull narrative?
RISKS & RAMIFICATIONS	What are the tail risks — financing, depreciation, frontier erosion, energy constraints — and how do they cascade across the AI investment universe?
MODELS & ANALYSIS	How should portfolio managers frame position sizing, sector allocation, and hedge construction given the evidence presented?

EXEC

Executive Summary

What hedge funds need to know — right now

Exponential View's June 2026 State of the AI Economy report is the most rigorous bottom-up deconstruction of AI demand yet produced. Drawing on 1,000+ firms, source-graded revenue models, and deduplication across three stack layers, EV puts annualized GenAI ecosystem revenue at $175 billion — growing at a $65 billion annualized clip over the trailing 12 months. The sector is scaling 3x faster than any prior IT platform wave. The investment question is no longer whether demand is real. It is whether that demand can compound fast enough to service $2 trillion in committed infrastructure spend before the depreciation clock catches up.

BOTTOM LINE FOR PORTFOLIO MANAGERS

The AI economy is revenue-real and demand-elastic. The core risk is not a demand bust — it is a margin compression cycle driven by open-weight commoditization of foundation models, rising depreciation burdens, and a shift in value capture toward infrastructure (hosting, energy) and eventually the application layer. Nvidia remains the clearest near-term beneficiary. Frontier labs face a time-limited pricing premium. Hyperscalers are in a CapEx supercycle that is currently — barely — clearing the depreciation hurdle.

S

SITUATION

The structural facts on the ground as of mid-2026

1.1 Demand Is Real, Validated, and Accelerating

The EV dataset resolves the central debate of the last two years. AI revenues are not intercompany transfers, buy-side enthusiasm, or accounting fiction. They are genuine end-customer payments traced to primary filings, audited accounts, and cloud attribution (e.g., OpenAI revenue surfaced through Azure, Anthropic through Bedrock).

Metric	Current Reading	Context / Benchmark
Trailing 12-month revenue (deduplicated)	$110 billion	3x faster ramp than mobile apps or cloud at equivalent age
Annualized run rate (June 2026)	$175 billion	1.6x spread over TTM = growth rate is still accelerating
Quarterly revenue growth (recent average)	~35% QoQ	Equivalent to 3.2x annual compounding
Time to generate incremental $1B revenue	<2 days	Was 180 days in early 2023; 90x faster
AI revenue as % of US GDP	0.42%	IT sector is 9.4% — implies 20x+ runway if AI penetrates broadly
Combined hyperscaler backlog (RPO)	$2.0 trillion	Microsoft $633B, Oracle $553B, Google $468B, Amazon $364B
Global semiconductor revenue (2026E, WSTS)	$1,511 billion	From $792B in 2025; near-doubling driven by AI silicon

1.2 The Compute Supercycle in Numbers

Demand has re-accelerated US electricity generation growth from flat (2008–2024) to +9 TWh/month annual growth. Data centers account for ~55% of the incremental US grid load forecast through 2030. The largest AI data centers have grown 50x in power draw in four years (from ~21 MW frontier supercomputers to planned ~2.3 GW Rainier complex). Memory's share of data center build cost has jumped from a rounding error (~2%) to ~18% of total cost — the single biggest shift in the cost stack.

ANALYST NOTE — MEMORY

The memory share shift from 2% to 18% of total data center build cost is the most under-appreciated structural driver in the report for HBM and advanced DRAM pricing. This validates the Supply Commitment Agreement thesis for Micron (MU). AQI has covered this dynamic extensively in prior MU earnings briefings.

T

THESIS

Where the structural bull case is strongest

2.1 The Demand-Elasticity Engine

The single most investable insight in this report is the token price elasticity finding. EV demonstrates that across Google, OpenAI, and ByteDance, price elasticity of demand for AI tokens runs approximately -1.2 to -1.8. Every 10% price cut generates 12–18% more token volume — meaning total token revenue rises even as unit prices fall. This is a Jevons Paradox dynamic: efficiency improvements increase consumption rather than reduce it.

Provider	Price Change	Volume Response	Net Revenue Direction
Google (Gemini API)	-97% over period	+50x volume	Total spend grew substantially
ByteDance (Doubao)	-50% in 12 months	+12x volume (4T → 50T daily)	Rapid revenue expansion
OpenAI (API)	-90% cumulative	Significant volume uptick	Revenue compounding confirmed
Industry blended	Tokens: $17 → $2/MTok	30Q+ tokens/month (14x YoY)	Gross revenue +3.2x/year

2.2 Infrastructure: The Current Value Capture Champion

Despite hosting's position at the bottom of the value chain, it currently captures ~82% of deduplicated quarterly GenAI revenue. This reflects a fundamental reality: the buildout era concentrates cash flows in the infrastructure layer before application-level monetization matures. The analog is telecom in 1998 — the pipes captured most of the economic rent before content did.

Nvidia's position is structurally exceptional. Its data center revenue is not capex-accounted by customers — it is the bill they pay for the capex itself. Supply commitments from Nvidia's customers have grown from $31B to $95B in twelve months, with no evidence of demand softening. GPU rental rates (H100 spot) have recovered from $1.70/hr overbuild fears to $2.40/hr on surging inference demand.

2.3 Early Adopter Revenue Advantage Is Measurable

EV's cross-section of 70,000 US businesses (Ramp data) shows a 92% revenue growth gap since November 2022 between firms in the top 25% of AI spend intensity versus firms with no AI spend. This mirrors the automation adoption premium documented by Bessen et al. (2018) across the 2000–2016 period, where automating firms outgrew non-automating peers by 37 percentage points. The signal is early but real.

THESIS STATEMENT

The AI bull case rests on three pillars: (1) elastic demand that converts price declines into volume acceleration, maintaining revenue growth even as unit economics compress; (2) a depreciation curve that current revenues are just clearing — creating a narrow but positive margin window for infrastructure operators; and (3) a winner-take-most dynamic in chips (Nvidia), frontier models (OpenAI/Anthropic), and eventual application monetization that rewrites the enterprise software P&L.

O

OPPOSITION

Where the bear case is strongest — and what we do not yet know

3.1 The Depreciation Overhang

The most important near-term risk the EV data surfaces is not demand — it is the depreciation math. The 2026E annual depreciation charge from the AI buildout approaches $111 billion. At the standard 6-year chip life assumption, Q1 2026 hyperscaler/neocloud revenue headroom after depreciation is only 19%. Depreciation currently absorbs 81% of infrastructure revenue before any other cost line.

This is thin coverage — and it gets worse before it gets better. As committed CapEx from 2025–2026 enters service, the depreciation base will rise faster than revenue in the near term. Revenue must continue to compound at 35%+ QoQ or margin headroom collapses.

Chip Life Assumption	Infra Headroom (Q1 2026)	All GenAI Headroom	Bear/Bull Signal
3 years (rapid obsolescence)	-40% (negative)	-18%	Deep bear — revenues can't cover costs
4 years	-12%	6%	Bear — thin or negative
5 years	7%	21%	Caution — marginal coverage
6 years (standard)	19%	32%	Baseline — barely adequate
8 years (T4 analog)	36%	46%	Bull — meaningful margin
9 years	42%	51%	Strong bull — CapEx fully recovered

BEAR CASE TRIGGER

If reasoning model capability improvements slow — reducing the pace at which new use cases unlock — OR if open-weight models catch the closed-weight frontier faster than anticipated, token volume growth could decelerate to below the 35% QoQ pace needed to outrun the depreciation curve. This is the single scenario where the CapEx story breaks.

3.2 Open-Weight Commoditization of the Foundation Model Layer

EV's OpenRouter data shows the combined Google/OpenAI/Anthropic token share falling from 72% to 33% among sophisticated model-routing users. While OpenRouter is not a representative cross-section of enterprise AI spend, it is a leading indicator of where price-sensitive, technically capable buyers route workloads. The frontier premium — currently well-documented and real — is time-limited.

The commoditization timeline compresses with each major open-weight release. DeepSeek-R1 achieved frontier-class performance at a fraction of the inference cost. DeepSeek V3.2 extended this. GPT-5 nano and Qwen Turbo demonstrate that lightweight open-weight models are closing the GPQA Diamond performance gap that justified closed-weight pricing.

3.3 GDP Measurement Gap — The 'Productivity Paradox' Risk

The economy section of the EV report surfaces a structural bear case for the macro AI thesis: despite $175B in annualized revenues, AI represents only 0.42% of US GDP. Corporate profits remain 32x larger than all GenAI revenues. Even among S&P 500 firms, only 33% made AI impact claims on earnings calls as of Q1 2026 — and only 20% provided quantified impacts.

Seven in ten GenAI claims on earnings calls focus on cost savings or efficiency, not revenue generation. This is rational (cost savings have better margin leverage) but it implies the AI economy is primarily a cost-restructuring technology thus far, not yet a revenue-generative force. The productivity paradox — where technology investment precedes measurable GDP impact by a decade or more — is a real risk to time-sensitive AI equity valuations.

3.4 External Financing Risk

Neocloud CapEx is primarily debt-funded. As the marginal AI infrastructure dollar becomes increasingly externally financed, downside scenarios move outside the balance sheets of cash-rich hyperscalers and into credit markets and third-party capital providers. A revenue shortfall scenario that merely dents hyperscaler profits could trigger covenant stress for highly-levered neoclouds.

R

RISKS & RAMIFICATIONS

Tail risks, cascades, and second-order effects

4.1 Risk Matrix

Risk Factor	Probability	Severity	Affected Exposure	Hedge / Monitor
Open-weight frontier parity	High (12–18 months)	High	Frontier lab revenues, FM licensing fees	Long infrastructure; short application wrappers
Depreciation overhang tightens	Medium-High	Medium-High	Hyperscaler FCF, neocloud solvency	GPU rental rate as real-time signal
Revenue growth deceleration below 25% QoQ	Medium	Very High	All AI equity valuations	Token volume metrics; quarterly cadence
Energy/grid constraint	Medium	Medium	Data center expansion pace, US utilities	Utility capex announcements; PPA pricing
Chip useful life shortens (<5 yr)	Low-Medium	Very High	Hyperscaler depreciation math breaks	Accelerated depreciation disclosures
AI GDP measurement remains opaque	High (structural)	Low (near-term)	Policy response, valuation multiples	Brynjolfsson-style consumer surplus metrics
Neocloud credit stress	Low-Medium	Medium	CoreWeave, Nebius creditors; spillover	Credit spreads; GPU collateral values

4.2 The Stack Compression Scenario

EV identifies four compression scenarios that would shift value from producers to consumers. The most likely near-term version is a hybrid of Scenario 1 (open-weight catch-up) and Scenario 3 (reduced compute needs via distillation and efficiency). In this hybrid, the application layer and consumers capture most of the surplus — at the expense of foundation model labs' token pricing power and, ultimately, hosting margin.

The cascade runs: open-weight capability parity → lab licensing fees compress → inference cost per token falls → demand volume increases (elasticity) → aggregate token revenue holds but margin compresses → hosting layer must improve utilization or face margin squeeze → Nvidia remains insulated because its customers keep buying regardless of margin.

RAMIFICATION FOR PORTFOLIO CONSTRUCTION

The risk structure of AI equity exposure is asymmetric by layer. Chips (Nvidia, TSMC) are the best risk-adjusted position: they capture CapEx as revenue regardless of which application wins. Foundation model labs (OpenAI, Anthropic — private) face a time-limited pricing premium. Hyperscalers benefit from the buildout but compete with their own customers. Pure-play application wrappers without proprietary data or workflow moats are the most exposed to compression.

M

MODELS & ANALYSIS

Positioning framework, scenario outputs, and watch variables

5.1 Scenario Framework: The Three Paths

Scenario	Revenue Growth (QoQ)	Token Price Decline	Infra Headroom	Likely Winners
Bull: Demand absorbs all	>35% QoQ sustained	-20%/yr	Expands to 30–40%	NVDA, TSMC, hyperscalers, vertical AI apps
Base: Growth decelerates gradually	20–35% QoQ	-30%/yr	Holds 15–25%	NVDA, cloud infra; pressure on FM labs
Bear: Growth misses, overbuild	<20% QoQ or deceleration	-40%+ yr	Compresses <10%	Energy utilities, edge computing; short neoclouds

5.2 Key Real-Time Variables to Monitor

The EV framework implies a compact set of forward-looking variables that will determine which scenario materializes. AQI recommends tracking the following on a quarterly or monthly basis:

Variable	Current Reading	Bull Signal	Bear Signal
H100 spot rental rate ($/hr)	$2.40/hr (recovering)	>$2.75 sustained	<$1.80 (overbuild fear return)
Hyperscaler QoQ revenue growth	~35% (AI segment)	Acceleration or hold	Deceleration below 25%
Hyperscaler RPO growth	$2.0T cumulative	Continued backlog additions	RPO growth slowing or cancellations
Open-weight benchmark parity vs. frontier	Closing gap on GPQA Diamond	Frontier holds 2+ tier lead	Open-weight matches frontier within 6 months
Token price elasticity signal	Beta ~-1.5 (robust)	Volume grows faster than price falls	Volume growth stalls despite price cuts
Nvidia supply commitments	$95B (Q1 2026)	>$120B in next disclosure	Flat or declining commitments
Memory share of DC build cost	~18% (vs. 2% in 2021)	Further increase — HBM demand surge	Reversal below 15%
US electricity load growth	+9 TWh/month annualized	Sustained or accelerating	Regulatory constraint or demand shortfall

5.3 Investment Positioning Implications

Synthesizing the STORM analysis, AQI frames AI equity exposure across four positioning tiers:

Tier	Category	Rationale	Key Names
1 — Core Long	Semiconductors (Chips + Memory)	Revenue = customer CapEx; insulated from who wins at app/model layer. Memory repricing from 2% to 18% of build cost is structural.	NVDA, TSMC, MU (HBM)

Tier	Category	Rationale	Key Names
2 — Overweight	Energy & Grid Infrastructure	AI is reigniting a flat 16-year US power market. Data centers = ~55% of new load growth. Secular re-rate underway.	Utilities, natural gas generators, nuclear operators
3 — Selective	Hyperscalers with AI-native revenue	Benefit from buildout but compete on multiple fronts. Valuation must discount both the upside (RPO backlog) and the CapEx burden (depreciation headroom thin at 19%).	MSFT, GOOGL, AMZN — selectivity based on AI revenue mix
4 — Hedge / Avoid	Pure-play FM application wrappers (no moat)	Open-weight commoditization + general-purpose model expansion squeezes pricing power for generic AI feature companies. Defend with proprietary data or domain depth.	Generic SaaS AI add-ons; wrapper-only apps

5.4 The Token Economy's 'Pay-Per-Click Moment'

One of the most structurally important observations in the EV report is the analogy between token-based pricing and Google AdWords' 2002 shift to cost-per-click. Untracked banner ads (free AI access, subscription bundles) had a limited, crash-prone market. Attribution-based pricing (CPC in advertising; tokens in AI) unlocked a massive expansion of addressable spend because buyers could directly link cost to measurable return.

The implication for investors: the shift from flat-fee AI subscriptions to consumption-based token pricing is an inflection point in AI revenue addressability. It enables enterprises to scale AI budgets proportional to demonstrated ROI — which is how IT budgets grow from rounding error to material P&L line. This is the mechanism by which AI's 0.42% of GDP ratio eventually looks like the IT sector's 9.4%.

AQI SYNTHESIS

The EV report makes the demand case convincingly. The investment edge is in the second-order implications: the depreciation math creates a narrow but real margin window that favors infrastructure operators today; open-weight commoditization will redistribute that margin to applications and consumers over the 2–4 year horizon; the token economy's pay-per-click moment enables budget expansion that could double or triple AI's GDP share within 5 years. Position accordingly — with chips as the structural core, energy as the secular re-rate, and hyperscalers as tactical rather than strategic exposure.

PORT

Thematic Portfolio Construction

Long/Short basket ideas derived from the STORM framework

The following baskets are organized by the five structural themes the EV report identifies as primary value drivers. Each basket includes a rationale anchored to specific EV data points, names across market cap and liquidity tiers, and explicit short ideas where the data supports the trade. This is a thematic framework, not a recommendation. All positions require independent fundamental diligence.

IMPORTANT DISCLOSURE

These baskets are constructed for internal research and discussion purposes only. They are not investment recommendations. AlphaQuery Intelligence does not hold positions in the securities listed below on behalf of clients without separate engagement. Past performance of thematic baskets does not predict future results.

Basket 1 — Silicon Supercycle (Long Bias)

EV ANCHOR: AI CapEx is $535B above the pre-AI trend in 2026E. Chip/logic share of data center build cost rose from 40% to 60%; memory alone went from 2% to 18%. Nvidia supply commitments grew from $31B to $95B in twelve months. GPU rental rates are recovering to $2.40/hr on surging inference demand. This basket owns the physical infrastructure of the AI economy — entities that receive revenue regardless of which model or application wins.

Name	Ticker	Mkt Cap Tier	Rationale (EV-linked)	L / S	Risk
Nvidia	NVDA	Mega-cap	Supply commitments $95B and growing. GPU rental rate recovery confirms demand absorption. Only supplier at scale for frontier training AND inference.	L	Valuation / China export
TSMC	TSM	Mega-cap	Manufactures silicon for every major AI chip including Nvidia, AMD, Broadcom, Apple, and custom hyperscaler silicon (TPUs, Trainium). CapEx beneficiary regardless of who wins.	L	Taiwan geopolitical
Micron Technology	MU	Large-cap	Memory repricing: 2% → 18% of DC build cost is the most under-appreciated driver. HBM3E/HBM4 sole or primary supplier alongside SK Hynix. SCA dynamics supportive.	L	DRAM cycle timing
SK Hynix	000660.KS	Large-cap	Co-dominant in HBM alongside Micron. Primary Nvidia HBM supplier. Beneficiary of the memory supercycle with less US export risk than US peers.	L	KRW / EM volatility
Broadcom	AVGO	Mega-cap	Custom silicon (XPUs) for Google, Meta, Apple. ASIC revenue growing as hyperscalers diversify from Nvidia. Networking (Tomahawk, Jericho) is essential AI fabric.	L	Integration execution (VMware)
Marvell Technology	MRVL	Large-cap	Custom AI ASIC design wins at Amazon (Trainium) and others. Data center interconnect. Meaningful exposure to the DC networking build.	L	Revenue concentration
AMD	AMD	Large-cap	MI300X gaining traction in inference workloads. Open ecosystem vs. CUDA is a long-term diversification play for hyperscalers. EV notes AMD as major chip revenue contributor.	L	CUDA ecosystem moat

Name	Ticker	Mkt Cap Tier	Rationale (EV-linked)	L / S	Risk
Intel	INTC	Large-cap	Gaudi AI accelerator has struggled to gain traction vs. Nvidia/AMD. Foundry business turnaround uncertain. AI chip market share loss is structural near-term.	S	Turnaround optionality

Basket 2 — Energy & Grid Infrastructure (Long Bias)

EV ANCHOR: US electricity generation growth reignited from flat (2008–2024) to +9 TWh/month annual growth. Data centers account for ~55% of new US grid load expected by 2030. The incremental US grid load forecast has grown 7x since 2022 (24GW → 166GW). This is a secular demand re-rate for a sector that was priced for structural stagnation. No other sector has a more durable AI demand catalyst with less valuation risk from AI model competition.

Name	Ticker	Mkt Cap Tier	Rationale (EV-linked)	L / S	Risk
Constellation Energy	CEG	Large-cap	Nuclear baseload is the preferred 24/7 carbon-free power source for hyperscaler PPAs. Microsoft Three Mile Island deal is the template. Data center demand secular tailwind.	L	Regulatory / nuclear license
Vistra Energy	VST	Large-cap	Diversified generation including nuclear. Texas grid exposure directly benefits from data center load growth in Austin/Dallas corridor. Highest beta to power price upside.	L	ERCOT price volatility
Talen Energy	TLN	Mid-cap	Amazon signed direct PPA for Susquehanna nuclear plant adjacent to AWS data center campus. Embedded co-location model is a structural premium.	L	Single-asset concentration
GE Vernova	GEV	Large-cap	Gas turbines, grid equipment, and wind. GE Vernova is the primary beneficiary of the US grid hardening and new generation buildout. Transformer and switchgear backlog multi-year.	L	Supply chain / execution
Eaton Corporation	ETN	Large-cap	Power management, switchgear, UPS systems. Every data center requires Eaton's products. Backlog visibility is high; lead times support premium pricing.	L	Cyclical input costs
Quanta Services	PWR	Large-cap	Electrical infrastructure construction. Transmission line buildout is the bottleneck for connecting data center campuses to the grid. EPC exposure to the constraint.	L	Labor / project execution
NextEra Energy	NEE	Mega-cap	Largest US renewable developer. Signed data center PPAs and long-duration storage investment. Valuation stretched; benefit is real but priced.	L/Neutral	Rate sensitivity; valuation
Traditional coal utilities (no gas/nuclear)	Various	Various	AI load growth primarily benefits dispatchable baseload (nuclear, gas) and grid infrastructure. Legacy coal utilities without these assets are structurally disadvantaged.	S	Regulatory /utility commission risk

Basket 3 — AI Infrastructure / Hyperscalers (Selective Long / Tactical)

EV ANCHOR: Hyperscaler and neocloud quarterly revenues first exceeded CapEx depreciation in Q4 2025, but headroom remains thin at 19% for infrastructure-only revenues. The combined hyperscaler RPO backlog is $2.0T. Infrastructure is the dominant layer capturing 82% of deduplicated GenAI revenue — but the depreciation clock is ticking. These are tactical rather than structural positions; sizing should reflect the thin margin coverage and CapEx risk.

Name	Ticker	Mkt Cap Tier	Rationale (EV-linked)	L / S	Risk
Microsoft	MSFT	Mega-cap	Largest hyperscaler RPO at $633B. Azure AI (OpenAI distribution via Bedrock equivalent) is the primary enterprise AI access point. Copilot monetization still early but real.	L	CapEx intensity; OpenAI dependency
Alphabet / Google	GOOGL	Mega-cap	Google Cloud $468B RPO. Gemini token volume 50x in 18 months (price -97%, volume +50x) is the cleanest elasticity proof point in the dataset. TPU efficiency is a moat.	L	Search disruption risk; antitrust
Amazon / AWS	AMZN	Mega-cap	AWS RPO $364B. Bedrock (Anthropic distribution) + Trainium custom silicon + EC2 GPU compute. Most diversified AI infrastructure exposure.	L	Logistics drag on multiple
CoreWeave	CRWV	Large-cap	Pure-play neocloud; debt-funded CapEx model. Higher upside in bull scenario (capacity fully contracted) but meaningful credit risk if demand softens. EV flags neocloud as primarily debt-funded.	L (small)	Leverage / refinancing risk
Nebius Group	NBIS	Small/Mid-cap	European neocloud with significant GPU capacity build. Early-stage, higher risk/reward. EV includes in hyperscaler/neocloud CapEx basket.	L (speculative)	Execution / geopolitical
Oracle	ORCL	Large-cap	RPO $553B — largest single RPO in the dataset after Microsoft. AI infrastructure contracts with OpenAI and others. Multi-cloud GPU compute is the growth driver.	L	Legacy software mix; valuation

Basket 4 — Application Layer (Long Moats / Short Wrappers)

EV ANCHOR: Value is moving up the stack toward apps and models — the app layer grew from 3% to 7% of deduplicated quarterly GenAI revenue in one year (+2.95x). BUT: open-weight commoditization of the FM layer means generic application wrappers without proprietary data or workflow depth face severe margin compression. The long/short split within this basket is the clearest fundamental pair trade in the AI economy.

Name	Ticker	Mkt Cap Tier	Rationale (EV-linked)	L / S	Risk
Palantir Technologies	PLTR	Large-cap	AIP (AI Platform) is a proprietary data ontology layer — the definition of workflow-specific, non-replicable AI integration. Government + enterprise. Moat is the data access, not the model.	L	Valuation (expensive)
ServiceNow	NOW	Large-cap	Agentic AI workflows embedded in enterprise IT operations. Switching costs are structural. AI is accelerating, not threatening, the ServiceNow moat.	L	Valuation; macro IT spend
Salesforce	CRM	Large-cap	Agentforce is the most mature enterprise agentic deployment platform. CRM data is the proprietary moat. Token consumption per enterprise workflow is 1,200x a chat task per EV data.	L	Agentforce adoption pace
Veeva Systems	VEEV	Large-cap	Vertical AI in life sciences — proprietary clinical and regulatory data. Exactly the domain-specific workflow EV identifies as defensible against general-purpose model compression.	L	TAM constraint; competition
Tyler Technologies	TYL	Mid-cap	Vertical AI in government — permitting, courts, utilities. Proprietary government data and workflow depth. Less analyzed AI application moat story.	L	Government budget cycles
Cursor / Anysphere	Private	Private	EV explicitly names Cursor as a high-growth standalone GenAI app with real revenue. Agentic coding is the highest token-intensity use case (1,200x chat). Watch for IPO.	L (private/pre-IPO)	Funding rounds only
Generic 'AI copilot' SaaS add-ons	Various	Various	Thin-wrapper AI features bolted onto legacy SaaS without proprietary data. Open-weight models + general-purpose frontier will disintermediate these features. EV Scenario 2 is the mechanism.	S	Takeover premium
Chegg	CHGG	Small-cap	Tutoring platform structurally disrupted by ChatGPT. Revenue decline is documented. No proprietary AI moat. Represents the clearest case of AI substitution eroding incumbent TAM.	S	Short squeeze; pivots
Shutterstock / Getty	SSTK	Small-cap	Stock photography TAM under direct assault from generative image models. Licensing revenue headwinds are structural. AI content generation is a direct substitute.	S	Licensing deal optionality

Basket 5 — Foundation Model Layer (Primarily Private / Proxy Plays)

EV ANCHOR: FM revenues are concentrated with OpenAI and Anthropic (HHI 1,594 — lowest concentration of any layer, reflecting open-weight competition). Frontier labs earn a time-limited premium on the current capability frontier. Last year's frontier commoditizes within 12–18 months per the GPQA Diamond price-per-capability data. The primary public market exposure to frontier model labs is through their cloud distribution partners.

Name	Ticker	Mkt Cap Tier	Rationale (EV-linked)	L / S	Risk
OpenAI	Private	Private	Largest FM revenue share. GPT-5 family capturing enterprise API and Claude Code / Codex agentic coding wave. Frontier premium is real but eroding. Pre-IPO secondary market.	L (private)	Valuation; competition; governance
Anthropic	Private	Private	Second-largest FM revenue. Claude Code is the highest-profile agentic coding deployment. AWS/GCP distribution provides Bedrock/Vertex revenue attribution. Pre-IPO secondary.	L (private)	Valuation; compute costs
Microsoft (OpenAI proxy)	MSFT	Mega-cap	Azure OpenAI Service is the most direct public-market proxy for OpenAI revenue. Microsoft's RPO growth is partly downstream of OpenAI enterprise adoption.	L	OpenAI relationship dependency
Amazon (Anthropic proxy)	AMZN	Mega-cap	Bedrock Anthropic is the primary cloud distribution for Claude. AWS benefits from Anthropic token consumption regardless of competitive dynamics between models.	L	Diffuse across large business
Baidu	BIDU	Large-cap	Ernie Bot is the leading Chinese frontier model with real enterprise traction. China market is excluded from EV scope but represents a parallel demand curve. Value stock with AI optionality.	L (selective)	China regulatory / geopolitical
Pure-play AI ETFs (non-discriminating)	BOTZ, ROBO, AIQ	ETF	These ETFs hold diversified AI-adjacent exposure without distinguishing between chip suppliers (high conviction), infrastructure (thin margin), and application wrappers (compression risk). No moat selectivity.	S/Underweight	Diversification reduces downside

Pair Trade Summary — Highest Conviction Long/Short Combinations

Derived from the STORM analysis, the following pair trades represent the cleanest fundamental divergence implied by the EV dataset:

Long Leg	Short Leg	Thesis	EV Data Anchor
Micron (MU)	Intel (INTC)	Memory repricing from 2% to 18% of DC build cost is structural tailwind for MU HBM. Intel has lost AI silicon share and foundry turnaround is uncertain.	Memory share of DC cost; chip market concentration HHI
Constellation Energy (CEG)	Legacy coal utilities	AI load growth favors 24/7 dispatchable baseload (nuclear). Coal economics structurally disadvantaged even with grid tightening.	9 TWh/month US electricity growth; data center = 55% of incremental load
Palantir (PLTR)	Generic AI SaaS copilots	Proprietary data ontology vs. bolt-on AI features. Open-weight compression hits wrappers; AIP moat is workflow depth not model access.	EV Scenario 2 — general-purpose models replace integration layer
Nvidia (NVDA)	Non-discriminating AI ETFs	NVDA captures CapEx as revenue regardless of model/app winner. Broad AI ETFs hold exposure at every layer including compression-exposed FM and wrapper plays.	CapEx $535B above pre-AI trend; supply commitments $31B → $95B
Quanta Services (PWR)	Generic data center REITs (no power)	Transmission buildout is the grid bottleneck. Data center REITs without secured long-term power contracts face stranded asset risk as grid capacity constrains expansion.	166GW US load growth by 2030; 7x increase in grid forecast since 2022
ServiceNow (NOW)	Chegg (CHGG)	Agentic workflow moat vs. direct AI substitution of core product. NOW's switching costs are additive to AI; CHGG's tutoring TAM is being substituted.	Chat-to-agent transition; 1,200x token multiplier for agentic tasks

PORTFOLIO CONSTRUCTION NOTE

These baskets are designed to be used as a research starting grid, not as equal-weighted positions. Basket 1 (Silicon Supercycle) and Basket 2 (Energy & Grid) carry the highest EV-data conviction given their insulation from model competition dynamics. Basket 4 (Application Layer) contains the highest alpha potential but also the most binary risk — the long/short split requires continuous monitoring of open-weight benchmark parity and enterprise AI ROI disclosures. Position sizing should reflect the depreciation headroom math: at 19% coverage, the margin of safety is thin, and any demand deceleration compresses it rapidly.

WATCH

AQI Research Watch List

Follow-on questions and monitoring priorities

The following open questions should drive AQI's next research cycle based on EV's framework:

- What is the actual GPU useful life distribution across deployed H100 and B200 fleets? The difference between 6-year and 8-year assumptions produces a 17-percentage-point swing in infrastructure headroom — this is the most consequential unknown in the entire AI CapEx math.
- How quickly does open-weight capability close the agentic task horizon gap? EV cites METR task horizon as the relevant benchmark. Tracking this quarterly is essential.
- What is the enterprise AI ROI measurement cadence? The shift from 50–60% unquantified claims to fully quantified will be the catalyst for AI budget acceleration in the S&P 500.
- Where does Anthropic's revenue sit within Bedrock? The cloud-attribution methodology EV uses implies Anthropic's revenue is partially visible through AWS disclosures. This warrants dedicated analysis.
- How will energy constraint risk materialize? The 7x increase in US grid load expectations (24GW to 166GW by 2030) implies permitting, transmission, and generation capacity will be the rate-limiting variable within 18–24 months.
- Is consumer surplus capturing the AI value that GDP misses? EV cites Stanford data showing $3–4B monthly consumer surplus above actual revenue. Tracking this welfare metric could provide early warning of mainstream adoption acceleration.