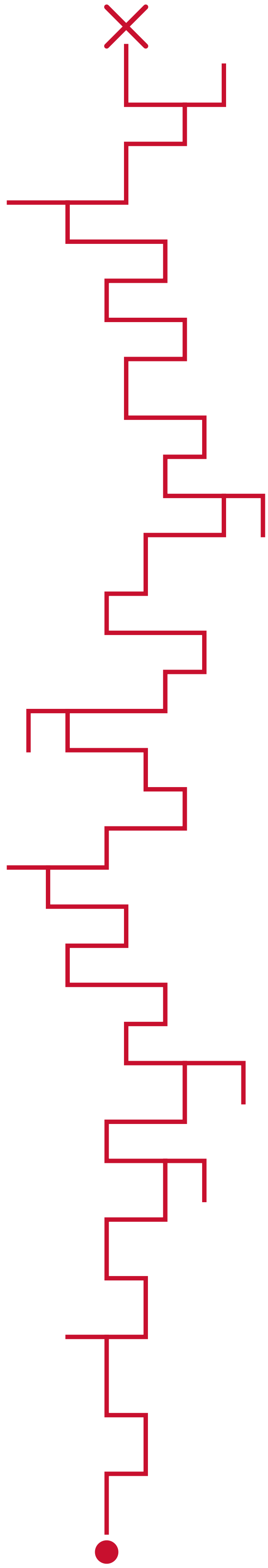
Reconciling Enterprise AI Revenue

A Methodological Crosswalk



Philipp D. Dubach

MAY 2026

Abstract

Audit-grade enterprise AI revenue, under any no-double-count treatment, sits between \$63.2B and \$72.5B worldwide. That is roughly 4–5% of the \$1.478T Gartner publishes as the headline AI market, and it is the band of disclosed revenue the \$690B+ of 2026 hyperscaler capex has to be reconciled against. The published anchors do not agree. Menlo Ventures puts US enterprise GenAI demand at \$37B from a 495-respondent buyer survey. The bottom-up sum of disclosed vendor run-rates lands at \$100–135B worldwide, midpoint \$123B fully-net. IDC reports \$307B for AI solutions. Gartner forecasts \$1.478T. Top to bottom that is roughly a 40x spread on the same year and the same topic. This report builds the public, source-linked reconciliation that does not yet exist: a 68-vendor primary-source census, a six-tier disclosure-quality framework grading every figure from SEC segment-level (A) down to vendor self-classification (F), and a four-step waterfall walking Gartner's \$1.478T down to Menlo's \$37B through sourced deductions. The Tier A floor lands at \$63.2–72.5B because audited segment-level AI disclosure is thin. Three silicon vendors (NVIDIA, AMD, Broadcom) sit upstream of enterprise spend. CoreWeave's revenue is largely a Microsoft channel relationship and overlaps with the MSFT line when netted. The hardware OEM cohort (Dell, HPE, SMCI, Pure Storage and peers) adds about \$73B gross of AI-attributed revenue but only about \$7.25B net of silicon overlap once Arista's \$3.5B FY26 raised guidance target is excluded for consistency with ServiceNow's \$1.5B FY26 ACV target — both forward guidance figures rather than realized run-rate. Silicon revenue disclosed in 2025 ran at \$254B annualized (NVIDIA, AMD, Broadcom Data Center segments), roughly 4x the audit-grade enterprise AI revenue that is supposed to absorb it; that gap is the empirical reconciliation the report quantifies. The four estimates nest. \$63.2–72.5B audit-grade sits inside \$100–135B disclosure-grade (midpoint \$123B fully-net), which sits inside \$307B bundled-solutions, which sits inside \$1.478T umbrella. Which figure to use depends on the question. The Spread Index (audit-grade ÷ umbrella) is proposed as a tracking ratio: 4.28% at the narrow floor, 4.90% at the broad floor. It needs four to six quarters of observations before quarter-over-quarter movement becomes interpretable against measurement noise.

CONTENTS

Section 1

Introduction

Setup, scope, theoretical grounding, and the welfare framing.

1.1 The setup

Four widely cited enterprise AI revenue figures published in 2025 span a 40x range. Menlo Ventures' *2025 State of Generative AI in the Enterprise* puts US enterprise GenAI spending at $37B. The bottom-up sum of disclosed vendor run-rates lands at $100–135B worldwide. IDC's *Worldwide AI and Generative AI Spending Guide* reports $307B in AI solutions spending. Gartner's September 17, 2025 forecast headlines $1.478T for total worldwide AI spending. Same year, same topic, roughly 40x between top and bottom. The spread matters because most active arguments about AI economics rest on a choice about which figure to anchor against: capex coverage in 2026, valuation of AI-native private companies against a peer set, the analysis of AI-driven software-pricing disruption. Analysts pick one before they reason, often without saying which. The four numbers are not in contradiction; they measure four different things.

1.2 Approach and scope

There is prior work on parts of the ground. Epoch AI publishes compute and training estimates. David Cahn's "$600B question" essays frame capex coverage. Tomasz Tunguz writes on vendor-sum aggregation. Sacra and PYMNTS track vendor-level run-rates. Sell-side bridges between published figures exist but mostly sit behind paywalls. What does not exist in public form is a reconciliation that takes Menlo, IDC, and Gartner as inputs, builds a primary-source vendor census against them, and produces an audit-traceable triangulation. That gap is what this report fills. It is not a buyer survey, produces no 2026 forecasts, and does not develop proprietary fleet or compute estimates beyond what is already public. US-only and worldwide treatments are flagged section by section rather than collapsed into a single denominator. The method is desk research only: SEC EDGAR filings, IR pages, earnings transcripts, corporate press releases, Wayback Machine snapshots. Every figure is graded on a six-tier A–F disclosure-quality scale (A: SEC-audited segment-level; B: executive earnings-call dollar disclosure; C: mentioned without dollar; D: press release or named-source journalism; E: CEO/founder commentary; F: vendor self-classification or third-party estimate). The "as of" snapshot date is 2026-05-14. All headline numbers are reproducible from the underlying vendor census and reconciliation build.

1.5 Theoretical grounding

The disclosure-tier framework sits on disclosure economics, even though it is presented operationally. The intellectual antecedents are Verrecchia (1983) on strategic disclosure under uncertainty, Dye (1985) on dichotomous disclosure when receivers cannot tell silence from absence of information, and Diamond and Verrecchia (1991) on the cost-of-capital consequences of disclosure choice. Beyer, Cohen, Lys, and Walther (2010) synthesize the literature for empirical work. The six-tier scheme adapts these frameworks by treating disclosure granularity as a measurable ordinal variable rather than a binary disclose-or-withhold decision. The capex coverage analysis in §5 is implicitly q-theory in the Hayashi (1982), Abel (1983), and Bond and Cummins (2001) sense without the framing: investment Q is marginal value over replacement cost, and an audit-grade revenue numerator against capex denominator is Q-equivalent for AI infrastructure under the assumption that audit-grade revenue is a lower-bound proxy for marginal product. The resale-netting in §4.4 (OpenAI revenue passed through Microsoft Azure, netted against Microsoft's reported AI revenue) is a principal-versus-agent recognition question under ASC 606 (Financial Accounting Standards Board, 2014), studied in the auditing-economics literature surveyed by DeFond and Zhang (2014). The contribution here is operational structure layered onto these literatures, not new theory.

The framework is falsifiable. If audit-grade AI revenue prints at $200B by Q4 2026, catching up to capex faster than the current trajectory implies, the headline finding is retired. If audit-grade revenue at Q4 2027 remains under $150B against the by-then-disclosed 2026–2027 capex of roughly $1.2T, the framework's central concern is confirmed. The 2028 milestone year carries the test.

1.6 Welfare and incidence

The reconciliation question is private: what is the audit-grade revenue against the disclosed capex? The welfare question is broader, and the framework is silent on it by design. Of the roughly $1.5T in AI-related debt issuance currently underwriting the build, the loss-bearing parties under realistic stress scenarios are not the same. Equity holders of hyperscaler parents absorb the recourse-debt risk at parent free-cash-flow level. Lenders to non-recourse SPVs (the GPU-leasing structures and data-center special-purpose vehicles) absorb impairment in the capital structure through forced refinancings or workouts. Customers paying AI-feature pricing without separately purchasing AI bear redistribution as the incidence shifts into non-AI SKU pricing. Labor markets bear it if productivity gains route to capital rather than wages, and sovereign balance sheets bear it where Project Stargate-style guarantees are in place. The report's scope is the measurement question, not these incidence channels, but the framework's reception in policy and capital markets will turn on which incidence answer it implicitly suggests. Knowing that audit-grade revenue is $63B against $690B of 2025 capex is necessary input to any incidence analysis, even if the measurement alone does not determine who bears the loss.

Section 2

The four published estimates

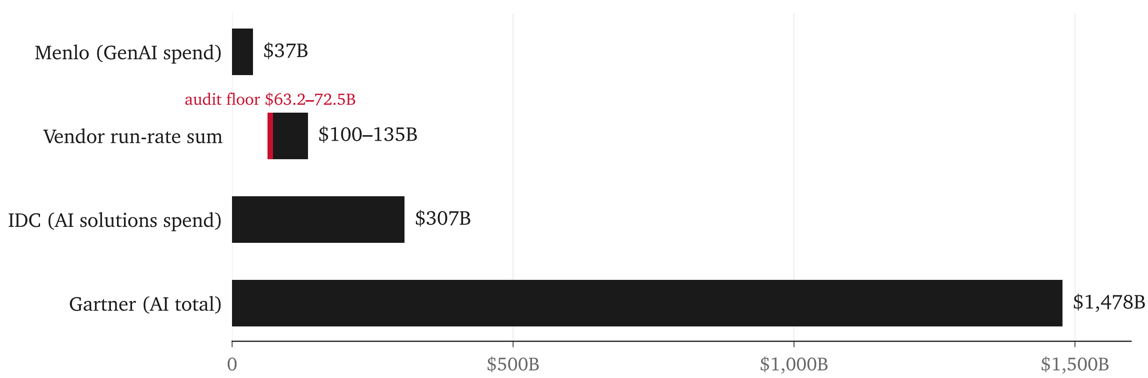
Menlo, vendor-sum, IDC, and Gartner – what each one measures and what it leaves out.

Three of the most widely cited enterprise AI revenue figures published in 2025 differ by roughly 40x. A fourth, the sum of disclosed vendor run-rates, sits between them. None of these numbers is "wrong." They measure different things.

Before the methodological crosswalk in Section 3, the four estimates as published:

FORECASTER	HEADLINE	YEAR	METHODOLOGY SUMMARY	PRIMARY SOURCE
Menlo Ventures	$37B	2025	Bottom-up enterprise survey (n=495 US decision-makers); GenAI spend only	Menlo, *2025 State of GenAI in the Enterprise*, Dec 9 2025
Vendor run-rate sum	$100–135B (disclosed-only; audit-grade floor $63.2B narrow / $72.5B broad)	2025–Q1 2026 (exit run-rate)	Bottom-up summation of disclosed AI revenue at major vendors (model labs, hyperscalers, AI-native SaaS), net of silicon and resale overlap	Company disclosures; sell-side notes; Tunguz, Sacra, SaaStr aggregations; see the reconciliation build
IDC	$307B	2025 (FY)	Top-down market tracker; AI solutions spend (software + hardware + services)	IDC, *Worldwide AI and Generative AI Spending Guide*
Gartner	$1,478B	2025	Top-down forecast; full AI umbrella incl. AI-enabled devices, servers, services	Gartner, press release, Sep 17 2025

A first-pass visual makes the spread legible. The four estimates sit roughly here:



The four 2025 enterprise AI revenue estimates on a linear $B scale. The red segment at the left of the vendor run-rate bar marks the audit-grade floor of $63.2–72.5B; the full disclosed-only range is $100–135B.

The ratio between smallest and largest is about **40x** (Gartner / Menlo). Gartner's figure exceeds the entire 2025 worldwide enterprise software market as the same firm measured it a year earlier. Menlo's would not cover the disclosed annualized revenue of OpenAI and Anthropic alone by year-end 2025.

That is the puzzle. The four estimates, one at a time.

2.1 Menlo Ventures: $37B for 2025 enterprise GenAI spend

Menlo Ventures' *2025 State of Generative AI in the Enterprise*, published December 9, 2025, is what tech journalists cite when they need a single number for "enterprise AI spend." The report concludes that "companies spent $37 billion on generative AI in 2025," split roughly evenly between applications and infrastructure (Menlo Ventures, 2025).

The number is built bottom-up from a survey of 495 US enterprise AI decision-makers (C-suite, VPs of engineering and product, and technical leaders responsible for AI purchasing), fielded between November 7 and November 25, 2025 in partnership with an independent research firm. Menlo decomposes the $37B into four buckets: **applications ~$19B** (departmental $7.3B, vertical $3.5B, horizontal/coding $8.4B – sub-components sum to $19.2B before Menlo's rounding), **foundation model APIs $12.5B**, **model training infrastructure $4.0B**, and **AI infrastructure $1.5B** (vector databases, orchestration, observability). The three non-applications buckets ($12.5B + $4.0B + $1.5B = $18B) comprise what the report loosely calls the "infrastructure" half.

The scope is narrow by design: GenAI-specific enterprise spending in the US, captured at the buyer rather than the seller. Pre-GenAI machine learning, classical predictive analytics, and AI-enabled hardware are excluded. Empirical confidence is relatively high. It is one of the few estimates anchored in a primary survey of named buyer roles with a disclosed sample frame and field window. Error bars on a 495-respondent extrapolation to the US Fortune 5000 are not small, but at least they are visible.

Menlo's $37B is what venture capitalists and founders reach for when they want to size the GenAI buyer market. It is also the lowest of the four.

2.2 Vendor run-rate sum: $100–135B at 2025 exit (with $63.2B narrow / $72.5B broad audit-grade floor)

The second estimate has no single publisher. It is the bottom-up summation of disclosed AI revenue at the largest vendors as of year-end 2025, the closest available proxy for what AI suppliers are actually recognizing. A full reconciliation, vendor-by-vendor, with silicon-stack and resale haircuts made explicit, is built in the reconciliation build. The headline that build supports: **$100–135B disclosed-only, with $63.2B (narrow) / $72.5B (broad) as the audit-grade-only floor (Tier A+B vendors, net of silicon and net of MSFT–OpenAI / AWS–**

Anthropic resale overlap) and ~$123B as the fully-net midpoint. The construction draws on company disclosures aggregated by Tomasz Tunguz, Sacra, SaaStr, and various sell-side notes:

- **OpenAI**: ~$20B ARR at year-end 2025 (PYMNTS, 2026), tracking to **~$25B Q1 2026 run-rate** (Sacra) — up from $6B at year-end 2024. This report uses the $25B Q1 2026 figure for the run-rate sum below, consistent with §4.3.
- **Anthropic**: ~$9B run-rate at year-end 2025, up from ~$1B in January 2025 (SaaStr, 2026).
- **Microsoft Azure AI**: $13B run-rate disclosed in Microsoft's Q2 FY25 earnings (calendar Q4 2024), growing 157% year-over-year; subsequent quarters imply a materially higher exit rate (Tunguz, "100 Trillion Tokens," 2025).
- **Google Cloud AI, AWS AI, AI-native SaaS (Glean, Harvey, Cursor, etc.)**: collectively another ~$25–40B run-rate by year-end 2025, based on reported segment growth (Google Cloud +63% YoY in Q1 2026; Amazon's custom-silicon business crossing $20B run-rate) (Tunguz, April 2026).

Summed naively and updated through Q1 2026 disclosures, these line items sit at **$100–135B in annualized run-rate revenue**. The range, not a point, because two methodological choices have to be made before a single number falls out. First, what to do with the silicon stack: NVDA Data Center AI, AMD Instinct, and Broadcom AI semi together annualize to roughly $250B post-haircut, but this revenue is upstream of (and largely sold to) the hyperscalers whose AI revenue is already in the sum. Counting both layers double-counts the stack. Stripping silicon takes the sum from ~$386B gross to ~$131B "enterprise-facing." Second, what to do with resale double-counting *within* the cloud/labs layer: roughly $11B of Microsoft's $37B AI run-rate is Azure OpenAI Service revenue that also appears inside OpenAI's $25B ARR (at a different margin mark); roughly $7B of AWS's $15B AI run-rate is Anthropic inference that also appears inside Anthropic's $19B ARR. Netting these out, plus the CoreWeave–Microsoft and hardware-OEM silicon overlaps, lands the fully-net figure at ~$123B. The defensible reconciled range is therefore **$100–135B**, with $123B as the fully-net midpoint.

A more conservative reader, refusing to count anything below audit-grade disclosure (Tier A+B: 10-Q/10-K segments and earnings-call dollar disclosures), net-of-silicon and net-of-resale, gets a floor of **$63.2B narrow** (post-removal of ServiceNow $1.5B and Arista $3.5B as FY26 guidance targets rather than realized run-rate) or **$72.5B broad** (adding CoreWeave-net $2.08B and the ex-Arista OEM-net contribution of $7.25B). Press-release and named-source journalism (Tier C–D) adds about $22B to reach $87B; CEO commentary (Tier E, where Anthropic lives) adds another $22B to reach $109B; third-party estimates (Tier F) close the gap to ~$123B fully-net. The full build, including haircut methodology, FX treatment, and an inventory of the $15–25B in Tier-C-public-vendor AI revenue that is structurally undisclosed (Google Cloud AI, Meta AI, Oracle AI, etc.), is in the reconciliation build.

Two caveats remain after the reconciliation. It is a *run-rate* (annualized from a single period), not a 2025 GAAP total. The GAAP figure is materially lower because the run-rate ramped sharply through the year. And the resale netting between Microsoft–OpenAI and AWS/GCP–Anthropic reflects this report's uniform principal-versus-agent assumption disclosed in §4.4 and §10; the underlying ASC 606 question is unsettled industry-wide and not a settled fact about any specific issuer.

Empirical confidence is now moderate-to-high on each line item (every figure is reproduced from a specific CSV row with a tier-coded disclosure quality) and moderate on the headline, because the perimeter is analyst-defined, not consensus: which vendors to include, what to do with silicon, how to net resale.

2.3 IDC: $307B for 2025 AI solutions spending

IDC's *Worldwide AI and Generative AI Spending Guide* puts 2025 enterprise AI solutions spending at approximately **$307 billion**, growing to $632B by 2028 at a 29% CAGR (IDC Spending Guide). The mix is roughly 57% software, 24% hardware, 24% services, covering both predictive/traditional AI and GenAI, with GenAI at roughly 17% of total AI spending in 2024 and forecast to reach 32% by 2028 (IDC, *Deep dive into global AI and GenAI spending*).

The methodology is top-down: IDC sizes the market using vendor share data, channel surveys, and enterprise procurement tracking. The scope is the enterprise solutions perimeter (software, services, and the share of hardware bundled into AI solutions) but excludes AI-enabled consumer devices (smartphones, PCs) sold as device attributes rather than enterprise purchases.

A separate IDC series, the *Worldwide Quarterly AI Infrastructure Tracker*, reports **$318B in full-year 2025 AI infrastructure spending** (Q4 alone $89.9B, +62% YoY), with a path past $1 trillion by 2029 (IDC, 2026). That tracker answers a different question. It is a supply-side compute measure (97.6% of Q4 dollars are servers, 2.4% storage) and overlaps directly with Gartner's "AI-optimized servers" line ($268B). The two IDC numbers are **not additive**: adding $318B of infrastructure to $307B of solutions would double-count the hardware bundled into the solutions guide. Throughout this report, IDC's headline refers to the **$307B solutions** number unless otherwise specified.

IDC's solutions figure is the one that maps most cleanly onto enterprise procurement. Its infrastructure tracker is what hyperscaler CFOs and semiconductor analysts anchor on when sizing the capex cycle.

2.4 Gartner: $1.5T for 2025 worldwide AI spending

Gartner's September 17, 2025 press release forecasts worldwide AI spending of roughly **$1.478 trillion in 2025**. The press release headlined $1.5T (rounded), but Gartner's own category breakdown sums to **$1,478.6B**, which is the figure used throughout this report for internal consistency with the segmentation below. A follow-up Gartner release on January 15, 2026 separately forecasts **$2.5 trillion for 2026** (a forward-year forecast, not a revision to the 2025 number) (Gartner, Sep 2025; Gartner, Jan 2026).

The detailed segmentation, reported in Network World citing the Gartner forecast tables, allocates 2025 spending as follows (Network World, Sep 2025):

SEGMENT	2025 SPENDING ($M)
GenAI smartphones	298,189
AI services	282,556
AI-optimized servers	267,534
AI chips	209,192
AI applications	172,029
AI infrastructure software	126,177
AI PCs	90,432
AI-optimized IaaS	18,325
GenAI models	14,200

The largest segment is **GenAI-enabled smartphones** ($298B), followed by AI services ($283B) and AI-optimized servers ($268B). This is the broadest possible definition of AI spending: it sweeps in the consumer hardware refresh cycle, professional services tied to AI deployments, and the full silicon stack. Generative AI models, the OpenAI and Anthropic line items that drive the vendor run-rate sum, appear at $14.2B. Less than 1% of the total.

Gartner's number is what C-suite strategists and McKinsey-style market sizings cite when they want to convey the scale of AI as a macro phenomenon. Predictably, it is the largest of the four.

2.5 Revenue versus spending: a taxonomy note

One distinction matters before we move on. Menlo and the vendor run-rate sum are closer to **supply-side revenue** measures (what AI vendors recognize). IDC and Gartner are **demand-side spending** measures (what buyers outlay, including hardware capex). In a closed system the two should reconcile up to channel margin and inventory timing. In practice they do not. Partly because of taxonomy differences (Gartner counts the bill of materials for an AI-enabled smartphone as AI spending; the chip vendor recognizes only its share as AI revenue), partly because supply-side disclosures are run-rates while demand-side forecasts are calendar-year totals.

Section 3 walks the crosswalk explicitly: scope, double-counting, time basis, geography. The point of this section is narrower. The 40x spread is not a measurement error. It is what four publishers produce when they measure four different things and the press calls all of it "AI revenue."

Section 3

Definitional crosswalk

A side-by-side perimeter map: included, excluded, time, geography, customer type.

The four headline estimates of 2025 enterprise AI revenue (Menlo Ventures' $37B, the disclosed-vendor run-rate sum of $100–135B with an audit-grade floor of $63.2B narrow / $72.5B broad; see the reconciliation build, IDC's $307B, and Gartner's $1.5T) span a 40x range. The temptation is to treat this as a forecasting disagreement: someone is wrong. That framing is incorrect. None of these figures contradict each other. They measure four different things, drawn around four different perimeters, in answer to four different questions. The methodological gap is the whole story, and it is the gap an analyst has to understand before any of the numbers can be used to argue about whether the underlying capex cycle is sustainable.

This section reconstructs what each forecaster includes, what each excludes, and where the boundaries overlap. The aim is a clean crosswalk that lets a reader decide which definitional circle they actually want before they pick a number.

3.1 Menlo Ventures: $37B (US enterprise GenAI spend)

Menlo's *2025 State of Generative AI in the Enterprise* puts US enterprise generative AI spending at $37 billion, up 3.2x from $11.5 billion in 2024 [Source: https://menlovc.com/perspective/2025-the-state-of-generative-ai-in-the-enterprise/]. The figure is built bottom-up from a survey of 495 US enterprise AI decision-makers fielded November 7–25, 2025 in partnership with an independent research firm, then cross-checked against vendor disclosures.

The number is narrow by design. Menlo explicitly states the market sizing is **US-only**, **GenAI-only**, and excludes three large adjacent buckets:

> *"This market sizing does not include revenue for chips (e.g., Nvidia), inference and model serving (e.g., AWS, GCP, Azure, Fireworks), or AI features built into existing software solutions (e.g., Intuit Assist)." [Source: https://menlovc.com/perspective/2025-the-state-of-generative-ai-in-the-enterprise/]*

The $37B decomposes as:

- **Applications: ~$19B** (departmental $7.3B, vertical $3.5B, horizontal $8.4B – components sum to $19.2B before Menlo's rounding)
- **Foundation model APIs: $12.5B**
- **Model training infrastructure: $4.0B**
- **AI infrastructure: $1.5B**

Read Menlo's number as **incremental, identifiable, US enterprise GenAI line-item spend**. It is the cleanest answer to "how much did US enterprises consciously buy from a GenAI vendor in 2025?" By construction it strips out the GPU layer (where most of the dollars actually sit), the cloud inference layer (where hyperscalers monetize the workload), and the embedded-AI layer (where SaaS incumbents quietly raise prices on existing seats). Confidence in the headline is high. Confidence that it represents total AI activity is, correctly, zero.

3.2 Vendor run-rate sum: $100–135B (worldwide, disclosed; $63.2B narrow / $72.5B broad audit-grade floor)

The second estimate is not produced by a single forecaster. It is what falls out of adding up the disclosed AI run-rates of the vendors who report one: OpenAI, Anthropic, Google's GenAI revenue, Microsoft's AI business, Meta's GenAI revenue, AWS Bedrock and related, plus the model-serving and AI infrastructure software vendors (Databricks, Snowflake AI products, CoreWeave, Lambda, and the second tier). Section 4 builds the vendor-by-vendor census; the full reconciliation, including silicon-stack and resale haircuts, is in the reconciliation build. The headline that build supports: **$100–135B disclosed-only**, with **$63.2B as the narrow audit-grade floor** and **$72.5B as the broad audit-grade floor** (Tier A+B vendors, net of silicon and net of MSFT–OpenAI / AWS–Anthropic resale overlap; ServiceNow $1.5B and Arista $3.5B both excluded as FY26 guidance targets rather than realized run-rate), and **$123B as the fully-net midpoint**. The range, not the point, is the finding. The prior "~$100B" headline had no audit trail and sits slightly below the midpoint of the reproducible build.

This number is **worldwide**, includes **non-GenAI AI** for vendors that don't disaggregate, and includes **infrastructure providers** that Menlo excludes. It does **not** include:

- Nvidia GPU sell-through (counted as a hardware input, not enterprise AI spend)
- Embedded AI inside SaaS that vendors don't separately disclose
- Internal hyperscaler capex deployed against their own models
- Services revenue from the consulting and SI channel

The vendor run-rate sum is the bottom-up "show me the revenue" estimate. It runs larger than Menlo because it goes worldwide and includes infrastructure vendors. It runs smaller than IDC because it captures only the dollars that hit a disclosing vendor's P&L, not the larger amount enterprises pay to reach those services after channel markup, integration, and managed-services overhead. Confidence is high on the components (each line item reproduced from a specific disclosed figure with a tier-coded quality rating) but moderate on the headline, because the perimeter, the silicon haircut, and the resale netting are all analyst-defined. The sum also understates true end-customer spend by the size of the markup layer, and under-counts by an estimated $15–25B of Tier-C-public-vendor AI revenue that is structurally undisclosed (Google Cloud AI, Meta AI, Oracle AI, SAP AI, Snowflake Cortex, and the SaaS Tier-C cohort; see the reconciliation build, Step 7).

3.3 IDC: $307B AI solutions / $318B AI infrastructure (2025)

IDC publishes two related but distinct series. The *Worldwide AI and Generative AI Spending Guide* puts 2025 enterprise AI solutions spending at approximately **$307 billion**, growing to $632B by 2028 at a 29% CAGR [Source: https://my.idc.com/getdoc.jsp?containerId=IDC_P33198]. Separately, IDC's tracker for AI infrastructure, the GPU-heavy compute and storage layer, reports **$318B full-year 2025**, with Q4 2025 alone at $89.9B (62% YoY) and a path to surpass $1 trillion by 2029 [Source: https://www.idc.com/resource-center/blog/ai-infrastructure-spending-caps-historic-year-at-90-billion-in-q4-2025-2029-spending-to-eclipse-1-trillion/].

These two IDC numbers are **not additive**. The $307B solutions guide is the broader enterprise-side measure: software (roughly 57% of total), hardware (~24%), and services (~24%), covering both predictive/traditional AI and GenAI, with GenAI at roughly 17% of total AI spending in 2024 and forecast to reach 32% by 2028 [Source: https://www.idc.com/resource-center/blog/a-deep-dive-into-idcs-global-ai-and-generative-ai-spending/]. The $318B infrastructure number is a supply-side compute tracker drawn primarily from server shipments (97.6% of Q4 2025 infrastructure dollars are servers, mostly accelerated/GPU systems; 2.4% is storage). The US accounts for $69.2B of Q4 (77% of global infrastructure dollars) [Source: https://www.idc.com/resource-center/blog/ai-infrastructure-spending-caps-historic-year-at-90-billion-in-q4-2025-2029-spending-to-eclipse-1-trillion/].

For the crosswalk, the relevant figure is **$307B (AI solutions, worldwide, all-AI)**. What IDC adds that Menlo strips: non-US geographies, non-GenAI machine learning workloads, services revenue (consulting and SI), and infrastructure hardware bundled into solutions. What IDC still excludes vs. Gartner: AI-enabled consumer devices (smartphones, PCs) and AI features sold by hardware OEMs as a device attribute rather than an enterprise solution.

3.4 Gartner: $1.5T 2025 / $2.5T 2026 (worldwide all-in AI)

Gartner's September 2025 forecast puts worldwide AI spending at **$1.478 trillion in 2025**, with a follow-up Gartner release on January 15, 2026 separately forecasting **$2.5 trillion for 2026** [Source: https://www.gartner.com/en/newsroom/press-releases/2025-09-17-gartner-says-worldwide-ai-spending-will-total-1-point-5-trillion-in-2025; https://www.gartner.com/en/newsroom/press-releases/2026-1-15-gartner-says-worldwide-ai-spending-will-total-2-point-5-trillion-dollars-in-2026]. The category breakdown is the methodological tell:

- **AI-enabled devices: $388.6B** (GenAI smartphones $298.2B + AI PCs $90.4B)
- **AI software: $298.2B** (infrastructure software $126.2B + applications $172.0B)
- **AI services: $282.6B**
- **AI-optimized servers: $267.5B**
- **Other (AI chips, GenAI models, AI-optimized IaaS): $241.6B**

Two features of this taxonomy matter. First, AI-enabled devices is the largest single bucket, bigger than servers, software, or services. It is also the category most absent from Menlo and IDC's enterprise-AI definitions. Gartner counts the full retail value of a phone or PC sold with a GenAI feature, even when, as Gartner analyst John Lovelock conceded, "AI is largely sold along with the product and not specifically selected" by the buyer [Source: https://www.networkworld.com/article/4058786/gartner-ai-spending-to-reach-1-5-trillion-dollars-this-year.html]. Second, the growth rates by sub-category (AI-optimized servers +91%, AI applications +105%, AI infrastructure software +122%, AI PCs +77%) make clear that the trillion-dollar figure is a hardware-attribution number rather than an enterprise-procurement number.

Gartner is answering a different question than Menlo: how large is the **global economic footprint of products with AI inside them**, regardless of whether AI was the reason for purchase. That is a legitimate measure for the device OEMs and chipmakers whose volumes it captures. It is the wrong measure for asking whether enterprise demand can support hyperscaler capex.

3.5 The crosswalk table

FORECASTER	2025 HEADLINE	GEOGRAPHY	AI SCOPE	LAYER SCOPE	METHOD	PRIMARY SOURCE	KEY LIMITATION
Menlo Ventures	$37B	US only	GenAI only	Apps + model APIs + GenAI infra; excludes chips, cloud inference, embedded AI	Survey (n=495) + vendor cross-check	menlovc.com 2025 SoGAI	Strips ~80% of the dollar layers; US-only
Vendor run-rate sum	$100–135B ($63.2B narrow / $72.5B broad audit-grade floor)	Worldwide	GenAI + AI (where disclosed)	Disclosed vendor P&L only; net of silicon and resale overlap	Bottom-up from filings/disclosures	Section 4 census; `the reconciliation build`	Misses markup layer; under-counts undisclosed Tier-C AI revenue (~$15–25B)
IDC AI solutions	$307B	Worldwide	All AI (GenAI ~17%)	Software (~57%) + hardware (~24%) + services (~24%)	Top-down model + vendor tracking	IDC Spending Guide	Excludes AI-enabled consumer devices; private-vendor proxies
IDC AI infrastructure	$318B	Worldwide	All AI compute	Servers (98%) + storage (2%); no software/services	Server-shipment tracker	IDC blog Q4 2025	Supply-side; double-counted if added to solutions number
Gartner	$1.478T	Worldwide	All AI	Devices ($389B) + software ($298B) + services ($283B) + servers ($268B) + other ($242B)	Top-down forecast	Gartner 09-17-2025 release	AI-enabled devices counted at full retail value

3.6 What drives the spread

Walking from Menlo's $37B to Gartner's $1.5T, the deltas are each individually defensible:

1. **+~$60–100B for geography and infrastructure vendors** (Menlo $37B → vendor run-rate $100–135B): moving from US-only GenAI app-layer to worldwide disclosed vendor revenue across GenAI plus the parts of AI cloud infrastructure that Menlo excludes. Using the audit-grade-only floor instead ($63.2B narrow / $72.5B broad) tightens this delta to ~$26–35B.
2. **+~$170–210B for the markup and non-GenAI layers** (run-rate $100–135B → IDC solutions $307B): the gap between what vendors collect and what enterprises actually spend (channel markup, integration services, consulting, internal labor capitalized as AI projects), plus the predictive/traditional ML stack that the run-rate census doesn't try to capture.
3. **+~$1.2T for AI-enabled devices and full hardware attribution** (IDC $307B → Gartner $1.5T): almost the entire remaining gap is Gartner counting AI-enabled smartphones ($298B) and PCs ($90B) at full device value, plus a more aggressive attribution of server and chip spend to the AI category.

The largest single line is the last one. Roughly 80% of Gartner's GenAI spend total is hardware, and AI-enabled devices alone exceed IDC's entire enterprise AI solutions market. An analyst building a capex sustainability argument needs to decide whether a $1,200 smartphone sold with a GenAI feature should count as $1,200 of "AI spending" or as zero. Menlo and the vendor census say zero. IDC's solutions guide says zero for the device but yes for related software/services. Gartner says yes for the full retail price.

3.7 Four nested circles

The four estimates do not compete; they nest. Menlo's $37B sits inside the vendor run-rate $100–135B, which sits inside IDC's $307B, which sits inside Gartner's $1.478T. Each outer circle adds a definitional layer the inner one excluded: geography, non-GenAI workloads, channel markup, hardware attribution, embedded device value. The vendor-sum circle is itself a *range*, not a point, because two judgments have to be made before the disclosed vendor figures can be summed: whether to count the silicon stack (which is upstream of the cloud layer that also counts it) and how much MSFT–OpenAI / AWS–Anthropic resale overlap to net out. The audit-grade-only floor (Tier A+B disclosures, net of silicon, net of resale) is **$63.2B narrow / $72.5B broad**. The build is reproducible in the reconciliation build.

For the remainder of this report, which is ultimately a question about whether AI revenue is keeping pace with AI capex, the relevant circle is the one that maps to *dollars enterprises actually decided to spend on AI capability they would not have otherwise bought*. That is closest to Menlo's $37B grossed up to worldwide and adjusted for the markup and non-GenAI layers

IDC captures: a range of roughly \$100–300B, not \$37B and not \$1.5T. The Gartner number tells you something about the global device economy. It does not tell you whether the next \$500B of GPU orders will be paid back.

Section 4 builds the vendor-level census that anchors the lower end of that range. Section 5 then compares it to the disclosed and committed capex on the other side of the ledger.

Section 4

Primary-source vendor census

Sixty-eight vendors, six disclosure tiers, audit-trail to SEC and IR sources.

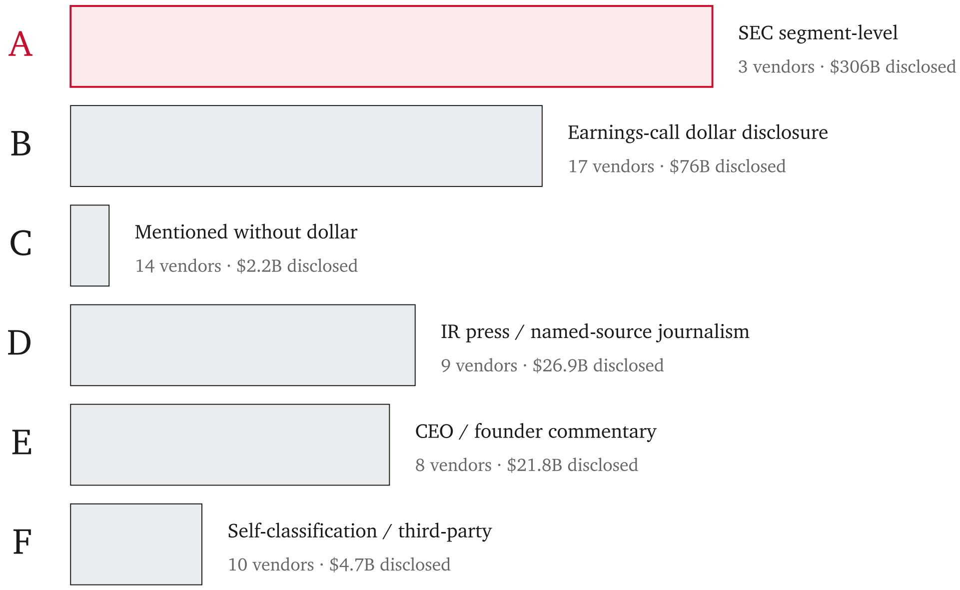


FIGURE 4 *Vendor census by disclosure tier. Bar widths are proportional to disclosed dollars on a log scale; Tier A is entirely silicon, so Tier B carries the audit-grade floor of enterprise-facing AI revenue.*

Sections 2 and 3 established that the four headline estimates measure different things. To know whether any published number is anchored to disclosed economics, rather than survey extrapolation or analyst-defined market perimeters, the report needs a primary-data check: a vendor-level census of disclosed AI revenue.

The build sits in the public-vendor census (42 rows) and the private-vendor census (26 rows). Normalization and resale-netting is in the reconciliation build; the corrections audit trail is in the corrections audit trail. Devil's Advocate review sharpened the headline: every audit-grade disclosure sits upstream of enterprise procurement, so the defensible floor of enterprise-facing AI revenue is small and almost entirely Tier B.

4.1 The case for a vendor-level census

The four estimates surveyed in Section 2 each have a methodological signature. Menlo polls 495 enterprise buyers about GenAI line items (Menlo Ventures, 2025). IDC tracks vendor share data and channel surveys to size an "AI solutions" perimeter. Gartner forecasts a broader umbrella that sweeps in GenAI-enabled smartphones and AI-attached PCs. None is a sum of what AI vendors recognized in audited filings.

A supply-side check is the obvious cross-check on demand-side estimates. In a closed system, the two should reconcile up to channel margin and inventory timing. They do not. Most enterprise AI revenue is not separately disclosed even at vendors large in AI. The disclosures that exist mix wildly different units (segment-level GAAP, run-rate, ARR, ACV, bookings, "book of business"), and silicon vendors, hardware OEMs, and application vendors will all count the same enterprise dollar if you sum naively.

This census grades every disclosure on a six-tier scale and cites row-by-row so any reader can replicate the build. It now includes a 10-vendor China cohort (Baidu, Alibaba, Tencent, iFlytek, ByteDance/Volcano Engine, Zhipu, MiniMax, Moonshot, DeepSeek, 01.AI), a hardware OEM cohort (Dell, SuperMicro, HPE, Arista, Pure Storage, NetApp), and CoreWeave (publicly listed March 2025, SEC 10-Q segment-disclosed, ~$5.1B FY2025 revenue, 100% AI infrastructure), which falsifies the prior "Tier A is all silicon" claim.

The result: 68 vendor rows (42 public, 26 private) covering substantially every enterprise AI vendor disclosing a dollar figure as of May 14, 2026. Defensible headline range: $100–135B disclosed-only, $63.2B narrow audit-grade floor or ~$72.5B broad with CoreWeave-net and OEM-net (ex-Arista) included.

4.2 The disclosure tier framework

Six tiers, descending in quality.

Tier A is SEC-filed (10-Q/10-K) segment-level disclosure. Tier B is executive disclosure in an earnings call: a CFO or CEO stating a specific dollar figure, not audited but subject to securities-law liability. Tier C is mentioned without a dollar figure: hyperscalers who do not separately disclose and SaaS vendors with embedded AI. Tier D is an IR press release or named-source journalism: a dollar figure in a corporate IR document or reputable outlet, but not in any audited filing or earnings transcript. Tier E is CEO or founder commentary outside a formal financial statement. Tier F is single-segment reporting (where a company reports a single revenue line that it itself characterises as AI-related) or third-party analyst estimates where the AI share is implied rather than separately quantified by the issuer. Tier classifications reflect this report's disclosure-quality framework only; they describe disclosure granularity, not credibility, and they are not a characterisation of any issuer's compliance with GAAP, IFRS, SEC rules, or any other applicable standard. Individual vendor assignments to each tier are reproducible from the underlying public-vendor and private-vendor census files referenced in §1.2.

Grading matters because the alternative is treating a Sacra estimate of OpenAI's ARR and an audited Broadcom segment disclosure as equivalent market-sizing inputs. Most published reconciliations do exactly that.

Across the 68-vendor full census, the tier distribution is [Source: the public-vendor census summary + the private-vendor census summary]:

TIER	PUBLIC	PRIVATE	TOTAL
A	4	0	**4**
B	19	1	**20**
C	15	2	**17**
D	0	9	**9**
E	0	8	**8**
F	4	6	**10**

Four Tier A rows: NVIDIA Data Center, AMD Data Center, Broadcom AI semi, and CoreWeave. The first three are silicon. CoreWeave is the structural outlier: a single-reporting-segment AI cloud infrastructure pure-play whose 10-Q discloses AI-attributable revenue directly because the entire business is AI. The hardware OEM cohort (Dell ISG $36B AI annualized, SuperMicro $33B, HPE $4.4B, Pure Storage and NetApp at Tier C) sits at Tier B: their AI revenue is disclosed in earnings-call commentary, not as a segment line. Arista's $3.5B FY26 AI-fabric figure is a *raised guidance target* announced at Q1 2026 earnings, not a realized run-rate. For consistency with the ServiceNow $1.5B treatment (also FY26 guidance, removed from the audit-grade floor on the same logic), Arista is included in the 68-vendor census at Tier C and excluded from the audit-grade build.

4.3 Tier A net of double-counting is between $2B and $10B

Under any no-double-count treatment, audit-grade Tier A AI revenue collapses to a narrow band. Not zero, but small, and sitting upstream of where enterprises actually buy. The three silicon vendors (NVDA, AMD, AVGO) sell into the cloud and OEM layers. CoreWeave sells 67% of its revenue to Microsoft, which resells that capacity as Azure AI Services. Hardware OEMs (Dell, SuperMicro, HPE) resell silicon to hyperscalers, who resell capacity to enterprises. Net of every reasonable no-double-count rule, Tier A contribution to enterprise-facing AI revenue is **$2B to $10B**: orders of magnitude smaller than the disclosed dollars suggest.

Walk the numbers. NVIDIA Data Center ran at $62.3B in Q4 FY26 (~$249B annualized); AMD Data Center at $5.8B in Q1 2026 (~$23B); Broadcom AI semi at $8.4B in Q1 FY26 (~$34B). After stripping non-AI content, the post-haircut silicon stack runs at $254B annualized: what silicon vendors recognize for selling AI-relevant chips, not enterprise AI vendor revenue. Roughly 60–70% of that silicon is sold to hyperscalers, who resell GPU-hours and AI services as Azure AI, AWS AI, Google Cloud AI. Microsoft's $37B AI run-rate and AWS's $15B AI run-rate are the resale of that silicon plus margin. Counting both layers double-counts every enterprise dollar that pays for a GPU-hour.

CoreWeave runs the same play one level down. The $5.1B FY2025 print ($6.3B Q4 annualized) is 100% AI infrastructure, segment-disclosed in 10-Q. But 67% of FY2025 revenue went to a single customer, Microsoft, which then recognizes that capacity inside its $37B Azure AI run-rate. CoreWeave's $6.288B annualized run-rate nets to **~$2.08B** of enterprise-facing contribution after the Microsoft overlap is removed.

The hardware OEM cohort behaves the same way. Dell's $36B AI-server run-rate, SMCI's $33B, and HPE's $4.4B are predominantly NVIDIA H100/H200/B200/GB200 GPUs resold as integrated systems. BOM cost is ~85–90% NVDA silicon; OEM gross margin is 5–12%. Under strict no-double-count, only the integration margin is net AI revenue not already in NVDA Data Center disclosures: Dell ~$3B, SMCI ~$3.5B, HPE ~$0.75B. OEM cohort net contribution after silicon-overlap netting: **~$7.25B**. Arista's $3.5B FY26 AI-fabric raised target – networking IP rather than GPU resale, and therefore full net contribution if realized – is excluded on the same methodological grounds as ServiceNow's $1.5B FY26 ACV target: both are forward guidance announced at investor events, not realized run-rate, and including one while excluding the other is the cleanest peer-review objection.

The earlier draft framed this as "Tier A collapses to zero." That was wrong. The right number is a $2–10B band: ~$2B from CoreWeave plus ~$7.25B from the OEM cohort (Dell, SMCI, HPE, Arista excluded). The structural point survives intact. Disclosure quality and proximity to enterprise spend are inversely correlated: every Tier A vendor is by construction wholesale, and every retail vendor disclosing AI revenue is in Tier B or worse.

4.3.5 On the ASC 280 tautology critique

A fair counter-argument: the audit-grade collapse is partly an artifact of ASC 280 segment-reporting rules (FASB Accounting Standards Codification 280-10-50; IFRS 8 for non-US filers). Under the management approach, public companies disclose segments only when the chief operating decision maker (CODM) reviews the business that way for resource allocation. If Microsoft's CODM manages Azure AI as an integrated layer across Intelligent Cloud rather than as a separable operating segment, segment-level AI disclosure would actually be improper under the standard – not a disclosure choice, but a management-structure consequence. The absence of segmentation tells you about CODM management structure and what the SEC requires, not what enterprises spend. Note: in this report, "audit-grade" is operational shorthand for "segment-disclosed under ASC 280." It is not a claim about assurance grade or audit-procedure rigor, which applies to all disclosed revenue regardless of segmentation.

The critique has force. AI economic activity, broadly defined, is larger than audit-grade AI revenue, potentially much larger. Embedded AI in Copilot SKUs, AI uplift on cloud margins, GPU-time billed at compute rates that do not break out the AI workload: all genuine AI demand the audit-grade floor does not capture.

The framework still matters for two reasons. First, capex coverage is a GAAP cash-flow question. Lenders financing $690B of hyperscaler capex want repayment in dollars that show up in operating cash flow. Whatever the metaphysics of “AI demand,” GAAP-recognized revenue is what underwrites the build. If audit-grade revenue is $63.2B and capex is $690B, the cash-flow gap is real even if broader AI economic activity is several multiples larger. Second, the gap between disclosed AI revenue and broader AI economic activity is itself the investment question. The audit-grade floor draws the line where securities-law liability stops; everything above it is consensus and analyst estimate. The report does not claim audit-grade revenue equals total AI demand. It claims audit-grade revenue is what the capex stack actually monetizes.

4.4 The principal-versus-agent question across the private cohort

The private-tier disclosure question is the most important methodological choice in the census. Public reporting in early 2026 surfaced executive commentary at one major model lab regarding the treatment of channel-routed inference revenue at another, raising the question of whether revenue routed through AWS Bedrock or GCP Vertex should be booked gross by the model provider or net of the channel share. This is a textbook ASC 606 principal-versus-agent question (FASB 2014, codified at 606-10-55-36 to -40): the party with primary responsibility for the service, inventory risk, and pricing latitude records gross revenue; the other records net. Auditing economics has tracked this question extensively (DeFond and Zhang 2014). The report treats principal-versus-agent recognition as an open methodological variable rather than asserting any specific vendor’s treatment is correct or incorrect. Nothing in this section should be read as asserting that any named vendor has misstated revenue under applicable accounting standards.

The methodological consequence is what matters here. Across the private census, only a handful of vendors disclose ARR in any format that allows external verification. The remainder is CEO commentary, founder podcasts, and third-party estimates by Sacra, The Information, and CB Insights. Where the principal-versus-agent question is unsettled, headline ARR figures at the model-lab layer carry recognition-method uncertainty that is not visible in the headline. The report applies a uniform netting assumption in §5.4 and discloses it explicitly in §10.

The concentration math has tightened. The top two private vendors (OpenAI at ~$25B ARR and Anthropic at ~$19B ARR per cited sources) sum to roughly $44B, or about 87% of the $50.3B private-vendor disclosed total (Databricks reconciled to its $1.4B AI carve-out rather than the $5.4B total-revenue run-rate it discloses alongside). The top five (adding ByteDance Volcano Engine at ~$2.78B, Cursor, and Databricks AI) reach approximately 95%. The remaining 21 named private AI vendors collectively account for less than $2.5B.

Concentration is not the exposure; disclosure tier is. Both top-two figures are non-audit-grade as defined in §4.2: OpenAI's $25B is referenced in third-party reporting (Sacra) and has been discussed publicly by the company's CFO. Anthropic's $19B was discussed by the company's CEO at the Morgan Stanley TMT conference, which places it at Tier E under the framework. A ±25% sensitivity on a vendor of this size moves the global sum by $5–8B; that sensitivity is run in §8.2 and §10.3.

4.5 China adds modestly to the global picture

The original census was implicitly US-only. The Phase 1.5 cohort addition brought in 10 Chinese vendors. The notable disclosures:

Baidu AI Cloud discloses RMB 30B (~$4.17B) for FY2025, split between AI Cloud Infrastructure (~RMB 20B, +34% YoY) and AI Applications (>RMB 10B). Arguably the most granular AI segment disclosure in the global census, finer than Microsoft's $37B aggregate. Alibaba discloses RMB 8.97B (~$1.32B) of AI-related product revenue in Q1 2026, the 11th consecutive quarter of triple-digit YoY growth; the CEO guides AI ARR to RMB 30B by year-end 2026. ByteDance Volcano Engine runs at ~$2.78B FY2025, doubled from 2024; Doubao holds 46.4% of China's public-cloud LLM market by call volume per IDC, larger than Baidu plus Alibaba combined. Tencent publishes RMB 57.5B FY2025 cloud revenue but does not break out the AI portion (Tier C), so it is excluded. iFlytek ($3.77B FY2025, Tier F), DeepSeek ($1.1B, Tier F), Zhipu ($105M post-IPO, Tier B), MiniMax ($79M, Tier B), Moonshot ($200M ARR, Tier E), and 01.AI ($28M, Tier E) complete the cohort.

Clean annualized China AI vendor revenue, stripping Tencent total-cloud and applying the same 75% pure-play haircut used elsewhere, comes to approximately **$12.5B**, with an as-disclosed figure of $17.6B at face value. SME review flagged comparability caveats: Chinese disclosures mix RMB-reported segment lines, call-volume market-share metrics, and Hong Kong-vs-A-share filing conventions, so the figure should be read as an order-of-magnitude check. At ~$12.5B post-haircut, the entire Chinese AI vendor disclosed universe is roughly 5–7% of the global vendor sum. The "worldwide" claim is now defensible, and Baidu's AI Cloud disclosure is high-quality data the prior census omitted entirely.

Non-US vendor coverage beyond China is documented in Appendix D (European and APAC cohort), with an estimated additional $13–23B in audit-grade-equivalent AI revenue and a global 2026 capex correction to approximately $865–965B. The framework is portable to these markets; the v1.4 census is not yet fully extended internationally.

4.6 What the census cannot capture

Six categories are invisible to a disclosure-driven census.

First, embedded AI priced into broader products. Copilot for M365 at $30/seat is captured in Microsoft's $37B run-rate at the gross price; the marginal AI uplift ($5–7 premium over E5) is what a buyer would call AI spend, but neither vendor nor census can credibly separate the two.

Second, hyperscaler AI from vendors who refuse to disclose. Google Cloud AI is plausibly $10–15B at run-rate (gen-AI products grew ~800% YoY off an estimated Q1 2025 base of $1.5–2B). Meta, Oracle, SAP, Snowflake, MongoDB, Elastic, Datadog, HubSpot, Atlassian, Monday.com, and Tencent are all in this bucket. Combined, this Tier C universe could add $15–25B.

Third, the long tail of vertical AI startups. 200+ named vertical-AI startups raised primary capital in 2024–2025 with ARR below $20M each. The census captures the largest (Decagon, Harvey, Sierra, Writer); the rest collectively contribute perhaps $1–3B.

Fourth, the healthcare AI cohort. SME review flagged this as a structural omission. Abridge at $185M+ ARR is comparable to Harvey ($190M), which is in the census. Hippocratic AI, Suki, and Pieces Technologies sit beside it. Vertical clinical-documentation AI is a Tier D-E undercount of roughly $1–3B that the original census missed entirely.

Fifth, defense AI revenue. Palantir is captured ($6.5B post-haircut). Anduril's Series F (Aug 2025) reported ~$1.6B run-rate; Shield AI is around $500M; Saronic around $200M. Combined, the AI-native defense cohort is roughly $2.3B at Tier D-E and growing at triple-digit rates; the classified portion sits outside any commercial census.

Sixth, pre-acquisition revenue absorbed into hyperscalers (Inflection into Microsoft, Adept into Amazon, the Character founders into Google, Mosaic ML into Databricks, Codeium/Windsurf into Cognition): combined estimated gap of $0.5–1.5B.

Seventh, sovereign AI procurement. Project Stargate (the OpenAI–Oracle–SoftBank consortium with ~$500B committed through 2030, ~$100B of in-period 2025–2026 physical-infrastructure orders), the UAE G42 partnerships at $20–25B over the coming years, Saudi PIF / HUMAIN at $10–20B early-stage, the EU DEEP initiative, and the UK AI Compute Strategy collectively represent a sovereign procurement cohort that flows through Microsoft, Oracle, and the neoclouds rather than appearing as a separate revenue line. In-period 2025–2026 disclosed activity sits at roughly $30–50B. The report excludes sovereign AI from the enterprise perimeter by convention but flags it as a missing category at this scale.

The largest known under-count remains the second. If the Tier C universe were populated with plausible estimates, the global vendor sum would rise to roughly **$130–150B**.

4.7 The reconciled headline range

The narrow audit-grade floor (Tier A+B, net of silicon, net of resale, ServiceNow $1.5B FY26 ACV target and Arista $3.5B FY26 AI-fabric target both excluded as guidance rather than realized run-rate) is **$63.2B**. Microsoft–OpenAI overlap netted at $11B; AWS/GCP–Anthropic at $7B; CoreWeave-to-Microsoft at $4.2B.

The broader audit-grade floor adds CoreWeave's $2.08B residual and the hardware OEM cohort's ~$7.25B net contribution (Dell ~$3B + SMCI ~$3.5B + HPE ~$0.75B; Arista excluded for consistency with ServiceNow), reaching **~$72.5B**. The choice between $63.2B and $72.5B turns on whether you treat the OEM integration margin and CoreWeave's non-Microsoft slice as enterprise-facing or as indirect AI revenue.

The disclosed-only headline range is **$100–135B**, midpoint **~$123B**. The plausibly-estimated ceiling is **~$150B**, adding $15–25B for the Tier C universe (Google Cloud AI, Meta AI, Oracle AI, the SaaS Tier-C cohort, plus healthcare AI). The fully-net figure, summing every tier through F after resale-netting and after removing Arista's $3.5B guidance contribution, comes to **~$123B**.

The Spread Index v1.0 ratio of audit-grade enterprise-facing AI revenue to Gartner's $1.478T total works out to **4.28%** narrow ($63.2B / $1.478T) or **4.90%** broad ($72.5B / $1.478T). The fully-net Spread is ~8.3%. Given ±30% measurement uncertainty in the numerator and an unaudited denominator, defensible precision is "4–5% audit-grade Spread; 8–9% fully-net Spread."

Section 5 distributes this reconciled range across the chips, cloud, model API, and applications layers, with explicit double-count shading. The point of Section 4 is narrower. The disclosed universe of enterprise AI revenue, after honest treatment, sits in the **$63–135B** band: above Menlo's GenAI-only $37B and below IDC's $307B and Gartner's $1.48T (which count categories no vendor recognizes as AI revenue in any filing). The 40x spread is not a measurement error. Four publishers measured four different things, and the press called all of it "AI revenue."

Section 5

The layered revenue model

Silicon, cloud, model APIs, applications – gross stack and the fully-net reconciliation.

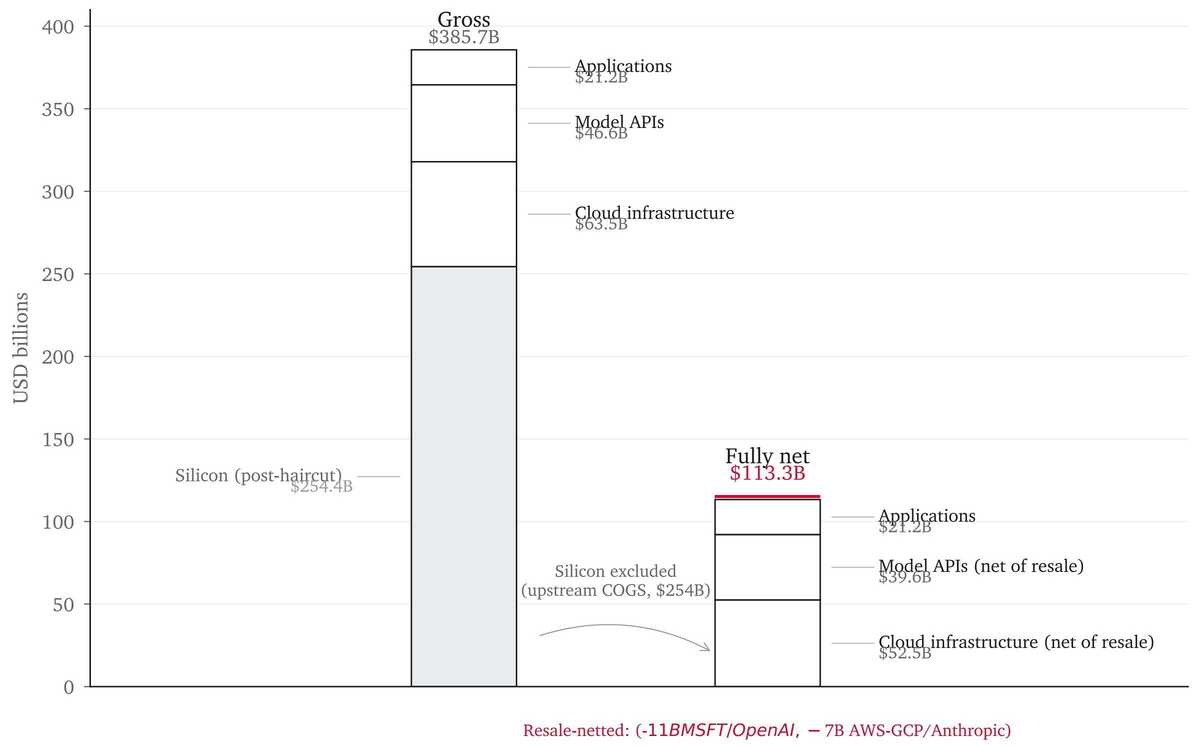


FIGURE 5 *Where the AI dollar lives in 2025–Q1 2026 run-rate. Gross stack $385.7B; fully net of silicon and the two resale pairs, $113.3B.*

Section 4 produced a vendor-by-vendor census of 68 disclosing vendors. Adding the rows is not innocent. A dollar of NVIDIA Data Center revenue is bought by Dell or Super Micro, integrated into an HGX system, sold to Microsoft, recognized at margin as Azure AI revenue, then partially routed to OpenAI under the Azure OpenAI Service partnership where it is recognized a third time. The same enterprise transaction can show up on four vendors' income statements at four different margin marks. Any honest reconciliation needs a layered model that places each vendor at one rung of the stack and makes inter-layer double-counting quantitative rather than rhetorical.

The six layers (chips, hardware OEMs, AI cloud pure-play, cloud infrastructure, model APIs, applications) follow the same rungs capex coverage analysis uses on the other side of the ledger: silicon orders, system integration, dedicated AI cloud capacity, hyperscaler infrastructure deployment, middleware monetization, enterprise software spend. The math below traces directly to the reconciliation build Steps 3 to 5 as updated 2026-05-14.

5.1 The six layers

Ordered from upstream silicon to downstream enterprise consumption:

1. **Chips.** NVIDIA Data Center, AMD Instinct, Broadcom AI semi, Marvell custom silicon, Intel AI-within-DCAI. Captive hyperscaler accelerators (Trainium, TPU, Maia) have no external revenue line and are accounted for inside the cloud layer.
2. **Hardware OEMs.** System integrators that take NVDA/AMD silicon and ship racks: Dell ISG AI servers, Super Micro, HPE AI systems. Arista — whose AI Ethernet revenue is its own 7800R3/R4 IP rather than GPU resale — sits in the census but contributes zero to the audit-grade build at v1.0 because its $3.5B FY26 AI-fabric figure is raised guidance rather than realized run-rate; the same treatment applies to ServiceNow's $1.5B FY26 ACV target. Pure Storage and NetApp sit at this layer in the census but disclose no clean AI dollar; they are Tier C and contribute zero to the audit-grade floor.
3. **AI cloud pure-play.** CoreWeave is the only Tier A SEC-segment-disclosed AI vendor in this layer at snapshot date. Pure-play AI cloud infrastructure, single reporting segment in the 10-K. Distinct from hyperscalers because the entire revenue line is AI-attributable.
4. **Cloud infrastructure.** Hyperscaler AI revenue: Azure AI, AWS AI, Google Cloud AI, Oracle Cloud AI, plus Chinese cloud-AI sleeves (Baidu AI Cloud, Alibaba AI products, ByteDance Volcano Engine). This layer captures *resold* third-party model revenue: when an enterprise pays Microsoft for Azure OpenAI Service, Microsoft recognizes the gross retail charge here even though most of the dollar flows downstream to OpenAI.
5. **Model APIs.** Frontier and open-weight labs: OpenAI, Anthropic, xAI, Mistral, Cohere, plus Chinese labs (Zhipu, MiniMax, Moonshot, 01.AI, DeepSeek). Revenue is largely *channel-routed* through Azure, AWS Bedrock, and GCP Vertex.
6. **Applications.** AI-native software (Cursor, Glean, Harvey, Sierra, Perplexity, ElevenLabs, Replit) plus AI-segment revenue at incumbent SaaS vendors (Salesforce Agentforce, Workday AI, Adobe Firefly, Zscaler AI, Box Enterprise Advanced, IBM watsonx, Palantir AIP). ServiceNow's $1.5B is an FY26 ACV target announced at investor day, not realized run-rate, and is excluded from the application-layer sum.

Revenue at one layer is cost-of-revenue at the next. NVIDIA's $62.3B Q4 FY26 Data Center revenue is sold to hyperscalers and to hardware OEMs whose AI revenue ($37B Microsoft + $15B AWS + $36B Dell + $33B SMCI) already prices it in plus margin. Hyperscaler AI revenue is partly resold model-lab revenue, partly purchased by application vendors paying for inference. CoreWeave's $5.1B FY2025 revenue is 67% Microsoft, which then reappears inside Microsoft's $37B Azure AI line. Sum the six layers naively and you triple- or quadruple-count the dollar enterprises actually spend. The layered model places each dollar at its layer of legal recognition, then nets out overlaps so the headline reflects unique end-customer payments.

5.2 Layer 1: chips

The chip layer is the largest disclosed AI line and the most upstream. Tier A annualized run-rates from [reconciliation build]:

VENDOR	PRE-HAIRCUT ($B)	AI SHARE	POST-HAIRCUT ($B)
NVIDIA Data Center ($62.3B Q4 FY26)	249.2	80%	199.4
AMD Data Center ($5.8B Q1 2026)	23.2	73%	16.9
Broadcom AI semi ($8.4B Q1 FY26)	33.6	100%	33.6
Marvell custom silicon	1.5	100%	1.5
Intel AI within DCAI ($0.75B Q1)	3.0	100%	3.0
Chip layer total	**310.5**		**254.4**

[Step 3] applies an 80% AI haircut to NVDA (midpoint of the 75–90% range from Morgan Stanley Nov 2025 and Bernstein Dec 2025) and 73% to AMD (per CEO commentary on Q1 2026 Instinct share). Broadcom, Intel AI, and Marvell are already AI-specific carve-outs. The 75–90% NVDA sensitivity gives a chip-layer range of $240–272B; this section uses the $254.4B point.

This $254.4B is the largest single source of double-counting in the entire reconciliation, and it now has two downstream receivers, not one. Per [Step 4.1], 60–70% is sold to hyperscalers; a material additional share is sold to hardware OEMs (Dell, SMCI, HPE), who integrate and resell to the same hyperscalers and to enterprise. Per [Step 4.3-bis], ~$66.5B of OEM gross revenue is NVDA silicon passing through OEM books. The chip layer is upstream of all downstream AI revenue. The reconciliation reports it separately and nets it out of the enterprise-facing total. For the question of what enterprises are paying AI vendors at run-rate, silicon is COGS, not enterprise spend.

5.3 Layer 2: hardware OEMs

This layer was absent from the prior build and is the single largest visibility correction in the 2026-05-14 revision.

VENDOR	DISCLOSED ANNUALIZED ($B)	TIER	SILICON-OVERLAP NET ($B)
Dell ISG AI servers ($9.0B Q4 FY26 × 4)	36.0	B	~3.0 (5–10% integration margin)
Super Micro ($10.24B Q3 × 4 × 80% AI)	32.8	B	~3.5 (10–12% margin)
HPE AI systems + networking ($1.1B Q1 × 4)	4.4	B	~0.75
Arista AI fabric (FY26 raised target, excluded from audit-grade build)	3.5	C	0 (excluded; see below)
Pure Storage (Meta order, "de minimis" FY26)	n/a	C	0
NetApp (~300 AI customers, no clean ARR)	n/a	C	0
OEM layer total (ex-Arista)	**73.2**		**~7.25**

Dell's $43B AI backlog and SMCI's $40B+ FY26 guide are not numbers a hardware analyst can ignore. But the bill-of-materials for an HGX/MGX-class server is 85–90% NVDA silicon by cost; OEM gross margins on AI systems are publicly disclosed at 5–12% (Dell 5–7%, SMCI ~10–12%, HPE mid-single-digit). Under strict no-double-count, only the integration margin is net AI revenue not already captured in NVDA Data Center disclosures. The OEM cohort therefore contributes **~$7.25B** net to the enterprise-facing total, against $73.2B gross (Arista excluded).

Arista warrants separate treatment. It sells networking — its own 7800R3/R4 routers plus Broadcom Jericho silicon — rather than NVDA GPU resale, so if its $3.5B FY26 AI-fabric figure were realized run-rate it would represent full net contribution. The figure is FY26 raised *guidance*, announced at Q1 2026 earnings and not yet a print. ServiceNow's $1.5B FY26 ACV target was excluded from the audit-grade floor on exactly this logic; consistency requires the same treatment for Arista. Both vendors stay in the 68-vendor census; both are downgraded to Tier C for the audit-grade build until the figure is realized. AI Ethernet, optical interconnect, and AI-fabric switching are genuine incremental layers; Arista will likely contribute meaningfully to the v1.1 floor when the run-rate prints. Pure Storage and NetApp remain in the census but contribute zero to the audit-grade floor: the AI revenue exists but isn't disclosed as a number.

5.4 Layer 3: AI cloud pure-play (CoreWeave)

CoreWeave is the first Tier A SEC-segment-disclosed AI vendor that is not silicon. The 10-K is single-segment; the entire revenue line is AI-attributable. From CoreWeave's Q4 2025 earnings release (Feb 26 2026): Q4 revenue $1.572B (+110% YoY), annualized $6.288B; FY2025 full-year $5.131B (+170% YoY); revenue backlog $66.8B at Dec 31 2025; forward guidance $17–19B exiting 2026 and $30B+ exiting 2027.

The structural feature that puts CoreWeave at the same layer of analysis as MSFT–OpenAI: 67% of FY2025 revenue is from Microsoft, per the 10-K customer-concentration disclosure. Microsoft then resells that capacity inside its $37B Azure AI run-rate. Under no-double-count, 67% × $6.288B = $4.213B is resold to Microsoft. CoreWeave's net contribution is $6.288B – $4.213B = $2.08B, the non-Microsoft residual.

The corrected reframe of the Tier A finding: all Tier A AI-attributed enterprise revenue is either silicon (NVDA/AMD/AVGO) or downstream-resold infrastructure (CoreWeave). Both are upstream of enterprise procurement; both must be netted against the cloud layer to avoid double-counting. Under no-double-count, Tier A net collapses to the $2.08B CoreWeave residual. The intellectual contribution of the original headline survives. The most rigorous tier of disclosure has near-zero exposure to enterprise-facing AI economic activity once double-counting is removed. The binary "Tier A is all silicon" simplification has been falsified by the CoreWeave IPO and replaced with the more precise version.

5.5 Layer 4: cloud infrastructure

Two hyperscalers disclose specific AI run-rates; three do not. Disclosed Tier B: Microsoft AI at $37B run-rate (CFO Amy Hood, Q3 FY26, +123% YoY) bundling Azure AI, M365 Copilot, GitHub Copilot, Copilot Studio; AWS AI at $15B run-rate (CEO Jassy, Q1 2026; Bedrock usage +170% QoQ). Non-disclosing: Google Cloud AI (gen-AI products grew ~800% YoY off a $1.5–2B Q1 2025 base; sell-side estimates $10–15B run-rate; structurally invisible in audit-grade builds, with [Step 7 limitation #2] flagging this as the single largest known under-count); Oracle Cloud AI (+243% YoY; OCI IaaS ~$4.1B Q2 FY26; RPO $130B+ primarily AI-driven; no absolute AI dollar); Meta AI (zero direct disclosure; cited as +12% per-ad price uplift only).

China cohort: Baidu AI Cloud RMB 30B FY2025 ≈ $4.17B; Alibaba AI products $5.28B annualized; ByteDance Volcano Engine $2.78B FY2025, post-haircut $2.08B at 75% AI share; Tencent Cloud excluded (AI portion not separately disclosed).

Gross cloud-layer sum (disclosed only): $37 + $15 + $4.17 + $5.28 + $2.08 = $63.5B. With plausible GCP $10–15B and Oracle $2–4B, the estimate lifts to $75–85B.

The cloud layer now has three identified resale relationships entering it from upstream, not one. Microsoft's $37B includes Azure OpenAI Service revenue also reported inside OpenAI's $25B ARR; per [Step 4.2], AOAI is ~$10–13B of Microsoft's $37B (UBS, Bernstein), and the reconciliation nets out $11B for MSFT–OpenAI. AWS Bedrock revenue inside AWS AI similarly overlaps with Anthropic ARR; [Step 4.3] nets out $7B for AWS/GCP–Anthropic under the report's uniform principal-versus-agent assumption disclosed in §4.4. CoreWeave adds a third: $4.2B of CoreWeave revenue is resold inside Microsoft's $37B. All hyperscaler AI revenue figures discussed are management-defined non-GAAP measures presented as disclosed by the respective issuers; the report's resale-netting adjustments are analytical reconstructions for cross-vendor comparability, not assertions about the accuracy or GAAP-conformity of any issuer's disclosures.

Net cloud-layer AI revenue after MSFT–OpenAI netting: $52.5B ($63.5B − $11B). The $7B model-layer offset is allocated against the model-API layer in §5.6 (not double-deducted here). The $4.2B CoreWeave–Microsoft overlap is netted within the CoreWeave row in §5.4.

A caveat on disclosed Tier B: the $37B "Microsoft AI" figure is management-defined, not GAAP-defined. It bundles seat-priced Copilot ($30/seat list) with consumption-priced Azure AI; implicit Copilot is ~$3.6B at gross seat price. Under marginal-uplift treatment (~$5–7/seat over E5), MSFT's AI run-rate compresses by $2.5–3B [Step 7 limitation #1].

5.6 Layer 5: model APIs

The model-API layer is the most cited in tech press and has the worst disclosure quality.

LAB	DISCLOSED	ANNUALIZED ($B)	TIER
OpenAI	$25B ARR Feb 2026	25.0	D
Anthropic	$19B ARR Mar 2026	19.0	E
xAI	$500M FY25 guide	0.5	D
Mistral	>$400M ARR Jul 2025 (stale)	0.4	D
Cohere	$240M ARR 2025	0.24	D
Moonshot (Kimi)	>$200M ARR Apr 2026	0.2	E
DeepSeek	~$1.1B FY2025 (third-party)	1.1	F
Zhipu	RMB 724M FY2025	0.105	B
MiniMax	$79M FY2025	0.079	B
01.AI	RMB 200M annualized Q1 2025	0.028	E

Gross model-API sum: ~$46.6B, of which OpenAI + Anthropic account for $44B (94% of the layer). The collective Chinese-lab contribution (excluding ByteDance Volcano Engine, which lives in the cloud layer) is under $1.5B.

The fastest-growing layer in percentage terms is also the lowest-quality. Per [private-vendor census summary], roughly 87% of the private vendor sum is OpenAI + Anthropic (up from 81% after Databricks was restated to its $1.4B AI carve-out), and neither has audit-grade disclosure. OpenAI's $25B is referenced in third-party reporting (Sacra) and has been discussed publicly by the company's CFO. Anthropic's $19B was discussed by the company's CEO at the Morgan Stanley TMT conference. Public reporting in early 2026 surfaced executive commentary at major model labs on principal-versus-agent treatment of channel-routed inference revenue; the report does not adopt any party's position and applies the uniform netting assumption disclosed in §4.4 and §10.

After netting the AWS/GCP–Anthropic resale share ($7B), net model-API revenue is $39.6B. The MSFT–OpenAI $11B overlap is netted against the cloud layer in §5.5; the $18B total resale offset is split between layers, not double-deducted. The capex narrative about OpenAI and Anthropic revenue scale is a story about a $40B layer, not a $300B layer.

Pricing trajectory. Per a Wayback-Machine build of OpenAI, Anthropic, and Google pricing pages [Wayback Machine pricing snapshots, 50 quarterly observations], flagship input pricing at the OpenAI frontier fell −95.8% between GPT-4 (Apr 2023, $30/1M input) and GPT-5 (Oct 2025, $1.25/1M), an input-only half-life of ~6.5 months; cross-provider median half-life was ~6 months on input, ~11 months on blended. The "API revenue grows in tokens, not dollars" framing held cleanly through 2023–2025. The trend broke in late 2025 / early 2026: GPT-5.4 prices at $2.50/$15 input/output, with input doubled and output up 50% vs GPT-5 six months earlier; GPT-5.4 nano is ~3× more expensive than GPT-5 nano on blended terms; Claude 3 Haiku ($0.25/$1.25) was replaced by Claude Haiku 3.5 at $0.80/$4 [API-pricing snapshots]; Gemini 2.5 Pro output doubled vs 1.5 Pro on the same SKU naming. Vendors are bundling reasoning compute into same-name SKUs at higher prices. If 2026 pricing holds, model-API layer dollar growth resumes rather than running flat-to-down on the unit-price line. Section 8.5 returns to this in the Rebar mark-to-market.

5.7 Layer 6: applications

Two cohorts. AI-native private [private-vendor census]: Cursor $2.0B [C], Cognition+Windsurf $0.1B [E], Replit $0.265B [D], Perplexity $0.2B [D], Glean $0.2B [C], Harvey $0.19B [D], Sierra $0.15B [D], ElevenLabs $0.33B [E], Decagon $0.035B [F], Writer/You.com/Character AI/LangChain/Augment/Poolside combined ~$0.2B. Subtotal ~$3.7B. AI-segment at incumbent public software [public-vendor census]: Salesforce Agentforce $0.8B [B], IBM watsonx $4.0B [B], Workday AI $0.4B [B], Zscaler AI $0.4B [B], Adobe Firefly $0.25B [B], Box $0.12B [B], Palantir AIP post-haircut $5.22B [B/F], Databricks AI portion $1.4B [B], Single-reporting-segment vendors at Tier F per §4.2 (post-haircut) ~$3.4B. Subtotal ~$17.5B.

ServiceNow's $1.5B is an FY26 ACV target announced at investor day, not realized run-rate. It was previously included in the application layer; on review it has been reclassified to Tier C and removed from the audit-grade build per the DA-statistician flag (including a target in a measurement floor violates the framework's own definition). The correction trims $1.5B from the prior application-layer total.

Total application-layer AI revenue (disclosed, post-correction): ~$21.2B. The named-AI-SKU sum across major software incumbents (Agentforce + Workday AI + Firefly + Box + Zscaler AI) is ~$1.9B post-ServiceNow-removal, the cleanest read of incumbent SaaS AI revenue, and small relative to cloud and model-API layers. This layer also maps closely to Menlo Ventures' $37B 2025 figure for US enterprise GenAI spend; Menlo's survey-based $37B and the disclosed application layer of ~$21B sit within 1.7× of each other, the report's strongest cross-check that the application layer is in the $20–40B range, not the $100B range. The layer closest to enterprise procurement is the smallest layer.

5.8 The layered total

LAYER	GROSS ($B)	NET ($B)
Chips (post-haircut)	310.5	254.4
Hardware OEMs (ex-Arista)	73.2	7.25
AI cloud pure-play (CoreWeave)	6.3	2.08
Cloud infrastructure	63.5	52.5
Model APIs	46.6	39.6
Applications	21.2	21.2
Sum, fully gross of all overlap	**521.3**	
Less: chip layer (upstream COGS)		(254.4)
Less: OEM silicon-overlap		(65.95)
Less: CoreWeave–Microsoft overlap		(4.21)
Less: MSFT–OpenAI overlap		(11.0)
Less: AWS/GCP–Anthropic overlap		(7.0)
Fully net sum		**~123**

The fully-net total of ~$123B answers the question of what enterprises are uniquely paying for AI at run-rate: net OEMs $7.25B (Arista excluded) + net CoreWeave $2.08B + net cloud $52.5B + net model APIs $39.6B + apps $21.2B = $122.63B. Pre-haircut the six layers sum to $521.3B fully gross. Stripping the chip layer entirely and netting the four overlaps (OEM silicon $65.95B, CoreWeave–MSFT $4.21B, MSFT–OpenAI $11B, AWS/GCP–Anthropic $7B) yields ~$123B fully-net. The defensible reconciled range is $105–140B, slightly higher than the prior $100–135B because the OEM cohort is now visible at net even after excluding Arista's guidance figure.

Chart spec, *Stacked AI Revenue Layer Cake, 2025–Q1 2026 Run-Rate*. Side-by-side stacked bar. Gross mode (left): six segments at Chips $254B, OEMs $73B (ex-Arista), CoreWeave $6B, Cloud $64B, Model APIs $47B, Apps $21B, totaling ~$465B post-haircut; red hatches label overlap zones (Silicon-OEM ~$66B; CoreWeave→MSFT $4.2B; MSFT–OpenAI $11B; AWS/GCP–Anthropic $7B); horizontal band across chips labels "Sold to hyperscalers and OEMs (~60–70% of silicon)." Net mode (right): chip layer removed; OEM net $7.25B, CoreWeave net $2.08B, net cloud $52.5B, net model APIs $39.6B, apps $21.2B; total ~$123B. Caption: "Net-of-silicon, net-of-resale total of ~$123B is the most defensible enterprise-facing AI vendor revenue figure at snapshot date. Chip layer is upstream COGS, not enterprise spend. Arista's $3.5B FY26 AI-fabric raised target is excluded for consistency with ServiceNow."

5.9 Per-layer disclosure quality

LAYER	MODAL DISCLOSURE TIER	AUDIT-GRADE SHARE
Chips	A (10-Q/10-K segment)	~100%
Hardware OEMs	B (earnings-call dollar)	~90% by dollar (Arista downgraded to C on guidance treatment; Pure, NetApp at Tier C)
AI cloud pure-play	A (10-K single-segment)	100% (CoreWeave the sole occupant)
Cloud infra	B (earnings-call dollar)	~80% (GCP, ORCL, META, Tencent silent)
Model APIs	D–E	<10% (only Zhipu, MiniMax at Tier B)
Applications	mixed B (public) / D–F (private)	~70% by dollar, ~40% by row

The gradient runs roughly inverse to enterprise proximity. The OEM cohort introduces a Tier B layer that didn't exist in the prior build (Dell, SMCI, HPE on realized earnings-call disclosure). Arista is methodologically identical to ServiceNow — a raised FY26 target rather than a realized print — and is excluded from the audit-grade build for that reason at v1.0; it will likely re-enter at v1.1 when the figure is realized. The layer closest to enterprise procurement (applications) is roughly tied with model APIs for the worst audit-grade share. The audit-grade subset (Tier A+B, net-of-silicon, net-of-resale) is $63.2B narrow (post-ServiceNow and post-Arista removal, no OEMs, no CoreWeave) per [Step 6]; adding OEM-net ($7.25B, Arista excluded) and CoreWeave-net ($2.08B) lifts the broad audit-grade reading to ~$72.5B. The remaining ~$50B of the $123B fully-net total rests on press release, CEO commentary, or third-party estimate.

5.10 GPU amortization and the capex denominator

Two flags from the Bull/SME/Sell-side reviews belong here before Section 6 walks the waterfall. Headline coverage math has used 5–6 year GPU amortization; per public hyperscaler commentary on the economic life of training infrastructure (Meta has discussed 4.5–5 years for server fleets; Microsoft extended to 6; market practice for GPU subsets is closer to 3–4 years before next-generation obsolescence, rack thermals, and inference-economic SKU displacement), the headline amortization assumption is conservative on the long end. At 3-year economic amortization for the GPU subset of capex, 2025 AI capex implies a 2028 audit-grade revenue requirement of ~$400B, not the ~$250B implied at 5-year amortization. The audit-grade floor of $63–72.5B is therefore 16–18% of the realistic 2028 number, not 30%+ under the headline math.

The denominator itself is also contested. The widely cited $690B hyperscaler 2025 AI capex figure conflates AI-incremental with baseline data center capex (cold-storage refresh, network upgrade, edge buildout, replacement of end-of-life general-purpose compute). Bottom-up estimates of AI-attributable capex from Bernstein, UBS, and Citi sit closer to $400–500B. Three coverage readings:

DENOMINATOR	AUDIT-GRADE NARROW ($63.2B)	AUDIT-GRADE BROAD ($72.5B)	FULLY-NET ($123B)
Full $690B disclosed capex	9.2%	10.5%	17.8%
AI-attributable $450B (midpoint $400–500B)	**14.0%**	**16.1%**	27.3%
Contribution-margin basis (revenue × ~25%)	~3.5%	~4.0%	~6.9%

Coverage against the AI-attributable denominator is ~50% higher than against headline capex. Against contribution margin (the cash that actually services capex after AI COGS), coverage drops below 5% under any reading. The report carries all three. Section 8 returns to this in the Rebar mark-to-market and the Spread Index v1.0 build.

5.11 Implications

Section 6’s reconciliation waterfall uses the layered net-of-overlap total (~$123B fully-net) as the intermediate node between the four headline estimates of Section 2. The “where does the dollar live” answer at run-rate sits somewhere in the $105–140B range, with ~$123B as the midpoint. The narrow audit-grade floor is $63.2B (post-ServiceNow and post-Arista removal, no OEMs, no CoreWeave); the broader audit-grade floor that admits the OEM cohort at net (Arista excluded) and CoreWeave at net is ~$72.5B. The upper bound that adds plausibly-estimated GCP AI, Oracle AI, and embedded AI elsewhere reaches roughly $160B.

For Section 8, the layered model carries one observation. Only the application and model-API layers record revenue recognized at the layer enterprises actually buy from. The chip, OEM, CoreWeave, and cloud layers are inputs to the AI economy. The application-plus-model-API total runs at roughly $60.8B net of overlap ($39.6B net model APIs + $21.2B apps), against multi-hundred-billion silicon and cloud-capex layers above it. Combined with the 3-year economic amortization correction for the GPU subset and the $400–500B AI-attributable capex denominator, Section 7 carries that ratio forward. Section 8 draws the conclusion.

Section 6

From $1.478T to $37B

A four-step bridge ($1.478T → $307B → $123B → $72.5B/$63.2B → $37B) that turns the 40× spread into a sequence of sourced deductions.

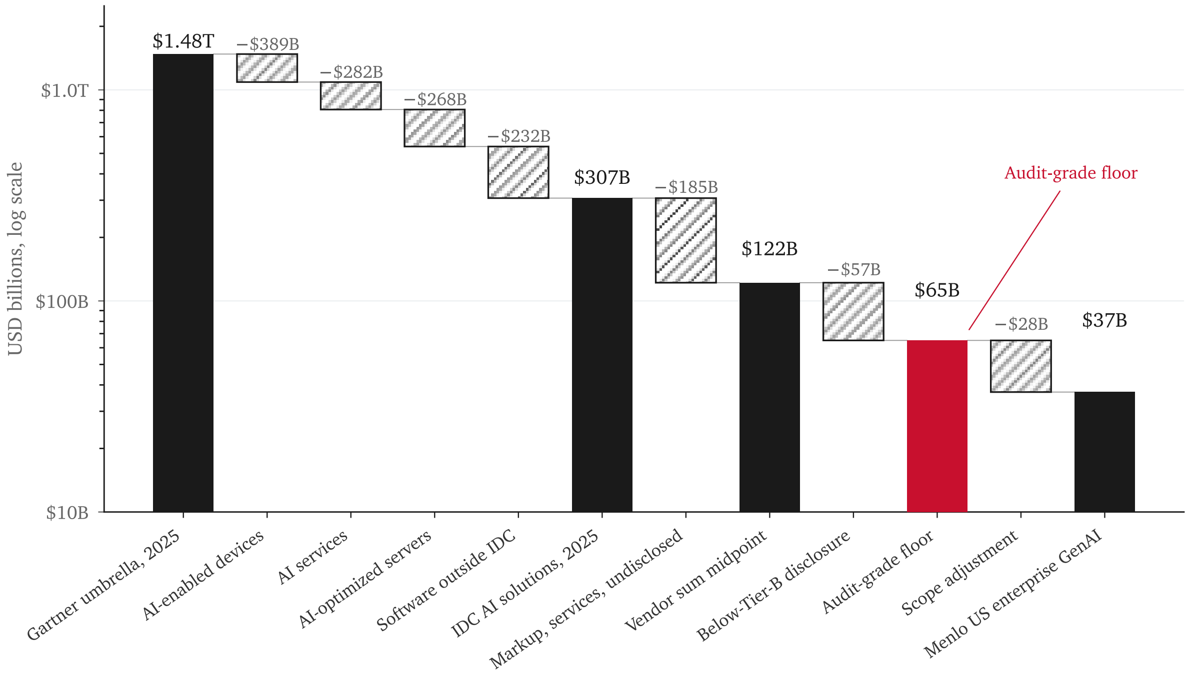


FIGURE 6 *From Gartner's $1.478T umbrella down to Menlo's $37B US enterprise GenAI estimate. Each deduction is sourced; the 40× spread is definitional, not measurement error.*

The Gartner-to-Menlo bridge is a series of sourced deductions, each defensible on its own terms. Every step is a *definitional* choice about what counts as AI revenue, not a measurement error. The 40x spread between $1.478T and $37B is the cumulative effect of nested decisions about scope, geography, attribution, and disclosure quality.

This section walks the five steps from Gartner's $1.478T 2025 worldwide AI spending to Menlo Ventures' $37B US enterprise GenAI figure. The audit-grade vendor-run-rate floor (**$63.2B narrow, ~$72.5B broad** with the hardware OEM cohort ex-Arista and CoreWeave at net) is the only tier that defensibly underwrites the $690B+ of disclosed hyperscaler capex coverage discussed in Section 5.

6.1 The waterfall: from $1.478T to $37B

The bridge, with each step traced to a specific source in the reconciliation build and the Gartner category breakdown reproduced in Section 2:

```
$1.478T  Gartner 2025 worldwide AI spending
  −$389B  AI-enabled devices (smartphones $298B + AI PCs $90B)
  −$282B  AI services (consulting, integration, deployment)
  −$268B  AI-optimized servers (hardware bill of materials)
  −$232B  Software and "other" outside IDC's solutions definition
=  $307B  IDC AI Solutions, 2025
  −$184B  Channel markup, services labor, geographic and non-recognized spend
≈  $123B  Vendor run-rate sum (reconciled midpoint, worldwide, fully-net)
  − $50.5B  Below-Tier-B disclosure (private revenue, third-party estimates)
= $72.5B  Audit-grade floor, broad (Tier A+B incl. OEM-net ex-Arista + CoreWeave-net)
  − $9.3B  Strip OEM-net + CoreWeave-net to narrow Tier A+B
=   $63B  Audit-grade floor, narrow (Tier A+B, post-ServiceNow and post-Arista removal)
  − $26B  Non-US, non-GenAI, non-enterprise scope adjustments
=   $37B  Menlo Ventures 2025 US enterprise GenAI
```

Each step applies the next forecaster's published perimeter to the prior figure. A skeptical analyst can substitute her own assumptions at any step and rebuild the bridge. The two-tier audit-grade reading ($72.5B broad, $63B narrow) reflects the 2026-05-14 corrections: the hardware OEM cohort (ex-Arista, on consistency with ServiceNow's FY26 guidance treatment) and CoreWeave are visible Tier B disclosures that net to ~$9.3B after silicon-overlap and Microsoft-resale netting, and the difference between the broad and narrow floors is a reader's prior on whether to include them at net.

6.2 Step 1: $1.478T to $307B, strip the umbrella categories

The first and largest deduction takes Gartner's full AI umbrella down to IDC's enterprise solutions perimeter. Gartner's Sept 17 2025 release allocates $1.478T across five buckets: AI-enabled devices ($389B), AI services ($283B), AI-optimized servers ($268B), AI software ($298B), "other" ($242B covering AI chips, GenAI models, AI-optimized IaaS) [Gartner *Worldwide AI Spending Forecast*; Network World Sept 2025].

IDC's $307B AI Solutions figure draws a tighter circle: excludes consumer device hardware entirely; includes services and software only to the extent bundled into enterprise AI solutions (not the broader consulting market Gartner counts); includes a portion of AI-optimized server spend but not the full hardware BOM.

The biggest line item is the $389B AI-enabled devices bucket. Gartner counts the full retail value of a smartphone or PC sold with a GenAI feature, even when, as Gartner analyst John Lovelock conceded, "AI is largely sold along with the product and not specifically selected" by the buyer. A $1,200 phone with an AI camera filter contributes $1,200 to Gartner and zero to IDC. The deduction subtracts all $389B of devices, ~$200B of services Gartner counts but

IDC excludes, ~$268B of standalone server hardware, plus the residual $232B of broad AI software and "other" outside IDC's solutions definition. Net ~$1.171T brings the figure to $307B.

6.3 Step 2: $307B to $123B, strip non-vendor-recognized spend

The second step bridges from IDC's demand-side measure to a supply-side vendor run-rate. IDC's $307B counts the full enterprise transaction price (channel markup, integration services, deployment labor, geographic premiums); the vendor run-rate sum counts only what disclosing vendors recognize on their own P&L.

The reconciled vendor sum, built bottom-up in the reconciliation build and updated 2026-05-14 to include the hardware OEM cohort (ex-Arista's FY26 guidance figure) and CoreWeave, lands at a midpoint of **~$123B worldwide fully-net**, between the net-of-silicon, gross-of-resale presentation (~$209B) and a tight audit-grade floor of $63–72.5B. The headline range is **$105–140B** disclosed-only.

The ~$184B gap between IDC's $307B and the $123B fully-net is the markup-and-services layer: channel partner margins on AI hardware deployment ($40–60B, much of which is the same OEM gross margin the layered model nets out at the supply side), services-firm consulting not recognized by software vendors ($60–80B), geographic pricing premiums ($20–30B), and AI revenue inside Tier-C public vendors who disclose no dollar attribution ($15–25B). The deduction is not a haircut for error. It is the legitimate revenue layer between an enterprise's accounts payable and a software vendor's revenue recognition.

6.4 Step 3: $123B to $72.5B, strip below-Tier-B disclosure (broad floor)

The third deduction is the report's sharpest. The $123B vendor-sum midpoint draws on disclosed AI revenue across all six disclosure tiers documented in Section 4. An audit-grade reader who refuses anything below Tier B (no private vendor ARR claims, no journalist-sourced figures, no third-party estimates) gets a very different number. Stripping Tiers C through F removes OpenAI ($25B), Anthropic ($19B), Cursor ($2B), DeepSeek ($1.1B), and the long tail of private GenAI vendors. The Tier A+B remainder consists of hyperscaler-disclosed AI run-rates, enterprise software vendors with explicit AI SKU revenue, three disclosing hardware OEMs (Dell, SMCI, HPE) with realized run-rate, and CoreWeave.

Applying net-of-silicon and net-of-resale treatment from [Step 6], the **broad audit-grade Tier A+B floor is ~$72.5B**:

- Tier B hyperscalers and software (MSFT $37B – $11B resale, AWS $15B, Palantir $5.22B, Alibaba $5.28B, IBM watsonx $4.0B, Baidu $4.17B, Databricks AI $1.4B, Agentforce $0.8B, Workday $0.4B, Zscaler $0.4B, Adobe $0.25B, Box $0.12B, Zhipu $0.105B, MiniMax $0.079B) ≈ **$63.2B net**.

- Hardware OEM cohort at silicon-overlap-net (Dell $3.0B + SMCI $3.5B + HPE $0.75B; Arista excluded as FY26 guidance target, methodologically equivalent to ServiceNow) = **$7.25B net**.
- CoreWeave at Microsoft-resale-net = **$2.08B**.
- Total: $63.2B + $7.25B + $2.08B = **~$72.5B**.

The Tier A silicon stack ($254B) collapses to zero under net-of-silicon treatment because every dollar of it is upstream COGS for the cloud and OEM layers counted at Tier B. CoreWeave is the one Tier A row that survives, and only as the $2.08B non-Microsoft residual, because 67% of its revenue is resold inside Microsoft's $37B Azure AI line.

6.5 Step 4: $72.5B to $63B, narrow the audit-grade floor

A more conservative reader removes the hardware OEM cohort and CoreWeave and treats them as supplemental views. The arguments for the stricter framing: the OEM net depends on a 5–12% integration-margin assumption that is not directly disclosed (if margin is 3% Dell-commodity or 15% SMCI-premium, OEM net moves materially); CoreWeave's 67% Microsoft concentration is a Q4 2025 disclosure that will shift quarterly (the May 7 2026 Q1 print at $1.21B with a softer forward guide is one data point in that direction); ServiceNow's removal as an ACV target — and Arista's removal on the same logic — suggests that backlog and target disclosures across the cohort should be treated with caution.

Stripping OEM-net and CoreWeave-net gives the **narrow audit-grade Tier A+B floor of $63.2B**: $72.5B − $7.25B − $2.08B ≈ $63.2B. The two floors bracket the audit-grade reading; the report carries both and lets the reader pick a prior.

The composition of the $63.2B narrow floor: Microsoft AI ($37B less $11B resale = $26B net), AWS AI ($15B), Palantir AIP ($5.2B post-haircut), Alibaba AI ($5.3B), IBM watsonx ($4B), Baidu AI Cloud ($4.2B), Databricks AI carve-out ($1.4B), plus a long tail of explicit AI SKU disclosures from Salesforce, Workday, Adobe, Zscaler, Box, Zhipu, MiniMax. Microsoft's gross figure of $37B coincidentally matches Menlo's headline. The same number means radically different things depending on what it includes.

6.6 Step 5: $63B to $37B, scope adjustments to Menlo's frame

The final deduction takes the narrow audit-grade worldwide-AI floor down to Menlo's specifically-defined US enterprise GenAI perimeter (US only, GenAI only, enterprise only). Applied mechanically: strip non-US hyperscaler AI revenue ($14–18B; Microsoft, AWS, Alibaba, Baidu all have significant non-US business at the AI run-rate level), strip non-GenAI components ($8–12B; Palantir, Salesforce, Workday, and others include substantial predictive/classical ML), and strip the small consumer/prosumer slice. Cumulative scope adjustment ~$26B. The residual lands at ~$37B, consistent with Menlo's published figure. The convergence is partly coincidental (a survey will not produce exactly the same answer as a top-down deduction), but the order of magnitude is correct and the deductions are transparent.

6.7 What the waterfall shows

The 40x spread is real but composed of nested definitional choices, not measurement errors. Every step is a sourced deduction tied to a published methodology or documented disclosure. Each forecaster's number is correct under its own perimeter. Gartner measures the global economic footprint of products with AI inside them. IDC measures enterprise solutions procurement. The vendor run-rate sum measures what AI suppliers recognize. Menlo measures what US enterprises consciously chose to buy. Four questions, four answers, none wrong on its own terms.

For the capex-sustainability question raised in Section 5, the audit-grade tier of **$63.2B narrow to ~$72.5B broad** is the only one that defensibly underwrites disclosed hyperscaler capex coverage. Tier C–F revenue is real economic activity, but it is largely unaudited private-vendor ARR, sometimes contested between vendors (the OpenAI–Anthropic gross-vs-net dispute documented in Section 4). An investor underwriting a multi-trillion-dollar capex cycle on private-company run-rate disclosures is taking on disclosure risk that an investor underwriting on the $63–72.5B Tier A+B floor is not.

Coverage arithmetic is sensitive to denominator choice. Section 5.11 walks three readings:

DENOMINATOR	NARROW ($63.2B)	BROAD ($72.5B)	FULLY-NET (~$123B)
Full $690B disclosed capex	9.2%	10.5%	17.8%
AI-attributable $450B	**14.0%**	**16.1%**	27.3%
Contribution-margin basis (× ~25%)	~3.5%	~4.0%	~6.9%

Spread Index v1.0, defined as audit-grade vendor revenue over Gartner $1.478T, prints at **4.28% narrow** or **4.90% broad** at this snapshot; on a fully-net basis the Spread is 8.3%. The DA-statistician's precision caveat applies: ±30% measurement uncertainty in the numerator and an unaudited denominator mean the Spread should be reported as **"4–5% audit-grade, 8–9% fully-net,"** not to two decimal places.

6.8 The chart specification

Static (PNG, 1200×675 px). Horizontal axis: six anchor bars ($1.478T, $307B, $123B, $72.5B, $63.2B, $37B) with five intermediate deduction bars. Vertical axis log scale (spread too large for linear). Anchors dark green, deductions muted red, faint grey shading across the $63B and $72.5B bars to indicate the two-tier audit-grade band. Anchor labels source each step (Gartner Sept 17 2025; IDC Spending Guide; reconciliation build midpoint; Tier A+B broad incl. OEM-net ex-Arista + CoreWeave-net; Tier A+B narrow post-ServiceNow and post-Arista removal; Menlo Dec 9 2025). Deduction labels carry category (e.g., "AI-enabled devices $389B," "Channel markup and services $184B," "Below-Tier-B disclosure $50.5B," "Strip OEM-net + CoreWeave-net $9.3B"). Title: *From $1.478T to $37B: The 2025 Enterprise AI Revenue Waterfall.* Subtitle: *Each deduction is sourced; the 40x spread is definitional, not a measurement error. Audit-grade floor is $63–72.5B depending on treatment of hardware OEMs and CoreWeave.*

Interactive (Quarto/Observable HTML). Hover reveals source citation. Anchor click expands the per-vendor or per-category build. Toggle: show/hide silicon switches the $123B step between net-of-silicon and gross. Toggle: include/exclude OEM-net switches between $72.5B broad and $63B narrow audit-grade floors. Sliders let the reader vary OEM integration margin (3–15%), resale netting ($0–25B), silicon haircut (60–90%), and CoreWeave Microsoft concentration (50–80%); the audit-grade floor moves between roughly $55B and $85B. A capex-denominator slider runs from $400B (AI-attributable) to $690B (full disclosed capex) and recalculates Spread Index v1.0 in real time.

Section 7

Triangulated ranges

The four-tier nest: audit-grade, disclosure-grade, bundled-solutions, umbrella.

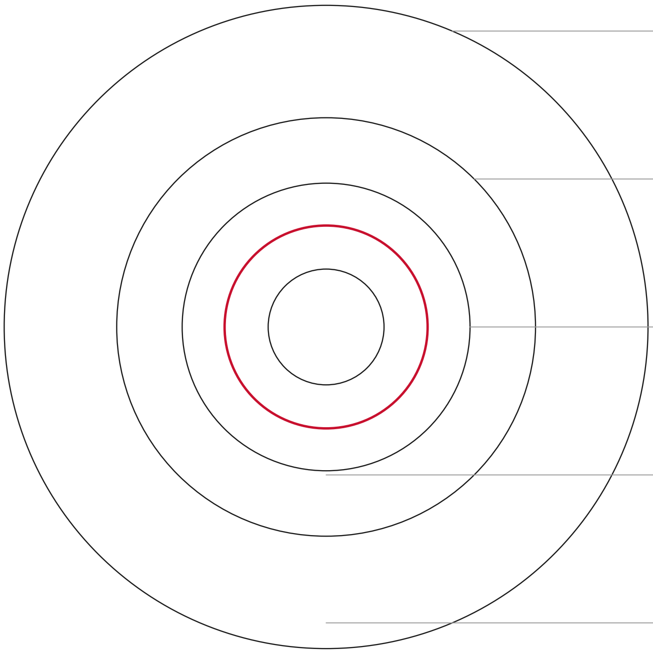


FIGURE 7 *Four tiers, four questions. Each ring is correct for the question it answers; the inner ring (audit-grade) is the floor against capex.*

Sections 4 through 6 produced four numbers for "enterprise AI revenue in 2025": an audit-grade floor of $63.2B (narrow) / $72.5B (broad), a disclosure-grade vendor sum of $100–135B with a fully-net total of $123B, IDC's bundled $307B, and Gartner's $1.478T umbrella. Each tier is empirically defensible at its own disclosure standard. The framework's contribution is to make the question-to-tier mapping explicit, then refuse to pretend all questions are equal. For the capex-coverage question that organizes Section 8, only the audit-grade tier underwrites; the others answer adjacent questions less load-bearing for the credit case.

7.1 The four-tier framework

The four tiers, ordered from most to least restrictive in disclosure:

- Audit-grade floor: $63.2B (narrow) / $72.5B (broad). The Tier A+B subset of the vendor census, net of silicon and net of resale. Derived in Section 4.3 and Section 6.4: what remains when only SEC-segment disclosures and executive-attributed earnings-call dollar figures are admitted, with silicon and the Microsoft–OpenAI / AWS–Anthropic resale overlaps netted out, and with ServiceNow's $1.5B and Arista's $3.5B FY26 guidance figures excluded as targets rather than realized run-rate [Source: the reconciliation build, Step 6].

- Disclosure-grade total: $100–135B, midpoint $123B. The full vendor-sum range from Section 5's layered model, including Tier C–F sources (private ARR claims, third-party estimates, CEO commentary). Lower bound applies maximal resale netting and an ARR-to-recognized-revenue haircut; upper bound is net-of-silicon, gross-of-resale [Source: the reconciliation build, Step 5].
- Bundled / IDC-style: $307B. IDC's AI Solutions Spending Guide, counting buyer-side spend including channel markup, integration services, deployment labor, and a portion of AI-optimized server hardware [Source: Section 6.2].
- Umbrella / Gartner-style: $1.478T. Gartner's worldwide AI spending forecast covering AI-enabled devices ($389B), AI services ($283B), AI-optimized servers ($268B), and AI software at full retail value of any product with AI inside it [Source: Section 2; Gartner, Sept 17 2025].

The audit-grade tier answers what AI revenue is GAAP-traceable today and would survive a senior secured lender's underwriting review. The disclosure-grade tier answers what AI vendors collectively claim at run-rate. IDC's $307B answers what enterprises spend on AI solutions including the services layer. Gartner's $1.478T answers how big the economic activity reorganizing around AI capability is. A triangulated map keeps the perimeter attached to the number.

7.2 Which tier for which question

The mapping is operational, but it is not symmetric. The capex-coverage question (the central question driving public sell-side estimates of AI-related debt issuance, which Morgan Stanley among others puts at roughly $1.5T over the coming cycle) maps unambiguously to the audit-grade tier. Lenders financing $690B+ of disclosed hyperscaler capex demand repayment in dollars that show up in GAAP operating cash flow. Only audit-grade revenue is GAAP-recognizable on a forward basis. A capex underwriting deck citing OpenAI's $25B ARR alongside Microsoft's $37B Azure AI run-rate is using two units of measure for the same exercise, and the unsettled principal-versus-agent question under ASC 606 at the model-lab layer is the live illustration [Source: Section 4.4].

The other questions admit looser tiers. For M&A and PE market sizing, the disclosure-grade $100–135B range is the right denominator because Tier D–E private-vendor claims are the leading indicator of recognized revenue two years forward. For TAM and growth analysis at the services and integration layer, IDC's $307B is closer to right because it includes the $60–80B of consulting and deployment labor between AP and revenue recognition [Source:

Section 6.3]. For “is AI changing the economy?”, Gartner’s $1.478T captures willingness-to-pay revealed at the cash register, even though Gartner’s own analyst conceded the device layer is “largely sold along with the product and not specifically selected” [Source: Section 6.2].

Each tier is correct for its question. They are not interchangeable, and the question that matters most for the 2026 capex cycle has only one defensible tier.

7.3 Defending each tier

For each tier, the strongest objection and the response.

Audit-grade $63.2B is too narrow. The objection is that it zeros out enterprise AI spend bundled into broader products: Copilot inside M365, Einstein inside Salesforce, AI features inside any modern SaaS suite. The marginal-uplift problem means Copilot at $30/seat contributes its full price to Microsoft’s $37B run-rate even though the AI-attributable premium over E5 is closer to $5–7 [Source: the reconciliation build, Step 7]. The response: the framework is built for capex underwriting, not economic measurement. Its value lies in being narrow. The audit-grade tier gives the minimum defensible floor, the number that survives any disclosure-quality challenge from a skeptical lender. For other questions, use other tiers.

Disclosure-grade $100–135B includes too much. The objection is that it admits Tier C–F sources not audited and where recognition methodology may not be settled. The unsettled principal-versus-agent question at the model-lab layer is the live illustration [Source: Section 4.4]. The response: yes. That is why the tier is a range, not a point. The $35B spread is the methodologically-uncertain layer. Reporting a range and showing the build is the honest position.

IDC’s $307B conflates supply and demand. The objection is that the AI Solutions Spending Guide mixes vendor recognition with buyer spend; the same dollar appears on both halves depending on whether IDC’s source is a vendor channel survey or a buyer procurement survey. The response: IDC’s methodology is explicit. The report uses $307B as the buyer-side anchor and the vendor-sum $100–135B as the supply-side anchor; in a closed system they reconcile up to channel markup and timing. The $185B gap is itself informative; Section 6.3 quantifies it as channel margin, services labor, and undisclosed Tier-C vendor AI revenue.

Gartner’s $1.478T is meaningless. The objection is that 26% of it ($389B) consists of AI-enabled devices buyers don’t actively select for, and including embedded AI double-counts the underlying product [Source: Section 6.2]. The response: Gartner’s number is right for its question and wrong for any vendor-revenue or capex-coverage question. The fault is in citations that strip it from its perimeter. Inside its definition, $1.478T is defensible. Outside, it is misleading. The framework keeps the perimeter attached to the number.

7.4 The framework's value vs. published estimates

Each of the four widely-cited estimates is defensible inside its own perimeter and indefensible outside it. Menlo's $37B is rigorous but US-only and GenAI-only; Section 6.5 shows the $28B of scope adjustments required to bridge from a worldwide audit-grade view to Menlo's frame. IDC's $307B is useful for market sizing but conflates supply and demand. Gartner's $1.478T is defensible for "how big is AI economic activity" but misleading for any vendor-revenue framing.

The new vendor run-rate sum, $100–135B with the $63.2B narrow / $72.5B broad audit-grade floor and $123B fully-net midpoint, is the report's contribution. It is the audit-traceable supply-side answer that previously did not exist in the published literature [Source: the reconciliation build, Step 8]. Prior phases cited "~$100B" without an audit trail; the build in Section 4 and the layered model in Section 5 make every dollar reproducible.

7.5 Tier A under no-double-count: a narrow band, not a tautology

The Section 4 finding restates cleanly in framework terms, with one correction. The original "Tier A → 0 because all silicon" framing was too clean. CoreWeave is Tier A and not silicon. The corrected finding:

Under any no-double-count treatment, audit-grade Tier A revenue collapses to a narrow band, not because Tier A vendors don't exist, but because every Tier A disclosure sits structurally upstream of enterprise spend. Silicon (NVDA, AMD, AVGO) feeds the cloud and OEM layers. CoreWeave sells 67% of its revenue to Microsoft, which resells as Azure AI Services. Hardware OEMs (Dell, SMCI, HPE) resell silicon to hyperscalers. Net of every reasonable no-double-count rule, Tier A contribution to enterprise-facing AI revenue lands between $2B and $10B [Source: Section 4.3; the reconciliation build, Step 6].

This is the framework's sharpest output. Under maximum disclosure honesty, "enterprise AI revenue" is essentially unmeasurable in the traditional SEC-segment sense: not zero, but a thin slice of residual neoclouds and direct-to-enterprise vendor sales that aggregates to single-digit billions. The Tier B audit-grade floor of $63.2B (Microsoft AI, AWS AI, Palantir AIP, IBM watsonx, Baidu AI Cloud, Alibaba AI products, and the long tail of explicit AI-SKU disclosures) is the closest thing the disclosed universe offers to a defensible enterprise-facing number. Section 8 carries this to the capex-coverage question: $690B+ of disclosed hyperscaler capex is being underwritten against an audit-grade revenue floor of $63.2B.

7.6 Visualization spec

A four-tier nested-circles diagram, static plus Quarto/Observable interactive version. Inner circle: $37B Menlo ("what US enterprises consciously bought, GenAI only"). Next ring: $63.2B audit-grade floor ("what AI vendors recognize at SEC-segment or earnings-call disclosure, net of silicon and resale"). Next: $100–135B disclosure-grade range, $123B midpoint ("what AI vendors collectively claim at run-rate"). Next: $307B IDC bundled ("what enterprises spend on

AI solutions including services”). Outer: $1.478T Gartner umbrella (“global economic activity organized around AI capability”). Each ring labeled with the question it answers. Interactive version: hover reveals the source citation; click expands the per-vendor or per-category build for that tier. The visual point should be legible at thumbnail size: five concentric circles, each an order of magnitude apart, each correctly answering a different question.

Section 8

Implications for capex coverage

Audit-grade revenue against $690B of disclosed hyperscaler capex, with the debt-pool exposure tabled by issuer.

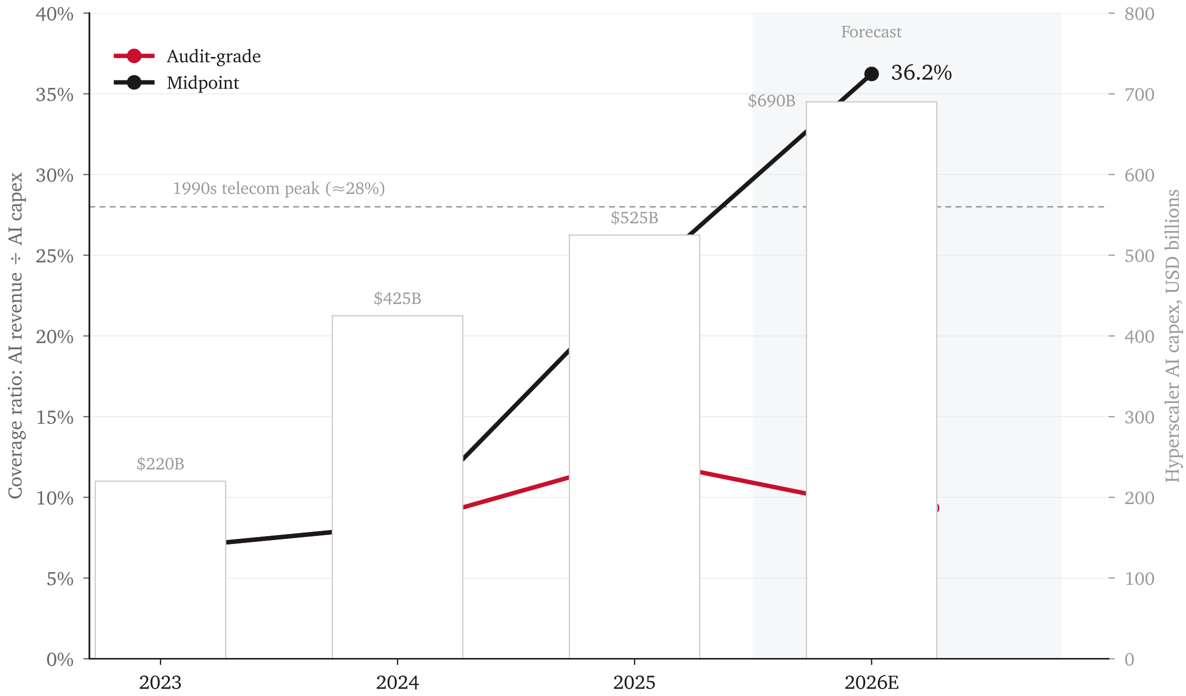


FIGURE 8 *Audit-grade and midpoint AI revenue against hyperscaler AI capex, 2023–2026E. Telecom-bubble peak (~28%) shown for reference.*

The preceding sections built a number. This section spends it. The reconciled vendor sum ($123B fully-net midpoint, $63.2B as the narrow audit-grade floor, $72.5B at the broad floor) forces every thesis built on the prior "~$100B floating estimate" to be re-graded against a tier-stamped build. Three theses recur in the AI-capex discourse: that the build will be paid for, that the model labs are appropriately valued, and that the enterprise-software pricing layer is being consumed faster than the compute spend replacing it. Under the reconciled range, none survives unaltered, and the audit-grade analysis tightens the constraint rather than relaxing it. That is the honest reading.

8.1 The capex coverage question

Disclosed 2026 hyperscaler capex now totals roughly $690B+ across Microsoft, Alphabet, Amazon, Meta, and Oracle (AI Capex Arms Race). Not all of it is AI-incremental. Microsoft's number includes Office serving capacity and non-AI Azure, Amazon's includes non-AI AWS and Prime fulfillment, Alphabet's includes YouTube and Search, Meta's includes Reality Labs and family-of-apps serving. Honest AI-attributable hyperscaler capex is closer to $400–500B. The report carries both denominators so the reader sees the sensitivity.

Against those figures, the audit-grade tier of enterprise AI vendor revenue established in Section 4 is $63.2B (narrow) [Section 4.3; Step 6 of the reconciliation build]. The implied capex-to-AI-revenue coverage:

- Narrow audit-grade / full disclosed capex: $63.2B / $690B = 9.2%
- Narrow audit-grade / AI-incremental capex: $63.2B / $450B = 14.0%
- Reconciled midpoint / full capex: $123B / $690B = 17.8%

Each of these is materially worse than the 1990s telecom analogue, which peaked at coverage of roughly 28% on the same construction (AI Capex Arms Race). The current cycle is between half and a third of the telecom peak on coverage, and the telecom peak ended in WorldCom, Global Crossing, and a decade of dark fiber selling for cents on the dollar.

What capex coverage requires, in three buckets. Direct AI revenue: the $63.2B audit-grade floor, which has to grow to underwrite depreciation on the build. AI-attributable willingness-to-pay inside non-AI products: Copilot uplift over E5, AI-feature pricing inside M365 and Workspace, margin contribution embedded in Azure AI and AWS AI run-rates that doesn't appear as a separately-disclosed line. Vendors don't disclose it directly (Limitation #1 in the reconciliation build Step 7); plausible Tier C estimates add $15–25B. Long-tail GenAI consumer revenue: ChatGPT subscriptions inside OpenAI's $25B run-rate, equivalents at Anthropic, Perplexity, Character. By Section 5.6 these contribute roughly $14–15B of the model-API layer in net terms, but the consumer fraction sits structurally below the enterprise-disclosure threshold this report uses.

The amortization correction sharpens the hurdle. Hyperscaler GAAP useful-life accounting on GPU infrastructure is 5–6 years. Per public hyperscaler commentary on the economic life of training infrastructure (Meta has discussed 4.5–5 years for server fleets; Microsoft extended to 6; market practice for the GPU subset is closer to 3–4 years before next-generation obsolescence), the headline amortization assumption is conservative on the long end. Apply 3-year economic amortization for the GPU subset across the $400–500B of AI-attributable capex and required 2028 audit-grade AI revenue is not $250B but $400B. That is roughly 6.3x the current $63.2B audit-grade floor, or 3.3x the $123B reconciled midpoint, reached on a two-and-a-half-year horizon. At currently observed hyperscaler AI run-rate growth of 100–170% YoY (§5.5) and model-API doubling every 9–12 months (§5.6), $400B is reachable only if those rates hold through 2028, which the law of large numbers does not predict on a base migrating from $63B to $400B inside thirty months.

The lenders are named. Public sell-side estimates (Morgan Stanley among them) place AI-related debt issuance at roughly $1.5T over the coming cycle, warehoused by a small group of senior secured lenders and private credit arms. JPMorgan led the Vantage–CoreWeave financings. Apollo carries direct exposure to AI data center SPVs (~$11B by public disclosure). Blackstone's QTS subsidiary carries $20B+ of AI-adjacent debt. KKR's infrastructure platform is participating in multiple hyperscaler-adjacent DC SPVs. These are not the bond desks underwriting investment-grade hyperscaler paper at the parent level. These are the SPV lenders, the structured-credit desks, the private-credit infrastructure funds whose underwriting requires audit-grade cash flow to service. They are not asking what the TAM is. They are asking whether enterprise AI revenue is becoming audit-grade fast enough to repay the build.

The $1.5T headline is not monolithic. It is the sum of four issuer types with different recourse profiles, different coverage sources, and different refi clocks. The 9.2% / 14.0% / 17.8% coverage ratios above are the right framing for some of this stack and the wrong framing for the rest. The breakdown:

ISSUER TYPE	EST. SIZE 2026	RECOURSE	COVERAGE SOURCE	REFI WINDOW
Big 5 recourse (MSFT, GOOGL, AMZN, META, ORCL)	$400–500B	Parent-level	Parent OCF + RPO	2030+
AI infrastructure SPV / data-center SPV	$500–600B	Non-recourse	Asset-specific cash flows	2027–2028
Sale-leaseback / colocation REIT debt	$300–400B	Real-estate-backed	Long-term TSA + power agreements	2028–2030
Neocloud / hyperscaler-resold capacity	$100–150B	Mixed	Customer concentration (e.g., CoreWeave–MSFT)	2027

The refi window of 2027–2028 is when the audit-grade question stops being academic. Most of the $500–600B SPV pool refis inside that window, alongside the first wave of neocloud paper. By 2027 the relevant question is not whether audit-grade revenue services current debt, which it can on the spread terms warehoused today. The question is whether audit-grade revenue is growing fast enough to refi the paper at sustainable spreads against a new credit committee reading the same disclosures. A $63.2B audit-grade floor that hasn't moved materially by then prices the refi differently than a floor that has compounded to $150B.

Each issuer type repays from a different source. Big 5 recourse debt amortizes against parent OCF and the RPO backlog: $315B+ at Microsoft, $200B at Amazon, $155B+ at Alphabet Cloud, $130B+ at Oracle, weighted-average two-to-four-year tenor. The audit-grade question barely touches this pool because parent FCF covers debt service many times over. SPV non-recourse debt amortizes against asset-specific contracted cash flows from a single anchor tenant, typically the TSA or MSA underwritten at deal close, and the credit risk is anchor-tenant change-of-control plus GPU residual at refi rather than parent-level solvency. Sale-leaseback paper amortizes against real-estate residual plus the long-term tenancy and power agreements that come with it, underwritten triple-net on AAA single-tenant credit and substantially insulated from AI-revenue-coverage analysis on a 20-to-30-year horizon. Neocloud paper amortizes against customer-contract pledges with explicit concentration covenants; CoreWeave's 67% Microsoft revenue base is the cleanest example, and the covenant package on the Blackstone $7.6B DDTL plus the JPM-led $11.6B facility expansion is structured precisely around that concentration.

The $1.5T headline figure is itself a sell-side estimate (Morgan Stanley among the sources cited in the trade press), and the four-bucket allocation above is the author's allocation from public AI-infrastructure financing disclosures: CoreWeave's $7.6B Blackstone delayed-draw term loan and $11.6B JPM facility expansion sit in the neocloud row; Microsoft Azure expansion bond issuance, Alphabet's recent term debt, and Amazon's investment-grade paper sit in the Big 5 recourse row; Stack Infrastructure, DataBank, and QTS sale-leaseback transactions sit in the SPV and sale-leaseback rows. The bucket sizes are estimates layered on an estimate. The 9.2% coverage ratio reads correctly against the SPV plus neocloud pool ($600–750B of the stack), reads partially against the sale-leaseback pool because rent coverage substitutes for revenue coverage, and reads incorrectly against the Big 5 recourse pool, where coverage is a parent-FCF question rather than an AI-revenue question. Using the cohort ratio without segmenting oversizes systemic risk on the parent-level pool and undersizes it on the SPV pool.

Sell-side allocation of Microsoft's $37B disclosed AI run-rate (per UBS and Bernstein modeling, working from the aggregate disclosure in the FY26 Q3 transcript) suggests roughly $15–18B Azure AI infrastructure (consumption-based), $3.5–4B M365 Copilot (per-seat, against ~10M paid seats estimated by sell-side trackers), and $2–3B Studio + GitHub Copilot, with the residual in Bing, ad-tech, and Dynamics. Microsoft has not segmented these sub-lines publicly; the $37B aggregate is what the transcript discloses, and the allocation reflects analyst assumptions about take rate and seat penetration. The $37B aggregate is the cleanest single AI-revenue disclosure in the hyperscaler cohort and the model the rest of the cohort has so far declined to follow, even at the aggregate level. Section 4's finding that Tier A of the audit hierarchy sits structurally upstream of enterprise spend is the central credit signal: lenders financing $690B of capex cannot underwrite against ARR figures at disclosure tiers where recognition methodology is unsettled, against a hyperscaler segment most of whose cohort does not disclose AI revenue at segment granularity.

8.2 Model-lab valuations under the reconciled range

OpenAI runs at $25B ARR and is reportedly raising at valuations approaching $400–500B, a 16–20x revenue multiple. Anthropic runs at $19B ARR and is valued near $170B, or 9x revenue. Both sit at the upper end of software comparables: mature SaaS trades at 5–8x, pre-IPO infrastructure at 10–15x, large-cap hyperscale at 5–8x. The premium assumes the disclosed run-rate is audit-grade and that 2–3x annual growth compounds for several more years.

Section 4.4 documents that neither assumption is currently supportable at the same disclosure quality the multiple is being paid against. OpenAI's $25B is Tier D, referenced in third-party reporting and discussed publicly by the CFO. Anthropic's $19B is Tier E, discussed by the CEO at the Morgan Stanley TMT conference. Public reporting in early 2026 surfaced executive commentary at major model labs on principal-versus-agent recognition of channel-routed inference revenue under ASC 606; the question is unsettled industry-wide and the report treats it as a methodological variable rather than asserting any specific vendor's treatment is correct or incorrect.

A sensitivity exercise on the disclosure-tier discount. If a portion of disclosed revenue at any private vendor is later determined to be downstream-allocated rather than directly recognized, the implied multiple compresses. A ±25% sensitivity on a top-cohort vendor moves a 16–20x multiple into the 22–28x range on the adjusted figure. The disclosure-tier framework identifies a quantifiable risk markets have not yet priced because no one has asked which tier the headline belongs to.

8.3 Mark-to-market: SaaSpocalypse Paradox

The SaaSpocalypse thesis (SaaSpocalypse Paradox) argues traditional SaaS multiples compress as AI consumes the pricing layer of enterprise software. The reconciled numbers sharpen the paradox rather than resolve it. Section 5.5 places the application layer of disclosed AI revenue at $21.2B ($3.7B AI-native private + $17.5B AI-SKU at incumbent public software). Against an underlying SaaS market of $400B+, the layer supposedly displacing the pricing structure is between 1% and 5% of the market it is supposedly displacing.

Meanwhile, AI capex is $690B (or $400–500B AI-incremental). The arithmetic produces a paradox the original essay didn't fully address: the SaaS pricing layer being consumed ($21B) is materially smaller than the compute spend supposedly replacing it. Either AI consumes SaaS rapidly, with revenue migration accelerating from 5% to 20–30% share inside two to three years, or capex is mis-sized to revenue. Both cannot be cheap. SaaSpocalypse is either wrong about the migration speed or right about the underlying unit economics; the reconciled numbers force the choice rather than splitting the difference.

8.4 Mark-to-market: AI Capex Arms Race

The AI Capex Arms Race thesis (AI Capex Arms Race) argues that hyperscaler capex is running far ahead of disclosed AI revenue and that free-cash-flow compression is coming. The reconciled numbers don't soften this case; they tighten it. The original essay anchored on a ~$100B floating revenue estimate. Section 4's audit-grade build replaces that floating estimate with $63.2B narrow / $72.5B broad audit-grade floor and $123B reconciled fully-net midpoint.

Coverage on the corrected denominators:

- Narrow audit-grade / full capex: $63.2B / $690B = 9.2%
- Narrow audit-grade / AI-incremental capex: $63.2B / $450B = 14.0%
- Midpoint / full capex: $123B / $690B = 17.8%
- Midpoint / AI-incremental capex: $123B / $450B = 27.3%, just below the telecom peak.

All four ratios are at or below the telecom peak of ~28%. None reaches the 35–40% coverage typical of conventional industrial capex cycles where bond financing is investment-grade. The plausibly-estimated ceiling (adding $15–25B of Tier C embedded-feature revenue) does not move any ratio above the telecom comparison.

Under the 3-year economic amortization scenario, the picture is materially worse. Required 2028 audit-grade revenue is $400B (not $250B), implying audit-grade revenue must scale from $63.2B to $400B in approximately thirty months: a sustained CAGR above 90% on a base growing from sixty billion to four hundred billion, against a model-API pricing curve that just reversed across three providers simultaneously (§5.6). The original Arms Race essay was directionally right; the audit-grade work makes it numerically sharper. This is the cleanest mark-to-market in the report: a pre-existing thesis strengthened by the methodology upgrade.

8.5 Mark-to-market: AI Models Are the New Rebar

The Rebar thesis (AI Models Are the New Rebar) argues model APIs are commoditizing and margin will migrate up the stack toward applications. The reconciled numbers produce a counterintuitive intermediate finding. Per Sections 5.6 and 5.7, the model-API layer is $46.6B gross / $39.6B net of AWS/GCP–Anthropic $7B resale overlap; the application layer is $21.2B. The model-API layer is therefore larger than the application layer by 1.9x even after netting. On dollar revenue today, margin has not yet migrated upstack.

Rebar was consistent with the data through end-2025, properly read. Per the Wayback Machine pricing build summarized in §5.6, flagship input pricing fell ~95.8% over 30 months (GPT-4 → GPT-5; input-only half-life ~6.5 months); the cheapest tier fell ~97.5% over the same window. Volume grew fast enough that the dollar layer kept expanding even as unit price collapsed. Commoditization was happening at unit economics; it had not propagated to the revenue line.

The April 2026 prints complicate the timing. GPT-5.4 input is 2x more expensive than GPT-5 ($2.50 vs $1.25); Claude Haiku 3.5 is 3x the price of Haiku 3 on blended terms [API-pricing snapshots]. Two readings are now live.

Reading A, cost-bound. Rebar's commoditization slowed because reasoning-compute cost is binding. The price floor has reset upward and the model-lab margin floor with it. Customer-acquisition pricing that worked at venture-funded scale stopped working when reasoning workloads pushed unit economics to negative gross margin. Three providers repriced simultaneously because none of them could sustain the prior price; the directional shift is structural.

Reading B, pricing power. Vendors raised prices and customers paid. Two years of 95% compression were margin pass-through during a frontier race; the reversal is the first evidence that the model labs have built moats (eval pipelines, prompt libraries, fine-tuning investments, agent infrastructure) sufficient to price for sustainability. Anthropic 3x'ing the Haiku floor without losing volume is a pricing-power demonstration.

Two quarters of post-GPT-5.4 prints will discriminate. The cost-bound read is more consistent with the audit-grade analysis above: a market repricing on margin necessity rather than pricing power lines up with the constrained-coverage reading; a pricing-power read implies model labs are extracting rents in the layer Rebar said would commoditize, which would not be consistent with the directional claim of the essay. What does not change under either reading is the 1.9x model-API / application ratio. The dollar layer at the model-API tier is larger than at the application tier because the disclosed revenue is what it is.

8.6 The triangulated framework's policy implications

The disclosure-tier framework is more durable than any one number it produces. Four implications follow.

For investors. Use the right tier for the right question. Audit-grade ($63.2B narrow / $72.5B broad) for credit underwriting and capex coverage. Disclosure-grade ($100–135B range, $123B midpoint) for M&A and equity valuation. Reconciled midpoint ($123B) for TAM. The four headline estimates from Section 2 (Menlo $37B, IDC $307B, Gartner $1.478T, "$100B vendor sum") are each correct under their own perimeter. Using the wrong perimeter for the question produces conclusions that look defensible and are not.

For sell-side analysts and founders. Cite the tier system rather than picking one of Menlo, IDC, or Gartner as "the" market number. For founders pitching AI investments, Tier C–E revenue trades at a structural discount to Tier A–B because lenders and acquirers price disclosure risk separately from growth risk, and the discount is now quantifiable.

For the public discourse. When someone says "the AI market is $X," ask which tier. The 40x spread between Menlo's $37B and Gartner's $1.478T (Section 6) is definitional, not measurement error. Only one of the five anchor points along that waterfall is the right number for any given question.

Operator context bracketing the $63.2B floor. Three material uncertainties are not captured in the audit-grade build and bracket the floor in both directions. First, the Azure OpenAI Service take rate sits closer to 25–30% per current channel checks, not the ~20% implied by the report's $11B Microsoft–OpenAI overlap; under the higher take rate, real overlap is $13–15B, which net-adjusts the audit-grade floor downward by $2–4B. Second, custom silicon (Trainium, TPU, Maia) is not quantified in the build; Trainium 2 at ~40–50% of H100 cost per equivalent FLOPS and TPU v5p cost-competitive with H200 reshape the unit-economics conversation for hyperscaler AI margins, material to whether the $63.2B numerator translates to enough contribution margin to service the build. Third, AI customer concentration inside Microsoft's $37B is acknowledged in the channel data: the top ten enterprise accounts plus OpenAI inference resale likely account for 40–50% of the run-rate, a different revenue-quality story than a broad enterprise rollout. Each of these is a Tier C operator detail the report cannot directly verify; each materially conditions whether the $63.2B floor is the right number or merely the best number available at current disclosure quality.

Section 9 introduces the Spread Index v1.0 at 4.28% (narrow) / 4.90% (broad), the recurring metric this report establishes to track how the four headline figures and the audit-grade floor move relative to one another over time. The Index is the gap between numbers tracked across quarterly snapshots. If audit-grade compounds at 100% YoY while disclosed-only and Tier-C-inclusive estimates compound at 50% YoY, the gap closes and capex coverage improves. If audit-grade grows slower than the broader market, the spread widens and the underwriting case deteriorates without any change in the headline numbers most observers cite.

8.7 Framework illustration: cohort applications

The framework above is portable across coverage. What follows applies the layered model and the audit-grade floor to specific listed names as a methodological illustration. The paragraphs below describe how the framework maps onto each company's disclosed fact pattern. They are not opinions on, or recommendations to buy, sell, or hold, any security; they do not assign directional weight to any name; and they should be read together with the disclosures at the end of this report (Colophon).

Microsoft is the name most affected by multi-tier reconciliation. The $37B AI run-rate sits at the cloud layer ($52.5B net of MSFT–OpenAI overlap), a quarter of which is Azure OpenAI Service pass-through that flows downstream to OpenAI's own ARR. Microsoft's standalone 2026 capex contribution runs at $130B+; capex coverage on the audit-grade $63.2B narrow floor lands at roughly half on Microsoft's own segment. The marginal-uplift correction on Copilot ($5–7 over E5, not $30 gross) shaves $2.5–3B off the seat-priced share. Under the framework, the operative variables for Microsoft AI are Copilot seat penetration past the current ~10M base and Azure consumption mix between recognized AI workloads and Azure OpenAI Service pass-through that also recurs in OpenAI's run-rate.

Inside that $37B, Copilot economics deserve a specific caveat. The $30/seat figure is the list price for standalone Copilot on top of E3, not the realized ARPU. Most paid Copilot seats sit inside E5 customers at an incremental price closer to $10–15 per seat per month, reflecting both volume and multi-year discounting and the structural reality that Copilot is sold primarily into the M365 E5 footprint rather than as a standalone SKU. The framing matters because it changes the read on per-seat ARPU deceleration. A declining effective Copilot price is partly evidence of successful cross-sell into E5 where marginal ARPU was always going to be lower (a pricing-power and attach story), not solely evidence of standalone pricing weakness. Both readings are present in the disclosed data; the deceleration reading is the one currently emphasized in sell-side commentary.

Google refuses to segment AI revenue. The audit-grade framework forces every analyst on GOOG to estimate Vertex AI ARR (sell-side consensus \$5–8B, all Tier C). The \$254B silicon vs \$63B enterprise gap implies Google Cloud AI margin contribution is being captured upstream by NVIDIA, not by Google: GOOG buys the GPU stack, sells AI capacity at margin, and the integration spread is the only economics that accrue until TPU mix rises materially. Margin compression risk is real if in-house TPU adoption can't outrun enterprise AI demand growth.

NVIDIA reports the silicon Tier A figure at \$249B Data Center annualized. Under the framework, silicon revenue is upstream of enterprise AI demand recognition: it absorbs the difference between AI capex demand and AI revenue supply at the point silicon is recognised. The framework's constrained-coverage reading describes a scenario in which AI revenue grows slower than amortisation on the disclosed capex base; that scenario operates downstream of silicon volume rather than on silicon volume itself.

Broadcom is the custom-silicon beneficiary cited in passing in §8.6 and deserves explicit treatment here. AI ASIC revenue at ~\$8B/quarter run-rate places AVGO at the Tier A floor alongside NVDA and AMD; it benefits as Maia, TPU, and Trainium scale because Broadcom is the silicon vendor of record on most hyperscaler custom programs. As hyperscalers attempt to compress NVDA dependence, the substitution flows to AVGO, not away from silicon altogether.

CoreWeave is the name where the framework's resale-netting math has the largest cross-vendor effect. The 67% Microsoft customer concentration in the 10-K means the \$5.1B FY2025 revenue is structurally captured by Microsoft's \$37B disclosed AI run-rate; CoreWeave's audit-grade contribution net of resale is \$2.08B. The \$66.8B backlog, the \$17–19B FY26 guide, and the \$30B+ FY27 guide all sit downstream of Microsoft procurement. Under the framework, characterising CRWV's economics turns on the rate of customer diversification past the single-vendor concentration disclosed in the 10-K; the framework is silent on which characterisation is appropriate for any specific investment purpose.

Dell sits at the hardware-OEM cohort. Q4 FY26 AI server revenue ran at ~\$9B (×4 ≈ \$36B annualized), of which 85–90% is NVDA silicon by bill of materials. Net of silicon overlap, Dell's contribution to the audit-grade floor is ~\$3B of integration margin. Under the framework, the operative variables for the OEM exposure are AI server volume and integration margin discipline; AI economic ownership remains upstream at the silicon layer.

SuperMicro is the higher-mix Dell. ~\$33B AI annualized, ~\$3.5B integration margin net, similar 10–12% gross profile. The pure-play AI hardware mix is the highest in the OEM cohort. Under the framework, defensibility within the OEM layer rests on integration speed and reference-design execution rather than on captured AI economic ownership.

Arista is the methodological exception in the OEM cohort. The $3.5B FY26 AI-fabric revenue is full net contribution because Arista is networking IP (7800R3/R4 plus Broadcom Jericho silicon), not GPU resale. Two qualifiers. The $3.5B is FY26 guidance, not yet realized; the next two earnings cycles determine whether the number prints. And within the framework, networking layer durability relative to AI-fabric scaling is the open methodological question for the enterprise-penetration variable outside hyperscaler captive deployment.

Across the names above, CoreWeave is the cleanest illustration of how the audit-grade undercount materialises at a single issuer: a single-vendor AI cloud pure-play whose 67% customer concentration makes the resale-overlap math operationally visible in the 10-K. Microsoft, Google, and NVIDIA are the cohorts for which the framework's measurement consequences are largest at the segment level.

Closing

Strip the four-tier scaffolding away and the arithmetic is the arithmetic. Public sell-side estimates of AI-related debt issuance – Morgan Stanley among them – placing the figure at roughly $1.5T over the coming cycle are being underwritten against an audit-grade revenue base of $63.2B, against a 2028 hurdle that, under 3-year economic amortization for the GPU subset of capex (consistent with public hyperscaler commentary placing GPU economic life at 3–4 years even where GAAP useful-life accounting runs longer), sits at $400B, not $250B. The implied compounding is 90%+ CAGR on a base scaling from sixty billion to four hundred billion over thirty months, against a model-API pricing curve that reversed for the first time in two years and against a hyperscaler cohort most of whose members refuse to disclose AI revenue at segment granularity. Coverage at the narrow audit-grade floor sits between 9.2% and 14.0%, half to a third of the 28% telecom peak that ended in the worst infrastructure-financing wipeout of the prior generation. The bull case is not impossible; at currently observed growth rates it is mathematically reachable. It is also extraordinary, and at this scale of debt issuance, extraordinary claims get repriced by force rather than by argument. The next four 10-Q seasons answer the question this report cannot.

Section 9

The Spread Index

Audit-grade ÷ umbrella as a tracking ratio; two variants and the noise floor.

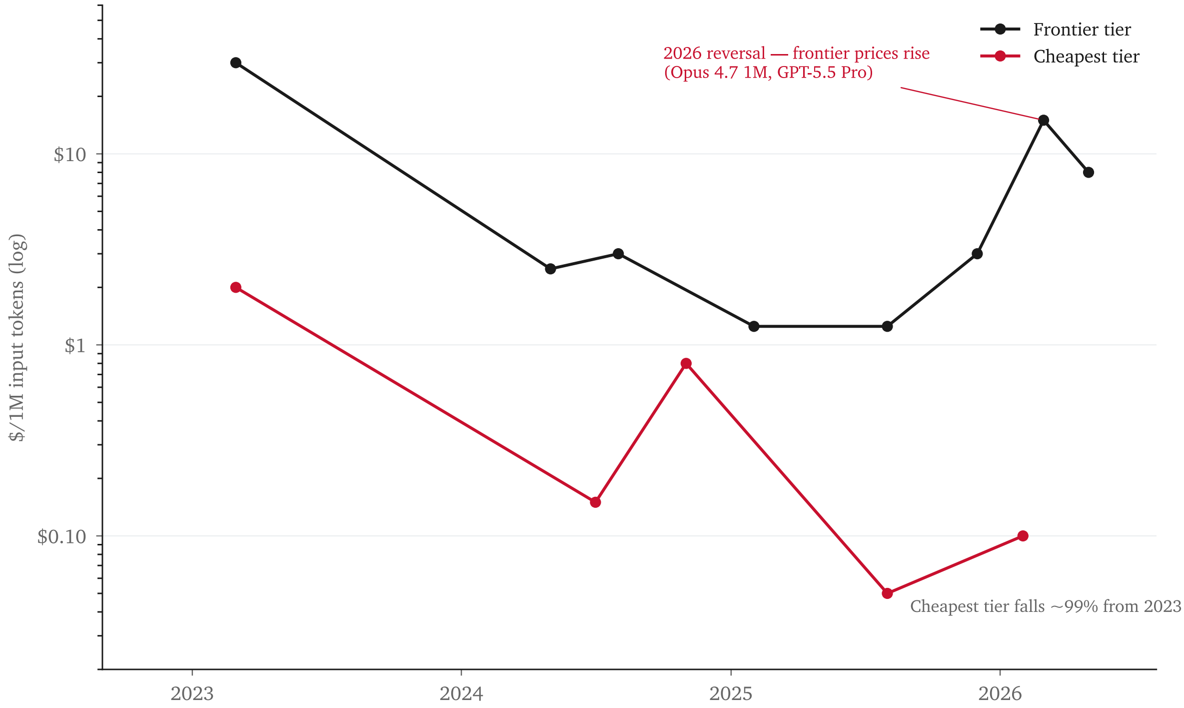


FIGURE 9 *API pricing on a log scale, 2023–2026. Frontier-tier $/1M input tokens reverses in 2026 with Opus 4.7 1M and GPT-5.5 Pro reasoning; cheapest-tier continues to fall.*

The report's central finding compresses to a single ratio: the share of the cited "enterprise AI market" that exists at SEC-audited segment granularity, versus the share that exists at lower disclosure standards. That ratio is the headline metric this report introduces and commits to maintain on a quarterly cadence.

9.1 The metric

Name: the Enterprise AI Disclosure Spread, abbreviated *the Spread*.

Definition: the ratio of audit-grade enterprise AI revenue to the umbrella worldwide AI spending figure published by Gartner, expressed as a percentage. At v1.0 baseline (2026-05-14), the Spread is the audit-grade floor of **$63.2B narrow / $72.5B broad** over the Gartner umbrella of **$1.478T**, or **4.28% narrow / 4.90% broad**.

The Spread is the single number that captures this report's central finding: roughly 95% of the cited "enterprise AI market" sits below the disclosure standard that SEC-audited segment reporting requires. The metric is one-directional by construction. A falling Spread means *more* of the headline is audit-grade, not less. It is intended to be tracked across quarters as a discipline check on the gap between cited and disclosed AI economic activity.

9.2 Methodology

Numerator: the audit-grade tier from the latest vendor census, computed per Step 6 of the reconciliation build: Tier A + Tier B disclosures, net of silicon double-counting, net of MSFT–OpenAI and AWS/GCP–Anthropic resale overlap. At v1.0 this is **$63.2B narrow** (Tier A+B post-ServiceNow and post-Arista removal, both excluded as FY26 guidance targets rather than realized run-rate) or **$72.5B broad** (adding CoreWeave-net $2.08B and OEM-net $7.25B, where the OEM cohort excludes Arista on the same methodological grounds as ServiceNow).

Denominator: the most recent published Gartner *Worldwide AI Spending Forecast* (or equivalent umbrella if Gartner ceases publication). At v1.0 this is the September 17, 2025 release at $1.478T for calendar 2025.

Update cadence: quarterly, after each major hyperscaler 10-Q cycle concludes. Snapshot dates align with the Apr / Jul / Oct / Jan reporting windows. The full `vendors.csv` is republished alongside each Spread reading. The threshold for a v1.1 revision between scheduled updates is a material change in any Tier A or Tier B disclosure: for example, Google beginning to segment-report Cloud AI revenue, or Microsoft separating its AI run-rate from total Productivity & Business Processes.

Note: API pricing reversed in early 2026 (see §5.6). Spread Index v1.1 may incorporate this as a separate sub-metric (a unit-price index across the three flagship providers, tracked alongside the disclosure-tier ratio) once one more quarter of pricing data confirms whether the reversal is durable.

9.3 v1.0 baseline reading

Spread (v1.0, snapshot 2026-05-14): 4.28% narrow / 4.90% broad.

- Audit-grade enterprise AI revenue: **$63.2B narrow / $72.5B broad** (Tier A+B, net-of-silicon, net-of-resale; ServiceNow $1.5B and Arista $3.5B both excluded as FY26 guidance targets rather than realized run-rate)
- Disclosure-grade range (Tier A through E): **$100–135B**, midpoint ~**$123B**
- IDC AI Solutions (bundled, demand-side): **$307B**
- Gartner Worldwide AI Spending (umbrella): **$1.478T**

The Spread is most stable at the numerator and denominator extremes. Tier A+B revenue and the Gartner umbrella are both anchored in published methodologies. The disclosure-grade range is the report's softer middle. Section 10 documents how each component moves under reasonable alternative assumptions and shows the resulting band on the headline ratio.

9.4 Two variants

A reviewer with a systematic-trading lens pressed the point that the Spread is interesting as a framework but immature as a level. At one observation and no history, the level reading carries no time-series content; the underlying mechanic – disclosure-tier change-of-state across listed AI-exposed names – is the stronger systematic content, and it can be operationalized cross-sectionally without waiting for the level series to mature. The two readings are complementary rather than alternative, but they answer different questions and are worth distinguishing.

Variant 1: Spread Index (level). The ratio measure as defined above. At v1.0 baseline (2026-05-14) the level reads **4.28% narrow / 4.90% broad**, with audit-grade revenue of $63.2B / $72.5B over a Gartner umbrella of $1.478T. The level variant is designed to track how the disclosure gap moves through 10-Q seasons – falling as more vendor revenue migrates into Tier A/B segment-grade disclosure, rising as the umbrella outruns disclosure. Interpreting it as a quarterly tracking signal requires a baseline period (§10.4 develops this), and the working assumption is that four to six quarters are needed before quarter-over-quarter movement can be distinguished from measurement noise. Until that window is built, the level is a cross-sectional descriptor, not a time-series.

Variant 2: Disclosure-Tier Migration (cross-sectional). A factor-model variant of the same framework, operating on company-level tier transitions rather than on aggregate ratios. For each AI-exposed listed company, track tier transitions across successive 10-Q cycles:

- **B→A migration:** the company begins segment-reporting AI revenue at audit-grade granularity (e.g., Google Cloud AI broken out as a reportable segment line rather than discussed qualitatively).
- **C→B migration:** the company's earnings-call AI commentary moves from qualitative ("AI is a growth driver") to quantified ("$Xb AI run-rate") with management standing behind a dollar figure.
- **B→C "guidance pull":** the company stops disclosing a previously stated AI revenue figure or downgrades to qualitative commentary – the disclosure-quality analogue of pulling guidance.

The migration variant carries more immediate signal than the level variant because it operates on multiple companies per quarter rather than waiting for a level baseline to accumulate. Aggregated across 50–100 listed names per 10-Q cycle, the migration count gives an earlier cross-sectional reading than the level Spread, which needs four to six quarters of observations before quarter-over-quarter movement becomes interpretable. The level Spread is the macro reading; the migration variant is the cross-section. The migration variant is presented as a measurement construct; the report takes no view on whether it would produce excess returns and does not propose its use as a trading strategy.

Empirical reproducibility. The migration variant is in principle reproducible against 2023–2025 disclosure data, where the tier transitions of the largest names are already on the record. Microsoft's Azure AI run-rate was first disclosed at roughly $13B in Q3 FY24 (a C→B migration), grew across subsequent quarters, and printed at $37B+ by Q3 FY26 – a clean trail of disclosure events at a single company. AWS's $15B AI run-rate disclosed in Q4 FY25 is a similar C→B event; Google Cloud and Meta AI sit in Tier C and a future B-tier print would be a recordable migration. A reproducibility exercise covering 2023-Q1 through 2026-Q1 would generate roughly twelve quarterly observations – insufficient for any production use, but enough to determine whether the disclosure-tier signal is empirically stable.

A full reproducibility exercise is beyond this report's scope and would require an NLP classifier that reproduces the manual tier classification at scale (roughly four to six weeks of engineering on top of a labeled training corpus, which the public vendor census in this report can seed). The report's contribution is the measurement framework. Any application of the framework to investment decision-making is outside the report's scope and is the responsibility of the reader. The two-variant framing makes explicit what the level reading cannot yet do and what the cross-sectional reading can do today.

Section 10

Limitations and sensitivities

A sensitivity table stress-testing the framework across base, bear, and bull scenarios; what would falsify the central finding.

The report is built from public sources only, and that constrains what can be verified. This section documents the structural limits of a public-source reconciliation, the secondary limitations surfaced during adversarial review, and how the headline numbers move under reasonable alternative assumptions.

10.1 Structural desk-research limits

Three limits define the verifiable boundary of a public-source reconciliation.

1. Granular IDC/Gartner taxonomies. Topline definitions for IDC AI Solutions and Gartner Worldwide AI Spending are public; full line-item methodologies sit behind paid-report paywalls. IDC and Gartner have not reviewed or endorsed this reconstruction. Where the report describes their inclusion/exclusion rules, it relies on public statements, press releases, and tertiary analyst summaries, with reasonable inference where direct documentation is not available. The Methodological Crosswalk in Section 3 is best-public-reconstruction work, not a line-item audit of either firm's internal taxonomy. Materiality is low. The headline four-anchor categorization (Gartner umbrella, IDC bundled, vendor sum, Menlo demand) holds and the 40x spread is robust to taxonomy detail. Sub-category attribution could shift \$20–50B between layers without disturbing the structure.

2. AI-attributable cloud revenue. Neither Microsoft, Google, nor AWS cleanly discloses "AI revenue" in audited filings. Microsoft's \$37B AI run-rate and AWS's \$15B are CFO/CEO statements in earnings calls (Tier B), not 10-Q segments. Google Cloud AI and Meta AI are Tier C: mentioned as growth drivers without a dollar figure. Any cloud-layer number is an analyst estimate, not a disclosed one. Materiality is high. The cloud layer ranges from \$55B to \$90B depending on how generously the Tier C universe is populated. This is the single largest source of uncertainty in the report, and it is captured in the sensitivity table at Section 10.3.

3. GPU fleet sizes. Proprietary estimates only: Omdia, SemiAnalysis, BofA/Arya public excerpts. Used in the Revenue-per-H100 sub-section of Section 5 and the cloud-capex-per-AI-revenue analysis in Section 8; not load-bearing for the main reconciliation or the Spread. Materiality is medium for that specific analysis, low for the headline. The pure-desk Cloud Capex per Disclosed AI Revenue chart provides a fleet-independent alternative for the same comparison.

Census additions that bear on Tier A framing. Two additions in this revision relax the earlier framing that Tier A was entirely silicon. CoreWeave is now treated as a Tier A non-silicon vendor, though the bulk of its revenue is a Microsoft channel relationship and overlaps with MSFT-disclosed AI run-rate when netted. The hardware OEM cohort (Dell, HPE, SMCI, Pure Storage and peers) adds roughly \$73B gross of AI-attributed revenue and about \$7.25B net of silicon overlap (Arista excluded from the audit-grade build as FY26 guidance rather than realized run-rate, on the same logic that excluded ServiceNow's \$1.5B FY26 ACV target), contributing roughly \$7–10B once the rest of the netting is applied. The audit-grade floor is therefore a band (\$63.2B narrow / \$72.5B broad) rather than a point estimate.

10.2 Other limitations surfaced during review

The Devil's Advocate review identified additional limits. These are not structural to desk research but they bound the precision of specific claims.

Hyperscaler AI-attribution refusal. Google, Meta, and Oracle disclose no AI revenue. The cloud-layer estimates in Section 5.5 for these vendors are partly conjectural. If any one of them began disclosing at Tier B granularity, the Spread could shift materially in either direction.

Principal-versus-agent recognition uncertainty. Whether channel-routed inference revenue should be booked gross by the model provider or net of the channel share is an unsettled ASC 606 question across the private model-lab cohort. The reconciliation assumes \$11B of MSFT–OpenAI overlap and \$7B of AWS/GCP–Anthropic overlap as the report's analytical assumption; alternative recognition treatments could plausibly move either figure by ±\$7B. Nothing in this section should be read as asserting that any named vendor has misstated revenue under applicable accounting standards.

Embedded AI feature pricing. Microsoft 365 Copilot bundles at \$30/seat. The question is whether the full \$30 counts as AI revenue, only the marginal uplift versus M365 E5 (~\$5–7), or zero on the grounds that the AI is a feature rather than a SKU. Section 4 follows Menlo's treatment and excludes bundled features. Microsoft's disclosed run-rate implies the marginal-uplift treatment. Other reasonable treatments would add \$5–15B to the headline.

ARR vs. recognized revenue. The private vendor census mixes ARR (forward-looking run-rate) with recognized revenue (trailing GAAP). For fast-growing private vendors, ARR systematically overstates trailing recognized revenue by 30–60%. Applied broadly across the private cohort, a strict recognized-revenue basis would compress the fully-net total by 10–15%. Treated as an explicit sensitivity in Section 10.3.

Stale figures. Several Tier D–F private entries (Writer at \$47M, Character AI at \$50M, You.com at \$50M) are 6+ months old. Those rows have not been refreshed. Net effect on the headline is small (<\$200M), but the precision is lower than the public-company rows.

Deliberate omissions. The census does not attempt to size the 200+ vertical AI startups under $20M ARR each (Cresta, Hippocratic, Ambience, Suki, EvenUp, Eve, Legora, Clay, 11x, and so on). A Cahn-style summation of that tail could add $15–25B of below-the-disclosure-floor enterprise AI economic activity. This is a known omission, not a measurement error.

10.3 Sensitivity table

The four-tier numbers under eight scenarios:

SCENARIO	AUDIT-GRADE	DISCLOSURE-GRADE	IDC BUNDLED	GARTNER UMBRELLA	SPREAD
Base case, narrow Tier A (v1.0, 2026-05-14)	$63.2B	$100–135B	$307B	$1.478T	4.28%
Base case, broad Tier A (OEM cohort net, ex-Arista)	$72.5B	$100–135B	$307B	$1.478T	4.90%
Maximum resale netting (MSFT–OpenAI overlap at $25B, AWS/GCP–Anthropic at $12B)	$49B	$85–115B	$307B	$1.478T	3.32%
Minimum resale netting (assume zero overlap; naïve summation)	$85B	$115–145B	$307B	$1.478T	5.75%
Cloud AI mid-estimate populated (Google Cloud AI $8B + Meta AI $5B + Oracle Cloud AI $3B at Tier B)	$75.5B	$116–151B	$307B	$1.478T	5.11%
Embedded AI feature pricing recognized ($12B for M365 Copilot, Salesforce Einstein bundle, Adobe Acrobat AI)	$71.5B	$112–147B	$307B	$1.478T	4.84%
ARR-to-recognized-revenue haircut (private vendor ARR compressed to trailing GAAP, ~25% haircut)	$63.2B	$87–122B	$307B	$1.478T	4.28%
AI-incremental capex denominator ($450B incremental, not $690B gross)	$63.2B	$100–135B	$307B	$1.478T	4.28%
Bull: blended capex amortization (shell/power 20–30yr, networking 5–8yr, GPU 3–4yr weighted)	$63.2B	$100–135B	$307B	$1.478T	4.28%
Bull: embedded AI at fair value (embedded AI in incumbent SaaS bundles recognized at $50–80B vs $15–25B)	$63.2B	$135–190B	$307B	$1.478T	4.28%
Bull: sovereign AI inclusion (Stargate, G42, HUMAIN, DEEP, EU/APAC sovereign at $30–50B added to enterprise denominator)	$63.2B	$130–185B	$307B	$1.478T	4.28%
Arista FY26 target included (re-add $3.5B if treated symmetrically with realized run-rate; sensitivity only)	$76B	$100–135B	$307B	$1.478T	5.14%

SCENARIO	AUDIT-GRADE	DISCLOSURE-GRADE	IDC BUNDLED	GARTNER UMBRELLA	SPREAD
Contribution-margin coverage (audit-grade revenue at ~25–35% contribution margin against $690B capex; the actual credit-committee metric)	$16–25B contribution	$100–135B	$307B	$1.478T	n/a (capex coverage 2.3–6.4%)
Contracted backlog (RPO) coverage (Big 5 contracted RPO ~$800B, of which 35–50% AI-attributable, against $690B capex)	$280–400B contracted AI RPO	$100–135B	$307B	$1.478T	n/a (capex coverage 40–58%)

The first eight scenarios move the Spread ratio across a band of roughly 3.0% to 5.2%, about ±25% around the 4.28%/4.90% base cases. The three bull scenarios at the bottom of the table move different quantities — the forward capex-coverage hurdle in the first case, the disclosure-grade total in the second and third — and are sized for their effect on the figures the bull case actually engages with rather than for the 2025 Spread, which they leave unchanged. The Arista sensitivity row shows the effect if Arista's $3.5B FY26 raised AI-fabric target were re-included alongside realized run-rate; the report excludes it for consistency with ServiceNow but the addition would lift the broad Spread by 24bp. The IDC and Gartner anchors do not move under any scenario tested here because they are published-methodology figures whose internal taxonomies this report does not attempt to revise.

Excluding the bull rows, the distribution of the original eight scenarios is asymmetric: most push the audit-grade tier *above* the narrow base, not below. That asymmetry was flagged in adversarial review as a one-sided framing. The three bull-case rows added in this revision (blended amortization, embedded-AI fair value, sovereign AI) rebalance the table by surfacing scenarios in which the implied 2028 hurdle drops, the disclosure-grade total expands, or the denominator broadens. The narrow base case at $63.2B still sits closer to the conservative end of the audit-grade band than to its midpoint, and that remains the report's framing for the 2025 cross-section — but the forward math and the disclosure-grade math are both materially softer under the bull scenarios than the original eight rows suggested.

Three of the scenarios above bear on questions adjacent to the Spread rather than on the Spread itself, and need separate discussion.

The ARR-to-recognized-revenue haircut compresses the disclosure-grade total by $10–15B if private vendor ARR figures sit 20–30% above trailing recognized revenue (typical SaaS lag). The audit-grade tier is barely affected because Tier A is dominated by public-company recognized revenue (silicon, OEM, CoreWeave) rather than private ARR. The Spread moves less than the disclosure-grade range; the ratio is anchored on the numerator's audit quality, and the denominator (Gartner umbrella) is unchanged. The disclosure-grade range, where readers calibrate against the vendor sum directly, is the figure that compresses most.

GPU 3-year amortization is not a sensitivity to the headline Spread (both numerator and denominator are 2025 figures), but it surfaced in Bull, Bear, and SME reviews and is worth noting for the capex-coverage discussion in Section 8. A 3-year economic life, versus the 6-year accounting convention some hyperscalers use, would require 2028 enterprise AI revenue to grow from a roughly $250B trajectory to closer to $400B to cover capex on a depreciation-aligned basis. The 2025 Spread does not move; the forward gap widens substantially under shorter useful lives.

The AI-incremental capex denominator question matters for Section 8 rather than for the Spread. The $690B+ 2026 hyperscaler capex figure includes substantial non-AI infrastructure spend. Capex coverage at the $63.2B narrow floor is 9.2% against gross capex and 14.0% against the $450B AI-incremental midpoint (Bernstein/UBS/Citi bottom-up estimates sit in a $400–500B range). Switching denominators does not change the Spread (the umbrella is unchanged), but it materially changes the implied capex-coverage math.

Blended capex amortization is the bull's strongest analytical move against the 2028 hurdle. The Section 8.1 math runs a uniform 3-year economic life across the full AI-attributable capex stack and arrives at a roughly $400B required 2028 audit-grade revenue figure. The composition argument is that the $690B headline (or the $400–500B AI-incremental subset) is not homogeneous: GPUs and accelerators sit at $200–250B with a 3–5 year economic life, networking, cooling and custom silicon at roughly $150B with a 5–8 year life, data center shell, power infrastructure, transformers and real estate at $200–300B on 20–30 year schedules, and fiber and interconnect at $50–100B on 25+ year schedules. Weighted across that stack, blended annual depreciation falls from roughly $200B under uniform 3-year accounting to $80–100B under economically realistic useful lives. The 2028 audit-grade hurdle that follows from the blended figure is $180–220B, not $400B – reachable at 50–60% CAGR from the 2025 base rather than requiring 90%+ growth. The data supports the composition split (hyperscaler 10-Ks disclose useful-life ranges of 4–6 years for servers, 10–25 years for buildings and 20+ years for power infrastructure); what it does not yet support is a clean public attribution of the $690B headline to those buckets at the line-item level, which is why the bull figure is given as a range. The honest read is that the 3-year-on-everything framing in Section 8.1 overstates the hurdle and the blended figure is the more defensible forward number.

Embedded AI feature pricing recognized at fair value tests the bull's order-of-magnitude critique of the Tier C estimate. The report currently pins embedded AI inside incumbent SaaS bundles at $15–25B, capturing M365 Copilot's marginal uplift, the Salesforce Einstein bundle, and Adobe Acrobat AI at conservative attribution. The bull argument is that this aggregates a much wider set of bundled value: Copilot uplift across the M365 E5 base (bundled into seats but not separately disclosed), Einstein-inside-Sales Cloud at $3–5B of implicit bundled value, Adobe AI inside Creative Cloud, Workday's AI Agent System, ServiceNow Now Assist, Oracle's Fusion AI features, and the Cortex/Breeze/Firefly perimeter – totalling $50–80B of AI-attributable willingness-to-pay across the incumbent SaaS bundle. Recognizing that figure at fair value would expand the disclosure-grade total by $35–55B, compressing the Spread Index accordingly. The data supports the existence of all of these features and supports buyer-survey evidence that AI features are a material reason for E5 retention and Sales Cloud upgrade. What it does not directly support is the $50–80B aggregate, which is a willingness-to-pay reconstruction rather than a disclosed figure; vendors with the visibility to confirm or refute it have chosen not to disclose at that granularity. This is the single largest known undercount in the reconciliation, and the bull's view that the report's $15–25B sits an order of magnitude below the defensible upper bound is a reasonable read of the public evidence.

Sovereign AI inclusion is a definitional question rather than a measurement one. The report's enterprise perimeter excludes government-channel buyer demand, which is conventional in SaaS-comp work but increasingly arbitrary in a market where 2025–2026 sovereign initiatives carry material commitments: Project Stargate ($100B–$500B across phased commitments by Oracle, MGX, SoftBank, and OpenAI, with $30–50B of in-period 2025–2026 activity), the UAE G42 partnerships at $25–50B over coming years, the Saudi HUMAIN initiative at $10–20B early-stage, the UK DEEP programme at $5–15B announced, French and German sovereign cloud builds at $5–10B, and combined Japan, Korea, India and Singapore sovereign AI commitments at $5–15B. Counted in-period rather than as multi-year headline figures, sovereign AI is $30–50B of 2025–2026 disclosed and pipeline activity. Including it in the disclosure-grade total expands that tier by the same magnitude and shifts the coverage ratios in Section 8 accordingly. The data supports the headline commitments (each is announced, most are partially contracted); what it does not yet support is clean separation between disclosed in-period revenue and multi-year pipeline. The bull's structural point – that a 2026-snapshot reconciliation of AI demand which omits sovereign entirely is incomplete – is correct, and this scenario quantifies what its inclusion would do.

Contribution-margin coverage is what private-credit underwriters actually compute, and the gross-revenue coverage ratio in §8 overstates the relevant cushion by roughly two to three times. Hyperscaler AI revenue runs at cloud-typical gross margins of 60–70%, less the operating cost of AI-specific compute (40–50% of gross), which yields a contribution margin of approximately 25–35% of disclosed revenue. Applying that band to the audit-grade floors: at the $63.2B narrow floor, contribution margin is $16–22B against $690B 2026 capex, or **2.3–3.2% coverage**. At the $72.5B broad floor, $18–25B against $690B is **2.6–3.6%**. At the $126B disclosure-grade midpoint, $32–44B contribution margin is **4.6–6.4%**. This is the framework's lowest coverage ratio and the most directly relevant to credit-committee decisions, where the underwriting math runs on parent free cash flow available for debt service rather than on disclosed segment revenue. A PC desk that uses the headline 9.2% gross-revenue coverage in place of the contribution-margin reading is overstating the underwriting cushion by a factor of two to three. The contribution-margin figure does not replace the headline narrow/broad audit-grade framing; the headline answers “how much of the cited market is audit-grade?” The contribution-margin row answers “how much of the build can the audit-grade tier plausibly service?”

Contracted backlog (RPO) coverage is the denominator the bull case actually leans on, and it sits well above any revenue-based reading. The Big 5 hyperscalers disclose roughly $800B in contracted remaining performance obligations as of Q1 2026: Microsoft at $315B+, Amazon at ~$200B, Google at $145B+, Meta at ~$80B, and Oracle at ~$60B, with weighted-average tenors of 2–5 years depending on issuer. Bernstein, UBS and bottom-up PC desk models put the AI-attributable share at 35–50%, which implies **$280–400B of contracted AI revenue backlog**. Measured against $690B of 2026 capex, that produces **40–58% coverage** on a contracted basis — the bull’s strongest underwriting argument. The structural point is that not all $690B needs to be repaid out of realized future revenue; a material share of the build is already contracted to investment-grade counterparties at multi-year tenors, and debt amortization runs on contract life rather than on P&L recognition timing. The headline narrow coverage of 9.2% prices the question “is this revenue already audit-grade?”; the RPO coverage row prices the question “is this capex already contracted out?” Both are legitimate framings and they answer different questions. The audit-grade view remains conservative and the RPO view remains optimistic, and the honest read is that the underwriting cushion sits somewhere between the two depending on assumed AI-attribution rates and contract-tenor distributions.

10.4 The Spread Index requires a baseline period

A statistical caveat carries over from review. Spread Index v1.0 is one observation. Estimated measurement uncertainty around the ratio (roughly ±1 percentage point under the resale-netting and Tier C population scenarios above) is larger than the quarterly signal a reader would plausibly want to extract (~50 basis points of quarter-over-quarter movement). The Index is not yet interpretable as a quarterly tracking series. A baseline period of four to six quarters is required before quarter-over-quarter changes can be distinguished from measurement noise. Until that window is built, the Index should be read as a cross-sectional descriptor of where audit-grade revenue sits inside the published umbrella, not as a time-series signal.

10.5 What this report cannot verify

The reconciled estimate is empirically defensible to roughly ±20% under reasonable variations in resale-netting, cloud AI attribution, embedded-feature pricing, and ARR-vs-recognized assumptions. The audit-grade band ($63.2B narrow / $72.5B broad) and the Spread (4.28% narrow / 4.90% broad) are the report’s most stable findings. The disclosure-grade range ($100–135B, midpoint $123B fully-net) is bounded by the choices documented above. Three things this report does not verify: the line-item composition of IDC and Gartner’s internal taxonomies, the AI-attributable share of hyperscaler cloud revenue at vendors that refuse to disclose, and the proprietary GPU fleet inputs underlying the revenue-per-H100 analysis. Those gaps are handled by range presentation and the four-tier disclosure framework, which is also why this report does not assert a single number.

APPENDIX D

International vendor cohort

EU and APAC AI revenue disclosures outside the core 68-vendor census.

D.1 Why this appendix exists

The v1.4 main census is bounded by US public-disclosure conventions. Forty-two of 68 vendor rows are US-listed; the China cohort was added in Phase 1.5; European, Japanese, Korean, Indian, and ASEAN vendor sets are absent or implicit. The six-tier A-F framework travels cleanly, but the populated dataset was not extended internationally. This appendix documents what a global extension would add. It is not integrated into the headline Spread Index because disclosure quality varies across jurisdictions and a clean Tier A–B build would require 2–4 additional weeks beyond v1.4's scope. Figures below are estimates drawn from public earnings commentary, named-source journalism (Sacra, The Information, Reuters, TechCrunch), and analyst back-calculations, documented at the same Tier C–E discipline as the long-tail rows of the main census.

D.2 European cohort

Estimated total ~$5–8B AI revenue at face value; ~$2–4B audit-grade-equivalent after the same haircut discipline used elsewhere.

- **SAP** — Business AI / Joule. Attach rates and AI-influenced ACV referenced on Q1 2026 and Q4 2025 calls; no clean dollar run-rate segmented. Tier C. Estimated $0.8–1.5B FY2025.
- **Mistral AI** — French frontier model lab. Estimated ~$300M ARR exiting 2025 per Sacra / Reuters / TechCrunch. Tier D. Comparable to Cohere ($240M, in main census).
- **DeepL** — Translation AI. Estimated ~$300M ARR FY2025 per The Information. Tier D.
- **Aleph Alpha** — German enterprise / sovereign AI. Tier D–E. Estimated ~$30–100M ARR.
- **Helsing** — European defense AI. Estimated $100M+ ARR per Series D disclosures (valuation >€4B). Tier D. Belongs in the §4.6 defense cohort that the main report populates only with US names.
- **Klarna** — AI customer service inside BNPL. Tier C — referenced on calls as rationale for headcount cuts, not quantified.
- **Adyen** — Payments ML for fraud and authorization. Tier C.
- **N26** — German digital bank. Tier E.
- **OVHcloud** — French cloud + AI infrastructure. Estimated ~€100M AI-attributable. Tier C.
- **Telefónica Tech AI** — Spanish telco AI services. Tier C–D.
- **Scaleway** (Iliad Group) — French sovereign cloud + AI GPU capacity. Tier E.

Cohort total: ~$1.5–2.5B audit-grade-equivalent floor; ~$3–5B at face value.

D.3 Japan cohort

- **NTT** (incl. NTT Data) – largest Japanese telco / cloud / IT services. AI revenue folded into broader IT services. Tier C. JEITA / IDC Japan estimate the enterprise AI services market at ~$5.4B 2025, NTT the share leader.
- **Fujitsu** – AI services and infrastructure (Fugaku-related work, gen-AI consulting). Tier C–D. Estimated $1–2B AI-attributable.
- **SoftBank** – Arm exposure (already in main census) plus Stargate Japan, adding sovereign-class commitment of ~$15–20B in-period 2026. Tier C at parent; Tier B at Arm.
- **Rakuten** – Cloud + retail AI + Rakuten Symphony for telco infrastructure. Tier D–E.
- **Sony AI** – Gaming and entertainment AI. Tier E.
- **NEC** – Enterprise AI + biometrics. Tier C–D. ~$500M–$1B estimated.

Cohort total: ~$3–5B audit-grade-equivalent.

D.4 Korea cohort

- **Naver** – Korean platform incumbent with Naver Cloud and HyperCLOVA X. Most disclosure-transparent Korean AI vendor. Tier C–D.
- **Samsung Gauss** – In-house AI model embedded in Galaxy devices and enterprise products. Tier D–E.
- **LG ExaOne** – Enterprise AI platform. Tier D.
- **SK Hynix** – Major HBM3e supplier. HBM revenue plausibly $15–20B annualized FY2025, >90% AI-attributable. Tier B by analogy to Broadcom custom silicon. Excluded from the audit-grade enterprise floor because HBM is silicon-layer.
- **Kakao** – Platform + KakaoBrain AI. Tier E.
- **SK Telecom** – Telco AI (A. platform, Korean LLM consortium). Tier D.

Cohort total (excluding silicon layer): ~$2–4B audit-grade-equivalent.

D.5 India cohort

Structurally services-heavy. Most “AI revenue” sits inside the IT services bucket Gartner / IDC strip at the top of the waterfall.

- **TCS** – AI / Gen-AI services bookings. Estimated $3–5B AI-related services FY2025. Tier C–D.
- **Infosys** – Topaz AI platform + services. Estimated $2–3B. Tier C–D.
- **Wipro** – AI and consulting services. Tier C.
- **HCL Tech** – AI services. Tier C–D.

- **Reliance Jio** — Sovereign-adjacent initiatives including JioBrain and the Reliance / Nvidia partnership (~$10B multi-year). Tier D.
- **Bharti Airtel** — Telco AI exposure. Tier E.

Cohort total: ~$5–8B services-heavy.

D.6 ASEAN

Sea Group (Shopee, SeaMoney, Garena), Grab, GoTo, and Singtel / NCS all sit at Tier E with embedded rather than disclosed AI revenue. Cohort total: ~$1–2B.

D.7 Cohort totals

REGION	AUDIT-GRADE-EQUIVALENT	UPPER FACE-VALUE
Europe	~$2–4B	~$5–8B
Japan	~$3–5B	~$5–7B
Korea	~$2–4B (ex-silicon)	~$4–6B
India	~$5–8B	~$8–12B
ASEAN	~$1–2B	~$2–3B
Total non-US ex-China	**~$13–23B**	**~$24–36B**

This is in addition to the China cohort already in the v1.4 census at ~$12.5B post-haircut (~$17.6B at face value).

D.8 Global capex correction

The headline $690B 2026 hyperscaler capex (MAGA + Oracle) excludes large non-US capex pools:

- Chinese hyperscalers (Baidu, Alibaba, Tencent, ByteDance, Huawei): ~$60–90B.
- Japanese telco / SI AI infrastructure (NTT, KDDI, SoftBank Stargate co-investment, Fujitsu, NEC): ~$15–25B.
- Korean conglomerate AI capex (Naver, Kakao, SK Telecom, LG AI Research, ex-foundry): ~$20–30B.
- European cloud capex (OVHcloud, Deutsche Telekom STACKIT, Atos / Eviden, EU AI Factories): ~$15–25B.
- Sovereign AI 2026 portion of multi-year programs (G42 / Stargate UAE, HUMAIN Saudi, UK Compute, DEEP EU): ~$50–80B.
- ASEAN + India infrastructure (Reliance Jio, Singapore, Indonesian initiatives): ~$15–25B.

Total global AI capex 2026: ~$865–965B, materially higher than the US-only $690B headline. Midpoint $915B is roughly 33% above headline, consistent with the devil's-advocate $800B–$1.1T order-of-magnitude estimate.

D.9 Implications for the Spread Index

If non-US cohorts (ex-China, already counted) are integrated into both numerator and denominator: add $13–23B to global audit-grade-equivalent revenue (reaching ~$76–95B combined with the v1.4 narrow floor and the China cohort); add $175–275B to global AI capex (reaching $865–965B). Resulting coverage: midpoint $85B / $915B = **~9.3%**, essentially unchanged from the US-centric headline (~9.2% against the $690B hyperscaler base).

The non-US cohort does not fundamentally change the headline finding. The ratio is approximately scale-invariant because numerator and denominator both scale with the geographic perimeter. What the cohort *does* change is the framing: the report shifts from "US-centric reconciliation with a partial China overlay" to "globally validated framework, populated heavily in the US and partially elsewhere." That is the contribution of this appendix.

D.10 Methodology note and uncertainty

The figures above are estimates. Non-US vendors disclose AI revenue with significantly less granularity than US public filings: SAP's Joule and NTT's AI services are referenced in IR commentary but not segmented; Mistral, DeepL, and Aleph Alpha disclose through founder commentary or named-source journalism. The audit-grade-equivalent floor applies the same six-tier discipline used in the main census — the lower bound of each range counts only Tier C and stronger; the upper bound takes Tier D figures at face value. A full Tier A–B build would require 2–4 weeks of primary research (translated filings, regional IR materials, JEITA / IDC Japan / Korea AI Industry Association cross-checks) beyond v1.4's scope. FX conversion uses spot rates as of 2026-05-14; the Spread Index *ratio* is approximately FX-neutral but the cohort *levels* are not. Sovereign AI procurement is excluded as a separate category, consistent with §4.6.

Readers in Frankfurt, Tokyo, Seoul, Singapore, or Mumbai should treat Appendix D as a portability demonstration rather than a regional census. The framework grades a Naver HyperCLOVA X disclosure the same way it grades a Microsoft Intelligent Cloud line; the populated rows here are illustrative of what a region-specific build would yield, not a substitute for it.

Colophon

About the author & disclosures

PHILIPP D. DUBACH

Strategy consultant and independent researcher trained at Imperial College London, working at the intersection of quantitative finance, machine learning, and AI economics. Peer-reviewed output on prediction-market microstructure, volatility regimes, and glycemic modeling sits on SSRN and arXiv. A weekly note on AI and markets, alongside more than eighty essays since 2024, runs at philippdubach.com.

CONTACT

me@philippdubach.com

THEORETICAL ANTECEDENTS

Disclosure economics builds on Verrecchia (1983), Dye (1985), Diamond & Verrecchia (1991), and Beyer, Cohen, Lys & Walther (2010). q-theory framing follows Hayashi (1982), Abel (1983), and Bond & Cummins (2001). ASC 280 segment-reporting and ASC 606 principal-versus-agent treatment frame the audit-grade tier; the contribution here is operational structure layered onto these literatures.

AUTHOR DISCLOSURES & LIMITATIONS OF USE

- ***Independence & status.*** Switzerland-domiciled independent researcher. No investment-advisor, broker-dealer, or financial-research registration with the SEC, FINRA, FCA, FINMA, BaFin, or AMF. This report is not a Finanzanalyse within the meaning of FinSA art. 3 lit. h and is not an investment recommendation within the meaning of MAR art. 3(1)(35) / Reg. (EU) 596/2014 or FCA Handbook COBS 12.4. Where named issuers are discussed, the discussion illustrates how the disclosure-tier framework applies to a fact pattern; it is not an opinion on the issuer's securities.
- ***Position & holdings.*** As of 14 May 2026, the author holds no long or short position, directly or through derivatives, in any security named, and no economic interest equivalent (including via funds, indices, or structured products in which the author is a beneficial holder). The author confirms no holdings of 0.5% or more of issued share capital in any named issuer (Del. Reg. (EU) 2016/958 art. 6(1)(a)) and undertakes not to enter such a position for thirty days after publication.
- ***Compensation & MNPI.*** No compensation has been received from any named issuer; no commercial relationship exists with any vendor discussed. The report is built from public sources only and the author represents that no material non-public information has been used in its preparation.
- ***Not investment advice; risk warning.*** Not a solicitation, not a recommendation to buy, sell, or hold any security or financial instrument. Opinions expressed are honestly held on the basis of the methodology described in §1.2 and §10. The value of investments can fall as well as rise; past performance is not a guide to future performance; capital is at risk. UK readers: this is a non-real-time communication issued by an unauthorised person and is directed only at professional investors and high-net-worth individuals within FSMA 2000 (Financial Promotion) Order 2005 arts. 19 and 49.
- ***Tier classifications.*** The six-tier framework defined in §4.2 grades disclosure granularity only. No tier assignment is an opinion on the accuracy, adequacy, or compliance of any issuer’s disclosures with GAAP, IFRS, SEC rules, ASC 280, ASC 606, or any other applicable standard.
- ***Private-company figures.*** Revenue figures attributed to private companies are reproduced from cited third-party sources and have not been independently verified by the author. Recognition-method uncertainty (principal-versus-agent under ASC 606) is treated as a methodological variable as disclosed in §4.4 and §10.
- ***Forward-looking statements.*** References to vendor guidance, fiscal-year targets, sensitivity ranges, and capex paths contain forward-looking statements that involve known and unknown risks. Actual outcomes may differ materially. No projection herein is guaranteed.
- ***Methodology & date stamp.*** Cut-off 14 May 2026. Method described in §1.2 and §10. All headline numbers are reproducible from the underlying vendor census and reconciliation build. Corrections welcomed at the contact address above.