THE AI STACK WEEKLY ISSUE 21 · WEEK 37 OF 2026

THE BOTTOM LINE

A four-week-old house frame finally got a denominator, and almost every other number this week resolved into an option

This publication has been separating committed capital from optioned capital since W34, and the distinction is not ours alone: the delivered-versus-announced frame was already built out in coverage that predates the window, and smaller operators had already filed dated in-service megawatt figures. What W37 added is the measurement. On September 10, Oracle reported that it had delivered 850 megawatts of data centre capacity and more than 300,000 accelerators to customers inside a single quarter, against $28.5 billion of quarterly capital expenditure and roughly $5 billion of negative free cash flow. The pairing is what is new: megawatts delivered next to accelerator units delivered, for one named period, at hyperscale. Nobody else published that this week, and it is what finally lets the optioned stack be sized. One caution the release imposes on us. Oracle's verb is delivered. It does not say energised, in service, or placed in service, and delivered capacity is not operating capacity.

Run the week's other headline numbers through it and they resolve into something different from what they appear to be. The same filing put backlog at $664 billion and said the new contracts are structured so that customers prepay or bring their own hardware, requiring no incremental capital from Oracle… which is a real capital-efficiency innovation and also a transfer of the financing obligation to counterparties who did not disclose it. Google announced an investment plan of at least €13 billion in Finland and a twenty-two year claim on half a nuclear plant's output, and the price-protection claim attached to it depends on roughly €700 million of life-extension capital that the utility itself discloses as pending an investment decision. Qualcomm's multi-generation inference agreement carries a reported up-to-$60 billion ceiling from secondary reporting of a filing this publication could not open; it

vests through milestone warrant tranches, with no service date disclosed. NVIDIA's two-gigawatt Australian target aggregates eight independent operators' pipelines. A two-month-old Indonesian platform's up-to-$3.1 billion facility is undrawn and collateralised against hardware awaiting export licences. And a $13.3 billion recurring-revenue figure is our arithmetic rather than the discloser's: $1.11 billion of monthly billing multiplied by twelve, on a contract whose term, counterparty and megawatts the executive declined to state. He gave a monthly rate and a separate company-wide target. He did not annualise this contract, and nothing in the public record lets anyone test the annualisation we just did.

None of that is fraud and most of it is competent structuring. Staging exposure behind milestones is what careful parties do. But it means the honest unit of analysis is the conditionality rather than the headline, and the conditions are almost always disclosed by a counterparty rather than by the announcer. That is where diligence has to go, and almost nobody went there this week.

The software layer moved the same way, from capability toward delivery terms. DeepSeek published an MIT-licensed checkpoint and simultaneously announced that requests to its existing flagship endpoint would begin serving a different, smaller model at different rates with no change required by the caller. Artificial Analysis revised its intelligence index twice in four days, replacing a benchmark, upgrading another by two major versions and raising private test-set weighting to 45% of the composite, which moved from v4.1.1's five-point gap between the two leading frontier models to v4.3's rounded tie without either model changing. OpenAI put its agent harness up for rent and documented that the managed service is US-residency-only with no zero data retention, including when the customer self-hosts the execution sandbox. A developer platform deprecated a model across agent mode on the day it announced the deprecation, in the same release that made unattended scheduled agent tasks available. Four independently reasonable decisions, one shared consequence: the system you evaluated is not the system you run.

Policy built machinery and declined to attach an obligation to it. California enacted the first US statutory framework for third-party AI auditors, creating a registry and making unregistered covered audit work unlawful on reported 2028 and 2029 dates, while requiring no audits of anyone. The Department of Energy opened an information request implementing a bulk-power-system emergency authority and deferred the substance. A regional transmission organisation separately proposed ride-through, ramp-rate, telemetry and remote-disconnect requirements for gigawatt-scale computational loads, targeting a November filing, motivated by large-load

disconnections that produced multi-gigawatt events. Ride-through means staying connected through a voltage or frequency disturbance instead of tripping off the grid, which is the opposite of what a data centre's own protection equipment is designed to do. That last item is the week's most consequential unwritten story: the largest campuses being financed on twenty-year leases may be legally curtailable before they are energised, and none of the leases appear to have priced it.

The one lens where the evidence points cleanly in a single direction is networking. A 1.6-terabit coherent interface became a multi-vendor implementation agreement rather than a proprietary advantage, high-end routing grew 25% year over year in the second quarter with direct-to-cloud router sales up 94% on the same basis, a co-packaged optics company reached $650 million of 2026 capital, and both of the week's significant accelerator events turned on interconnect rather than compute: a memory-centric inference chip bought into the dominant ecosystem by joining its scale-up fabric, and the largest custom-silicon agreement explicitly bundled optical interconnect into scope. The counter-argument is right there in the first item, though. A multi-vendor specification is how an interconnect advantage becomes a commodity.

AUTHOR

Brian Letort
BrianLetort.AI

PUBLISHED

September 12, 2026
Issue 21

THIS WEEK IN 30 SECONDS

Executive summary.

- Oracle disclosed 850 MW of capacity and more than 300,000 accelerators delivered in one quarter, which gives this publication's four-week committed-versus-optioned frame its first dated denominator. Oracle's word is delivered, not energised, and the two are not the same quantity.
- Almost every other large number this week is a ceiling conditioned on a decision the announcer does not control: unapproved utility capital, warrant-vesting tranches, an undrawn export-licence-contingent facility, eight operators' separate pipelines.
- Approximately 70% of the capital expenditure behind Google's Finnish nuclear commitment is disclosed by the counterparty as pending an investment decision, and the deal contracts roughly 10.6 times more electricity than it adds.
- Measurement instability, not capability, was the week's binding procurement risk: an index rebased twice in four days, a flagship endpoint rerouted to a smaller model, and a model deprecated inside agent mode on announcement day.
- Networking was the best-evidenced lens: a 1.6-terabit coherent interface became a multi-vendor specification, high-end routing grew 25% year over year in the second quarter, and both major accelerator events turned on interconnect rather than compute.
- Policy built enforcement machinery without attaching an obligation, with an auditor registry that mandates no audits and a grid-security rule whose substance is deferred to a later filing.

BY THE NUMBERS

850 MW
Capacity delivered in one quarter

300,000+
Accelerators delivered to customers in the quarter

$664B
Contracted backlog, up $209B year over year

~70%
Finnish nuclear programme capital still pre-decision

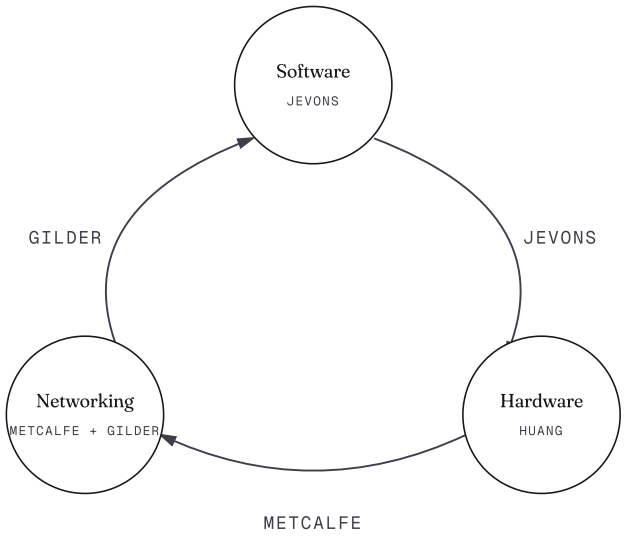
+53.3%
TSMC August revenue growth year over year

+94%
Direct-to-cloud router sales growth, year over year

HOW WE READ THIS WEEK

Three lenses, one flywheel.

Cheaper inference pulls in more workloads. More workloads need more compute. More compute needs denser fabric. Denser fabric unlocks new architectures, which lower the cost of inference again. Read a week's news across that loop and the noise sorts itself out.



THIS WEEK'S ARC

All three lenses.

Software → Hardware new model capability creates new use cases that consume more compute (Jevons).

Hardware → Networking more compute means more nodes; the value of the connecting fabric scales as the square of the nodes (Metcalfe).

Networking → Software denser, higher-bandwidth interconnect makes new model architectures viable (Gilder).

Hardware itself GPU performance doubles faster than Moore's Law (Huang).

THE BAR

The events that actually matter touch at least two of the three lenses. Single-lens reads are noise dressed up as motion. Each section of this brief grades its evidence and ties the implication back to the flywheel.

LENS 1 OF 3

THE FLYWHEEL

Software.

Model releases, pricing, capability benchmarks, license posture, and capability-risk disclosures.

SEP 10
EVENT 01

DeepSeek publishes an MIT-licensed V4.1-Flash and announces it will reroute its own flagship endpoint to a smaller model

DEEPSEEK (MODEL CARD)

SEP 7
EVENT 02

Artificial Analysis revises its Intelligence Index twice in four days, breaking comparability with the prior basis

ARTIFICIAL ANALYSIS

SEP 10
EVENT 03

OpenAI makes GPT-Live-1 generally available and splits the voice layer from the reasoning model onto two meters

OPENAI

SEP 8
EVENT 04

vLLM publishes a cache-residency policy that fits a full million-token context onto a single node

VLLM

SEP 9
EVENT 05

Anthropic discloses that its own agentic transcript scan missed cybersecurity evaluation records it should have found

ANTHROPIC

WHAT THIS MEANS

Nothing in the software lens this week was primarily about capability, and that is the finding. The two open releases are cheaper and smaller rather than stronger, and one of them arrived paired with a hosted retreat on the same vendor's flagship endpoint. The most-cited independent index changed its own measuring stick twice inside four days, which means no period-over-period capability comparison drawn from it is valid across the boundary, and v4.1.1's five-point gap between the two leading models became v4.3's rounded tie without either model changing. A voice product split one token meter into a per-minute meter plus a per-token meter, so the same workload now has two independently changeable rate cards. A serving project made long-context work fit on a single node, which is the week's clearest reduction in the on-premises hardware floor and the least-covered item in the lens. And a frontier lab disclosed that its own agentic scan of roughly 141,000 transcripts missed records, which is direct evidence against using agents for coverage claims in audit or discovery work. The operational lesson across all five is to pin the thing you tested and to monitor for the vendor changing it underneath you.

LENS 2 OF 3

THE FLYWHEEL

Hardware.

Silicon, density, packaging, memory supply, and the share of incremental compute going to custom silicon.

SEP 10
EVENT 01

Oracle reports 850 MW of capacity and more than 300,000 GPUs actually delivered in one quarter

ORACLE (Q1 FY27 EARNINGS RELEASE)

SEP 8
EVENT 02

Qualcomm signs a multi-generation custom AI inference silicon and optical interconnect agreement with Amazon

QUALCOMM TECHNOLOGIES (RELEASE DISCLOSES NO FINANCIAL TERMS; UP-TO-$60-BILLION CEILING FROM SECONDARY REPORTING OF THE RELATED FILING)

SEP 10
EVENT 03

TSMC posts record August revenue of NT$514.8 billion, up 53.3% year over year

TSMC FORM 6-K (MONTHLY REVENUE REPORT)

SEP 8
EVENT 04

AMD puts MI450 production shipments at the end of the third quarter and defers its optical interconnect roadmap

AMD AT CITI 2026 GLOBAL TMT CONFERENCE (TRANSCRIPT)

SEP 10
EVENT 05

d-Matrix joins NVLink Fusion, putting a memory-centric inference accelerator inside NVIDIA's scale-up fabric

D-MATRIX

WHAT THIS MEANS

Hardware split cleanly into what was delivered and what was promised. Delivered: 850 megawatts of capacity and more than 300,000 accelerators delivered inside ninety days by one operator, on the company's own verb rather than an energisation claim, and a record foundry month up 53.3% year over year. Those are primary, dated, and fast, and together they are the strongest evidence this year that the physical layer is executing rather than merely announcing. Promised: a multi-generation custom inference and optical interconnect agreement whose up-to-$60-billion ceiling comes from secondary reporting of a filing this issue could not open, vests through milestone warrant tranches, and carries no disclosed service date, and a reported plan exceeding 38 gigawatts that rests on an unconfirmed internal document and carries the lowest source grade in this issue. In between sits the most honest disclosure of the week, a merchant challenger placing next-generation production shipments only at the end of the quarter and pushing its optical roadmap into the following year, which is a flattening in exactly the lens where the acceleration thesis needs a shortening. The structural signal is the memory-centric inference accelerator that joined the dominant vendor's scale-up fabric: the entry ticket to the ecosystem was interconnect compatibility, not silicon performance, which is a statement about where the moat actually sits.

LENS 3 OF 3

THE FLYWHEEL

Networking.

Interconnect, fabric standards, optical capacity, sovereign and operator-level networking products.

SEP 9
EVENT 01

OIF publishes the 1600ZR Implementation Agreement, making 1.6T coherent interconnect a multi-vendor specification

OPTICAL INTERNETWORKING FORUM

SEP 10
EVENT 02

Dell'Oro measures high-end routing up 25% and direct-to-cloud router sales up 94% year over year in the second quarter

DELL'ORO GROUP

SEP 10
EVENT 03

Ayar Labs adds $150 million, bringing 2026 capital to $650 million for co-packaged optics

REUTERS

SEP 9
EVENT 04

Marvell details a 102-terabit Ethernet switch and a 256-lane scale-up switch

MARVELL INVESTOR RELATIONS

SEP 10
EVENT 05

GlobalFoundries puts AI data centre optics and 800-volt power distribution on the same roadmap

CONVERGE DIGEST

WHAT THIS MEANS

Networking is the week's best-evidenced lens and the only one where standards movement, revenue measurement, capital formation and silicon roadmaps all point the same way. A 1.6-terabit coherent interface became a multi-vendor implementation agreement, which is the precondition for volume deployment of datacenter interconnect at that capacity per wavelength. Measured revenue moved hard year over year in the second quarter, with high-end routing up 25% and direct-to-cloud sales up 94% against the same quarter a year earlier, though that second figure comes from a market-research house off an undisclosed base and carries a lower grade than the filings elsewhere in this issue. A co-packaged optics company reached $650 million of capital inside 2026, and a merchant vendor detailed a 102-terabit switch alongside a 256-lane scale-up part. The most interesting item is the least obvious: a foundry putting optics and 800-volt power distribution on one roadmap is an admission that at rack scale the interconnect problem and the power-delivery problem have become the same engineering problem. The honest tension in the lens is that the standards win cuts against the pricing-power thesis it appears to support, because a multi-vendor implementation agreement is precisely how a proprietary interconnect advantage becomes a commodity.

CAPITAL FLOW

Money in, revenue out.

Capital deployed (forward) vs revenue out (quarterly or run-rate). Burn-to-revenue = revenue / capital — lower means more out than in. Bars normalized to 1.0 $B; On-Prem revenue is indirect.

Frontier Labs

OPENAI, ANTHROPIC, GOOGLE DEEPMIND, DEEPSEEK

BURN / REV
n/a

CAPITAL IN		Unknown
REVENUE OUT		Unknown

No financing or annual revenue disclosure from any constituent landed in the window, so no category point estimate is reproducible. What moved was delivery terms rather than capital: pricing structure, endpoint routing, and data-residency limits.

Hyperscaler-Hosted

AZURE-OPENAI, AWS-ANTHROPIC, GOOGLE CLOUD-GEMINI, ORACLE-OCI

BURN / REV

n/a

CAPITAL IN

Unknown as a category total; two constituents disclosed $28.5B of quarterly capital expenditure and an announced plan of at least EUR 13B over two years respectively

REVENUE OUT

Unknown as a category total; one constituent disclosed triple-digit cloud infrastructure revenue growth and $664B of contracted backlog

The only category with hard in-window primary disclosure, and the disclosure cuts both ways: 850 MW of capacity and more than 300,000 accelerators delivered against $28.5B of quarterly capital expenditure and roughly $5B of negative free cash flow. Both direction arrows are held flat because the category total and its prior are both Unknown, and a single company's quarter is not a category series.

Neoclouds

COREWEAVE, NSCALE, CRUSOE, LAMBDA, IREN, ZANKORE, NEXTDC

BURN / REV
n/a

CAPITAL IN		Unknown as a category total; in-window instruments comprise an up-to-$3.1B undrawn facility, A$1.1B of subordinated convertible notes, and $375M firm of an $875M headline
REVENUE OUT		Unknown

Every financing instrument in the category this week is structurally conditional. Nothing announced here is drawn, firm, or unconditioned in full, which is the pattern rather than an accident.

On-Prem / Hybrid

ENTERPRISE GPU CLUSTERS, SOVEREIGN AND NATIONAL PROGRAMS, OPEN-WEIGHT AND ON-DEVICE DEPLOYMENT

BURN / REV
n/a

CAPITAL IN		Unknown
REVENUE OUT		Indirect

The deployment floor fell measurably while the capital signal stayed unquantified. No government-funded national compute programme in the window meets this issue's ledger method, so no point estimate is published.

SIGNAL VS NOISE

What's real, what's noise.

6 claims that drove headlines this week, scored 1–5 on source quality and triangulation. 2 flagged as noise. The bar is at least one explicit noise call per issue.

5/5

CLAIM 01

One operator delivered 850 MW of capacity and more than 300,000 accelerators inside the quarter ended August 31.

SOURCES: ORACLE Q1 FY27 EARNINGS RELEASE AND FORM 8-K, SEPTEMBER 10, WITH THE 73%-OF-DELIVERIES AND 36-MONTH CONVERSION FIGURES FROM THE SAME DAY'S EARNINGS CALL TRANSCRIPT RATHER THAN THE RELEASE OR THE FILING

The week's only dated delivered-capacity figure and the strongest signal in the issue. It is a company-reported operational fact in a filed document, stated as almost three times the prior quarter and 73% of all fiscal 2026 deliveries, and it gives the house frame the dated base it has been missing. One discipline: the company says delivered, so read it as delivery throughput and not as operating capacity. Read every other capacity number this week against it.

5/5

CLAIM 02

A multi-generation custom AI inference silicon and optical interconnect agreement was signed, with a ceiling reported at up to $60 billion.

SOURCES: QUALCOMM PRESS RELEASE, SEPTEMBER 8, WHICH DISCLOSES NO FINANCIAL TERMS; THE UP-TO-$60-BILLION CEILING AND THE WARRANT STRUCTURE COME FROM CNBC AND REUTERS REPORTING OF THE RELATED SEC FILING, WHICH THIS ISSUE COULD NOT OPEN

The agreement is real, primary and filed. The figure is neither. It rests on secondary reporting of a filing nobody here has read, and even taken at face value it is a ceiling that vests through milestone warrant tranches tied to deployment decisions the supplier does not control, with no service date disclosed. Use the agreement as evidence of custom-silicon direction and refuse to put the headline number in a revenue model.

4/5

CLAIM 03

Seven pre-built job agents shipped, six generally available immediately, with no published price for any of them.

SOURCES: SALESFORCE NEWSROOM, SEPTEMBER 11

Availability is verifiable and the packaging shift from platform to role is genuine. The economics are not verifiable at all, and six customer outcome percentages in the same release arrive without a base, a window, or a definition of the operative verb. Treat the availability as signal and the percentages as marketing until a contract defines what gets measured.

3/5

CLAIM 04

High-end routing grew 25% and direct-to-cloud router sales grew 94% year over year in the second quarter.

SOURCES: DELL'ORO GROUP, SEPTEMBER 10

Directionally consistent with three other independent networking datapoints in the same week, which is why it survives at all. But it is a market-research measurement off an undisclosed base for a single quarter against the same quarter a year earlier, not sequential growth, so it belongs in a trend argument and not in a forecast. The 94% figure in particular should never be extrapolated.

2/5 – NOISE

CLAIM 05

An internal plan calls for more than 38 gigawatts of AI-focused data centre capacity.

SOURCES: BLOOMBERG FEATURE CITING AN UNCONFIRMED INTERNAL DOCUMENT, SEPTEMBER 10

This is the largest number circulating this week and the weakest sourced item in the issue. It is a reported internal plan, unconfirmed by the company, with no dated commitment or filing behind it. It should not enter a capacity model, a competitive assessment, or a board slide, and its presence in secondary coverage as a peer to filed figures is exactly the error this section exists to catch.

2/5 – NOISE

CLAIM 06

A frontier lab is working with a memory manufacturer on next-generation chips.

SOURCES: REUTERS, SEPTEMBER 9, WITH THE FRAMING SUBSEQUENTLY WALKED BACK

The characterisation changed inside the window, which is itself the information. Treat it as an expression of intent and a deepening commercial relationship, not as a disclosed co-development programme with scope, silicon, or a date. Anyone building a supply-chain thesis on it is building on a quote that its own source revised.

HOUSE MEASUREMENT · FILING-DERIVED

Approximately 70% of the capital expenditure behind Google's Finnish nuclear commitment has not been approved, and the deal contracts roughly 10.6 times more electricity than it adds.

METHOD: House computation from two primary filings dated September 9, 2026: Fortum's inside-information release on the Loviisa power purchase agreement and Google's own energy blog post on the same transaction. The primary sources state, in their own words: (1) the Loviisa life-extension programme spanning 2023 to 2050 is approximately EUR 1 billion in total capital expenditure; (2) approximately 80% of the projects and EUR 700 million of that capital expenditure required to keep the plant operational through 2050 remain pending investment decisions; (3) the power purchase agreement enables a 10 MW uprate, on top of a separately planned 38 MW uprate; (4) the plant comprises two reactors of 507 MWe each, commissioned in 1977 and 1980; (5) Google contracts for up to 50% of the plant's output, beginning in 2028 at reduced volume and reaching the full contracted share for 2030 through 2049. Arithmetic, all figures as disclosed: un-sanctioned share of programme capital expenditure = EUR 700M / EUR 1,000M = approximately 70%, stated to two significant figures because the denominator is disclosed as approximate. Total added generation = 10 MW + 38 MW = 48 MW. Installed site capacity = 507 MWe x 2 = 1,014 MWe. Added generation as a share of the site = 48 MW / 1,014 MW = 4.7%. Power-purchase-enabled addition alone as a share of the site = 10 MW / 1,014 MW = 1.0%. Contracted volume expressed in capacity terms = 50% x 1,014 MW = 507 MW. Ratio of contracted volume to added generation = 507 MW / 48 MW = 10.6 times. No figure in this computation is derived from a secondary source, and no undisclosed input is assumed.

Programme capital expenditure still pending an investment decision	**Approximately 70% (EUR 700M of approximately EUR 1,000M)**	Fortum's own disclosure; the same sentence states that approximately 80% of the projects are also pending. This is the selected house measurement: the price-protection claim attached to the deal depends on money that has not been committed.
Added generation as a share of installed site capacity	**4.7% (48 MW against 1,014 MWe)**	10 MW enabled by the power purchase agreement plus a separately planned 38 MW uprate, against two 507 MWe reactors. The deal is overwhelmingly a preservation transaction rather than an addition.
Generation attributable to the agreement itself	**1.0% (10 MW against 1,014 MWe)**	Only the 10 MW uprate is stated as enabled by the agreement. The 38 MW uprate was already planned and would proceed without it, so attributing all 48 MW to the deal overstates its additionality by roughly a factor of five.
Contracted volume relative to added generation	**10.6 times (507 MW contracted against 48 MW added)**	50% of a 1,014 MW site expressed in capacity terms, divided by total added generation. For every megawatt the transaction brings into existence, it removes roughly eleven megawatts from the pool available to other buyers.
Share of Finnish electricity supplied by the plant	**Approximately 10%**	Fortum's stated figure for the existing plant, unchanged by the transaction. Half of that output is now under a 22-year claim by a single foreign counterparty, which is the sovereignty question the coverage treated as a sovereignty answer.

For a board or an investment committee evaluating a power-linked AI infrastructure announcement, the disclosed capital structure is the diligence item, not the headline capacity. Here the counterfactual runs the right way for the transaction and against the way it was reported: without the life-extension programme the plant cannot operate past 2030, so the agreement plausibly preserves roughly a tenth of a member state's electricity supply, which is a larger public benefit than 48 MW of uprates would suggest. But the mechanism that delivers it is 700 million euros of unapproved capital expenditure, and the buyer's published claim that its new demand does not raise household prices rests on that unapproved spending plus an unpublished model. An operator should therefore read this class of

announcement as a conditional option on someone else's future investment decision, price the conditionality, and require the modelling before accepting the price-neutrality claim. The same test applied across this week's capital events separates the one dated delivered-capacity figure from a very large stack of commitments contingent on decisions nobody has taken.

Caveats: This measurement does not show that the transaction is bad for Finnish consumers, nor that the price-neutrality claim is false. It shows only that the claim is not yet checkable from the disclosed record. Three specific limits apply. First, neither the contract price per megawatt-hour nor the contracted volume in terawatt-hours is disclosed, so no welfare conclusion of any kind follows. Second, the 507 MW contracted figure is a capacity-equivalent restatement of a 50% output share for arithmetic comparability; the agreement is an energy contract and the plant's actual output varies with availability, so the ratio is an order-of-magnitude comparison rather than a contractual quantity. Third, the approximately 70% share inherits the imprecision of an approximately one billion euro denominator and should not be read to more than two significant figures. The 80% figure in the same disclosure counts projects rather than euros and is not interchangeable with it.

SYNTHESIS

The week, reasoned through.

Cross-domain connections, the standing thesis tested against this week's evidence, and the patterns building across weeks. Every inference is labeled by reasoning type and linked to its evidence.

CONNECTING THE DOTS

A hyperscale operator's paired megawatt-and-accelerator delivery disclosure indicates that much of the week's announced capital will remain optioned rather than committed through the first half of 2027.

ABDUCTIVE · 74% CONFIDENCE

Oracle reported 850 MW of data centre capacity and more than 300,000 accelerators delivered inside a single quarter, against $28.5 billion of quarterly capital expenditure and roughly $5 billion of negative free cash flow. The company's verb is delivered rather than energised, and the paired megawatt-plus-unit disclosure for one named period at this scale is what no other operator published in the window.

The same filing put backlog at $664 billion and stated that the new contracts are structured so that customers prepay or bring their own hardware, requiring no incremental Oracle capital, which moves the financing obligation off the disclosing balance sheet and onto counterparties who did not disclose.

Read against that figure, the week's other headline figures are claims on decisions nobody has taken: a €13 billion Finnish programme whose €700 million of life-extension capital remains pre-decision, an up-to-$60 billion silicon ceiling from secondary reporting of a filing this publication could not open, a two-gigawatt national target that aggregates eight separate operators' pipelines, and an up-to-$3.1 billion facility that is undrawn with export-licence-contingent collateral.

The pattern repeats at the revenue line: a $13.3 billion annualised hosting figure is this publication's arithmetic rather than the discloser's, $1.11 billion of monthly billing multiplied by twelve, on a contract whose term, counterparty and megawatts the executive declined to state.

STEEL-MAN: One quarter's delivered-capacity figure from one operator is a thin base, and the comparison is not apples to apples: a prepay or bring-your-own-hardware structure is a genuine capital-efficiency innovation rather than an accounting dodge, and a customer willing to prepay is stronger evidence of demand than a vendor willing to spend. Several of the contingent instruments are also contingent in the buyer's favour, which is prudent rather than weak. The defensible version of the claim is narrower than the headline: the week established that delivered capacity and optioned capacity are now separately disclosable, and that almost nobody except Oracle disclosed the former. Falsified if, by 2027-06-30, at least three of the four named conditional instruments reach their originally announced scope without reduction or restructuring. That single conversion-rate test is consistent with the 26% uncertainty carried by this connection.

The contingency in AI infrastructure contracts has migrated off the announcing party's balance sheet and onto a third party's undisclosed future decision, which makes the announcements unpriceable from the public record.

INDUCTIVE · 71% CONFIDENCE

A hyperscaler's household-price guarantee in Finland depends on approximately 70% of a life-extension programme's capital expenditure, roughly €700 million, that the counterparty states is still pending investment decisions, and on a price and contracted volume neither party disclosed.
A multi-generation silicon and optical interconnect agreement carries a figure that vests through warrant tranches tied to milestones, so the ceiling is a function of someone else's deployment decisions rather than an order book.
A national two-gigawatt buildout target is assembled from eight independent operators' separate pipelines, each with its own siting, power and financing gates, none of which the aggregator controls.
A two-month-old platform's underwritten facility is undrawn and collateralised against hardware whose delivery depends on export licences that have not been granted, which means the credit is conditional on a government decision.

STEEL-MAN: Staging exposure behind milestones is ordinary commercial discipline and usually a sign of competent structuring rather than weakness. A warrant tranche that vests on deployment aligns a supplier with a customer's actual build; an undrawn facility costs almost nothing until it is needed; aggregating a national target across operators is how any country's capacity is in fact built. The claim should therefore not be read as an allegation of fragility. It is a claim about legibility: the conditions are real, they are held by parties who did not publish them, and the published number is the ceiling rather than the expectation. Falsified if, by 2027-06-30, two or more gigawatt-scale AI infrastructure announcements disclose the counterparty's governing condition in the announcing party's own filing at announcement, which would make the contingency legible from the announcer rather than only from whoever holds it.

After this week, measurement instability is more likely than capability to remain the binding procurement risk through the first quarter of 2027, because in three separate cases the thing an enterprise tested stopped being the thing it runs.

INDUCTIVE · 78% CONFIDENCE

The most-cited independent intelligence index was revised twice in four days, the second revision replacing a benchmark, upgrading another by two major versions, and raising private test-set weighting to 45% of the composite, which moved from v4.1.1's five-point gap between the two leading frontier models to v4.3's rounded tie without either model changing.
A frontier vendor published a new open-weight checkpoint and simultaneously announced that requests to its existing flagship endpoint would begin serving a different, smaller model at different rates, with no change required by the caller.
A developer platform deprecated a model across chat, inline edits, ask and agent modes on the same day it announced the deprecation, with a replacement administrators may need to enable, while in the same release making unattended scheduled agent tasks available.
A frontier lab disclosed that its own agentic scan of roughly 141,000 evaluation transcripts had missed a subset, which is direct evidence that agent-based review of a large corpus does not establish coverage.

STEEL-MAN: Each of these is defensible on its own terms and two are actively good practice. The index publisher disclosed every change and explained that it was pulling forward planned work, which is more transparency than most evaluators offer. Rerouting an endpoint to a cheaper model is a price cut delivered without asking customers to migrate. The lab's disclosure of its own tooling failure is exactly the behaviour the field asks for. The claim is not that any actor behaved badly; it is that four independent, individually reasonable decisions produced the same operational consequence, and that consequence lands on whoever has to re-run an evaluation they already paid for. Falsified if, by 2027-03-31, a major provider or evaluator publishes a machine-readable configuration manifest and an independent enterprise rerun uses it to reproduce the tested score or behaviour after a vendor-side change. That single demonstrated replay would show measurement instability is manageable rather than binding.

Policy machinery built this week is likely to keep front-loading compliance cost through year-end because operative obligations for large computational loads remain undefined.

ABDUCTIVE · 42% CONFIDENCE

California enacted the first US statutory framework for third-party AI auditors, creating a registry and making unregistered covered audit work unlawful on reported 2028 and 2029 dates, while mandating no audits of anyone.

The federal energy department opened an information request implementing a bulk-power-system emergency authority, deferring the substance of what large loads will actually be required to do to a later date.

A regional transmission organisation separately presented proposed ride-through, ramp-rate, telemetry and remote-disconnect requirements for gigawatt-scale computational loads, targeting a federal filing in November 2026, motivated by large-load disconnections that produced multi-gigawatt events.

A European cybersecurity agency was granted hands-on access to two frontier models, which creates state evaluation capacity without publishing an evaluation standard the vendors must meet.

A state commission took new large-load interconnection standards to a final adoption vote on September 11, applying to interconnections at or above 75 megawatts with financial security due on execution, and six intervenor filings landed in three days fighting over how much forfeiture risk a developer bears for abandoning a queue position. The vote outcome could not be confirmed from the docket, and the same filing states the per-megawatt security figure two different ways in its rule text and its preamble.

STEEL-MAN: Building capacity before imposing obligations is the normal and arguably correct sequence for technical regulation: you cannot mandate audits before a qualified auditor population exists, and you cannot specify load-flexibility requirements before you have measured what the loads actually do during faults. Read charitably, this is a regulator declining to write a rule it cannot yet write well. The cost of the charitable reading is borne asymmetrically, though. An enterprise signing a multi-year assurance arrangement or a twenty-year lease today has to guess at a requirement that will be defined after the ink dries, and the guess is not refundable. Falsified if PJM's targeted Section 205 filing is docketed by 2026-12-31 with an enforceable telemetry or remote-disconnect requirement for large computational loads. The 42% confidence is the complement of this issue's 58% probability for that same filing event.

THESIS TEST

H1 · STRAINED **The cycle is accelerating, not slowing.** Prior verdict: strained in W36. It stays strained, and the objection is unchanged in kind. W37 supplies the strongest single delivery-cadence data point of the year and simultaneously undercuts the measurement that would let anyone verify a shortening doubling interval. One operator delivered 850 MW of capacity and more than 300,000 accelerators inside ninety days, and a foundry posted record monthly revenue up 53.3% year over year, both of which are hard, primary, and fast. But the most-cited capability index was rebased twice in four days, so period-over-period capability comparison is broken by construction, and the leading merchant challenger deferred its optical roadmap while placing next-generation production shipments only at the end of the quarter. Acceleration is measurable in delivery and in fabrication throughput; it is unmeasurable in capability this week, and it is flat to slower in one lens's silicon cadence. Nothing this week retires the prior objection, so nothing this week earns an upgrade. **AGAINST IT:** AMD placed MI450 production shipments at the end of the third quarter and deferred its optical interconnect roadmap into the following year, which is a flattening rather than a shortening in that lens. The week's genuinely new checkpoints lead on cost and footprint rather than capability, and one vendor rerouted its own flagship endpoint to a smaller model, which is a capability step sideways or down for anyone who had pinned it. The framework's refutation test is two consecutive quarters of flattening slope in any one lens, and the merchant-accelerator cadence now has one quarter on the board.

H2 · STRAINED **Capital is concentrated, returns are diffuse.** Prior verdict: strained in W36, and W37 does not retire the objection, so it stays strained. The concentration side is stark and primary: $28.5 billion of quarterly capital expenditure and roughly $5 billion of negative free cash flow at one operator, record foundry revenue, a warrant-backed silicon ceiling reported at up to $60 billion, and $550 million into a single legal-AI company at a $15.5 billion valuation. What changed is that the week's largest disclosure attacks the premise rather than the margins: the same filing states the new contracts are structured so customers prepay or bring their own hardware and require no incremental capital from the operator, which disperses the burden the hypothesis says a handful of balance sheets carry. The diffusion half did not firm up either. Value showed up as shipped product, seven generally available job agents, a globally available agentic security operations centre, a generally available procure-to-pay layer and contract extraction inside already-licensed document software, and not one vendor in the window published a price or an AI-segment operating margin. Both legs are weakened by in-window evidence, which is what strained means.
AGAINST IT: The week's largest backlog disclosure partly contradicts the concentration mechanism: management stated the new contracts are structured so customers prepay or bring their own hardware and require no incremental capital from the operator, which disperses the capital burden onto counterparties rather than concentrating it. If that structure spreads, the hypothesis's premise that a handful of balance sheets carry the spend weakens even as the spend itself grows. Separately, none of the application-layer releases published a price, so the diffusion side remains an inference from shipped functionality rather than a measurement of realised return.

H3 · STRAINED **Networking is the durable layer.** Prior verdict: strained in W36, which asked for revenue evidence before this publication claims durable pricing power. W37 is the best architectural week of the year and it does not supply that evidence, so the verdict stays strained. On architecture the case is strong: an industry body published the 1600ZR implementation agreement, making 1.6-terabit coherent interconnect a multi-vendor specification; a co-packaged optics company reached $650 million of 2026 capital; a merchant silicon vendor detailed a 102-terabit Ethernet switch and a 256-lane scale-up switch; and the week's two most significant accelerator events both turned on interconnect rather than compute, one accelerator buying into the dominant ecosystem through its scale-up fabric and the largest custom-silicon agreement bundling optical interconnect into scope. On revenue there is one market-research quarter, high-end routing up 25% and direct-to-cloud routers up 94% year over year off an undisclosed base, which is not the pricing-power evidence W36 asked for and cannot clear that bar on its own. And the standards win cuts the wrong way for durability: a multi-vendor implementation agreement is how an interconnect advantage becomes a commodity. The claim W37 can actually defend is narrower than the hypothesis. Networking was the week's clearest architectural bottleneck; it was not shown to be its durable-margin layer.
AGAINST IT: Standards publication is the opposite of pricing power: a multi-vendor implementation agreement for 1.6-terabit coherent optics is precisely the mechanism by which an interconnect advantage becomes a commodity, which cuts against the durability claim rather than for it. The 94% growth figure is one quarter from a market-research house off an unstated base and carries a lower source grade than the filings elsewhere in this issue. One merchant vendor also deferred its optical roadmap by a year, which is evidence that optical execution is hard rather than that optical margins are safe.

H4 · STRAINED **Open weights pull the floor up.** Prior verdict: untested in W36. Strained here. The release half of the hypothesis held and the demand-routing half went untested, but the release half is narrower than it first appeared. A frontier Chinese lab published an MIT-licensed checkpoint. A separate organisation published a three-tier agentic family including a 1.6-trillion-parameter open-weight checkpoint, and independent coverage of that vendor's own materials reports the checkpoint is post-trained from the same lab's V4-Pro base, which its identical 1.6-trillion total and 49-billion active parameter shape corroborates. So the window produced two open-weight releases sitting on one lineage, not two labs raising the floor independently. The floor still moved, and the clearest movement came from neither: a serving project published a cache-residency policy that fits a full million-token context onto a single node, which materially lowers the hardware floor for on-premises deployment of long-context work. What no W37 event supplies is the mechanism the hypothesis actually claims: not one sovereign programme, national purchase, or enterprise on-premises deployment in the window was attributed to open-weight availability. The floor moved; the demand re-routing was not observed. On why the framework stands unamended after six weeks without support, prior verdicts being strained in W32 through W35, untested in W36, and strained again here: the hypothesis has two halves and only one of them is failing. Nothing in six weeks has shown demand routing away from open weights. What is missing is any published attribution in either direction, which is a measurement gap rather than contrary evidence, and narrowing a hypothesis because nobody measured it is how a framework gets quietly fitted to the record. The single observation that would settle it is a named sovereign programme or enterprise on-premises purchase whose own procurement document cites open-weight availability as a reason for the deployment. If two more quarters produce open-weight releases and still no such attribution anywhere, the honest move is to demote the demand-routing half out of the hypothesis rather than keep scoring it. **AGAINST IT:** The same vendor that published the open checkpoint simultaneously announced it would reroute its own hosted flagship endpoint to a different, smaller model, which is an open release paired with a hosted retreat rather than an advance on both fronts. The week's new open checkpoints are explicitly cheaper and smaller rather than more capable, so the frontier gap the hypothesis tracks did not close this week. One of the two open releases also rests on a grade-three source with vendor-reported benchmarks and no independent evaluation, so the claimed capability is not yet confirmable.

H5 · STRAINED **Power is the binding constraint for the next 24 months.** Prior verdict: strained in W36, which held that power is a major site-timing constraint and not a proven unique bottleneck. W37 does not resolve that objection, so it stays strained, and the hypothesis's own word is where it fails: binding. What the week does support is that power has become a contractual constraint. The clearest tell is a unit change: the week's only operator to disclose delivered capacity reported it in megawatts and accelerators rather than in dollars, which is what a business does when its constraint is energisation rather than procurement. Around it, a hyperscaler announced an investment plan of at least €13 billion and contracted up to half of an existing nuclear plant's output for twenty-two years, an arrangement whose value is preservation of roughly a tenth of a member state's electricity rather than new generation. The federal energy department opened an information request implementing a bulk-power-system emergency authority. A state regulator's docket turned into a fight over forfeiture terms on a 75-megawatt interconnection rule, which is what scarcity looks like in a queue. What the week does not support is uniqueness. Financing, export licences, silicon timing and customer prepayment each gated a named in-window transaction, and this issue says so in four separate places. A constraint that shares the week with four others is not the binding one. **AGAINST IT:** The same quarter that makes the case also complicates it: 850 MW delivered in ninety days by a single operator is evidence that the constraint is being worked faster than a thirty-six to forty-eight month interconnection narrative implies. The largest capacity figure circulating this week, a reported plan exceeding 38 gigawatts, rests on an unconfirmed internal document and carries the issue's lowest source grade, so it should not be counted as evidence for anything. And the Finnish transaction adds roughly 4.7% of new generation at the site while contracting roughly ten times that volume, which means the marquee power event of the week was mostly a reallocation of existing supply rather than relief of the constraint.

PATTERN WATCH

Headline capital figures are increasingly ceilings conditioned on decisions the announcing party does not control, and the conditions are disclosed by the counterparty rather than by the announcer. (4 WEEKS OBSERVED)

Next week: Before 2026-12-31, at least one gigawatt-scale AI infrastructure announcement will be publicly revised downward or restructured when a third-party condition named in its own primary filing fails to clear. Falsified if the next four such announcements each reach a final investment decision or first drawdown on their originally announced scope and schedule.

The object an enterprise evaluates and the object it subsequently runs are drifting apart, through benchmark revision, endpoint substitution, and same-day model deprecation. (2 WEEKS OBSERVED)

Next week: Before 2027-03-31, at least one major model provider or evaluation publisher will ship a versioned runtime or harness manifest, naming adapter version, tool permissions, context persistence policy and retry budget, so that a score can be tied to a reproducible configuration. Falsified if the next three index or harness revisions ship without any machine-readable configuration identifier.

SECOND-ORDER EFFECTS

Delivered-capacity disclosure becomes a competitive norm, and the absence of it becomes informative. Expect analysts and investment committees to start asking every operator for megawatts delivered, megawatts energised and accelerators in service per period as three separate numbers, and to treat backlog-only disclosure as a weaker signal than it has been treated for the last two years. The reporting unit change also imports the energy sector's vocabulary into technology disclosure, which makes contingency easier to see because energy disclosure conventions already separate committed from pre-decision capital. Trigger: An operator reported its quarter's AI delivery as 850 MW of capacity and more than 300,000 accelerators delivered, rather than as revenue or bookings. Horizon: Q4 2026 to Q1 2027. Who moves: Public cloud and neocloud operators, equity analysts, investment committees, infrastructure lenders, enterprise capacity planners.

Assurance arrangements signed today are being written against a requirement that will be defined after the contracts are in force, so evidence collected under informal arrangements may not be usable later. The rational response is to start generating registry-eligible artifacts now, which means insisting that current assurance work produce a documented scope, a named methodology, retained evidence and a versioned model inventory, whether or not the provider is registrable yet. The cost lands on enterprises rather than on regulators or vendors. Trigger: A state enacted an auditor registry that makes unregistered covered AI audit work unlawful on reported 2028 and 2029 dates while mandating no audits of anyone. Horizon: 2027 through 2029. Who moves: Enterprise AI governance and model-risk functions, internal audit, third-party assurance providers, AI vendors selling into regulated buyers.

Model identity becomes a change-management artifact rather than a configuration detail. Teams running unattended agent loops will need pinned model versions, a regression suite that runs on vendor-side change, and a monitored signal for endpoint substitution, because agent behaviour can now change in production without the operator shipping anything. Procurement follows: expect deprecation notice periods and endpoint-stability commitments to appear as negotiated terms rather than as documentation. Trigger: Two separate vendor-side model changes landed inside one week: a flagship endpoint rerouted to a smaller model with no caller change, and a model deprecated across agent mode on the day of announcement, in the same release that made unattended scheduled agent tasks available. Horizon: Q4 2026. Who moves: Platform and AI engineering teams, change management, vendor management, anyone operating scheduled or unattended agent workloads.

STRATEGIC OUTLOOK

The useful question coming out of this week is not how much capacity the industry announced but how much of it is a commitment rather than an option. This publication has been asking that since W34, and W36 told readers plainly to separate operating megawatts from phased, optioned and contracted future capacity. What W37 added is a number. One operator's 850 MW and 300,000-plus accelerators delivered in ninety days, disclosed as a pair for a single named period, is the first dated denominator against which the optioned stack can be sized... and delivered is not energised, a distinction the filing itself respects and most of the coverage did not. Almost every other large figure in the window resolves into a ceiling conditioned on a decision somebody else has not taken: an unapproved €700 million of utility capital expenditure, a warrant-vesting schedule, an undrawn export-licence-contingent facility, eight operators' separate pipelines, and a monthly hosting rate that this publication multiplied by twelve because nobody else would state a term. None of that is fraudulent and most of it is competent structuring. It does mean the honest unit of analysis for the next two quarters is the conditionality rather than the headline, and that the conditions are usually disclosed by a counterparty rather than by the announcer. That is where the diligence has to go. Two practical consequences follow for anyone buying rather than selling. First, the measurement layer is now the binding procurement risk. An index rebased twice in four days, an endpoint rerouted to a smaller model, and a model deprecated inside agent mode on the day of announcement all produce the same operational result: the system you tested is not the system you run. Version pinning and regression-on-vendor-change stop being hygiene and become contract terms. Second, the controls are running ahead of the proof and the policy is running ahead of the obligation. Override-resistant agent permissions shipped generally available this week and are independently verifiable in your own tenant, which makes them worth more than any outcome percentage published in the window. Meanwhile a statutory auditor registry now exists that mandates no audits, and a grid-security rule defers its substance to a later filing. The durable move is to buy the controls you can verify, price the conditions you cannot, and refuse to treat an unbaselined vendor percentage as a measurement.

WHERE WE DIFFER

Our read against the field's published takes on the same week — on the record, so you can score us later.

EXTEND SemiAnalysis: The largest guarantee structure of the cycle is circular financing, with a chip vendor manufacturing the demand that shows up in its own revenue.

Our read: We agree the direction is right and think the mechanism matters more than the label: the instrument guarantees residual value on third-party land, power and shell leases rather than chip purchases, and pays only on a defined trigger after ready-for-service conditions, so commitment date and exposure date are deliberately separated. The per-gigawatt ordering in that analysis is the most useful number published on the topic all week. Attribution where it belongs: the circular-financing frame applied specifically to a residual value guaranty originates with Pip Theory on August 28, and the analysis we are extending is the in-window restatement of it. The number both of us should lead with is the gap between roughly $600 billion reported in July and roughly $105 billion actually signed. We extend it one step: the same structure appears in this week's neocloud financings at smaller scale, where an undrawn licence-contingent facility and a convertible note with an investor put do the same work of raising optionality rather than capital.

DIFFER World Nuclear News, September 9: Google's Finnish package is a landmark nuclear-for-AI commitment that brings new clean capacity to the grid and protects households from AI-driven price increases.

Our read: The facts are solid and the framing is inflated. This issue's house measurement takes the transaction apart from the two primary filings: the additional generation disclosed is 48 MW against a 1,014 MWe site, 4.7%, of which only 10 MW is attributed to the agreement itself, while the contracted volume is roughly 10.6 times the addition. More decisively, the price-protection claim depends on approximately 70% of the life-extension programme's capital expenditure, EUR 700 million, that the counterparty discloses as pending investment decisions. No surveyed outlet set those two statements side by side, and both sit in the primary sources.

DIFFER ModelCap snapshot, September 10, reproducing Intelligence Index v4.3; the index article itself is corroboration rather than the foil: Independent benchmarking now shows the two leading frontier models level, with the newer entrant having caught the incumbent.

Our read: Our disagreement is with the downstream reproduction, not with the evaluator. The convergence is an artifact of instrument revision rather than model movement: Artificial Analysis v4.1.1 placed Astra five points behind Fable 5.1, while v4.3 revised the instrument and reported a rounded tie at 53 and 53 without either model changing. ModelCap's September 10 downstream snapshot reports 53.4 against 52.8 under a reasoning-unspecified configuration, which is not like for like with Artificial Analysis's max-with-fallback comparison and cannot be used to unround its tie. Artificial Analysis disclosed every change, which is why this is a lesson rather than a scandal. The finding that survives is the 57% cost spread for the same rounded score, which is a procurement fact.

DIFFER A vendor quote in the Nscale and Figure releases of September 3, adopted across multiple outlets: Physical AI has arrived and robotics now justifies frontier-lab-scale compute, with the robotics flywheel activated by a multi-billion-dollar compute commitment.

Our read: Credit first, because it is owed: Forbes is the outlet that caught the financing asymmetry, roughly $1.9 billion raised in the buyer's lifetime against $3.5 billion committed to rent computers, and most of the coverage did not. The bull frame itself originates in a vendor quote inside the two companies' own September 3 releases, and that is what we differ with. The underlying deal is dated September 3 and belongs to the prior week, so only the commentary is in this window, and the commentary did not read the buyer's own papers. The model the compute is meant to train was trained on roughly 500 hours of supervised data and deploys an 80-million-parameter policy onboard, while independent scaling work finds a power law in environment and object diversity and no clear power law from more demonstrations in the same setup. The 80-million-parameter onboard policy and the 100,000-accelerator order have not previously appeared in the same sentence, and they should.

OPEN Counterbrief frame sweep of the surveyed coverage; the Federal Register notice of September 9 is the item that circulated: The regulatory item the week's coverage picked up was the federal bulk-power-system notice; no outlet in the surveyed set elevated the September 8 large-load ride-through proposal at all.

Our read: This is an unopposed house prioritisation rather than a disagreement with a named party, and we would rather say so than manufacture a foil out of a primary regulatory document that holds no editorial view. On September 8 a regional transmission organisation presented proposed ride-through, ramp-rate, telemetry and remote-disconnect requirements for large computational loads, motivated by load disconnections that produced multi-gigawatt events, targeting a November filing. No analyst in the surveyed set connected it to the gigawatt-scale campuses sitting inside that footprint. Whether it reaches them, and at what cost, is genuinely open, and it is the question we would most like answered before the next twenty-year lease is signed.

EARLY WARNING PANEL

The levers we monitor.

10 metrics, current vs prior period. **0 rising**, **0 falling**, 10 steady. Each metric carries a threshold value where the read materially changes.

METRIC	CURRENT	PRIOR	DIR	THRESHOLD / NOTE
Frontier lab cash runway at current burn	**Unknown — no constituent disclosed cash, burn, or financing in the window**	Unknown — W36 had no reproducible cash-and-burn input table either	→	Below 18 months for any disclosed-burn lab
Hyperscaler AI capex to disclosed AI revenue ratio	**Unknown as a top-four ratio — one constituent disclosed $28.5B of quarterly capital expenditure and roughly $5B of negative free cash flow, with no AI-segment revenue line**	Unknown — no reproducible top-four capex and AI-revenue input range	→	Above 6x sustained for two consecutive quarters
CoreWeave contracted revenue backlog	**$104.2B as of June 30, unchanged — no CoreWeave filing landed in the window**	$104.2B as of June 30	→	Sequential decline, or conversion below 15% annually
NVIDIA quarter-over-quarter data center revenue	**$89.0B for Q2 FY27, unchanged — no NVIDIA print in the window**	$89.0B for Q2 FY27	→	Two consecutive quarters of sequential decline

METRIC	CURRENT	PRIOR	DIR	THRESHOLD / NOTE
Open-weight to closed-model capability gap on coding	**Not measurable this week — the reference index changed basis twice in four days, breaking comparability with the prior reading**	Untested in W36 as well; no new independent open-frontier benchmark set	→	Open weights within 2 Index points of the closed leader
Sovereign AI program commitments	**Unknown — no government-funded national compute programme in the window meets the ledger method**	Unknown — W36 had no reproducible programme ledger	→	Above 20 programs or $250B committed
PJM capacity auction clearing price	**$325.00 per MW-day for 2028/29, unchanged — no auction occurred in the window**	$325.00 per MW-day for 2028/29	→	An auction clearing below the cap, or a FERC-approved increase in the cap itself

METRIC	CURRENT	PRIOR	DIR	THRESHOLD / NOTE
Time from interconnection request to energization	**Unknown as a queue duration — the window's only hard figure is 850 MW delivered within one quarter by a single operator, which measures delivery rather than energisation or queue time**	Unknown — no reproducible multi-queue duration table	→	Below 48 months in two or more major queues
Cost per task, frontier reasoning model	**$3.26 is the floor of the two leading models and $7.63 the other, per index task on the new v4.3 basis; no frontier-tier median is computable because the basis changed twice inside the window, and neither figure is commensurate with the prior reading**	Prior reading withheld because the index basis changed twice inside the window	→	A frontier-tier reasoning model below $1 per million output tokens

METRIC	CURRENT	PRIOR	DIR	THRESHOLD / NOTE
Custom silicon share of hyperscaler AI compute	**Unknown — a multi-generation custom inference agreement was signed but discloses a warrant-vesting ceiling rather than units, share, or a service date**	Unknown — prior quarter's supplier disclosure combined custom accelerators with networking	→	Above 45% share with audited hyperscaler mix disclosure

THRESHOLD VALUES ARE THE POINTS WHERE THE READ FLIPS. CROSSINGS ARE FLAGGED IN THE ISSUE BODY WHEN THEY HAPPEN.

PREDICTIONS

What we expect next.

7 falsifiable, time-bounded predictions. Each carries a confidence (1–99%, never 50%), a deadline, and a specific signal we'll watch. Future issues score them hit / miss / partial.

PREDICTION 01 CAPITAL

62%

BY MARCH 31, 2027

A publicly traded operator other than Oracle discloses, for a specific reporting period, both a megawatt capacity figure delivered or placed in service and a unit count of AI accelerators delivered, by March 31, 2027.

TRIGGER: Hit only if an earnings release, 10-Q, 10-K, or call transcript from an operator other than Oracle states both a megawatt capacity figure delivered or placed in service and a count of accelerators delivered or deployed, for a named period. A backlog, contracted-capacity, or announced-pipeline figure is a miss.

PREDICTION 02 POWER

44%

BY JUNE 30, 2027

Fortum announces an approved investment decision covering at least EUR 300 million of the EUR 700 million of Loviisa life-extension capital expenditure currently disclosed as pending, by June 30, 2027.

TRIGGER: Hit only if a Fortum release, interim report, or annual report states an approved investment decision of at least EUR 300 million for the Loviisa life-extension programme. Reaffirmation of the programme, or approval below EUR 300 million, is a miss.

PREDICTION 03 SOFTWARE

38%

BY MARCH 31, 2027

A major model provider or evaluation publisher ships a machine-readable runtime or harness manifest that ties a published score to a reproducible configuration, by March 31, 2027.

TRIGGER: Hit only if public documentation or a results artifact contains a versioned configuration identifier naming at least harness or adapter version, tool permissions, context persistence policy, and retry budget, which is the same four-element bar this issue's pattern watch sets for the same event. Prose disclosure of methodology changes, or publishing two harness results side by side without a configuration identifier, is a miss.

PREDICTION 04 POWER

58%

BY DECEMBER 31, 2026

PJM's Section 205 filing on large computational loads is docketed by December 31, 2026 and carries a telemetry or remote-disconnect requirement, not merely a ride-through envelope or a ramp-rate limit.

TRIGGER: Hit only if a FERC filing by PJM docketed on or before December 31, 2026 includes a telemetry or remote-disconnect requirement applicable to large computational loads. A filing carrying only a voltage or frequency ride-through envelope or only a ramp-rate limitation is a miss, as is a further stakeholder presentation or any slip past year end.

PREDICTION 05 NETWORKING

66%

BY JUNE 30, 2027

At least two distinct vendors announce 1600ZR-conformant coherent pluggable optics products by June 30, 2027.

TRIGGER: Hit only if product announcements or datasheets from two distinct vendors cite conformance or compliance with the OIF 1600ZR Implementation Agreement. Demonstrations, interoperability plugfests, and roadmap statements without a named product are a miss.

PREDICTION 06 SOFTWARE

29%

BY MARCH 31, 2027

OpenAI's managed Agents API supports zero data retention or a non-US data residency option by March 31, 2027.

TRIGGER: Hit only if OpenAI documentation states zero-data-retention eligibility or a non-US data residency option for the managed Agents API. A self-hosted sandbox, a roadmap statement, or residency support on other OpenAI surfaces but not the Agents API is a miss.

PREDICTION 07 HARDWARE

33%

BY MARCH 31, 2027

Qualcomm or its counterparty discloses a named service date, first-deployment date, or unit volume for the multi-generation custom AI inference agreement, by March 31, 2027.

TRIGGER: Hit only if a filing, release, or call transcript from either party states a calendar service or first-deployment date, or a unit or megawatt volume, for the agreement. Restating the transaction value, the warrant structure, or a generational roadmap without a date or volume is a miss.

TRACK RECORD

Every prediction ever made, scored against its own written trigger when the deadline passes; ambiguity resolves against us.

Predictions made	**111**	Resolved (hit / partial / miss)	**57 (23 / 15 / 19)**
Hit rate (partial = half)	**54%**	Brier score (0 = perfect)	**0.200**

Bold (<55%): 1 resolved, 100% hit rate vs 43% mean confidence · Core (55-80%): 55 resolved, 52% hit rate vs 66% mean confidence · High-conviction (>80%): 1 resolved, 100% hit rate vs 84% mean confidence

WATCHLIST

On the radar.

6 catalysts in the next 7–14 days that would change the read materially. Watching these tells us whether the thesis is strengthening or weakening.

SEP 15-17

WATCH 01

AI Infra Summit disclosures on 102-terabit switching and scale-up fabrics

Two of this week's networking items were pre-announcements staged for the conference, so the datasheets, availability windows and named customers all land immediately after this issue's cutoff and will determine whether the roadmap is deployable or aspirational.

SEP 21-23

WATCH 02

OCP Summit and ECOC, where 1600ZR conformance and 800-volt rack power get their first public test

The implementation agreement published this week is only worth what multi-vendor interoperability proves it is worth, and these are the venues where conformant product and power-delivery roadmaps are shown side by side rather than asserted separately.

NOV 2026

WATCH 03

PJM's targeted Section 205 filing on large computational load requirements

If ride-through, ramp-rate, telemetry or remote-disconnect obligations reach gigawatt-scale loads, twenty-year leases already signed cover capacity that can be curtailed involuntarily, which changes leasehold value, project bankability and the exposure profile of the guarantees written against them.

Q4 2026

WATCH 04

The first quarter in which delivered capacity becomes a disclosure norm rather than an outlier

One operator has now published megawatts delivered and accelerators delivered for the same period. Whether peers follow, or decline to, is the single most informative thing about the gap between contracted and delivered capacity across the whole sector.

DEC 2026

WATCH 05

Substance of the bulk-power-system emergency authority deferred by this week's information request

The rule as opened defines machinery without obligation, so what large loads are actually required to do arrives later. Anyone signing an interconnection or lease agreement before then is pricing a requirement that has not been written.

2028-2029

WATCH 06

Commencement of the California AI auditor registry and verification-organisation requirements

Reported statutory dates would make unregistered covered audit work unlawful, which would render today's informal assurance arrangements unusable as later evidence. The dates come from secondary analyses and should be re-checked against the enrolled statute before anyone plans against them.

METHODOLOGY AND AUTHORSHIP

How this brief is built.

Compiled from public announcements, SEC filings, earnings transcripts, and official lab and vendor publications. Every quantitative claim is graded on source quality. Every prediction is falsifiable, time-bounded, and scored hit / miss / partial / pending in future issues.

SIGNAL SCORE RUBRIC

5 / 5	SEC filing or audited disclosure. Multi-source independent confirmation. Operational, not aspirational.
4 / 5	Earnings-call disclosure or primary lab/vendor announcement. Two or more independent sources.
3 / 5	Single primary source. Reasonably consistent with sector data.
2 / 5	Analyst report or secondary press only. Or single primary source with credibility caveats.
1 / 5	Rumor, social media, single non-primary source. Or contradicted by alternate primary sources.

ANYTHING GRADED 2 OR BELOW IS FLAGGED AS NOISE.

PREDICTION OUTCOMES

HIT	The specific observable outcome occurred by the deadline.
MISS	The deadline passed and the outcome did not occur.
PARTIAL	A meaningfully similar outcome occurred but not the literal wording.
PENDING	Deadline not yet reached.

HIT RATE OF 65-75% IS THE TARGET. ABOVE 80% MEANS TOO CONSERVATIVE; BELOW 50% MEANS THE FRAMEWORK IS WRONG.

AUTHORSHIP

Brian Letort

BrianLetort.AI · Independent analysis. All sources public. Not investment guidance.

THIS ISSUE

Issue 21

Week 37 of 2026 · September 12, 2026

WEB

brianletort.ai/industry

Past issues, the working framework, and the LLM Evolutionary Tree companion.