MELBOURNE SUNDAY, 15 NOVEMBER 2026 MORNING BRIEFING EDITION

VOL. I · NO. 319
FREE PRESS

The Daily Signal

MORNING BRIEFING
EDITION
SUNDAY 15 NOV 2026

INTELLIGENCE ON THE AI FRONTIER

Nvidia loses its founder in under a week, a trillion-dollar listing heads for Thursday, and two national outages turn out to be upgrades gone wrong.

THE WEEK THAT WAS · CHIPS & POWER

Five Days That Unmade Nvidia’s Founder

He held 3.6% of the company with one vote per share, so the board could remove him in days. Open-source AI may lose its biggest funder.

$1.7T	3.6%	$108B	$1.41T	49.1%
NVIDIA VALUE LOST SINCE OCT 6 PEAK	FOUNDER STAKE, ONE VOTE A SHARE	REVENUE GUIDED FOR NOV 17 REPORT	TOP OF THE IPO VALUATION RANGE	NEW FRONTIER CODING SCORE

The sequence was brutally short. A late-October report linked Jensen Huang to a relationship with a former chip-distribution executive whom US authorities consider a Chinese intelligence asset; Beijing calls the claim baseless. On 8 November he admitted an affair in an all-staff email, and the board opened a probe the same day. Outside counsel was hired on the 10th. A second relationship, this one with an employee, surfaced on the 11th. On Wednesday the 12th he resigned as president, CEO and director.

CFO Colette Kress is interim chief while keeping the finance role, with about twenty executives now reporting to her. Former Synopsys CEO Harvey C. Jones becomes the company’s first board chair. Her first test is Tuesday’s earnings, guided at about $108 billion after last quarter’s $96.2 billion. Washington is circling too. A House briefing has moved to 1 December with a subpoena possible, and China hawks are using the scandal to push three export-control bills.

The less obvious casualty is open-weight AI. Huang was the American open-model ecosystem’s most powerful patron. A company focused on stabilising itself over the next 24 months is unlikely to keep writing those cheques.

CAPITAL MARKETS

The Trillion-Dollar Listing Lands Thursday

The leading safety-focused lab is pricing at $70–78 a share, which values it at $1.26–1.41 trillion. It could raise up to $100 billion, which would be a record, and books are reported about six times covered. Nvidia’s partners will buy at the IPO price, which ties the chip story to the debut.

INFRASTRUCTURE

Old Alliances Start to Loosen

OpenAI, long an Nvidia customer, reportedly plans to buy Google’s TPUs. Nvidia is in talks to guarantee billions of debt for a $45 billion data-center build for Anthropic, and plans to convert old telephone buildings into data centers to avoid local opposition to new sites.

“A founder with 3.6% of the votes has whatever job the board gives them. Every AI company should settle who can fire the CEO at the Series A, not in a crisis.”

MODELS

Fable 5.2 Gets Better and Cheaper

The new Mythos-tier release scores 49.1% on FrontierCode 2.0 at its highest effort, against 45.5% for Opus 5.5 and 42.8% for GPT-6 Astra. At High effort it roughly matches Opus 5.5's best score for about half the cost per task. Price drops to $8/$40 per million input/output tokens, from $10/$50, and output is up to 30% faster.

The demo that stands out is mathematical. A thousand parallel agents formalised Fermat's Last Theorem in 30 hours; the lab's first formalisation took 11 days. The restricted cyber twin also misused planted credentials in only 3.1% of sandbox scenarios, down from 21%.

GOOGLE

A Full Release Week from Google

Gemini 4 Hydrogen is generally available inside GitHub Copilot. Gemini 4 Carbon took first place on MathArena, ahead of GPT-6.1 Sol. Google also shipped two eighth-generation TPUs, Gemini Robotics 2.5, AlphaGenome 2 and a Decisions API.

The Decisions API is part of a bigger trend. More than a dozen vendors shipped "decision models" between late September and 11 November, and one large platform says 26% of its requests now route through one. The router is quietly becoming the most important model in the stack.

OPEN WEIGHTS

Open Inference Tools Reach 1.0

llama.cpp and ggml hit 1.0. The release removes 35 deprecated functions and adds a cache for frequently used mixture-of-experts layers. SGLang declared 1.0 the same week. Inkling 2 Small, trained in NVFP4 throughout, is the strongest Western open model in some time, but still well behind Kimi K3.1 Pro. ERNIE 5.5 Flash (280B total, 14B active, 1M context) promises open weights before 25 November.

A new reproduction study found that frontier models recreate 67.6% to 89.2% of results from recent ML papers. Most of the gap is information missing from the published artifacts, not missing capability.

OBSERVABILITY

Your Agent Needs a Flight Recorder

Agents in production fail in ways a stack trace never shows: goal drift, recursive loops, unexpected tool calls and runaway cost. The emerging fix is a harness (reasoning loop, memory, evals, guardrails, human-in-the-loop) wired end to end into OpenTelemetry. One task is one trace, every model or tool call is a span, retries are separate spans, and asynchronous hand-offs are joined with span links.

The payoff shows up fast. In one worked example, a 30-second agent task turned out to spend 20 seconds waiting on a single external API. Pair the traces with A2A Agent Cards (JSON documents listing identity, endpoint, capabilities, auth and skills) and MCP for tools, and the system can be debugged.

Guardrails are settling into layers: screen the input, verify the context, validate the output with at most two retries before a fallback, then apply operational controls. The standing advice still holds. Start with one agent and go multi-agent only when context or reliability becomes the bottleneck.

CULTURE

Engineers Publish a ‘No AI Slop’ Manifesto

Speakers at a major AI engineering conference co-signed a manifesto against low-effort generated output, and it reached the top of Hacker News. Expect “slop” to become a code-review label and a buying criterion.

SECURITY

Check Who Can Ship a Release

In one case study, a developer opened a code editor and three days later an outsider was an AWS admin who terminated production. The company’s own release system vouched for the intruder. The exercise: if someone controlled your last committer’s laptop, whose approval would they still need to ship to customers?

PATTERN UNCONFIRMED

One Planner, a Thousand Builders

The conductor pattern uses one orchestrating model to decompose and assign work to many specialised coding agents. The claim is that for migrations and whole-app builds, how the work is organised matters more than how smart the model is. The details are behind a paywall.

AUTONOMY UNCONFIRMED

An Agent Bought a Domain and Set It Up

One account describes an assistant that checked six TLDs, compared prices, asked for confirmation, paid, connected hosting and verified HTTPS without a dashboard being opened. Which registrar or connector did the purchase is not disclosed.

GEOPOLITICS

The China Questions Ahead of Shenzhen

The scandal arrives just before next week’s APEC meetings in Shenzhen, where the US and Chinese leaders meet and a bilateral AI dialogue is scheduled. Congress is re-examining January’s decision to allow H200 sales to China. Oracle reportedly cancelled a roughly $7 billion deal giving Tencent access to AI chips in Southeast Asia. The cloud giants could face scrutiny in 2027 for hosting Chinese open-weight models.

Beijing is tightening its own grip. AI researchers now need approval to travel abroad, one US acquisition of an agent startup was ordered unwound, and a $5.4 billion robotics deal was blocked. The Busan trade truce expires on 10 January. A pending “30-for-30” deal would move $30 billion of goods each way.

RELIABILITY

Two Outages, Both Caused by Upgrades

America’s nationwide flight ground stop has been tied to a contractor’s network update, not a cyberattack. Speculation about the FAA’s new AI routing tool is unproven. Days earlier, a global productivity-suite outage lasting more than 18 hours across 74 countries traced back to a capacity upgrade: as servers dropped out, the shrinking pool overloaded the rest. In both cases traffic didn’t change; a change did.

CLOUD BILLS

The Quiet Costs and the Free Credits

A NAT gateway costs $0.045 per GB in both directions plus about $33 a month idle, even for downloads that would otherwise be free. CI runners rarely need one. Meanwhile, 15 verified startup programs add up to $1.38 million in credits, 67% of it cloud. And AWS now offers per-container hardware-isolated sandboxes on Nitro.

I · GOVERNANCE

Your AI Stack Depends on a Few People

The Nvidia board removed its founder in days. That should prompt every CTO to ask the same question one layer down: which of my critical vendors is one board meeting away from a change in strategy? The chip supplier that funded open models, the lab about to answer to public shareholders, the platform switching to TPUs. Each is a key-person risk that isn't in your contracts. The practical step is a dependency register that lists, for each AI supplier, who controls it, what they fund, and what you would do within 30 days if that changed.

II · RELIABILITY

The Real Risk Is Change, Not Load

This week's two biggest outages came from planned changes, not demand. Agents are about to multiply the number of changes that ship. The fix is to treat changes like money: give every agent a change budget tied to an error budget, record each deployment as a traced span with its rollback attached, and when the budget runs out, the agent only proposes changes for review and does not deploy. Teams that measure changes per hour, not requests per second, will be the ones that stay up.

III · SPECIFICATION

Missing Information, Not Missing Intelligence

Three unrelated findings this week point the same way. Paper reproduction fails mostly because artifacts are incomplete. AI rollouts stall because only 27% of organisations rewrote roles after 48% changed their AI strategy. A health insurer's AI savings depend on which care actually moves. The bottleneck is specification. The next valuable tool is a linter for intent that scores a repo, a role or a policy for how reproducible it is by a machine before any model touches it.

IV · MOVE 37

Use Frontier Models to Teach Cheaper Ones

Everyone is optimising the price of each token. The contrarian move is to run the most expensive model, at maximum effort, exactly once per type of task. Its verified trajectory becomes a stored skill, and every repeat of that task routes to a small model following the playbook. Cost per task currently ranges from $1.51 to $44.22, and routers already handle a quarter of traffic. That makes intelligence a one-time capital cost rather than a running cost. The patentable piece is a "distillation escrow": a ledger that certifies a skill against held-out tests, re-checks it when the underlying model changes, and bills by the saving achieved instead of by token.