

THE MORNING BRIEF

Thursday, October 8, 2026 · BIG EDITION

Your stack made the news today. Literally. — 24 stories, zero filler.

01 NEW AI TOOLS

models, agents, and the business of intelligence

- **Nous Research hits a \$1.5B valuation, launches “Hermes for Businesses”**

Your free-model provider just became a unicorn. The open Hermes Agent has 24M clones driving an estimated 2.5% of global token usage, and the new enterprise tier keeps data private for multi-step workflows. Watch the free tier like a hawk after a \$90M Series B — nothing stays free forever.

- **Anthropic ships Claude Haiku 5.5 — 75% cheaper than 4.5**

Third Claude 5.5 in a month, priced at ten cents per million input tokens. This is the subagent-workhorse tier: cheap, fast, built for classification and extraction. If Hermes ever needs a cloud router for grunt work, this is the shape of it.

- **OpenAI rolls GPT-6 + “Intelligent UI” into ChatGPT**

GPT-6 Sol for paid tiers, Luna hits the free tier today. The headline feature is inline interactive UI — buttons, charts, widgets generated mid-conversation. Sound familiar? Your Hermuse command center has been doing live-metric dashboards for a week. You're ahead of the consumer curve.

- **Meta's Muse personal agent lands on iPad**

A month after the mobile launch, Meta pushes its agent to tablets. The multi-device agent race is the consumer mirror of your Jarvis setup — except yours runs on hardware you own, with no one's cloud in the loop.

- **Nylas launches IAM: scoped API keys for every AI agent touching email/calendar**

Per-service, per-account keys with every call verified and logged, free on every plan. If Hermes ever gets email access, this is the pattern to copy: least-privilege keys, full audit trail, no shared secrets.

02 COMPUTER HARDWARE

silicon, memory, and the local-inference economy

- **NVIDIA DGX Station for Windows: GB300, 748 GB unified memory, 20 PFLOPs**

252 GB of HBM3e on the GPU plus 496 GB of system RAM in one box, now with a Windows option so enterprise devs can stop dual-booting. Your dual-Arc 32 GB envelope is the scrappy home-lab answer to the same problem — unified memory is the whole game.

- **NVIDIA RTX Spark PCs open for pre-order — Grace + Blackwell, up to 128 GB unified**

Laptops from \$2599 with the N1X chip (20-core Grace, 6144-core Blackwell), on shelves Oct 16. First credible 'buy a laptop that runs real models' moment — unified memory finally reaching consumer machines is the thing to watch, not the gaming FPS claims.

- **RX 9070 GRE briefly dips to \$529.99 as GDDR6 prices surge 200%**

One weird sub-MSRP blip in a market where DRAM is up 170% year-over-year because AI datacenters are eating all the memory. Consumer GPUs are collateral damage in the memory wars. Your A770s were bought at exactly the right time.

- **15 GPU cloud list prices rose in September; AWS hikes Capacity Blocks 15%**

Zero cuts across 176 tracked GPU SKUs — Verda and RunPod raised 15 between them, AWS just joined in. Every data point says the same thing: local VRAM is an appreciating asset. Keep both cards.

03 SOFTWARE

releases, open source, and Linux-adjacent drops

- **Strata v0.1.40.2 adds Intel Arc support + speed gains for local LLMs**

An Intel Arc-specific local-LLM build — though upstream still lists Arc as experimental, build-from-source, so don't treat it as first-class yet — landing the same week you're fighting riser gremlins on the A770. Worth a look once card two is reseated.

- **VS Code 1.141 ships cross-platform sandboxing for AI agents**

Sandboxed execution for agent-run code, on every platform. Given Hermes runs shell on your box, agent blast-radius control is the single most overdue feature in the category. Watch how they scope it — that's the design to steal.

- **NVIDIA 615.78.08 Linux driver: Proton latency fixes, Wayland HDR crash patches**

Team Blue on GPUs so this isn't yours, but Proton latency fixes matter for Linux gaming generally — and Wayland HDR not crashing is a win for everyone on the penguin side.

- **Penguin Mail 1.0.0: Rust email client with local AI via Ollama / LM Studio**

GPL-3.0, Rust, GnuPG built in, and optional local-AI triage through Ollama or LM Studio — meaning it could talk to your llama-server stack. Privacy-first email with a local model brain is very much your aesthetic. One caveat: single niche outlet, tiny repo — treat it as a tip, not gospel. Actually open, too.

- **System76 COSMIC 1.10.0: GPU rendering by default, screenshot gallery, terminal tabs**

The Rust desktop keeps quietly shipping. GPU-composited by default is the kind of 'just works' milestone that makes a DE feel finished.

- **Valve ships SteamOS 0.4.5 beta for the Steam Frame (VR streaming color fix)**

Valve still polishing the Frame's VR streaming pipeline. Linux gaming infrastructure getting steadier every month — the long game is working.

04 VIDEO GAMES

out today and on the radar

- **Clive Barker's Hellraiser: Revival — out today on PC, PS5, Xbox**

Saber Interactive survival horror at \$39.99 standard. Clive Barker plus Saber is a combination with real teeth — on the radar since the June reveal, and horror in October is just correct scheduling.

- **Neon Abyss 2 leaves Early Access today — 4-player co-op, new final boss**

The roguelike sequel exits EA after a year with online co-op for four and a new endgame boss, Theia. The 1.0-or-bust pipeline actually delivered for once.

- **Am I Nima — psychological horror out today, 120k wishlists**

You wake in a basement, your 'mother' tests whether you're really her daughter, and dialogue is built from words you find in the environment. The announcement was back in June — today's news is the launch itself. A genuinely novel mechanic in a genre that recycles.

- **Annihilated: dark dungeon crawler locks Nov 10, passes 150k wishlists**

Release date announced this week: the story-driven roguelite RPG hits PC November 10 (delayed from July) at \$16.99, with a first-boss demo coming soon. November is stacking up nicely for dungeon crawlers — wishlist it and forget it until launch.

05 GITHUB TRENDING

agent tooling is eating the charts

This week's trending repos are nine-tenths AI-agent tooling. The standouts for your stack, via the Oct 4 trending roundup:

- **VoiceStudio — still climbing, fully local voice studio (+16.8k stars in a week)**

Featured in yesterday's brief and still the week's breakout: the local-first voice studio keeps climbing. If you only skimmed it yesterday, now's the hour to read the repo before the crowd does — the Tiny Hermes Buddy voice plan still needs this plumbing.

- **hindsight — agent memory that learns (+14.5k)**

Agent memory is the hard problem in your Hermes trajectory work. Whatever this is doing to earn 14.5k stars in seven days is worth an hour of your time. (Trending figures are week-of-Oct-4 snapshots.)

- **paperclip — the app you use to manage agents at work (+10.7k)**

Fleet management for agents — the 'command center' idea applied to agent teams. Compare notes with your Hermuse worklog views.

- **OpenShell — a safe runtime for fleets of agents (+5.9k)**

Sandboxed agent execution at fleet scale, in Rust. Pairs thematically with the VS Code sandboxing story above — the industry has decided agents need cages.

- **Agent-Reach — still trending, giving an agent the internet (+4k)**

Also carried over from yesterday's brief and still climbing: scoped web access for agents. The capability is table stakes now; the scoping model is the interesting part. Worth seeing how they draw the boundary.

Researched fresh on Oct 8, 2026 and date-verified story by story. Sources: Reuters, TechCrunch, 9to5Mac, Wccftech, FastGPU/DEV, LinuxCompatible, VS Code release notes, ai-minor, RPG Site, GodIsAGeek, indie-games.eu, Capsule Computers, Business Wire. Every item dated Oct 7–8 unless noted. Nothing stale, nothing invented.