

BOTS FIRED

FRIDAY, OCTOBER 9, 2026 | READ TIME: ~7 MIN

Liquid opens edge decision models while enterprise agents gain more plumbing and scrutiny.

Today is less about a single brilliant model than the surrounding machinery: local hardware, testing, governance, hospital review queues, and a little money for the people making very small machines. The useful question is not whether the agent can act. It is who can inspect what happened afterward.

FRONTIER MODELS

Anthropic Claude Haiku 5.5 becomes available through Amazon Bedrock

Amazon Bedrock now offers Anthropic's Claude Haiku 5.5 through regional and global inference profiles. AWS says the model is also available through Claude Platform on AWS, combining Anthropic access with AWS billing and authentication. The company

positions Haiku 5.5 for subagents and high-volume work, with effort controls intended to tune cost against capability per task. AWS says teams can keep data within its infrastructure and use existing IAM, CloudTrail, CloudWatch and Bedrock Guardrails controls. Anthropic calls Haiku 5.5 its fastest and most efficient model in the family, while AWS says it costs around 75 percent less than Haiku 4.5 for most tasks. That pricing and performance framing comes from the providers, so a real workload still needs its own bill and stopwatch. Regional availability still varies by profile. The [AWS Haiku 5.5 availability announcement](#) lists supported inference profiles and access requirements.

⚙️ **HARDWARE & CHIPS**

NVIDIA and Microsoft announce RTX Spark systems for Windows

NVIDIA and Microsoft announced RTX Spark systems for Windows laptops and compact desktops. NVIDIA says the platform combines a Blackwell RTX GPU with an NVIDIA Grace CPU, offering up to 128GB of unified memory and up to one petaflop of FP4 AI performance. It says laptop preorders opened on

announcement day, with compact desktops planned for a later release rather than immediate delivery. Microsoft also announced general availability of Microsoft Execution Containers, which Microsoft describes as operating-system infrastructure for agents running persistently in the background. The companies are pitching the combination as a way to run larger local models and agents on familiar Windows hardware. Hardware claims, model fit and performance figures come from NVIDIA and Microsoft, not a neutral bake-off. Still, the choice to pair local compute with containment and governance is more interesting than a spec sheet trying to bench-press a cloud region. The [NVIDIA RTX Spark and Windows announcement](#) gives the stated specifications and rollout dates.

ENTERPRISE & APPS

Google Cloud announces a Gemini agent for work systems

Google Cloud announced the Gemini agent, a work-focused agent that connects to customer business systems. Google says it can plan work, use skills and tools, choose a model for the task, and return results inside

documents, inboxes and developer environments. The announcement frames the product as a single prompt window for knowledge work, content creation, coding and questions grounded in an organization's business context. It also highlights built-in cost controls plus enterprise security, administration and governance. That is a large job description, even by agent standards. The important qualification is that this is an announcement of Google's intended product behavior, not proof that every connected system will be equally ready or useful. Buyers will still need to examine access boundaries, connector coverage, accuracy and review paths before handing it the keys to the office. The [Google Cloud Gemini agent announcement](#) outlines the company's stated capabilities.

OPEN SOURCE

Liquid AI releases two open-weight decision models for edge use

Liquid AI released d1-3B and d1-omni-600M as open-weight decision models. Rather than generate prose, the models return a decision in one forward pass. The 3B model accepts text and images; the experimental 600M checkpoint

also accepts text with images or audio. Both are available on Hugging Face, and Liquid says llama.cpp support arrived on day one. The company reports 48.57 on its public Decision Index split for d1-3B and says the model answered a single question in 8 ms on an RTX 4090 in its tests. Those are vendor measurements, not a universal speed limit, naturally. Liquid also demonstrated camera-loop demos and a Jetson-based Isaac Sim setup. The release is a useful reminder that some workloads want a bounded choice, not another paragraph with opinions. The weights provide a concrete test bed. The [Liquid AI open d1 announcement](#) has the model details and test setup.

SECURITY & PRIVACY

Vijil introduces adaptive red teaming for enterprise AI agents

Vijil released DART, an automated red-teaming system for enterprise AI agents.

DART uses multiple adversarial agents to probe a target through multi-turn attacks, changing tactics across turns and engagements instead of repeatedly sending a fixed prompt list. Vijil says the system tests an agent's tool use,

memory and multi-turn behavior in its own environment. It also says teams can use predefined risk taxonomies based on OWASP, MITRE and its research, or bring an organization-specific catalog. DART is part of Vijil Diamond, with the vendor describing related modules for discovering agents, applying policy controls and suggesting code changes. The promise is wider than a chatbot jailbreak check, which is sensible once agents hold tools and data. The claim that it finds more issues than static testing is Vijil's, and requires independent use to settle. The [Vijil DART release report](#) describes the product's stated scope.

DEALS & FUNDING

Atomic Machines emerges with \$250M for code-defined micro-machines

Atomic Machines emerged from stealth saying it has raised \$250 million for code-defined micro-machines. The company calls its manufacturing system the Matter Compiler and says it can produce devices with moving parts, multiple materials and features down to single-digit microns without fixed tooling. Its first product is PrimeSwitch PS-150, a power relay

aimed at AI data centers moving toward 800-volt direct current. Atomic says the relay opens in 50 microseconds, carries 150 amps continuously, and is shipping to unnamed early-access customers for evaluation. It did not disclose a valuation or say how the \$250 million was divided across rounds. The company also says its system measures and corrects steps in real time, while its AI runs experiments on hardware. These performance and manufacturing claims are Atomic's, with public demonstrations still ahead. Programmable manufacturing is an ambitious phrase; a relay is at least a refreshingly concrete place to start. The [Next Web Atomic Machines report](#) details the announcement and first product.

Healthleap raises \$38M for hospital record review software

Healthleap raised \$38 million across seed and Series A funding for hospital record review software. TechCrunch reports an \$8 million seed round co-led by Sequoia and First Round, followed by a \$30 million Series A led by Hummingbird Ventures. The company did not disclose a valuation. Healthleap says its platform reads structured records and clinicians'

notes to flag patients who may need additional review for conditions including malnutrition and delirium. It says the software is deployed in more than 50 hospitals and writes risk scores into existing care workflows. These are company deployment claims, not independent evidence of clinical accuracy. A flag still needs review by the care team. Healthleap says it plans to use the money for engineering, product, sales and customer success while adding conditions and expanding care settings. The [TechCrunch Healthleap funding report](#) contains the financing and deployment claims.

Manus raises more than \$500M after its Meta acquisition is reversed

Manus parent Butterfly Effect raised more than \$500 million after Meta's acquisition was forced to unwind. CNBC reports the round was led by Boyu Capital and IDG Capital, with follow-on investment from Tencent, HSG and ZhenFund. The company did not disclose its post-funding valuation. Bloomberg had previously reported that the financing could value Manus at \$4 billion, but that figure is not a company disclosure and should not be treated as the price of a finished deal. Manus said

earlier this month that it had resumed independent operations after separating from Meta. CNBC reports Chinese authorities forced the reversal of Meta's short-lived acquisition after the companies had begun integration work. The funding demonstrates investor willingness to back the company as an independent agent developer; it does not establish profitability or a particular valuation. Funding news has a habit of arriving with a very expensive fog machine. The [CNBC Manus financing report](#) provides the disclosed investors and transaction context.

KEY TAKEAWAYS

Claude Haiku 5.5 is available on Amazon Bedrock - AWS presents it as a configurable, lower-cost model option for defined workloads.

NVIDIA and Microsoft announced RTX Spark Windows systems - availability and performance details are company statements.

Google Cloud announced a Gemini agent for workplace systems - the stated scope depends on customer integrations and governance choices.

Liquid AI released two open-weight decision models - reported performance and latency are the company's own measurements.

Vijil introduced DART for multi-turn agent testing - the product's broader detection claims remain vendor claims.

Atomic Machines says it raised \$250 million for micro-machine manufacturing - its first relay is in early customer evaluation.

Healthleap raised \$38 million for clinical review software - deployment figures and planned expansion are company claims.

Manus raised more than \$500 million as an independent company - its post-funding valuation was not disclosed.

[BotsFired · Bots Keep Firing](#) 🔥