

Cheaper frontier models, and the first checks on AI agents

GPT-6 Sol and Claude Opus 5.5 cut prices, open weights reach a trillion parameters, and supervisors start asking what AI agents are allowed to do.

Welcome to the first issue of AI Weekly, our short read on what changed in AI and what it means for banks, insurers, manufacturers and energy companies. This week two frontier model releases cut prices sharply. Meanwhile, supervisors and industry began working on a less exciting but more important question: how to check what an AI agent is allowed to do.

MODELS & RELEASES

OpenAI halves prices with GPT-6 Sol and Luna

OpenAI released two new GPT-6 models on 22 September. Sol is aimed at everyday professional work and Luna at fast, high-volume tasks; both bring methods from the flagship GPT-6 Astra to cheaper tiers. Sol costs \$2 per million input tokens and \$10 per million output tokens, half the price of the previous tier.

Why it matters: For banks and insurers running document-heavy work such as claims intake or KYC review, a 50% price cut changes which use cases clear the business case. It is also a good moment to re-benchmark existing model contracts.

Our view: We build agent workflows so the model can be swapped. A client can move to a cheaper model by re-running its own evaluation set, without rebuilding the pipeline.

Source: [OpenAI](#)

Anthropic's Claude Opus 5.5 brings top-tier results at a lower price

Anthropic released Claude Opus 5.5 on 22 September, the first model in its 5.5 family. On the company's own benchmarks it performs at the level of its larger Claude Fable 5.1 on most work, at \$4 per million input tokens and \$20 per million output tokens, 20% less than Opus 5. It is available on AWS, Google Cloud and Microsoft Azure.

Why it matters: Availability on all three major clouds matters for regulated firms: they can use the model inside a cloud region and contract they have already approved, instead of onboarding a new vendor.

Our view: Two frontier releases on the same day show how fast the model layer moves. Keep the model behind a stable interface, and invest in data access, evaluation and audit trails, which outlast any single model.

Source: [Anthropic](#)

AGENTS & TOOLING

Baselayer launches 'Know Your Agent' credentials for AI agents that transact

US start-up Baselayer raised a \$35 million Series A on 22 September and launched an Agentic Identity Suite. Its 'Know Your Agent' credential lets a bank or merchant check who deployed an AI agent, whom it represents and what it may do. FIS, Prove and Socure are among the partners.

Why it matters: As customers start sending AI agents to open accounts, pay bills or file claims, banks and insurers need to tell an authorised agent from an impersonator. KYA extends the familiar KYC logic to software acting on someone's behalf.

Our view: In the agent workflows we build, every action is tied to a named person or system identity and logged. Shared agent identity standards will matter once agents start acting across company boundaries.

Source: [Baselayer \(PR Newswire\)](#)

ENTERPRISE ADOPTION

Survey: most companies running AI agents have already had an agent-related incident

A survey published by Guild.ai on 22 September found that 96.4% of IT decision-makers believed they had a complete, accurate inventory of their AI agents. Yet 66.7% of organisations using agents reported an operational consequence caused by an agent in the past 12 months. Guild.ai sells agent management tools, so the figures are worth reading with that in mind.

Why it matters: Supervisors will expect firms to know which agents run where, with which permissions. An agent inventory is becoming as basic as an application inventory, and the gap between perceived and actual control is a risk in itself.

Our view: Every agent we put into production has an owner, a defined permission scope and monitoring. Those three answer most of the questions an auditor will ask.

Source: [Guild.ai \(GlobeNewswire\)](#)

REGULATION & GOVERNANCE

US state bank supervisors publish an AI examination framework

On 16 September the Conference of State Bank Supervisors released an AI Supervisory Framework for examiners of state-chartered banks and licensed non-banks. It sets out the questions examiners can ask, the documents they can request and how to decide when an AI system needs a closer look, with an optional worksheet for tiering AI use cases by risk. CSBS stresses it is a discretionary tool, not a new rule.

Why it matters: Although it is a US document, it is a practical preview of what supervisors ask: an inventory of AI uses, risk tiering, third-party and model risk, and customer impact. The same themes run through European and Turkish guidance on AI in finance.

Our view: We deliver AI systems with the documentation examiners ask for: use-case description, risk tier, data sources, points of human oversight and monitoring. Preparing it from day one costs far less than reconstructing it before an inspection.

Source: [National Law Review](#)

SOVEREIGN AI & INFRASTRUCTURE

Xiaomi publishes a trillion-parameter open-weight model under the MIT licence

On 21 September Xiaomi released its MiMo-V2.6 models on Hugging Face, including a Pro model with about one trillion parameters and a one-million-token context window, and a smaller Flash model. The weights are under the MIT licence, which allows commercial use, fine-tuning and self-hosting. The models accept text, image, video and audio.

Why it matters: A permissive licence gives institutions that cannot send data to a foreign API a credible option to run a strong model in their own data centre. The Pro checkpoint alone is around 570 GB, so the smaller variants are the realistic starting point for most.

Our view: Sovereign, on-premise deployments are a large part of our work. Before adopting any open model we check its licence and provenance, and test it on the client's own data and languages, including Turkish and Dutch.

Source: [Xiaomi MiMo on Hugging Face](#)

19 EU countries design Europe's first joint AI infrastructure project

On 16 September the European Commission welcomed the design of IPCEI-AI, the first Important Project of Common European Interest dedicated to AI, notified by 19 member states and coordinated by Germany. Eleven of those states, which will support participants with state aid, said they would start pre-notifying their projects to the Commission in September.

Why it matters: The aim is AI infrastructure and applications whose data stays under EU jurisdiction. For European enterprises, including in the Netherlands, this could add EU-governed options for compute and models alongside the US hyperscalers, though reaching production will take years.

Our view: We already build on-premise and EU-hosted deployments for clients who need data residency. Projects like this widen the set of European building blocks we can use.

Source: [European Commission](#)

REGIONAL WATCH

Türkiye's COP31 presidency announces an 'Antalya Pledge' on AI and energy

On 21 September in New York, COP31 President and Environment Minister Murat Kurum announced the Antalya Pledge on AI, a political commitment on how AI is designed, procured, powered and governed in support of climate action. It is being developed with the International Telecommunication Union, and a draft will be shared before the summit opens in Antalya in November. Kurum called on companies to be open about their energy use and to run on clean energy.

Why it matters: AI's energy use is moving onto Türkiye's political agenda. Companies running large AI workloads, and energy companies supplying data centres, should expect questions about the source and amount of energy behind their AI.

Our view: Sizing models to the task, using smaller models where they are enough and measuring inference cost are part of how we design systems. The same measurements will support energy reporting.

Source: [Climate Home News](#)

HUMAIN's Horizon Ultra AI PC goes on sale to enterprises

Saudi Arabia's HUMAIN and Qualcomm made their Horizon Ultra laptop available for enterprise purchase from 20 September. The Windows device runs on a Snapdragon X2 Elite chip with a dedicated neural processing unit, built to run AI tasks on the device itself.

Why it matters: On-device AI keeps sensitive documents on the laptop instead of sending them to a cloud service, which suits Gulf data-localisation rules and the needs of banks and public bodies. It also shows HUMAIN building products beyond data centres.

Our view: Many of our conversations in the Gulf start with data residency. Running smaller models on the device or on premises, and keeping cloud models for the tasks that need them, is a pattern we use often.

Source: [Qualcomm](#)

WHAT WE'RE WATCHING

Anthropic has said Sonnet 5.5 and Haiku 5.5 will follow within weeks. We also expect the first IPCEI-AI pre-notifications to reach the Commission, and a draft of the Antalya Pledge before COP31 opens in November.

Want to apply this in your organisation?

We build production-grade AI agents for banking, insurance, manufacturing and energy. A working POC usually takes one to two weeks.

[Read on agenticdynamic.com](https://agenticdynamic.com)

[Talk to us →](#)