

GPT-6 Astra What it means for the next leg of AI

Saxo Bank · 2026-09-07 · English original

Key points:

- AI is moving from chatbots towards digital workers. GPT-6 Astra points to models doing more multi-step professional work across coding, research, software and cybersecurity. The opportunity is significant, but reliability, security, cost and governance remain important constraints.
- The infrastructure story is broadening beyond GPUs. More capable models and potentially wider adoption can increase demand for compute, memory, networking, storage and power, although the risk is that capacity and spending expand faster than end-demand or returns.
- The next phase is increasingly about AI economics. More capable models can also be more expensive to run. The market is likely to focus less on who spends the most and more on who can translate AI into revenue, productivity and better margins after accounting for the cost of AI itself.

OpenAI launched GPT-6 Astra on September 3, describing it as its most capable model yet. The headline benchmark numbers are striking, but the most important development for investors is arguably less glamorous: Astra is much better at using computers and completing multi-step professional tasks.

It can browse, write software, analyse data and work across applications rather than simply responding inside a chat window. OpenAI describes improvements across computer use, research, coding and complex end-to-end work.

It also brings a significant step-up in cybersecurity capabilities. OpenAI says Astra is its first model to reach the “Critical” level for cybersecurity capability under its Preparedness Framework, potentially allowing it to identify previously unknown vulnerabilities and work through complex security problems. That creates potential applications in cybersecurity, but also introduces obvious misuse risks, which is why OpenAI has added stronger monitoring and safeguards around the model.

The direction of travel looks increasingly like:

And that is more important for markets than whether we officially call it AGI (artificial general intelligence).

The economic opportunity potentially expands considerably if AI can complete meaningful parts of professional workflows rather than simply assist with them. But technological capability is only one side of the equation. Models still need to prove they can operate reliably, securely and at a cost that businesses can justify.

Our thoughts on the launch

- Capability is still moving quickly — and that supports continued AI spending. GPT-6 Astra represents another step forward in computer use, coding, cybersecurity and multi-step professional workflows. Staying at the frontier remains highly compute- and capital-intensive, keeping the broader AI infrastructure build-out in focus.
- Better capability could unlock broader adoption. If AI can complete more useful parts of professional workflows, usage could spread further across enterprises and industries. But adoption will still depend on reliability, security, integration and businesses being willing to redesign workflows around AI.
- Better AI also comes at a higher cost. GPT-6 Astra's standard API pricing is \$10 per million input tokens and \$50 per million output tokens, compared with \$4 and \$20 respectively for GPT-5.6 Sol — 2.5x higher on both measures. A model can therefore be technically impressive without being economically sensible for every workload.
- That raises the hurdle for monetisation. Companies will increasingly need to show that more capable AI generates enough additional revenue, labour savings or productivity to justify the higher cost. Cheaper models may still make more economic sense for many routine tasks.
- The next phase becomes more selective. Stronger models can sustain infrastructure demand, but the debate is shifting from simply measuring AI spend to assessing the returns on that spend. The risk is that capital and operating costs rise faster than adoption or monetisation.

1. Compute: the infrastructure cycle is not finished

Frontier AI remains extremely compute-intensive.

Nvidia CEO Jensen Huang said Astra was trained using more than 100,000 Nvidia Grace Blackwell NVLink72 systems, highlighting the scale of resources required to keep pushing model capability forward.

That keeps demand for GPUs and accelerators, data centres, cooling and power in focus. More capable agents could also increase recurring compute needs if each user instruction triggers multiple steps of reasoning, research and software interaction.

But the key question is increasingly whether utilisation and monetisation can justify the capital being deployed. If adoption disappoints or models become more efficient faster than expected, returns on some infrastructure investments could come under pressure.

2. Memory and networking: the bottlenecks broaden

More complex models and workflows require large amounts of data to be moved, accessed and stored. That raises the importance of high-bandwidth memory, advanced packaging, networking, optical connectivity and storage.

Simply put, increasingly powerful processors are less useful if data cannot reach them quickly enough.

That broadens the infrastructure story beyond GPUs. But it also introduces a familiar semiconductor-cycle risk: rapid capacity expansion can eventually produce oversupply if expected AI demand does not materialise.

3. Software: opportunity and disruption

More capable agents could expand AI adoption across coding, finance, legal services, design, cybersecurity, customer support and other knowledge-work industries.

But the impact on software is unlikely to be uniformly positive.

AI could make some applications significantly more valuable by improving what users can achieve and allowing companies to charge for better outcomes. At the same time, an AI layer capable of operating across multiple applications could reduce the need for certain products or software seats.

That puts greater emphasis on valuable workflows, proprietary data, customer relationships and distribution — and on whether AI functionality ultimately translates into sustainable revenue or productivity gains.

Bottom line: Productivity becomes the real test

So far, AI capital expenditure has been much easier to observe than AI productivity.

The next phase is likely to require clearer evidence that all this spending is translating into measurable economic outcomes, whether through:

- higher revenue per employee
- faster product development
- lower servicing costs
- stronger margins
- greater output from existing resources

The companies spending the most on AI will not necessarily generate the strongest economic returns. Some businesses may ultimately benefit by using increasingly capable models to improve their own productivity without having to build the technology themselves.

That leaves the AI investment story with two forces moving in parallel: capability continues to improve, supporting further infrastructure and adoption, while rising costs make the economics of that adoption increasingly important.

Frontier models and agentic workloads require significant processing power

Capex runs ahead of utilisation and returns

More complex AI increases data movement and storage needs

Capacity expansion creates future oversupply

Training and broader AI usage support infrastructure demand

Higher costs, grid constraints and overbuilding

AI adoption can create recurring compute demand

Price competition and higher model costs pressure returns

AI can automate larger parts of professional workflows

AI may also disrupt seats, pricing and existing products

Potential for higher productivity and lower costs

Benefits may take longer to materialise or fail to justify AI costs

GPT-6 therefore reinforces the familiar question for AI: not simply how intelligent the models become, but whether that intelligence can be turned into sustainable economic value.

Source: <https://www.home.saxo/content/articles/equities/gpt-6-astra-what-it-means-for-the-next-leg-of-ai-07092026>