

# Stop ‘tokenmaxxing’ and deploy AI sensibly instead



**Companies, tech workers and researchers are in a frenzy to embed agentic AI into their workflows, locked in a self-imposed race not to fall behind. There must be a better way to make use of AI technology.**

It is only a few years ago that large language models (LLMs) emerged and transformed artificial intelligence (AI) technology. By now, many AI users have moved on to agentic AI approaches, in which one or many LLMs tackle multi-step tasks by accessing various tools and databases, and by reasoning, planning and collaborating with each other. AI agentic frameworks can display seemingly semi-autonomous capabilities in code writing, financial transactions, clinical decision making and even scientific discovery.

A recent viral news story highlights how keen companies are for employees to commit to agentic AI, encouraging the practice of ‘tokenmaxxing’, in which tokens are bits of text processed by LLMs<sup>1</sup>. In some cases, companies even introduced internal leaderboards that ranked employees by the number of tokens used. Although token usage is surely not a good metric for productivity, some believe it is an effective way to encourage a change in mindset, counting on agentic AI to be transformative in every area. In a widely reported recent quote, Jensen Huang, chief executive officer of Nvidia, said he expects a high-level engineer to consume US\$250,000 worth of tokens per month<sup>1</sup>.

Are there drawbacks to tokenmaxxing? Of course. Yet the frenzy seems hard to stop, with few willing to be left behind. Token usage has exploded this year, fuelled in part by the rise of OpenClaw, an open-source, user-friendly platform designed to run agentic AI applications continuously. The demand for tokens has reached fever pitch. However, although much of the tech industry seems to be in denial, hard limits are now being met<sup>2</sup>. There are not enough data centres, the ‘token factories’ that underpin the AI rush, or enough electricity, water and hardware to sustain them. The environmental costs of building data centres are substantial, and local communities are raising vocal opposition to further expansion<sup>3</sup>.

With the cost of AI continuing to rise, and tech companies struggling to make a profit<sup>4</sup>, several firms are beginning to restrict access to their AI tools or are shutting down platforms altogether. In March, OpenAI abruptly closed Sora, its video-generation model, only a few months after announcing a US\$1-billion partnership with The Walt Disney company. GitHub has paused new subscriptions to its Copilot programming assistant and will move to usage-based billing from June.

Beyond financial and environmental issues, agentic AI also raises questions about the potential human implications of outsourcing cognitive skills. Concerns about cognitive decline have already been raised, rightly or wrongly, in connection with the widespread use of chatbots to generate text and write essays, emails, cover letters and papers, among other tasks<sup>5</sup>. With the emergence of

agentic AI systems capable of accessing vast knowledge bases and specialized tools, expertise across many domains, including scientific research, can increasingly be outsourced. As a result, agentic AI may enable researchers to move faster or explore new directions, but it also risks weakening the development of users’ skills and domain knowledge.

Ultimately, human ingenuity will remain essential to deploy agentic AI usefully in scientific applications. Developers of end-to-end scientific AI workflows typically warn that human oversight is still required at every stage<sup>6</sup>. This step is much harder than it sounds. A central challenge will be how to sift through the output generated by these agentic AI frameworks – how to validate and decide what to trust. Prompting models and generating results is straightforward (until token budgets run out), but output evaluation is not. The verification of AI-generated outputs is imperative for reliable scientific progress and will almost certainly continue to require substantial human involvement and effort.

Published online: 18 May 2026

## References

1. Keary, T. *Forbes* <https://go.nature.com/4wk7scy> (13 April 2026).
2. *The Economist* <https://go.nature.com/4nml4id> (27 April 2026).
3. Plautz, J. & Tomich, J. *Politico* (2 May 2026).
4. Morris, C. *Fast Company* <https://go.nature.com/3QX4vi9> (6 April 2026).
5. Hogenboom, M. *BBC* <https://go.nature.com/4dfvqx7> (20 April 2026).
6. Jones, N. *Science* <https://doi.org/10.1126/science.zmr759f> (2026).