The latest artificial intelligence safety report

1. The capabilities of AI models are improving

A host of new AI models – the technology that underpins tools like chatbots – were released last year, including OpenAI's GPT-5, Anthropic's Claude Opus 4.5 and Google's Gemini 3. The report points to new "reasoning systems" – which solve problems by breaking them down into smaller steps – showing improved performance in maths, coding and science. Bengio said there has been a "very significant jump" in AI reasoning. Last year, systems developed by Google and OpenAI achieved a gold-level performance in the International Mathematical Olympiad – a first for AI.

However, the report says AI capabilities remain "jagged", referring to systems displaying astonishing prowess in some areas but not in others. While advanced AI systems are impressive at maths, science, coding and creating images, they remain prone to making false statements, or "hallucinations", and cannot carry out lengthy projects autonomously.

Nonetheless, the report cites a study showing that AI systems are rapidly improving their ability to carry out certain software engineering tasks – with their duration doubling every seven months. If that rate of progress continues, AI systems could complete tasks lasting several hours by 2027 and several days by 2030. This is the scenario under which AI becomes a real threat to jobs.

But for now, says the report, "reliable automation of long or complex tasks remains infeasible".

The Guardian, Dan Milmo Global technology editor - Tue 3 Feb 2026

Useful vocabulary

Word	Definition	Word	Definition
capability	ability to do something	hallucination	AI making false statements
to underpin	to support, be the basis of	autonomously	by itself, without human help
reasoning system	AI that thinks step-by-step	infeasible	impossible, not practical
jagged	uneven, irregular	breakthrough	major discovery or achievement
prowess	great skill or ability		