

First full International AI Safety Report, chaired by Bengio, is published

Source copy kept by wheresthe.ai for the story "First full International AI Safety Report, chaired by Bengio, is published" (<https://wheresthe.ai/d/intl-international-ai-safety-report-2025-01/>).

Source: <https://www.gov.uk/government/news/first-international-ai-safety-report-to-inform-discussions-at-ai-action-summit>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: fd84e062d30c0083539285223162bb5fcf6f4aa7701dc54fe7e6906f1d629291

Press release

First International AI Safety Report to inform discussions at AI Action Summit

First Independent International AI Safety Report to become the global handbook on AI safety, ahead of the France AI Action Summit.

From:

Department for Science, Innovation and Technology and AI Safety Institute

Published: 29 January 2025

2024 to 2026 Starmer Labour government ">

This was published under the 2024 to 2026 Starmer Labour government

First Independent International AI Safety Report to become the global handbook on AI safety, ahead of the France AI Action Summit

Inspired by the UN's IPCC Report, the publication sets a new standard for scientific rigor in assessing AI safety

Brings together input from 100 world-leading AI experts put forward by 30 countries including France, China, the USA and UK, as well as the UN, EU, and OECD

Ahead of the AI Action Summit hosted by France next month, the Independent International AI Safety Report published today sets out the first comprehensive, shared scientific understanding of advanced AI systems and their risks.

Spearheaded by Yoshua Bengio - a Turing Award-winning AI academic and the most cited computer scientist in the world - the report brings together insights from 100 independent international experts. Launched at the AI Safety Summit in November 2023, the report is mandated by more than 30 countries including France, China and the United States, with operational support provided by the Department for Science, Innovation, and Technology.

As policymakers worldwide grapple with rapid and unpredictable advancements in AI, today's report contributes to bridging the gap by offering a scientific understanding of emerging risks to guide decision making.

The report also highlights how quickly the technology has evolved in recent years and months, including how AI systems are increasingly capable of acting as AI agents - autonomously planning and carrying out complex tasks.

Its publication looks to plug the gaps by building up a scientific basis of evidence to support policymakers in advancing AI safety, while the full implications of advanced AI systems are still being discovered.

Report's Chair, Yoshua Bengio, Full Professor at Université de Montréal and Scientific Director of Mila - Quebec AI Institute, said:

The capabilities of general-purpose AI have increased rapidly in recent years and months. While this holds great potential for society, AI also presents significant risks that must be carefully managed by governments worldwide.

This report by independent experts aims to facilitate constructive and evidence-based discussion around these risks and serves as a common basis for policymakers around the world to understand general-purpose AI

capabilities, risks and possible mitigations.

Key areas identified for further research include how rapidly capabilities will advance, how general-purpose AI models work internally, and how they can be designed to behave reliably.

While there are still many challenges in mitigating the risks of general-purpose AI, the report highlights promising areas for future research and concludes that progress can be made. The report emphasises widespread agreement that improving our understanding of how AI works should be a top priority, as international governments and AI companies prepare to gather for the AI Action Summit.

Ultimately, the report emphasises that while AI capabilities could advance at varying speeds, their development and potential risks are not a foregone conclusion. The Report concludes by saying that the outcomes depend on the choices made by policymakers both today and in the future.

Secretary of State for Science, Innovation, and Technology, Peter Kyle said:

The transformative potential of AI is clear, which is why we have placed it at the heart of our government's Plan for Change. It will help us kickstart economic growth, transform public services, and boost the living standards of working people across the country, but I remain clear eyed that safety must be baked in from the outset.

The UK is already at the forefront of building the global consensus needed on responsible AI, and this report will go a step further as we prepare for the AI Action Summit. It will support decision-makers with the scientific evidence they need to seize the opportunities of AI, which is a charge we are already leading by putting the technology to work to deliver more jobs, more money in people's pockets, and transformed public services.

French Minister Delegate for Artificial Intelligence and Digital Technologies, Clara Chappaz said:

Artificial intelligence is a central topic of our time, and its safety is a crucial foundation for building trust and fostering adoption. Scientific research must remain the fundamental pillar guiding these efforts. I salute the work of Yoshua Bengio and the international team who produced this report, work which must be perpetuated in the long term in the general interest.

This first comprehensive scientific assessment provides the evidence base needed for societies and governments to shape AI's future direction responsibly. These insights will inform crucial discussions at the upcoming AI Action Summit in Paris.

Notes to editors

The UK government will continue to provide the Secretariat for the report until a suitable long-term international home is agreed, and Professor Yoshua Bengio will continue acting as chair for 2025. This will be informed by ongoing global dialogues on AI governance, including those within the UN Global Digital Compact, the Network of AI Safety Institutes, and other forums, along with ongoing stakeholder consultations.

DSIT media enquiries

Email press@dsit.gov.uk

Monday to Friday, 8:30am to 6pm 020 7215 3000

Share this page

The following links open in a new tab

Share on Facebook

(opens in new tab)

Share on X

(opens in new tab)