

UK turns its Frontier AI Taskforce into a permanent AI Safety Institute

Source copy kept by wheresthe.ai for the story "UK turns its Frontier AI Taskforce into a permanent AI Safety Institute" (<https://wheresthe.ai/d/gb-ai-safety-institute-launch-2023-11/>).

Source: <https://www.gov.uk/government/news/prime-minister-launches-new-ai-safety-institute>

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

SHA-256 of the saved text: 098ae4f55688aee4d992486dccc2bc6dd2e6f86d251b230065a870b9f41ffd30

Press release

Prime Minister launches new AI Safety Institute

World's first AI Safety Institute launched in UK, tasked with testing the safety of emerging types of AI.

From:

Prime Minister's Office, 10 Downing Street , Department for Science, Innovation and Technology , The Rt Hon Michelle Donelan and The Rt Hon Rishi Sunak MP

Published: 2 November 2023

2022 to 2024 Sunak Conservative government ">

This was published under the 2022 to 2024 Sunak Conservative government

Government's plan for the AI Safety Institute unveiled to cement the UK's position as a world leader in AI safety

new hub will help spur international collaboration on AI 's safe development, with leading AI companies and nations including the US, Singapore and Google DeepMind agreeing to partner with the institute

the AI Safety Institute puts the UK's leading Frontier AI Taskforce on a permanent footing, following two days of praise for its work on AI safety

A new global hub based in the UK and tasked with testing the safety of emerging types of AI has been backed by leading AI companies and nations, as the world's first AI Safety Institute launches today (2 November).

After four months of building the first team inside a G7 Government that can evaluate the risks of frontier AI models, it has been confirmed today that the Frontier AI Taskforce will now evolve to become the AI Safety Institute, with Ian Hogarth continuing as its Chair. The External Advisory Board for the Taskforce, made up of industry heavyweights from national security to computer science, will now advise the new global hub.

The Institute will carefully test new types of frontier AI before and after they are released to address the potentially harmful capabilities of AI models, including exploring all the risks, from social harms like bias and misinformation, to the most unlikely but extreme risk, such as humanity losing control of AI completely. In undertaking this research, the AI Safety Institute will look to work closely with the Alan Turing Institute, as the national institute for data science and AI .

In launching the AI Safety Institute, the UK is continuing to cement its position as a world leader in AI safety, working to develop the most advanced AI protections of any country in the world and giving the British people peace of mind that the countless benefits of AI can be safely captured for future generations to come.

World leaders and major AI companies have today expressed their support for the Institute as the world's first AI Safety Summit concludes. From Japan and Canada to OpenAI and DeepMind, the collective backing of key players will strengthen international collaboration on the safe development of frontier AI – putting the UK in prime position to become the home of AI safety and lead the world in seizing its enormous benefits.

Leading researchers at the Alan Turing Institute and Imperial College London have also welcomed the Institute's launch, alongside representatives of the tech sector in TechUK and the Startup Coalition.

Already, the UK has agreed two partnerships: with the US AI Safety Institute, and with the Government of Singapore to collaborate on AI safety testing – two of the world's biggest AI powers.

Deepening the UK's stake and influence in this transformative technology, it will also advance the world's knowledge of AI safety – with the Prime Minister committing to invest in its safe development for the rest of the decade, as part of the Government's record investment into R&D .

Prime Minister Rishi Sunak said:

Our AI Safety Institute will act as a global hub on AI safety, leading on vital research into the capabilities and risks of this fast-moving technology.

It is fantastic to see such support from global partners and the AI companies themselves to work together so we can ensure AI develops safely for the benefit of all our people. This is the right approach for the long-term interests of the UK.

Secretary of State for Science, Innovation, and Technology, Michelle Donelan said:

The AI Safety Institute will be an international standard bearer. With the backing of leading AI nations, it will help policymakers across the globe in gripping the risks posed by the most advanced AI capabilities, so that we can maximise the enormous benefits.

We have spoken at length about the Summit at Bletchley Park being a starting point, and as we reach the final day of discussions, I am enormously encouraged by the progress we have made and the lasting processes we have set in motion.

The launch of the AI Safety Institute marks the UK's contribution to the collaboration on AI safety testing agreed by world leaders and the companies developing frontier AI at a session in Bletchley Park this afternoon.

New details revealed today, as governments from across the globe gathered for a second day of talks, set out the body's mission to prevent surprise to the UK and humanity from rapid and unexpected advances in AI . Ahead of new powerful models expected to be released next year whose capabilities may not be fully understood, its first task will be to quickly put in place the processes and systems to test them before they launch – including open-source models.

From its research informing UK and international policymaking, to providing technical tools for governance and regulation – such as the ability to analyse data being used to train these systems for bias - it will see the government take action to make sure AI developers are not marking their own homework when it comes to safety.

AI Safety Institute Chair Ian Hogarth, said:

The support of international governments and companies is an important validation of the work we'll be carrying out to advance AI safety and ensure its responsible development.

Through the AI Safety Institute, we will play an important role in rallying the global community to address the challenges of this fast-moving technology.

Researchers are already in place to head up the work of the Institute who will be provided with access to the compute needed to support their work. This includes making use of the new AI Research Resource, an expanding £300 million network that will include some of Europe's largest super computers, increasing the UK's AI super compute capacity by a factor of thirty.

It follows the UK Government's announcement yesterday of additional investment in Bristol's "Isambard- AI " and a new computer called "Dawn" in Cambridge, that researchers will be able to access at the same time to boost their research and make AI safe. The AI Safety Institute will have priority access to this cutting-edge supercomputer to help develop its programme of research into the safety of frontier AI models and supporting government with this analysis.

It comes as government representatives were joined by CEOs of leading AI companies and a number of civil society leaders earlier today to discuss the year ahead and consider what immediate steps are needed - by countries, companies, and other stakeholders – to ensure the safety of frontier AI .

As the final day of talks come to a close at Bletchley Park, the AI Safety Summit has already laid the foundations for talks on frontier AI safety to be an enduring discussion with South Korea set to host next year.

Further information:

Read an overview of the AI Safety Institute .

A CEO for the Institute will be recruited in due course.

Notes for editors:

U.S. Secretary of Commerce Gina Raimondo said:

I welcome the United Kingdom's announcement to establish an AI Safety Institute, which will work together in lockstep with the U.S. AI Safety Institute to ensure the safe, secure, and trustworthy development and use of advanced AI . AI is the defining technology of our generation, carrying both enormous potential and profound risk. Our coordinated efforts through these institutes is only the beginning of actions to facilitate the development of safety standards, build testing capabilities for advanced AI models, and to expand information-sharing, research collaboration, interoperability, and policy alignment across the globe on AI safety.

Singapore Minister for Communications and Information Josephine Teo said:

The rapid acceleration of AI investment, deployment and capabilities will bring enormous opportunities for productivity and public good. We believe that governments have an obligation to ensure that AI is deployed safely. We agree with the principle that governments should develop capabilities to test the safety of frontier AI systems. Following the MoUs on Emerging Technologies and Data Cooperation signed by Singapore and the UK earlier this year, we have agreed to collaborate directly with the UK to build capabilities and tools for evaluating frontier AI models. This will involve a partnership between Singapore's Infocomm Media Development Authority and the UK's new AI Safety Institute. The objective is to build a shared understanding of the risks posed by frontier AI . We look forward to working together with the UK to build shared technical and research expertise to meet this goal.

Canadian Minister of Innovation, Science and Industry, the Honourable François-Philippe Champagne said:

Canada welcomes the launch of the UK's AI Safety Institute. Our government looks forward to working with the UK and leveraging the exceptional Canadian AI knowledge and expertise, including the knowledge developed by our AI institutes to support the safe and responsible development of AI .

The Government of Japan said:

The Japanese Government appreciate the UK's leadership in holding the AI Safety Summit and welcomes the UK initiative to establish the UK AI Safety Institute. We look forward to working with the UK and other partners on AI safety issues toward achieving safe, secure, and trustworthy AI .

The German Government said:

Germany is interestedly taking notice of the foundation of the AI Safety Institute and is looking forward to exploring possibilities of cooperation.

CEO of Amazon Web Services Adam Selipsky said:

We commend the launch of the UK AI Safety Institute. As one of the world's leading developers and deployers of AI tools and services, Amazon is committed to collaborating with government and industry in the UK and around the world to support the safe, secure, and responsible development of AI technology. We are dedicated to driving innovation on behalf of our customers and consumers, while also establishing and implementing the necessary safeguards to protect them.

CEO & co-founder of Anthropic Dario Amodei said:

While AI promises significant societal benefits, it also poses a range of potential harms. Critical to managing these risks is government capacity to measure and monitor the capability and safety characteristics of AI models. The AI Safety Institute is poised to play an important role in promoting independent evaluations across the spectrum of risks and advancing fundamental safety research. We welcome its establishment and look forward to partnering closely to advance safe and responsible AI .

CEO & co-founder of Google DeepMind Demis Hassabis said:

AI can help solve some of the most critical challenges of our time, from curing disease to addressing the climate crisis. But it will also present new challenges for the world and we must ensure the technology is built and deployed safely. Getting this right will take a collective effort from governments, industry and civil society to inform

and develop robust safety tests and evaluations. I'm excited to see the UK launch the AI Safety Institute to accelerate progress on this vital work.

CEO & co-founder of Inflection Mustafa Suleyman said:

We welcome the Prime Minister's leadership in establishing the UK AI Safety Institute and look forward to collaborating to ensure the world reaps the benefit of safe AI .

President of Global Affairs at Meta Sir Nick Clegg said:

Everyone has a responsibility to ensure AI is built and deployed responsibly to create social and economic opportunities for all. We look forward to working with the new Institute to deepen understanding of the technology, and help develop effective and workable benchmarks to evaluate models. It's vital that we establish ways to assess and address the current challenges AI presents, as well as the potential risks from technology that does not yet exist.

Vice Chair and President of Microsoft Brad Smith said:

We applaud the UK Government's creation of an AI Safety Institute with its own testing capacity for safety and security. Microsoft is committed to supporting the new Institute and to advancing the close collaboration that will be needed among governments, with industry, and with academic researchers and across civil society. These new steps will be vital to ensuring that innovation and safety move forward together.

CEO of OpenAI Sam Altman said:

The UK AI Safety Institute is poised to make important contributions in progressing the science of the measurement and evaluation of frontier system risks. Such work is integral to our mission – ensuring that artificial general intelligence is safe and benefits all of humanity – and we look forward to working with the Institute in this effort.

Dr Jean Innes, CEO of The Alan Turing Institute, said:

AI has immense potential to do good, but in order to realise the benefits our societies must be confident that risks are being addressed. We welcome the AI Safety Institute which will generate further momentum in this global endeavour, and we look forward to collaborating in the weeks and months ahead, helping to leverage the Turing's expertise alongside the science and innovation capabilities of the UK's universities, research community and wider AI ecosystem, building on the country's strong track record of delivering work on AI safety, ethics and standards.

Executive Director of Startup Coalition Dom Hallas said:

We're proud to see the UK take this critical step in its work on AI safety because a well-rounded approach to the issues at hand is vital to the AI ecosystem. When partnered with the UK's other initiatives - and hopefully future ones that tackle talent, compute, and investment - that all focus on safe scaling and AI adoption, the UK is well on its way to creating a state capacity unlike any of our international competitors. Nailing the fundamentals of AI safety and building the regulatory capacity to keep up with the rate of innovation are large steps. When coupled with a well-rounded approach that tackles the needs of our AI startups and scaleups, the AI Safety Institute will help ensure the UK and its businesses' places as global AI leaders.

Julian David, CEO of techUK, said:

techUK welcomes the establishment of the AI Safety Institute which will carry forward the UK's pioneering work on Frontier AI . We are pleased that the Institute will have three clear objectives: to develop and conduct evaluations on advanced AI systems; to drive foundational AI safety research; and to facilitate information exchange. These are important but complex tasks and it is vital that the Institute has access to the compute capacity and skills that it will need. The diplomatic effort invested in the AI Safety Summit should help to ensure that the Institute is well placed to build further international collaboration on frontier AI . We look forward to working with the AI Safety Institute to facilitate industry collaboration in this important area.

Professor Mary Ryan, Vice-Provost (Research and Enterprise), Imperial College London said:

The new UK AI Safety Institute is an important step in our understanding of AI risks. Universities will play a critical role with the new Institute in the UK AI ecosystem - accelerating innovation from foundational to applied AI . Only by combining deep technical and academic expertise together with that of industry and policymakers, can we

effectively develop frameworks that will ensure the safe, productive and accelerated deployment of AI .

Shahid Omer, Director of Policy at Universities UK said:

We welcome today's announcement that will help to cement the UK's status as a world-leader in AI research and AI safety. Backed by the world-leading research and innovation of UK universities, this important new Institute can help the UK to further understand and take advantage of AI , in a safe and secure manner.

UK universities are well placed to explore both the technological and societal impacts of AI . New funding announced earlier this week will also enable our universities to continue to carry out cutting-edge research into AI , boost the UK's own AI skills base, as well as help attract talented AI researchers from overseas.

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)