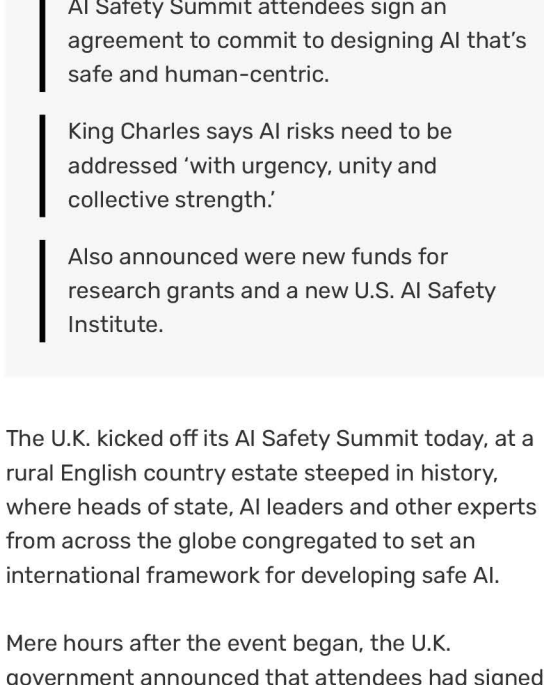


AI Safety Summit: 28 Nations and EU Sign the ‘Bletchley Declaration’

Each nation can classify and categorize AI risks based on their own circumstances and legal frameworks

Ben Wodecki, Jr. Editor
November 1, 2023 9 Min Read



Attendees during the International summit on Artificial Intelligence at Bletchley Park. JUSTIN TALLIS/AFP VIA GETTY IMAGES

At a Glance

AI Safety Summit attendees sign an agreement to commit to designing AI that’s safe and human-centric.

King Charles says AI risks need to be addressed ‘with urgency, unity and collective strength.’

Also announced were new funds for research grants and a new U.S. AI Safety Institute.

The U.K. kicked off its AI Safety Summit today, at a rural English country estate steeped in history, where heads of state, AI leaders and other experts from across the globe congregated to set an international framework for developing safe AI.

Mere hours after the event began, the U.K. government announced that attendees had signed the **Bletchley Declaration on AI Safety**, named after its venue Bletchley Park, the birthplace of modern computing and the site of the British code-breaking operation in World War II led by computer science pioneer Alan Turing.

The Bletchley Declaration on AI Safety

The agreement is a list of pledges to ensure AI is “designed, developed, deployed, and used, in a manner that is safe, in such a way as to be human-centric, trustworthy and responsible.”

Signatories commit to working together through existing international forums to promote cooperation on addressing the risks of AI across its lifecycle, including identifying risks of shared concern, building understanding, and developing policies. However, each nation can classify and categorize AI risks based on their national circumstances and legal frameworks.

The agreement also commits to holding more AI Safety Summits – reaffirming the need for more inclusive global dialogue on the topic.

The signatories listed on the **policy paper** were 28 nations including the U.S. and China, plus the EU bloc. They agreed to pay particular attention to the risks arising from so-called ‘frontier AI’ – defined as highly capable general-purpose AI models including foundation models – without naming the actual nations’ representatives.

Related: UK PM Sunak Urges Balanced Rules on AI, Invites China to Summit

“This is a landmark achievement that sees the world’s greatest AI powers agree on the urgency behind understanding the risks of AI – helping ensure the long-term future of our children and grandchildren,” said U.K. Prime Minister Rishi Sunak.

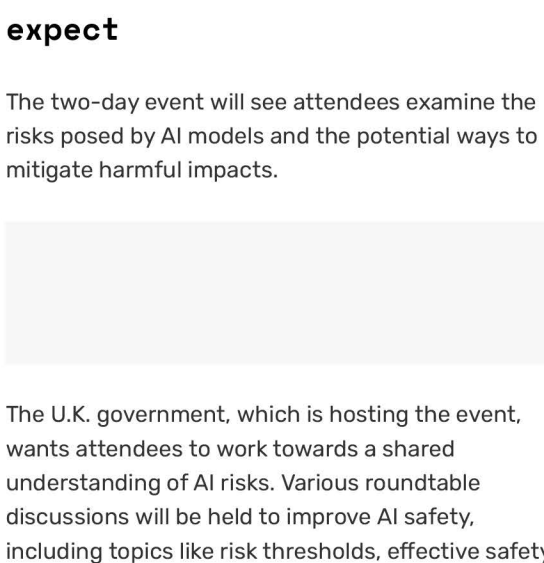
U.K. Tech Secretary Michelle Donelan said the event itself “marks the start of a long road ahead, and the Summit will kickstart an enduring process to ensure every nation and every citizen can realize the boundless benefits of AI.”

However, Robert F. Trager, director of the Oxford Martin AI Governance Initiative at Oxford University, said in an emailed comment that the declaration nevertheless is “short on details of how countries will cooperate on these issues. The Summit appears to have achieved a declaration of principles to guide international cooperation without having agreed on a roadmap for international cooperation.”

He also pointed out that the agreement said organizations developing powerful ‘frontier AI capabilities’ such as big tech companies, carry a “particularly strong responsibility” to ensure safety. However, “this suggests that governments are continuing down the road of voluntary regulation, which is very likely to be insufficient.”

Related: AI Leaders Warn About Existential Risks Again – Now Armed with Facts

U.S. AI Safety Institute



U.S. Commerce Secretary Gina Raimondo | Credit: LEON NEAL/POOL/AFP via Getty Images

During the opening plenary, U.S. Commerce Secretary Gina Raimondo announced that the U.S. will be launching an AI Safety Institute.

The U.S. AI Safety Institute (US AISI) will reside inside the National Institute of Standards and Technology (NIST). It will be tasked with creating guidelines, tools, benchmarks and best practices for evaluating AI risks.

US AISI will also develop technical guidance for regulators for drafting rules on issues like watermarking AI-generated content, transparency and privacy.

The new entity will also share information with similar institutes, such as the U.K.’s own version – UK AISI.

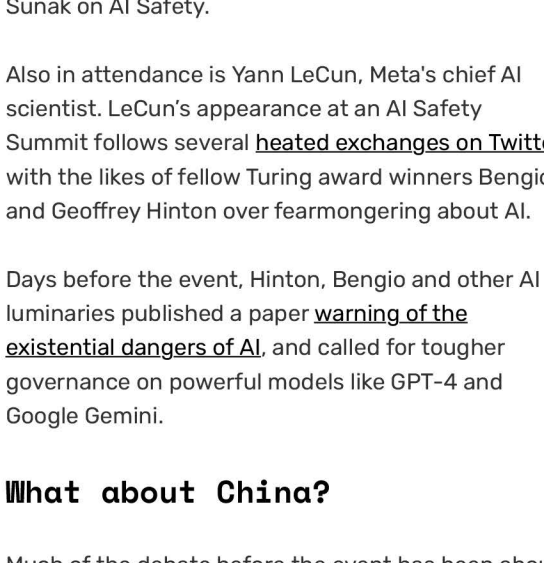
“I will almost certainly be calling on many of you in the audience who are in academia and industry to be part of this consortium,” Raimondo said. “We can’t do it alone, the private sector must step up.”

Royal input: King on ‘imperative’ to keep AI safe

During the opening plenary, King Charles III gave a pre-recorded speech where he said AI safety “demands international coordination and collaboration.”

Related: Biden Issues Executive Order on AI Regulation

His Majesty expressed his view that risks presented by AI need to be addressed “with a sense of urgency, unity and collective strength.”



King Charles III during his virtual address

“It is incumbent on those with responsibility to meet these challenges, to protect people’s privacy on livelihoods, which are essential to both economic and psychological wellbeing to secure our democracies from harm.”

The King thanked attendees for their “vital role” in “laying the foundation of a lasting consensus on AI safety and for ensuring that this immensely powerful technology is indeed a force for good.”

AI grants

The day before the event, the U.K. government announced a new £118 million (\$143 million) pledge to boost AI skills funding.

The cash will be used to fund 12 new Centers for Doctoral Training for AI as well as scholarships to “ensure the (U.K.) has the top global expertise and fosters the next generation of researchers needed to seize the transformational benefits of this technology.”

The fund also includes a £1 million (\$1.2 million) pot called the ‘AI Futures Grants,’ which will go towards helping AI researchers relocating to the U.K. The funds are in addition to the £117 million (\$141 million) that the government already pledged.

The U.S. has also earmarked cash for AI advancements. Vice President Kamala Harris, who is attending the event, is set to announce an investment of more than \$200 million in AI grants. Bloomberg reports that the funds will come from philanthropic foundations and aimed at initiatives focused on safeguarding democracy and improving the transparency around AI.

The AI Safety Summit: What to expect

The two-day event will see attendees examine the risks posed by AI models and the potential ways to mitigate harmful impacts.

What to expect

The U.K. government, which is hosting the event, wants attendees to work towards a shared understanding of AI risks. Various roundtable discussions will be held to improve AI safety, including topics like risk thresholds, effective safety assessments, and defining robust governance and accountability mechanisms. The discussion is set to largely focus on ‘frontier models,’ which pose the biggest risks to society.

Major models are set to release in the next year, such as Google Gemini and potentially GPT-5, and the U.K. government argues the capabilities of such models “may not be fully understood.”

In a **speech ahead of the event**, Sunak said he wanted attendees to agree to the first-ever international statement about the nature of AI risks and to establish a “truly global expert panel” for AI, nominated by the countries and organizations attending.

Stay updated. Subscribe to the AI Business newsletter.

Eden Zoller, chief analyst for applied intelligence at sister research firm **Omdia**, **wrote** in a commentary that the event will ultimately be about exploration – “identifying issues and suggesting processes rather than agreeing concrete actions or governance frameworks and milestones.”

She also warned that multilateral collaboration and agreement at an international level is “notoriously difficult.” The EU AI Act illustrates this at a regional level,” the analyst wrote. It was “first tabled in 2021 and has been dogged by disagreements, subject to multiple revisions and is still not completely finalized.”

Who is in attendance?

Some of the biggest names in AI are descending on Bletchley Park for the event. The government publicly confirmed the **guest list** mere days before the event.

Guests include representatives from the United Nations, UNESCO and the European Commission, as well as academics, think tanks and civic organizations such as Oxford and Stanford universities, Rand Corp. and the Algorithmic Justice League.

AI companies who sent representatives include OpenAI, Nvidia, Google DeepMind, IBM, Meta, OpenAI, Alibaba, Anthropic and XAI, tech billionaire Elon Musk’s new startup. Musk, CEO of Tesla and SpaceX, is hosting a livestreamed conversation with Sunak tomorrow after the event closes, on his social media site X (formerly Twitter).

Who is in attendance?

Elon Musk | Credit: Leon Neal/Getty Images

OpenAI CEO Sam Altman also flew in for the event, as has Yoshua Bengio, who is an independent advisor to Sunak on AI Safety.

Also in attendance is Yann LeCun, Meta’s chief AI scientist. LeCun’s appearance at an AI Safety Summit follows several **heated exchanges on Twitter** with the likes of fellow Turing award winners Bengio and Geoffrey Hinton over fearmongering about AI.

Days before the event, Hinton, Bengio and other AI luminaries published a paper **warning of the existential dangers of AI**, and called for tougher governance on powerful models like GPT-4 and Google Gemini.

What about China?

Much of the debate before the event has been about whether should China attend. Colleagues in the U.S. were reportedly at odds with inviting a Chinese delegation, but Sunak wants a diversity of voices to discuss regulating AI.

Deputy Prime Minister Oliver Dowden confirmed that China has accepted the invitation, with a delegation from China listed on the confirmed **guest list**. According to FT sources, China’s delegation will comprise members of the country’s Ministry of Science and Technology.

There will also be a presence from Chinese companies working on AI, including Alibaba, which has been developing open source models like **Qwen-ZB**, and Tencent which is working on its own **large language model**, Hunyuan.

Chinese academics will also be in attendance, including Turing award winner Andrew Yao who co-wrote the paper with Hinton, Bengio and others that called for regulations to avoid AI’s existential threat to humanity.

Summit snubs startups?

While some of the biggest names in AI are attending the event, the presence of startups is rather thin among the event’s 100 guests. The handful attending – besides AI leaders such as OpenAI and Anthropic – included **large language model** developers Cohere and Graphcore, the British startup designing chips for AI workloads. Also invited is Hugging Face, the open source repository platform many AI startups and developers rely on for access to models and systems.

Victor Botev, CTO and co-founder at **Iris.ai**, which uses AI to help researchers make sense of mountains of data from papers, said in an emailed comment that the Summit “missed a great opportunity by only including 100 guests, who are primarily made up of world leaders and big tech companies.”

“It is vital for any consultation on AI regulation to include perspectives beyond just the tech giants. Smaller AI firms and open-source developers often pioneer new innovations, yet their voices on regulation go unheard.”

Hector Zenil, chief visionary officer and founder of **Oxford Immune Algorithmics**, an Oxford University spinout, expressed concerns in an emailed comment that the event is “being heavily influenced by CEOs who are very focused on one branch of AI, namely **large language models** and generative AI.”

“If the AI [Safety] Summit at Bletchley Park and the AI Advisory committee are dominated by individuals with a particular research or commercial focus for AI, then it will make it harder to develop regulatory frameworks which reflect all the potential use cases.”

Read more about:

ChatGPT / Generative AI Applied Intelligence

About the Author

Ben Wodecki
Jr. Editor

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben Wodecki

See more from Ben W