



Bletchley Park: The UK AI Safety Summit and what you need to know

This is the first of our 3-part series on [Artificial intelligence \(AI\)](#) and safety. This briefing specifically covers the transformative insights and outcomes of the [AI Safety Summit 2023](#) held at Bletchley Park.

The UK AI Safety Summit 2023

On 1 and 2 November 2023, the UK hosted the world's first AI Safety Summit at Bletchley Park, the birthplace of modern computing, where Alan Turing famously cracked the enigma code and laid the foundations for a new digital age. The summit brought together leading AI nations, technology companies, researchers, and civil society groups to accelerate global efforts towards the safe and responsible development of frontier AI.

What is frontier AI?

At the summit frontier AI was characterised as highly capable general-purpose AI models that can perform various tasks and match or exceed the capabilities of today's most advanced models.

What happened at the AI Safety Summit?

At the summit, countries participated in a broad and inclusive discussion. To structure discussions, the UK Government published five objectives before the summit, as follows:

Objective 1: a shared understanding of the risks posed by frontier AI and the need for action.

Objective 2: a forward process for international collaboration on frontier AI safety, including how best to support national and international frameworks.

Objective 3: appropriate measures that individual organisations should take to increase frontier AI safety.

Objective 4: areas for potential collaboration on AI safety research, including evaluating model capabilities and the development of new standards to support governance.

Objective 5: showcase how ensuring the safe development of AI will enable AI to be used for good globally.

What were the outcomes of the AI Safety Summit?

The Bletchley Declaration

As part of the international AI safety summit, the UK government published the Bletchley Declaration, a landmark agreement recognising a shared consensus on the opportunities and risks of AI and the need for collaborative action on frontier AI safety. Twenty-eight countries from across the globe, including the US and China, as well as the EU signed the Declaration.

AI Safety Institutes

Rishi Sunak, UK Prime Minister, announced the creation of an AI Safety Institute, a new global hub based in the UK tasked with testing the safety of emerging types of AI. On 1 November, the US also announced that it will launch its own national AI Safety Institute.

AI 'State of the Science' report

The countries represented at Bletchley Park agreed to support the development of an international, independent and inclusive 'State of the Science' report on the capabilities and risks of frontier AI, led by the Turing Award-winning scientist Yoshua Bengio.

AI Safety Testing

In collaboration with AI Safety Institutes, state-led testing was agreed upon to ensure the safety and reliability of the next generation of AI, both pre and post-deployment of models.

Key takeaways

The AI Safety Summit was the first global event to address the risks of frontier AI and how to mitigate them through international cooperation. Whilst no binding commitments were reached at the summit for nations and AI developers, two days beforehand it was announced that US President Joe Biden had signed an Executive Order on Safe, Secure, and Trustworthy AI.

With US companies at the forefront of frontier AI development, US lawmakers will undoubtedly influence the creation of international frameworks addressing AI safety risks. The outcome remains uncertain, as it is still being

determined whether international collaboration will shape national AI strategies or if leading AI nations will establish standards through their own regulatory approaches.

In the context of developing new AI products and solutions, it would align with other similar global initiatives to develop and comply with an international framework, basic manufacturing criteria, safety standards and formatting and applicable regulations.

Meanwhile, in the EU, legislators are in the final stages of agreeing on the AI Act, which seeks to introduce new, comprehensive regulations for the majority of AI systems. In the UK, the Government has proposed that existing regulators develop rules and regulations tailored to their specific sectors and roles.

Looking ahead, the upcoming summits in South Korea and France in 2024 are expected to provide greater clarity and unity on AI regulation and policy at a global level.

As part of this 3-part series of articles on AI safety, we will be covering:

- [Cyber security and AI](#), specifically the consequences of cyber security breaches, relevant guidance and practical steps businesses should take to protect themselves in this landscape; and
- The risks relating to approved and frequent unapproved use of AI by employees within a company and the consequential need for AI policies to be implemented.

For further information on the AI Safety Summit and its impact on your business, please do not hesitate to contact dean.drew@LA-law.com or 0330 0539 759.