

International developments in AI governance: same goal, different paths

On the first two days of November 2023, the UK hosted the inaugural AI Safety Summit at Bletchley Park, which brought together government leaders from across the world to discuss the risks of artificial intelligence (AI) and how these can be mitigated through internationally co-ordinated action. Two days before the summit, on 30 October 2023, the US government issued an executive order outlining an expansive strategy to support the development and deployment of safe and secure AI technologies. Meanwhile, EU lawmakers are currently negotiating the final text of the AI Act, with hopes of approving it by early 2024, although many of the obligations would only begin to apply to regulated entities in 2026 or later.

The summit, the executive order and the AI Act stand as important developments shaping the future of global AI governance and regulation (see box “Timeline and weblinks”). While there are significant differences in approach, the developments share some common ground.

Differences in approach

Through the AI Act, the EU is seeking to implement a new legal framework, modeled on EU product safety legislation, that would impose a detailed set of technical and organisational requirements on providers and users of AI systems (see *News brief “Artificial intelligence: the dawn of a new legal era”*, www.practicallaw.com/w-031-0858 and *feature article “AI governance, risk and compliance: shaping an unknown future”*, www.practicallaw.com/w-040-0428). Providers of high-risk AI systems would bear the bulk of obligations, from data governance, training, testing and validation, to conformity assessments, risk management systems and post-market monitoring. The AI Act would also prohibit some uses of AI systems altogether and impose transparency obligations on others.

In contrast, the executive order does not create new legislative obligations. Rather, it introduces a number of directions for government agencies, including instructing the US Department of Commerce to develop rules requiring companies that develop or provide infrastructure for AI models to make disclosures under certain circumstances. The executive order is also broader in scope than

the AI Act in some respects; for example, it covers social issues such as advancing equity and civil rights, outlines requirements related to attracting and retaining highly skilled AI workers, and directs the US State Department to lead an effort to establish international frameworks governing AI.

Similarly to the US, a new legal framework dedicated to AI is yet to be established in the UK. In contrast to the EU’s approach in the AI Act of creating an overarching AI law, the UK government proposed a sector-specific vision in its March 2023 white paper on a pro-innovation approach to AI regulation (the white paper), and has not proposed introducing new legislation or establishing a new regulatory body to oversee the development or use of AI (see *News brief “White paper on regulating AI: is a pro-innovation approach enough?”*, www.practicallaw.com/w-039-2427). More recently, in the King’s Speech on 7 November 2023, the government outlined its legislative priorities for the parliamentary year ahead and confirmed that it does not plan to introduce AI regulation. Instead, it will require existing regulators to take responsibility for the establishment, promotion and oversight of responsible AI in their respective sectors.

In parallel to pursuing the initiatives set out in the white paper, the UK government convened the summit by setting out the following five key objectives:

- Developing a shared understanding of the risks posed by frontier AI and the need for action (see “Areas of common ground” below).
- Developing a process for international collaboration, including how to support national and international frameworks.
- Developing appropriate measures for individual organisations to increase frontier AI safety.
- Agreeing areas for potential international collaboration on AI safety research.
- Showcasing how the safe development of AI can enable it to be used for good globally.

As part of the summit, 28 attending countries and the EU agreed the Bletchley Declaration, which is a non-binding agreement on the opportunities and risks posed by frontier AI systems. In addition, the UK government launched the AI Safety Institute, which is a new global hub tasked with testing the safety of emerging forms of general-purpose AI. All of the signatories to the Bletchley Declaration have agreed to attend the second and third iterations of the summit, and the US and Singapore have also agreed to partner with the AI Safety Institute in the future.

Areas of common ground

While the UK, the EU and the US have taken varying approaches to regulating AI, there are some key areas in which they are aligned.

Focus on high-risk systems. The AI Act will create a risk-based approach and impose the most significant compliance requirements on providers of AI systems that it classifies as high risk. High-risk systems will be subject to a number of obligations, including requirements that are designed to enable record-keeping, allow for human oversight and achieve an appropriate level of accuracy, robustness and cybersecurity. Notably, the AI Act proposes introducing specific obligations for foundation models in addition to high-risk AI systems. Foundation models are defined as AI models that are trained on broad data at scale, are designed for generality of output and can be adapted to a wide range of distinctive tasks.

The executive order similarly focuses on high-risk AI systems by requiring developers of certain dual-use foundation models to share safety test results and other critical information with the US government. This includes results from red-teaming, which is where the accuracy and technical robustness of an AI model is tested to the limits of its capabilities. The red-teaming and reporting requirements apply to models that present a serious risk to security and meet certain technical requirements outlined in section 4.2(b)(i) of the executive order.

The UK government has also directed its focus towards high-risk AI systems, with the summit and corresponding announcements centering on the opportunities and potential risks stemming from frontier AI systems, which are

Timeline and weblinks

21 April 2021	The European Commission publishes a proposal for the first ever legal framework on artificial intelligence (AI) in the form of the AI Act (https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai).
2 December 2022	The EU-US Joint Roadmap for Trustworthy AI and Risk Management is published (https://digital-strategy.ec.europa.eu/en/library/ttc-joint-roadmap-trustworthy-ai-and-risk-management).
29 March 2023	The UK government publishes a white paper on its pro-innovation approach to AI regulation (www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach).
27 October 2023	The UK government publishes a document on emerging processes for frontier AI safety ahead of the AI Safety Summit (https://assets.publishing.service.gov.uk/media/653aabbd80884d000df71bdc/emerging-processes-frontier-ai-safety.pdf).
30 October 2023	The White House publishes an executive order on the safe, secure and trustworthy development and use of AI (www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/).
30 October 2023	The UK government announces that it will launch the AI-Airlock regulatory sandbox (www.gov.uk/government/news/mhra-to-launch-the-ai-airlock-a-new-regulatory-sandbox-for-ai-developers).
1 and 2 November 2023	The UK hosts the first global AI Safety Summit at Bletchley Park (www.aisafetysummit.gov.uk/).
1 November 2023	The Bletchley Declaration is published (www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023).
2 November 2023	The UK government launches the AI Safety Institute (www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute).
7 November 2023	The UK government confirms in the King's Speech that it does not intend to introduce legislation on AI (www.gov.uk/government/speeches/the-kings-speech-2023).

defined as highly capable general-purpose AI models that can perform a wide variety of tasks and match or exceed the capabilities present in today's most advanced models.

Transparency and labelling. The AI Act will require providers of AI systems that are intended to interact with natural persons to develop them in such a way that individuals know that they are interacting with the system. Similarly, users of AI systems that engage in emotion recognition and biometric categorisation must inform individuals who are exposed to them, and users of AI systems that generate or manipulate images, audio, or video deepfakes must disclose that the content is inauthentic (*see feature article*

"The rise of the deepfake: looking into a dystopian future?", www.practicallaw.com/w-026-8753).

The executive order also addresses transparency requirements for AI-generated content by requiring the US Secretary of Commerce, together with other relevant agencies, to submit a report within 240 days that identifies standards, tools, methods and practices for:

- Authenticating content and tracking its provenance.
- Labelling synthetic content, such as by watermarking.

- Detecting synthetic content.
- Preventing generative AI from producing child sexual abuse material or non-consensual intimate imagery of real people.
- Testing software for these purposes.
- Auditing and maintaining synthetic content.

Following the report, the Director of the US Office of Management and Budget, in coordination with the heads of other relevant US agencies, must issue guidance to agencies for the labelling and authenticating of synthetic content.

The UK has not yet proposed a specific transparency or labelling regime for AI content. However, the government noted in the white paper that regulators may need to look for ways to support and encourage relevant entities to implement appropriate transparency measures, including through product labelling. The government is also considering measures for identifying AI-generated content, including the use of watermarks, as outlined in the document on emerging processes for frontier AI safety that was published in advance of the summit on 27 October 2023. UK regulators are likely to make further announcements on this topic.

AI standards and sandboxes. The AI Act provides for the creation of AI regulatory sandboxes; that is, controlled environments that are intended to encourage developers to test new technologies for a limited period of time, with a view to complying with the AI Act.

The AI Act also provides for the development of harmonised technical standards for the design, development and deployment of AI systems.

Similarly, in the white paper, the UK government confirmed its commitment to supporting sandbox initiatives that are aimed at assisting AI businesses and the development of technical standards as a way of providing consistent, cross-sectoral assurance that AI has been developed responsibly and safely. On 30 October 2023, the government announced that it has begun work on establishing sector-specific AI testbeds, such as AI-Airlock, a sandbox that will assist in the development and deployment of software and AI medical devices.

The executive order also requires the creation of new standards by directing the US National Institute for Standards and Technology to

issue guidelines for AI development with the aim of promoting consensus with industry standards and requiring the US Secretary of Energy to implement a plan for the development of AI model evaluation tools and AI testbeds. As part of efforts related to international leadership, the executive order also directs the US Secretary of State to establish a plan for global engagement on prompting and developing AI standards. Collaboration in this area between the US and the EU will undoubtedly be facilitated through the US-EU Trade and Technology Council's joint Roadmap for Trustworthy AI and Risk Management of December 2022, which aims to advance collaborative approaches in international standards bodies related to AI.

Martin Hansen is of Counsel, Lisa Peets is a partner, and Marianna Drake is an associate, at Covington & Burling LLP.
