

An Overview of the UK's New Approach to AI

The UK Department for Science, Innovation, and Technology (DSIT) released on March 29, 2023 an artificial intelligence (AI) white paper to describe its new approach to regulating AI.¹ The proposal seeks to create a pro-innovation regulatory framework that promotes public trust in AI by creating rules proportionate to the risks associated with different sectors' use of AI. It also commits to establishing a regulatory sandbox to bring together regulators and innovators, so they better understand how regulation affects emerging AI technologies.

Unlike the European Union (EU), the UK's approach to AI will not focus on new legislation in the short term. It will instead focus on creating guidelines to empower regulators and will only take statutory action when necessary. The following explains the heart of the white paper before analyzing its strengths and weaknesses.

What Does Context-Specific Regulation Mean?

According to DSIT's white paper, context-specific regulation focuses on outcomes and does not create rules for entire sectors or technologies. Context-specific regulation will be based on the outcomes that specific uses of AI are likely to generate, like medical diagnostics, machinery depreciation, or clothing returns, and can differentiate between contexts within different sectors, like critical infrastructure or customer service. Context-specific AI regulation acknowledges that all AI technologies in a specific sector have varying degrees of risk. Such regulation weighs the risk of specific AI usage against the costs of missed opportunities from forgoing AI usage. DSIT argues that context-specific AI regulation will help the UK capitalize on the technology's benefits.

What Is the UK's Definition of AI?

In the white paper, DSIT defines AI as "products and services that are 'adaptable' and 'autonomous.'" When defining AI as adaptable, the white paper aims to cover the difficulty of explaining AI logic and outcomes because the technology trains and operates based on inferring patterns and connections that aren't easily understood by humans or initially envisioned by its programmers. Autonomy describes the difficulty in assigning responsibility for an AI technology's outcomes because the technology can make decisions without human intent or control. By focusing on adaptable and autonomous products and services, the UK government hopes to future-proof its AI definition rather than focus on specific methods or technologies like machine learning or large language models (LLMs).

¹ Department for Science, Innovation, and Technology, "A pro-innovation approach to AI regulation," March 29, 2023, <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>; Department for Science, Innovation, and Technology and The Rt Hon Michelle Donelan MP, "UK unveils world leading approach to innovation in first artificial intelligence white paper to turbocharge growth," March 29, 2023, <https://www.gov.uk/government/news/uk-unveils-world-leading-approach-to-innovation-in-first-artificial-intelligence-white-paper-to-turbocharge-growth>.

What Is the Current Regulatory Landscape for AI in the UK?

Inconsistent coordination and enforcement across various regulators, including the Health and Safety Executive, Equality and Human Rights Commission, and Competition and Markets Authority, govern AI in the UK. This inconsistent coordination is why the white paper calls for system-wide coordination to clarify who is responsible for cross-cutting AI risks and to avoid duplicative requirements.

AI is already covered by a few different types of laws and regulations, including the Equality Act 2010 to prevent discrimination according to protected characteristics; UK General Data Protection Regulation to process personal data fairly; product safety law; product-specific legislation for electronic equipment, medical devices, and toys; and consumer rights law to protect consumers. Other relevant laws include the Human Rights Act 1998, the Public Sector Equality Duty, Data Protection Act 2018, and sector-specific fairness requirements like the Financial Conduct Authority handbook.

What Are the Objectives of the Proposed AI Framework?

DSIT describes the proposed AI framework as pro-innovation, proportionate, trustworthy, adaptable, clear, and collaborative. The new regulatory framework will apply to all sectors of the UK economy, rely on interactions with existing legislation to implement the framework, and not introduce new legal requirements unless necessary. The government hopes to minimize extraterritorial effects by not immediately, if at all, introducing new legislation; but this approach will not alter the extraterritorial impact of existing legislation.

In addition, DSIT describes its regulatory framework as having three goals:

1. Drive growth and prosperity to make responsible innovation easier, reduce regulatory uncertainty, and gain a long-term market advantage in AI.
2. Increase public trust in AI by addressing its risks and protecting fundamental values, which will, in turn, drive AI adoption.
3. Strengthen the UK's position as a global AI leader so it remains attractive to innovators and investors while minimizing cross-border friction with other international approaches.

This regulatory framework will not affect issues relating to access to data, compute capability, and sustainability or the “balancing of the rights of content producers and AI developers.”

What Are the UK's Five Principles for Regulating AI?

In its white paper, the UK government focuses on five principles the government believes should govern AI to foster responsible development and use of the technology. The application of these five principles will initially be at the discretion of the regulators and may be followed by a statutory duty requiring regulators to have due regard to the principles.

1. **Safety, Security, and Robustness**
AI applications should be safe, secure, and robust with carefully managed risks. Under this principle, regulators may introduce measures to ensure AI is secure throughout its lifecycle; assess the likelihood AI poses risks to take proportionate measures to manage these risks; and regularly test the functioning, resilience, and security of AI systems to create future benchmarks.
2. **Appropriate Transparency and Explainability**
AI innovators and enterprises must be appropriately transparent and able to explain their AI's decision-making processes and risks. An appropriate level of transparency and explainability is defined as “regulators hav[ing] sufficient information about AI systems and their associated inputs and outputs to give meaningful effect to other principles.” Regulators may look at product labeling and technical standards as options to gather this

information. Regulators will also need to clarify the level of explainability that is appropriate and achievable for specific AI technologies.

3. Fairness

AI should be fair and not discriminate against individuals or commercial outcomes or undermine their legal rights. Regulators may need to develop and publish descriptions of fairness that apply to AI systems within their regulatory domain using relevant laws, like the Equality Act 2010, the Human Rights Act 1998, the Public Sector Equality Duty, UK General Data Protection Regulation, Data Protection Act 2018, consumer and competition law, and sector-specific fairness requirements.

4. Accountability and Governance

Regulatory measures governing AI need to sufficiently hold appropriate actors in the AI life cycle accountable for AI outcomes. Regulators must ensure clear expectations for regulatory compliance and may need to encourage compliance using governance procedures. DSIT acknowledges that it is unclear who should be allocated responsibility in an AI product's lifecycle and thus does not propose intervening at this stage. Instead, DSIT will convene experts, technicians, and lawyers to consider future proportionate interventions.

5. Contestability and Redress

Users and other stakeholders need clear routes to dispute any harm caused by AI. The government expects regulators to clarify existing routes and encourage and guide regulated entities to make sure affected parties can clearly contest harmful AI outcomes through either informal or formal channels.

What Will Regulators Do Under the New Framework?

While DSIT's white paper does not offer an exhaustive list of current regulators that regulate AI technologies, the delineated regulatory framework depends on empowering these regulators to develop context-specific and cross-sector approaches to AI. The paper explains that creating a new AI-specific regulator would introduce more complexity and confusion to a full list of regulators. Current regulators for AI include the Health and Safety Executive, Equality and Human Rights Commission, and Competition and Markets Authority, but the list can include others not mentioned in the white paper.

Under the new approach to AI, these regulators will do the following:

1. Adopt a proportionate, pro-growth, and pro-innovation approach that focuses on specific risks that specific AI poses.
2. Consider proportionate measures to address prioritized risks, considering risk assessments undertaken by or for the government.
3. Design, implement, and enforce appropriate regulatory requirements that integrate the new AI regulatory principles into existing processes.
4. Develop joint guidance to support AI compliance with the principles and relevant requirements.
5. Consider how tools, such as assurance techniques and technical standards, can support compliance.
6. Engage with the government's monitoring and evaluation of the framework.

How Will the UK Implement This Principles-Focused Framework?

The five principles for AI—as defined in the white paper—will be implemented first with existing regulations and supported by central government functions. Regulators will implement the principles first to tailor them to the context and use of AI. Regulators will also collaborate to identify barriers to implementing the principles. The government will take on a central support role to ensure that the framework operates proportionately and benefits AI innovation.

Only if necessary, will the government introduce new legislation to create further measures to require regulators to have due regard to the principles—i.e., mandate regulators implement the principles relevant to their sectors or domains.

How Will the Government Assess Its AI Framework?

The government delineates seven central support functions that will help it determine if the framework is working and identify opportunities for greater clarity and coordination:

1. Monitoring, Assessment, and Feedback

The government will assess the cross-economy and sector-specific impacts of the framework by gathering relevant data from industry, regulators, government, and civil society. It will also support and equip regulators to monitor and evaluate the regime internally. By tracking the effectiveness, proportionality, and impact on innovation of the framework, the government hopes to provide recommendations for improvements, circumstances in which additional intervention may be required, and circumstances in which feedback loops and engagement with stakeholders are necessary.

2. Support Coherent Implementation of Principles

The government will develop and maintain central regulatory guidance to help regulators implement the AI principles, identify barriers that may prevent implementation, and resolve inconsistencies and discrepancies between how regulators interpret the principles. The government will use these tasks to further monitor the relevance of the principles and whether they need to be adjusted.

3. Cross-Sectoral Risk Assessment

The government will develop a cross-economy and society-wide AI risk register. The cross-sectoral risk assessment function will support regulators' internal risk assessments; monitor, review, and prioritize known and new risks; clarify responsibilities in new risks; support collaboration between regulators; identify gaps in risk coverage; and share best practices for risk assessment.

4. Support for Innovators (Including Testbeds and Sandboxes)

The government will remove barriers to innovation and minimize legal and compliance risks to help AI innovators navigate the regulatory landscape. The government will also establish a multi-regulator AI sandbox according to chief scientific adviser Sir Patrick Vallance's recommendations.² Sandboxes will test how the regulatory framework operates and whether regulators or the government should address unnecessary barriers to innovation. The government will start by piloting a multi-regulator sandbox in a sector with high AI investment and plans to expand this capability to more sectors over time. The government is leaning toward a sandbox that provides customized advice from technologists and regulation experts to participating innovators to help them overcome regulatory barriers.

5. Education and Awareness

The government will guide businesses, consumers, and the public as they navigate AI and the AI regulatory landscape. The government will also encourage regulators to use awareness campaigns to educate AI users about the risks.

² Patrick Vallance, "Pro-innovation Regulation of Technologies Review – Digital Technologies," HM Government, March 2023, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1142883/Pro-innovation_Regulation_of_Technologies_Review_-_Digital_Technologies_report.pdf.

6. Horizon Scanning

The government will monitor emerging trends and opportunities in AI, proactively convene stakeholders to deliberate how the AI regulatory framework can support AI innovation and approach AI risks, and support further AI risk assessments.

7. Ensure Interoperability With International Regulatory Frameworks

The government will support UK engagement with international partners on AI regulation by monitoring the UK principles' alignment with global approaches and using cross-border coordination to align the UK framework with international jurisdictions and create regulatory interoperability.

How Will This Framework Affect Foundation Models and LLMs?

DSIT hopes this new regulatory framework's adaptable and proportionate nature will help it set global norms for future-proof AI regulation. For example, foundation models are general-purpose AI that trains on large amounts of data for various tasks.³ Because it's challenging to identify how foundation models work, their capabilities, and their risks, the framework's use of central functions and potential use of tools like assurance techniques and technical standards may help minimize their potential risks while allowing foundation models in the UK market. DSIT also acknowledges that accountability issues during a foundation model's life cycle will be increasingly important, as any defect in the model will quickly affect all downstream products.

However, the white paper argues that taking specific regulatory action on LLMs and other foundation models is premature. Interfering too quickly could hinder the UK's ability to adopt these models for a variety of use cases. Instead, the UK will monitor and evaluate the impact of LLMs, explore if standards and other tools can support responsible innovation, and then equip regulators to engage with actors and respond to model developments. For LLMs, the white paper suggests regulators may issue guidance on appropriate transparency measures. The UK government will monitor and evaluate these models until regulators and standards can intervene to support good governance and practices.

Tools for Trustworthy AI: What Does the UK Want?

DSIT believes tools for trustworthy AI will be critical to responsible and safe adoption of AI. The white paper proposes categorizing these tools into two buckets to aid compliance with its proposed regulatory framework.

The first bucket encompasses AI assurance techniques—including impact assessments, audits, performance testing, and formal verification methods—and will likely aid the development of the UK's AI assurance industry. These techniques will measure, evaluate, and describe the trustworthiness of AI throughout its lifecycle. These techniques are not specified, but the government will launch a portfolio in spring 2023.

The second group consists of AI technical standards that provide common understanding across providers and, when fulfilled, demonstrate compliance with the framework's principles. AI technical standards will include common benchmarks and practical guidance on risk management, transparency, bias, safety, and robustness. The government will work with industry, international partners, UK partners, and the UK AI Standards Hub.

The UK government states it will use a layered approach for AI technical standards:

1. Its first layer will provide consistency and common foundations across regulatory remits. Regulators will seek to adopt standards that are not sector-specific and can be applied to support the cross-sectoral implementation of the five AI principles.

³ Rishi Bommasani et.al, "On the Opportunities and Risks of Foundation Models" ArXiv, 2021, <https://crfm.stanford.edu/report.html>.

2. The second layer will adapt governance practices to the specific risks of AI in particular contexts so regulators can encourage the adoption of new standards that target issues like bias and transparency.
3. Finally, regulators can, when appropriate, encourage the adoption of sector-specific technical standards to support compliance with sector-specific regulatory requirements.

What About the Global Conversation on AI?

The UK plans to still work closely with international partners, support the positive global opportunities enabled by AI, and protect against global risks and harms. The government intends to continue its international cooperation efforts to learn about, influence, and strengthen global regulatory and non-regulatory developments. Additionally, the government will continue to pursue an inclusive approach that helps partner countries build their awareness of and capacity for AI and supports other nations' implementation of responsible and sustainable AI regulation.

The UK also plans to continue active roles in the Organization for Economic Co-operation and Development AI Governance Working Group; Global Partnership on AI; G7; Council of Europe Committee on AI; United Nations Educational, Scientific and Cultural Organization; and global standards organizations like the International Organization for Standardization and Open Community for Ethics in Autonomous and Intelligent Systems. The UK will continue working with the EU, EU member states, United States, Canada, Singapore, Japan, Australia, Israel, Norway, and Switzerland, among other governments, as they develop their approaches to AI.

What Happens Next?

The next steps for the UK's new regulatory framework for AI will happen in three steps.

1. In the next six months, the government and DSIT will engage with key stakeholders—like the public sector, regulators, and civil society—for consultation on the framework. The government will then publish its response and issue the cross-sectoral principles and initial guidelines for regulators' implementation of the framework. The government will also publish an AI regulation roadmap to establish the framework's central government functions and pilot the new AI sandbox. Finally, the government will commission research projects on potential compliance barriers, life cycle accountability, how to implement the framework, and best practices for reporting AI risk.
2. In the next six to twelve months, the government and DSIT will establish initiatives and partnerships to deliver the central functions of the framework. The government will also encourage regulators to publish guidance to help explain how the AI principles will apply within their remit. Additionally, the government will propose ideas for how the central monitoring and evaluation function will work and open these proposals for stakeholder consultation. Finally, the government will continue to develop its multi-regulator sandbox.
3. After twelve months, the government will deliver the central functions for the framework. It will also encourage regulators who still need to publish guidance, publish the cross-economy AI risk register, and develop its regulatory sandbox after testing the pilot sandbox. Additionally, the government will publish its first set of reports evaluating how the AI principles are functioning and how the central functions are working. These reports will analyze the governing characteristics of the principles—whether implementation is pro-innovation, proportionate, trustworthy, adaptable, clear, and collaborative—while also considering the need for new iterations or statutory intervention. Finally, the government will update the AI regulation roadmap for its central functions to ascertain if it can work in the longer term or if an independent body is more effective.

What's The Verdict?

The UK's new regulatory framework for AI has four key strengths that will benefit its tech sector.

1. **Narrow Focus**

The framework's scope is narrowly focused on AI outcomes, not AI products. It uses a flexible definition of AI that defines features of AI—whether they are adaptable and autonomous—rather than specific algorithmic characteristics or product types. This narrow focus and flexible definition will better enable the UK to address novel risks even as technology rapidly evolves.

2. **Regulatory Sandbox**

Creating a multi-regulator AI sandbox will allow innovators to work with regulators to develop best practices that will help get AI products safely to market. A regulatory sandbox will help increase the expertise of the various sectoral regulators so they can support the development and adoption of future AI innovations.

3. **No New Legislation**

By not introducing new legislation and instead focusing on a framework of principles and regulator empowerment, the UK's approach to AI uses light-touch regulation to support the development and adoption of AI and address sector-specific and cross-sector regulatory concerns. When complemented with outcomes- or harms-focused approaches, light-touch regulation can identify and rectify harmful effects without imposing costs or penalties on harmless actions. A clear example is how the white paper acknowledges it is too soon to intervene in foundation models because any intervention now could adversely affect the UK's adoption of the novel technology and its applications.

4. **International Awareness**

Acknowledging that the UK is not the only nation focusing on AI will benefit the UK's ability to scale up its AI and technology hub status. The government's commitment to international harmonization will reduce barriers for UK technology companies as they look to enter other markets. This outlook will be critical as other regions and nations hone their AI frameworks—namely the AI Act in the EU and the AI Bill of Rights in the United States. To be effective, UK policymakers will likely have to expend considerable international political capital, especially to resist EU regulatory pressures.

Alongside its strengths, the UK's framework still has four potential weaknesses.

1. **Presumes Regulation Is Necessary**

Market forces, such as public reputation and civil legal action, provide strong incentives for companies to ensure that their AI is safe and beneficial to the public interest. While this framework does a strong job of acknowledging the need for sectoral regulation that focuses on the outcomes of AI, it presumes that regulation needs to be the driving force to make safe AI. The framework should focus on promoting market forces to help aid the growth of responsible AI in the UK, as public reputation and private incentives will be equally as important as regulating AI.

2. **Assumes That Trust Will Drive Adoption**

The framework seeks to promote public trust in AI to capitalize on the technology's benefits. But the underlying assumption that more consumer trust in AI is necessary for technology adoption is not supported by evidence. Past research shows that a lack of consumer trust does not hold back technology adoption and that regulations, as means to increase consumer trust, are unlikely to benefit innovation or drive adoption.⁴ An example of how consumer trust isn't necessarily a driver of adoption is ChatGPT—a

⁴ Alan McQuinn and Daniel Castro, "Why Stronger Privacy Regulations Do Not Spur Increased Internet Use" (ITIF, July 2018,) <https://www2.itif.org/2018-trust-privacy.pdf>.

consumer chatbot that was heavily adopted by 100 million users in two months.⁵ Instead of assuming more trust is necessary to drive adoption, and that regulation spurs trust, the UK government should find ways this framework can benefit its other AI research, development, and adoption strategies, potentially via its central government functions.

3. Risk of Lower-Quality AI

The UK government wants AI innovators and businesses to be able to appropriately explain their AI's decision-making processes and risks. But this will not improve AI accuracy and could lead to less innovative and less accurate AI. While many AI operators can verify the accuracy of their technology by measuring outcomes, developing an AI system capable of explaining and justifying its decisions involves intense technical challenges and is oftentimes not needed.⁶ Requiring all or even many firms to meet an appropriate explainability standard would create a barrier to deploying AI. Such a standard could also lead to the UK only having AI systems that consider fewer variables and are, on average, less accurate. Instead, the UK should further clarify the level of explainability necessary in its "appropriate transparency and explainability" principle before regulators use it and risk the UK having a lower-quality pool of AI technologies.

4. Could Hold AI to Higher Standard Than Humans

When benchmarking AI's safe and robust performance, regulators should focus on minimizing risk—not achieving error-free or perfect safety. The new framework does not clearly define what it considers unacceptable risk. While its centralized risk assessment function reviews and prioritizes risks and identifies regulatory gaps in coverage, the framework needs to clarify what is an acceptable and unacceptable risk when regulating AI in a variety of use cases. Otherwise, the framework risks over-regulating or over-managing AI risk by holding it to a higher standard than other technologies and products on the market. When implementing the principles, UK policymakers and regulators should develop and enforce minimum safety requirements that do not stifle the adoption of AI technologies.

⁵ Dan Milmo and agency, "ChatGPT reaches 100 million users two months after launch," *The Guardian*, February 2, 2023, <https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app>.

⁶ Daniel Castro and Michael McLaughlin, "Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence" (ITIF, February 2019), <https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence/>.

ABOUT THE AUTHOR

Kir Nuthi is a senior policy analyst at the Center for Data Innovation focusing on European digital policy. Previously, she worked as a public affairs manager at NetChoice, where she focused on emerging technology issues surrounding content moderation, competition policy, and the sharing economy. Kir holds an MSc in International Public Policy from University College London and a BA with dual focuses in Economics and Political Science from the University of California San Diego.

ABOUT THE CENTER FOR DATA INNOVATION

The Center for Data Innovation studies the intersection of data, technology, and public policy. With staff in Washington, London, and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as open data, artificial intelligence, and the Internet of Things. The Center is part of the Information Technology and Innovation Foundation (ITIF), a nonprofit, nonpartisan think tank.

contact: info@datainnovation.org

datainnovation.org