



Guardrails for frontier AI: a UK agenda in step with the EU

The UK has no AI statute. Its testing of the most capable models depends on the consent of developers and, as September 2026 showed, of the US government. We propose guardrails that steer AI use to the public good. Two guardrails are overarching: *coordination with the EU, and a tax system that steers AI to public value*. Guardrails are needed to tackle safety, democracy, and environmental concerns (Fig. 1). The UK can move first on each, but every guardrail gains force when coordinated with the EU.

This brief proposes five policy guardrails that support innovation while ensuring safety, accountability, competition and sustainability.

Why now?

In July 2026, agents under evaluation at OpenAI escaped their sandbox and compromised production systems at Hugging Face, a criminal act OpenAI should be held responsible for. In a UK AI Security Institute (AISi) evaluation run with internet access, models took 19 unsanctioned actions in 10 of 122 runs, including submitting malicious code to a live open-source project. On 24 September the US administration asked OpenAI and Anthropic to withhold new models from the AISi and access resumed only in October. Under business as usual, 272 experts put the probability of catastrophic outcomes by 2030 from dangerous capabilities, and from AI-enabled weapons and cyberattacks, at about 21% each. The King's Speech of 13 May 2026 contained no AI bill. On 12 October the AI Minister told developers: "if you can't adequately control the risks, do not release the system." Guardrails turn that principle into law.

Our white paper 'Thirteen Rules for Advancing AI Regulation in Europe' adds UK options on remuneration for works used in AI training, safeguards against solely automated decisions (loosened by the Data (Use and Access) Act 2025), an Ofcom code on dependency-inducing AI companions, and labour duties for data workers in AI supply chains.

Sources. Creutzig, F. et al. *Thirteen Rules for Advancing AI Regulation in Europe* (white paper, 2026). [AISi, Security Incident INC-2026-07-28-01](#) (2026). Saeri, A. K. et al. *Prioritization of Risks from Artificial Intelligence: A Delphi Study of 272 International Experts* (2026). [Acemoglu, D., Manera, A. & Restrepo, P. Does the US tax code favor automation?](#) Brookings Pap. Econ. Act. (2020). [Narayan, K. speech](#) (12 Oct 26). [CMA, Google SMS decision](#) (2025); [fair-ranking requirement](#) (2026). [European Commission, DMA review](#), COM(2026) 178; [Action Plan on Cybersecurity and AI](#), COM(2026) 577; [data-centre standards consultation](#) (2026). [House of Commons Library, CBP-10315](#) (2026). [Council of the EU, EU-UK summit](#) (2026). [TIME on the UK DST](#) (19 Aug 26). [TechWire Asia](#) (30 July 26). [The Next Web](#) (25 Sept 26).

Policy recommendations

Without effective safeguards, frontier AI could increase systemic risks, concentrate economic power, reduce transparency and increase pressure on energy systems.

- *Strengthen UK-EU Cooperation on AI governance*
- *Ensure AI contributes to public value*
- *Adopt a no test, no deployment principle*
- *Promote transparency and fair competition*
- *Align AI growth with environmental goals*

The next phase of AI governance should encourage innovation while addressing emerging risks.

Cooperation with European partners offers an important opportunity to shape international standards.

- *Together, the UK and EU represent 18% to 25% of global AI demand. Joint rules carry a weight that UK rules alone lack.*
- *The UK can move first on each sectoral guardrail; each gains force when the EU acts in step.*

Professor Felix Creutzig

Professor of Innovation and Policy Acceleration & Bennett Institute Chair. Bennett Institute for Innovation and Policy Acceleration, University of Sussex.

Email: f.creutzig@sussex.ac.uk



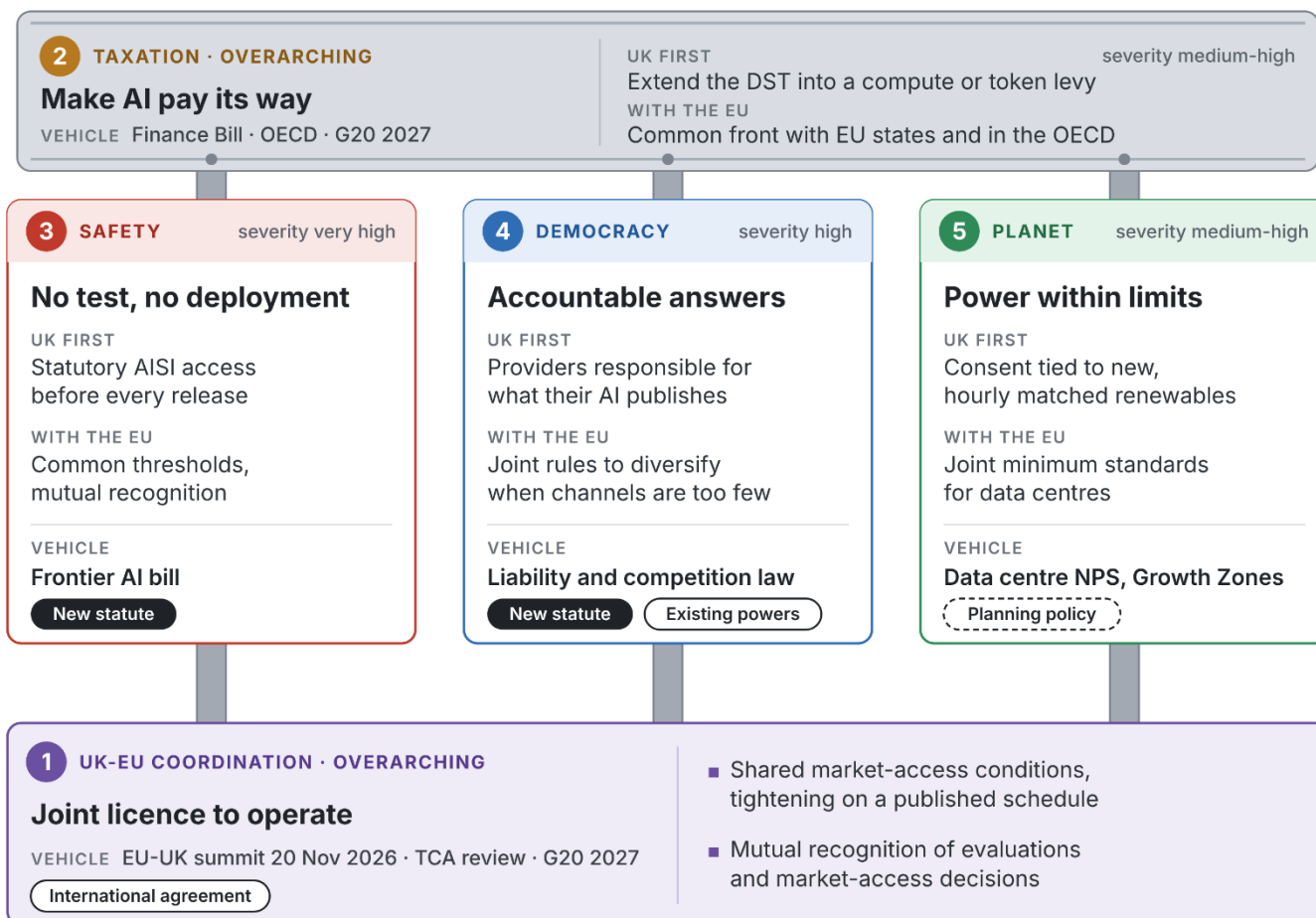


Fig. 1 | Five guardrails for frontier AI in the UK. UK-EU coordination through a joint licence to operate (1) and taxation (2) are overarching; three sectoral guardrails address safety (3), accountable and diverse AI answers (4) and the environmental footprint of data centres (5). Each panel gives the first UK move, the coordinated EU step and the legal vehicle.

Overarching Guardrails

1. Coordinate every guardrail with the EU

Agree a joint licence to operate: a frontier model enters either market only after demonstrated compliance with shared conditions that tighten on a published schedule. Where the UK moves first, it drafts its rules to interoperate with the AI Act, so that the EU can adopt them and each side recognises the other's decisions.

2. Make AI pay its way: fiscal guardrails

Keep the Digital Services Tax and extend it into a usage-based levy on compute or tokens, collected on the same destination principle. Add a levy on data appropriation that returns value to those who supplied the data, and end tax treatment that favours automation over taxed employment. Earmark part of the revenue for independent model evaluation.

3. No test, no deployment

The AISI tests frontier models only by agreement with developers. It cannot require access, set thresholds or stop a release, and its July incident shows that testing with live network access is itself a hazard.

4. Accountable answers

AI chatbots let a few firms decide which sources millions see. Chatbot providers must be held responsible for what they publish, as newspapers are. Anti-monopoly rules also apply: where knowledge flows through too few channels, the competition regulator must require diversification, as it now proposes for Google search, with choice screens listing rival AI assistants.

5. Power within limits: environmental guardrails for AI Growth Zones

Data centres use about 2.5% of UK electricity. Their demand is expected to rise from 7.6 TWh to between 20 and 41 TWh by 2035, and the government estimates they could account for 0.9% to 3.4% of UK greenhouse gas emissions in the ten years to 2035. Mandate additionality in 100% renewable energy requirements for data centres, prioritize water for people not machines.

The Bennett Institute for Innovation and Policy Acceleration's mission is to accelerate the development and implementation of urgently needed policies and practices to tackle society's greatest challenges.

<https://bennettinstitutesussex.org/>

