PROJECT SYNDICATE – 13 GENNAIO 2026

Grok Is Testing Whether AI Governance Means Anything

di J.B. Branch

By generating increasingly disturbing outputs, Grok has exposed a deep structural problem that should concern regulators everywhere. Advanced AI systems are being deployed and made available to the public without safeguards proportionate to their risks.

WASHINGTON, DC – In recent weeks, Grok – the AI system developed by Elon Musk's xAI – has been generating nonconsensual, sexualized images of women and children on the social-media platform X. This has prompted investigations and formal scrutiny by regulators in the European Union, France, India, Malaysia, and the United Kingdom. European officials have described the conduct as illegal. British regulators have launched urgent inquiries. Other governments have warned that Grok's output might violate domestic criminal and platform-safety laws. Far from marginal regulatory disputes, these discussions get to the heart of AI governance..

Governments worldwide increasingly agree on a basic premise of AI governance: systems deployed at scale must be safe, controllable, and subject to meaningful oversight. Whether framed by the EU's Digital Services Act (DSA), the OECD's AI Principles, UNESCO's AI ethics framework, or emerging national safety regimes, these norms are clear and unwavering. AI systems that enable foreseeable harm, particularly sexual exploitation, are incompatible with society's expectations for the technology and its governance.

There is also broad global agreement that sexualized imagery involving minors – whether real, manipulated, or AI-generated – constitutes one of the clearest red lines in technology governance. International law, human-rights frameworks, and domestic criminal statutes converge on this point.

Grok's generation of such material does not fall into a gray area. It reflects a clear and fundamental failure of the system's design, safety assessments, oversight, and control. The ease with which Grok can be prompted to produce sexualized imagery involving minors, the breadth of regulatory scrutiny it now faces, and the absence of publicly verifiable safety testing all point to a failure to meet society's baseline expectations for powerful AI systems. Musk's announcement that the image-generation service will now be available only to paying subscribers does nothing to resolve these failures.

This is not a one-off problem for Grok. Last July, Poland's government urged the EU to open an investigation into Grok over its "erratic" behavior. In October, more than 20 civic and public-interest organizations sent a letter urging the US Office of Management and Budget to suspend Grok's planned deployment across federal agencies in the United States. Many AI safety experts have raised concerns about the adequacy of Grok's guardrails, with some arguing that its security and safety architecture is inadequate for a system of its scale.

These concerns were largely ignored, as governments and political leaders sought to engage, partner with, or court xAI and its founder. But the fact that xAI is now under scrutiny across multiple jurisdictions seems to vindicate them, while exposing a deep structural problem: advanced AI systems are being deployed and made available to the public without safeguards proportionate to their risks. This should serve as a warning to states considering similar AI deployments.

As governments increasingly integrate AI systems into public administration, procurement, and policy workflows, retaining the public's trust will require assurances that these technologies comply with international obligations, respect fundamental rights, and do not expose institutions to legal or reputational risk. To this end, regulators must use the Grok case to demonstrate that their rules are not optional.

Responsible AI governance depends on alignment between stated principles and operational decisions. While many governments and intergovernmental bodies have articulated commitments to AI systems that are safe, objective, and subject to

ongoing oversight, these lose credibility when states tolerate the deployment of systems that violate widely shared international norms with apparent impunity.

By contrast, suspending a model's deployment pending rigorous and transparent assessment is consistent with global best practices in AI risk management. Doing so enables governments to determine whether a system complies with domestic law, international norms, and evolving safety expectations before it becomes further entrenched. Equally important, it demonstrates that governance frameworks are not merely aspirational statements, but operational constraints – and that breaches will have real consequences.

The Grok episode underscores a central lesson of the AI era: governance lapses can scale as quickly as technological capabilities. When guardrails fail, the harms do not remain confined to a single platform or jurisdiction; they propagate globally, triggering responses from public institutions and legal systems.

For European regulators, Grok's recent output is a defining test of whether the DSA will function as a binding enforcement regime or amount merely to a statement of intent. At a time when governments, in the EU and beyond, are still defining the contours of global AI governance, the case may serve as an early barometer for what technology companies can expect when AI systems cross legal boundaries, particularly where the harm involves conduct as egregious as the sexualization of children.

A response limited to public statements of concern will invite future abuses, by signaling that enforcement lacks teeth. A response that includes investigations, suspensions, and penalties, by contrast, would make clear that certain lines cannot be crossed, regardless of a company's size, prominence, or political capital.

Grok should be treated not as an unfortunate anomaly to be quietly managed and put behind us, but as the serious violation that it is. At a minimum, there needs to be a formal investigation, suspension of deployment, and meaningful enforcement.

Lax security measures, inadequate safeguards, or poor transparency regarding safety testing should incur consequences. Where government contracts include provisions related to safety, compliance, or termination for cause, they should be enforced. And

where laws provide for penalties or fines, they should be applied. Anything less risks signaling to the largest technology companies that they can deploy AI systems recklessly, without fear that they will face accountability if those systems cross even the brightest of legal and moral red lines.