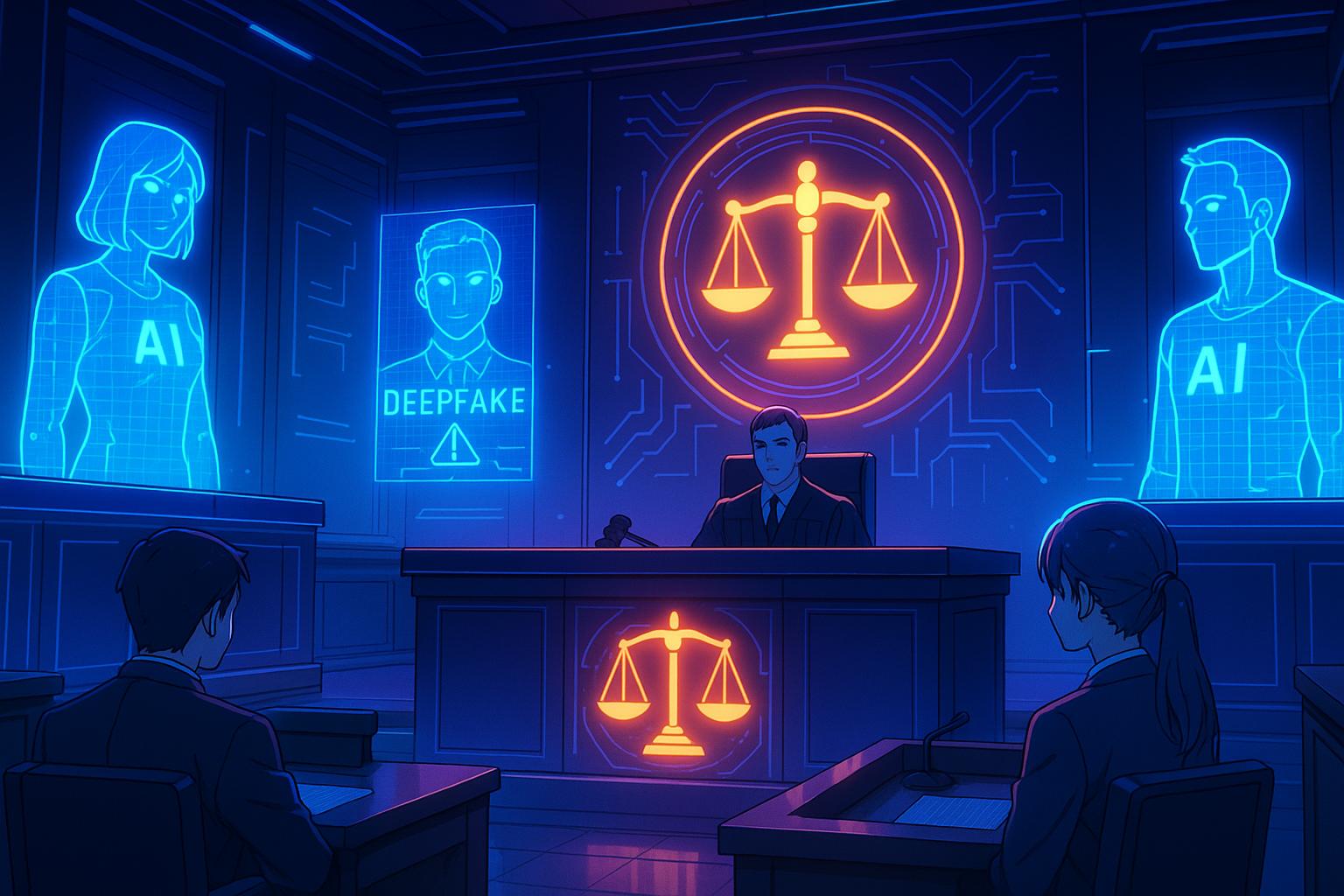
# Regulating AI’s dark side: lessons from climate policy to tackle deepfake abuse



The swift evolution of artificial intelligence (AI) has brought about numerous unintended consequences, one of the most alarming being the rise of technologies that facilitate sexual violence, such as deepfake pornography. While AI was not conceived for such malicious purposes, its rapid advancement has made it a tool that can be, and often is, misused in distressing ways.

The implications of AI extend far beyond the realm of explicit content. Current regulatory frameworks grappling with these issues tend to revolve around intended uses, largely ignoring the chaotic creativity that characterizes many AI environments. The emergence of what have been termed “underspheres” — loosely connected online communities where users freely experiment with AI technologies — illustrates this point clearly. In places like GitHub and Hugging Face, innovation thrives but often strays into dangerous territory. Users remix and repurpose AI models, creating applications that can serve harmful ends, a phenomenon urgently needing regulatory attention.

Regulatory efforts, such as the European Union's AI Act, established a framework to classify AI systems based on risk. This model has gained traction globally, with similar approaches emerging from the United Kingdom, the United States, and China. However, a common flaw in these strategies is their focus on intended use cases rather than the unintended, creative applications that often arise in these underspheres. For instance, deepfake technology, initially created for benign applications, has been weaponised in the form of non-consensual pornography. An analysis of deepfake videos found that a staggering 98% are pornographic, underscoring the need for more vigilant regulatory approaches.

Recent legislative initiatives in the United States highlight the growing urgency to combat AI misuse. The introduction of state laws, such as those proposed in Minnesota, aims to impose civil penalties on companies generating explicit imagery without consent. Supporters argue that such legislation is critical to preventing harm, although legal experts warn of potential free speech implications. Meanwhile, the newly passed Take It Down Act seeks to establish a framework for removing non-consensual deepfakes from social media platforms, again illustrating the difficulty of keeping pace with rapidly evolving technology.

As policymakers attempt to strike a balance between regulation and creative freedom, the path forward is fraught with complexity. One promising new lens through which to approach AI governance is climate policy. Both fields face similar challenges due to their inherently interconnected and unpredictable nature. Climate governance has matured over decades, developing frameworks that acknowledge uncertainty while maintaining public discourse. Adapting these principles to AI could facilitate proactive regulation, quickly addressing emerging threats rather than reacting after incidents arise.

However, caution is warranted regarding the pitfalls experienced in climate policy. Loopholes and competing interests have frequently stalled genuine progress, creating a sense of inertia that could easily translate into AI governance. To avoid these missteps, there must be concerted efforts to align public oversight with the self-regulatory behaviour of tech developers, fostering transparency and accountability.

Furthermore, the global dimension of AI misuse cannot be ignored. Effective regulation will require international cooperation, especially as AI technologies continue to transcend borders. Ultimately, the task is to adapt to ongoing technological developments, ensuring that societies can navigate the complexities of both AI and its consequences safely.

As the landscape of AI continues to evolve, the imperative for comprehensive and flexible regulatory frameworks has never been clearer. By learning from climate policy and addressing the unique challenges posed by generative AI, there lies the potential not just to manage risks but to harness AI’s capabilities responsibly for the broader good.

### Reference Map

1. Paragraphs 1, 2, 3, 4 — [[1]](https://theconversation.com/ai-is-moving-fast-climate-policy-provides-valuable-lessons-for-how-to-keep-it-in-check-255624)
2. Paragraphs 5, 6 — [[2]](https://apnews.com/article/3147de9d15d943421be32d4497d6f79a), [[3]](https://www.ft.com/content/e2fa34b2-6987-494d-a81a-1bdb6693671f), [[4]](https://time.com/7277746/ai-deepfakes-take-it-down-act-2025/)
3. Paragraphs 7, 8 — [[5]](https://www.reuters.com/technology/us-prosecutors-see-rising-threat-ai-generated-child-sex-abuse-imagery-2024-10-17/)
4. Paragraph 9 — [[6]](https://apnews.com/article/741a6e525e81e5e3d8843aac20de8615)
5. Paragraph 10 — [[7]](https://www.reuters.com/legal/legalindustry/report-deepfakes-what-copyright-office-found-and-what-comes-next-ai-regulation-2024-12-18/)

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://theconversation.com/ai-is-moving-fast-climate-policy-provides-valuable-lessons-for-how-to-keep-it-in-check-255624> - Please view link - unable to able to access data
2. <https://apnews.com/article/3147de9d15d943421be32d4497d6f79a> - Minnesota is considering legislation to combat the creation of explicit images using "nudification" AI technology without consent. This follows a case where Molly Kelley discovered that a known individual used AI to create explicit visuals from family photos she posted on social media. The proposed bill, supported by Democratic Sen. Erin Maye Quade, seeks to impose civil penalties on companies that run these technologies, with fines up to $500,000 per violation. The aim is to prevent the creation and spread of such material. However, legal experts caution that the bill may face constitutional challenges related to free speech. Other states and Congress are also working on legislations targeting the misuse of AI in creating non-consensual explicit content. If Congress fails to act, Maye Quade hopes state initiatives will provide protection. Victims like Megan Hurley have shared their distress, emphasizing the ease and devastation of such AI misuses, while experts urge precise legal language to balance protection and constitutional rights.
3. <https://www.ft.com/content/e2fa34b2-6987-494d-a81a-1bdb6693671f> - The article explores the growing issue of explicit AI-generated deepfakes and the challenging legal landscape surrounding their regulation. Omny Miranda Martone, founder of the Sexual Violence Prevention Association, becomes a victim of such deepfakes despite their expertise in the field. The article underscores the absence of federal mandates in the US requiring the removal of non-consensual explicit deepfakes and absent federal criminal legislation addressing their creation and dissemination. Different countries have adopted varied approaches, with some focusing on individual penalties (e.g., the UK and Australia criminalizing dissemination) and others on broader measures (e.g., South Korea criminalizing creation, sharing, and viewing). The US has introduced bills like the Defiance Act and the Shield Act, aiming to empower victims while focusing on individual perpetrators. There is bipartisan support but ineffective regulation currently. The rapid advancements in AI technology have made deepfake creation easier and more realistic, complicating prosecution and raising concerns about digital evidence. Activists and lawmakers advocate for robust legislation to mitigate this issue, although achieving consensus and technology catch-up remains challenging.
4. <https://time.com/7277746/ai-deepfakes-take-it-down-act-2025/> - On April 28, 2025, the U.S. Congress passed the Take It Down Act, a landmark bipartisan bill targeting the harms of AI-generated deepfake pornography. Spearheaded by Senator Ted Cruz following advocacy from teen victims Elliston Berry and Francesca Mani, the law criminalizes non-consensual deepfake porn and requires social media platforms to remove such content within 48 hours of notification. The bill garnered sweeping bipartisan support, passing the House 409-2, with endorsements from both progressive and conservative groups. It bypasses controversy over Section 230 by using FTC enforcement authority under deceptive trade practices. Tech companies like Meta and Snapchat supported the bill due to its balanced approach. Despite strong backing, civil liberties groups expressed concerns over potential misuse and free speech suppression, especially under the Trump administration’s weakened FTC. Notably, First Lady Melania Trump’s support helped galvanize House passage. Advocates hope this legislation will set the stage for broader child online safety reforms. For survivors like Berry and Mani, the bill represents a crucial step in reclaiming dignity and preventing future victimization.
5. <https://www.reuters.com/technology/us-prosecutors-see-rising-threat-ai-generated-child-sex-abuse-imagery-2024-10-17/> - U.S. federal prosecutors are intensifying efforts to address the rising threat of AI-generated child sex abuse imagery. The Justice Department has initiated two criminal cases this year, targeting suspects who used AI tools to create or manipulate explicit images of minors. James Silver, the deputy chief of the Justice Department's Computer Crime and Intellectual Property Section, predicts more cases to follow, emphasizing the need to prevent the normalization of such material. The surge in generative AI technology, including Stable Diffusion, has raised concerns about its potential misuse for creating illicit content and complicating law enforcement efforts to identify real victims. Despite challenges in applying existing laws to AI-generated imagery, cases like those against Steven Anderegg and Seth Herrera highlight the ongoing legal battles. Advocacy groups and major tech companies like Google, Meta, and OpenAI are also working to prevent AI systems from producing abusive material, stressing the urgency of addressing this issue proactively.
6. <https://apnews.com/article/741a6e525e81e5e3d8843aac20de8615> - The U.S. Congress has passed the bipartisan "Take It Down Act," aiming to combat the spread of non-consensual intimate images, including AI-generated deepfakes and revenge porn. Sponsored by Senators Ted Cruz and Amy Klobuchar and supported by First Lady Melania Trump, the legislation stipulates it is illegal to knowingly publish or threaten to publish such content without consent. It mandates websites and social media platforms to remove flagged material within 48 hours and eliminate duplicates. The act, which now awaits President Donald Trump's signature, has received backing from tech companies like Meta and advocacy organizations for its potential to protect victims, particularly minors. The initiative was partially inspired by the case of Elliston Berry, a teen targeted with a deepfake image. However, critics, including the Electronic Frontier Foundation and the Cyber Civil Rights Initiative, argue the bill’s broad language risks over-censorship, affecting legal and journalistic content, and lacks safeguards against misuse. Concerns also include potential pressure on platforms to over-monitor and rely on imprecise automated filters. Despite these objections, proponents hail the bill as a significant step to ensuring accountability and legal recourse for victims of online abuse.
7. <https://www.reuters.com/legal/legalindustry/report-deepfakes-what-copyright-office-found-and-what-comes-next-ai-regulation-2024-12-18/> - The Copyright Office released a report on July 31, 2024, addressing the issue of deepfakes, noting that current copyright and intellectual property laws are insufficient to tackle the harm posed by these AI-generated digital replicas. The technology, which creates realistic but false images, audio, or videos, can be used positively but also poses risks such as fraud, misinformation, and privacy violations. The report urges new federal legislation to protect individuals' likenesses and ensure accountability. The NO FAKES Act, introduced by a bipartisan group of Senators, seeks to create a national standard by granting individuals the right to control the use of their voice and likeness and holding platforms accountable for unauthorized replicas. While the passage of federal legislation is pending, states and local authorities are already taking action with laws and lawsuits against misuse.