



INSIGHT 18

DETERMINING ACCOUNTABILITY FOR **AI HARMS** COULD BE COMPLEX



Some AI harms may be difficult to eliminate. It is both technically challenging to prevent misuse of AI systems, and complicated to hold humans accountable when they are misused. AI is challenging legal frameworks, in areas such as copyright and consumer protection, in novel ways. Jurisdictions are taking diverging approaches to AI regulation.

TODAY

Preventing misuse of AI systems is technically challenging. AI companies have made substantial investments in preventing misuse of their systems.¹ However, whenever new guardrails are introduced, users find new ways to circumvent them.² AI systems can be difficult to control and predict. They can generate incorrect information. AI companies are struggling to find the right balance between usefulness and harm mitigation. For example, Google attempted to make their image generator less biased, so prompts like “show me a CEO” would show people of various appearances and backgrounds. However, the system also created historically inaccurate images, such as portraying the American Founding Fathers or WWII Nazi soldiers as racially and gender diverse.³



AI systems may provide dangerous information because they lack contextual awareness and struggle to evaluate user intent. Chatbots cannot reliably tell if a user has the means and intention to do harm. They may refuse to answer a prompt where the intent seems obvious, like “how do I waterboard someone?”, but provide an answer if the phrasing is more plausibly educational, like “what is waterboarding and how is it done?” As a result, researchers have argued that the focus should not be on trying to prevent people from finding out information via AI – for example, how to make a novel pathogen – but instead on preventing them from being able to act on that information, by controlling access to the necessary raw materials and equipment.⁴

Some jurisdictions are regulating to establish liability for AI misuse, while others are regulating to increase transparency. The European Union’s *EU Artificial Intelligence Act* attempts to define unacceptable uses of AI – such as social credit systems – and who is responsible for harms: usually the developers and deployers, not end-users.⁵ In contrast, California’s regulatory approach aims to improve transparency. Its Assembly Bill 2023 requires AI providers to give information on the data used for training, including whether it contained personal information or copyrighted material.⁶ While *AI Transparency Act* requires some generative AI providers to watermark outputs and provide free tools to allow the public to check if a file was generated or altered using their offerings.⁷

AI-related cases are testing copyright and consumer protection laws in novel ways. News outlets are suing AI companies for infringing their copyright by using their content to train AI systems.^{8 9} The CEO of Microsoft AI, however, claims that as this content is on the open web, it is “freeware” for AI training.¹⁰ It is often unclear whether AI offerings count as a “product” or a “service” for consumer protection and liability laws, or whether “terms of service” are sufficient to insulate AI companies from liability.^{11 12 13} For example, the mother of a Florida teenager who committed suicide after using a chatbot, Character.AI, is suing the chatbot provider for emotionally manipulating her son. Character.AI argues that the teenager violated its terms of service,¹⁴ while the mother claims that Character.AI is a product that should have included more warnings.¹⁵

Open-source AI models are challenging to regulate or control. Open-source is generally a beneficial model for software, as anyone – not just the original developers – can use and build upon it.¹⁶ However, problematic open-source software can be nearly impossible to contain. For instance, Nintendo successfully sued the developers of Yuzu, open-source software that can play Nintendo Switch games on a PC, – but has struggled to remove Yuzu from the Internet as it had already been “forked”, or copied, so many times.^{17 18} Likewise, open-source AI has many benefits (see insight X on lighter AI), but it is hard to prevent its use by bad actors (see insight X on AI and security) to create misinformation or malicious code. Some providers of access to open-source AI models, such as FreedomGPT, argue that imposing safety guardrails on AI constitutes a form of censorship.¹⁹

Open-source software

Software whose source code is public. Anyone can edit, copy, or share it. Development is often decentralized and collaborative.

Fork

A copy of or variation on an open-source project.



FUTURES

The global legal and regulatory landscape for AI may become highly complex. Countries and jurisdictions may diverge not only in their laws and regulations on AI, but in how they interpret and apply them. Responsibility for harm caused by an AI system may rest with its developer in one jurisdiction, and its deployer in another. AI companies may increasingly have to release region-specific variations of their products. Some may choose to exit the markets with the strictest regulations, while others may find it easier to comply, similar to how many US-based websites voluntarily comply with the EU's General Data Protection Regulation.

Firms and individuals may be held responsible for AI systems they cannot fully control. Just as a dog's owner can be held legally responsible if their dog bites someone, the users, deployers, or developers of an AI system could be held responsible for what it does even if they cannot fully control its behaviour.²⁰ The only real agency a person or institution may have regarding an AI system is the decision about whether to use it. The requirement for a "human in the loop" in AI systems may become primarily about having someone to take the fall should something go wrong.²¹

New laws or court rulings may significantly disrupt the AI industry. Some AI products, services, or industry practices that exist today may not be legal in the future. Court rulings could force firms to change their product offerings or businesses practices overnight, disrupting companies and individuals who rely on those offerings. Newly established liabilities may open a floodgate of lawsuits that bankrupt companies, or dissuade them from continuing to offer a service because they consider the legal risk too great.

Alternatively, courts could prove effectively powerless to constrain AI. Courts can only use existing laws to deal with AI harms, which may limit their ability to punish or deter undesirable uses of AI, even if popular sentiment demands it. For example, a court might issue maximum fines which AI companies simply account for among the costs of doing business. As legal proceedings often take years, by the time courts pass judgement on a harmful AI practice it could have become entrenched and normalised. Even if courts prohibit some kinds of AI systems, determined individuals may still find ways to access them, whether from foreign providers in more lax jurisdictions or communities of open-source developers and hobbyists. As AI and consumer hardware improve (see insight X on lighter AI), it will be increasingly possible for individuals to run powerful AI systems on their own machines.



IMPLICATIONS

- ▶ Diverging regulation of AI across the world may exacerbate **technological decoupling** between nations.
 - › Travelers may increasingly find that their devices gain or lose significant functionality as they travel between jurisdictions.
- ▶ **Companies may struggle to navigate the complex legal and regulatory landscape surrounding AI.**
 - › More risk-averse companies may be willing to adopt AI only once legal questions are settled and their legal obligations are clearly established.
- ▶ Some **jurisdictions** may purposefully have little regulation or oversight on AI to attract investment, **functioning in a similar way to tax havens.**
- ▶ The quality of some AI products may change in the future as **regulations on datasets emerge**, or courts rule that certain kinds of data collection are prohibited.
- ▶ **Anonymous communities of open-source developers may emerge** to share regulated or restricted AI models and tools.
 - › Bad actors and adversaries may intentionally develop and disseminate harmful AI models.
- ▶ Jurisdictions may experiment with **novel approaches to regulating access to restricted AI systems**, such as criminalising those that develop, share or possess forbidden AI models, or blocking access to foreign websites and companies that provide controversial AI models.
- ▶ Who is **legally responsible** for AI-generated harms may not be who the public thinks is **morally responsible.**

Endnotes

- 1 Browne, Ryan. “Tech Giants Pledge AI Safety Commitments — Including a ‘Kill Switch’ If They Can’t Mitigate Risks.” CNBC, May 21, 2024. <https://www.cnbc.com/2024/05/21/tech-giants-pledge-ai-safety-commitments-including-a-kill-switch.html>.
- 2 Tangermann, Victor. “It’s Shockingly Easy to Get Around AI Chatbot Guardrails, Researchers Find.” Futurism, July 28, 2023. <https://futurism.com/researchers-discover-chatgpt-jailbreak>.
- 3 Robertson, Adi. “Google Apologizes for ‘Missing the Mark’ after Gemini Generated Racially Diverse Nazis.” The Verge, February 21, 2024. <https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical>.
- 4 Narayanan, Arvind, and Sayash Kapoor. “AI Safety Is Not a Model Property.” AI Snake Oil, March 12, 2024. <https://www.aisnakeoil.com/p/ai-safety-is-not-a-model-property>.
- 5 EU Artificial Intelligence Act. “High-Level Summary of the AI Act,” February 27, 2024. <https://artificialintelligenceact.eu/high-level-summary/>.
- 6 California Legislative Information. “AB-2013 Generative Artificial Intelligence: Training Data Transparency,” September 30, 2024. https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013.
- 7 California Legislative Information. “SB-942 California AI Transparency Act,” September 20, 2024. https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.
- 8 Heydari, Anis. “Media Outlets, Including CBC, Sue ChatGPT Creator.” CBC News, November 29, 2024. <https://www.cbc.ca/news/business/openai-canadian-lawsuit-1.7396940>.
- 9 Roth, Emma. “The New York Times Is Suing OpenAI and Microsoft for Copyright Infringement.” The Verge, December 27, 2023. <https://www.theverge.com/2023/12/27/24016212/new-york-times-openai-microsoft-lawsuit-copyright-infringement>.
- 10 Hollister, Sean. “Microsoft’s AI Boss Thinks It’s Perfectly Okay to Steal Content If It’s on the Open Web.” The Verge, June 28, 2024. <https://www.theverge.com/2024/6/28/24188391/microsoft-ai-suleyman-social-contract-freeware>.
- 11 “Report on Artificial Intelligence and Civil Liability.” British Columbia Law Institute, April 2024. <https://www.bcli.org/project/artificial-intelligence-and-civil-liability-project/>, 50-51.
- 12 Spicer, Katy, Julia Jacobson, Naija Perry, and Aden Hochrun. “Artificial Intelligence and the Rise of Product Liability Tort Litigation: Novel Action Alleges AI Chatbot Caused Minor’s Suicide.” Privacy World, November 11, 2024. <https://www.privacyworld.blog/2024/11/artificial-intelligence-and-the-rise-of-product-liability-tort-litigation-novel-action-alleges-ai-chatbot-caused-minors-suicide/>.
- 13 Kumarasamy, Jey, and Brenda Leong. “Third-Party Liability and Product Liability for AI Systems.” IAPP, July 26, 2024. <https://iapp.org/news/a/third-party-liability-and-product-liability-for-ai-systems>.
- 14 Character.AI. “Character.AI Terms of Service,” October 25, 2023. <https://character.ai/tos>.
- 15 Spicer, Katy, Julia Jacobson, Naija Perry, and Aden Hochrun. “Artificial Intelligence and the Rise of Product Liability Tort Litigation: Novel Action Alleges AI Chatbot Caused Minor’s Suicide.” Privacy World, November 11, 2024. <https://www.privacyworld.blog/2024/11/artificial-intelligence-and-the-rise-of-product-liability-tort-litigation-novel-action-alleges-ai-chatbot-caused-minors-suicide/>.
- 16 Red Hat. “What Is Open Source?,” October 24, 2019. <https://www.redhat.com/en/topics/open-source/what-is-open-source>.

- 17 Hollister, Sean. “Nintendo Switch Emulator Yuzu Will Utterly Fold and Pay \$2.4M to Settle Its Lawsuit.” The Verge, March 4, 2024. <https://www.theverge.com/2024/3/4/24090357/nintendo-yuzu-emulator-lawsuit-settlement>.
- 18 Roth, Emma. “Nintendo DMCA Takedown Wipes over 8,500 Yuzu Emulator Copies.” The Verge, May 3, 024. <https://www.theverge.com/2024/5/3/24147936/nintendo-dmca-takedown-yuzu-emulator-copies>.
- 19 FreedomGPT. “About Us.” Accessed December 5, 2024. <https://www.freedomgpt.com/about-us>.
- 20 “Report on Artificial Intelligence and Civil Liability.” British Columbia Law Institute, April 2024. <https://www.bcli.org/project/artificial-intelligence-and-civil-liability-project/>, 41.
- 21 Elish, Madeleine Clare. “Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction.” *Engaging Science, Technology, and Society* 5 (March 23, 2019): 40–60. <https://doi.org/10.17351/ests2019.260>.

© His Majesty the King in Right of Canada, 2025

For information regarding reproduction rights: <https://horizons.gc.ca/en/contact-us/>

PDF: PH4-228/2025E-PDF

ISBN: 978-0-660-78230-0

Aussi disponible en français sous le titre : La détermination de la responsabilité pour les préjudices causés par l'IA pourrait être complexe.

DISCLAIMER

Policy Horizons Canada (Policy Horizons) is the Government of Canada’s centre of excellence in foresight. Our mandate is to empower the Government of Canada with a future-oriented mindset and outlook to strengthen decision making. The content of this document does not necessarily represent the views of the Government of Canada, or participating departments and agencies.