

Policy Approaches for Building a Responsible Ecosystem

*Contextualising AI Governance Challenges Within Other
Regulatory/Governance Sectors and Histories*

Bhargavi Ganesh



This fellowship ran from January - May 2023 as part of BRAID. It is based on the state of affairs in March 2023.

BRAID is a UK-wide programme dedicated to integrating Arts and Humanities research more fully into the Responsible AI ecosystem, as well as bridging the divides between academic, industry, policy and regulatory work on responsible AI. Funded by the Arts and Humanities Research Council (AHRC), BRAID represents AHRC's major investment in enabling responsible AI in the UK. The Programme runs from 2022 to 2028. Working in partnership with the Ada Lovelace Institute and BBC, BRAID supports a network of interdisciplinary researchers and partnering organisations through the delivery of funding calls, community building events, and a series of programmed activities. Funding reference: Arts and Humanities Research Council grant number AH/X007146/1.

Learn more at www.braiduk.org

This research was supported via UK Research and Innovation by the R&D Science and Analysis Programme at the Department for Culture, Media & Sport. Any primary research, subsequent findings or recommendations do not represent Government views or policy and are produced according to research ethics, quality assurance, and academic independence.

To request an alternative format of this report please email braid@ed.ac.uk

Table of Contents

2	Abstract	3
3	Key Outcomes	4
4	Executive summary	6
5	Report	9
6	Conclusion	39
7	References	40



● ● ● Abstract

AI has quickly come to underpin a wide range of systems in both the public and private sector. Initially flourishing in less regulated areas, such as social media and gaming, AI has the potential to further transform domains, including finance, health, and criminal justice. While the expansion of AI has enabled exciting new advances, with improved understanding of protein structures and autonomous driving, for example, there have also been well-publicised harms brought about through the use of AI-based systems, such as wrongful arrests, amplification of misinformation, and discrimination in hiring. As a result of these harms, a growing number of scholars, practitioners, and members of the public have called for the design and implementation of safeguards to protect those affected by algorithmic decision-making.

Although many Western countries have recognised the need for regulation, there are a number of challenges for effective governance, including, but not limited to: the difficulty of understanding the full scope of AI uses in such a rapidly developing field; the complex interorganisational nature of AI design and deployment, which often crosses sectoral and traditional boundaries between products and services; the proprietary nature of data and other information used by companies for the purposes of building algorithmic systems; and the difficulty of negotiating between individual-level harms and population-level risks in an increasingly algorithmic society. The unregulated nature of these technologies, despite increasing vulnerability to their impacts, has led to mistrust in AI systems and the institutions designing and deploying them. One way such mistrust has manifested is through activist efforts around the world, which have focused on issues such as banning facial recognition software and protecting children and other vulnerable populations from being exposed to harmful and manipulative content on social media.

Developing a regulatory framework for AI is imperative for ensuring that everyone can benefit from the economic development and social value brought about by these new advances, and that the benefits do not only accrue to a small group of individuals or

corporations at the expense of those with less power. Policymaking in this area thus has the opportunity to stimulate innovation and trust in technology, such that designers, deployers, and users of AI are incentivised to prioritise the safety and fundamental rights of those impacted by these systems. This report therefore addresses the following research questions: **In what ways can policymakers enable effective AI governance, accountability, and compliance?** and **What useful lessons can AI policymakers draw from other regulatory/governance sectors and histories?** Addressing these questions provides a starting point for policymakers to understand the relevant operational challenges of AI governance, and the landscape of potential interventions to consider in meeting these challenges.

● ● ● Key Outcomes

- “Responsible AI” has come to be defined largely by the design-level changes some companies have made in response to the algorithmic harms highlighted by a handful of successful activist efforts. However, the status quo does not enable the consistent relational practice of responsibility, in which the concerns of affected parties can effectively feed into the process of improving systems at scale, or issuing appropriate remedies.
- Historically, governance and regulation have played a key role in filling the responsibility gaps created by new innovations. Regulatory efforts in response to steamboat accidents, air pollution, airline safety concerns, the sale of harmful/misleading drugs, and financial crises all demonstrate the important role that governance and regulation play in establishing a responsible ecosystem. At the moment, many companies are externalising the risks associated with algorithmic systems, making governance intervention of paramount importance.
- Although governing bodies are typically reluctant to regulate technologies, the challenges posed by AI are not so unprecedented as to require delaying regulation. In addition, the process of governance is necessarily iterative, and

requires initial steps to be taken in order to work towards making innovations progressively safer.

- The United Kingdom of Great Britain and Northern Ireland (UK), the European Union (EU), and the United States of America (US) have built up a number of regulatory capacities in their history of technology governance. In particular, they have managed distorted information ecosystems caused by hype, false claims, and exaggerated/ill-defined risks emerging in the presence of new technologies. In addition, regulatory governance has traditionally been utilised to address the issue of assigning responsibility/liability in the presence of many different contributing actors (the issue of “many hands”).
- Many AI governance challenges recreate past challenges, and existing regulatory tools, such as registration/adverse event recording, licensing, inspection, risk monitoring, and prohibition of false advertising, remain robust mechanisms for meeting these challenges. In adapting these approaches to AI, a number of considerations emerge: (1) In addition to devolving some of these functions to appropriate sectoral bodies, governments should also consider creating one unified body to record AI-related adverse events, so as to understand the systemic risks posed by AI; (2) While efforts to prohibit illegitimate uses of AI remain important in the face of significant risks to citizen rights and safety, those pursuing this governance strategy should ensure that they do not inadvertently privilege certain types of actors in the AI supply chain over others; (3) Curbing false advertising in commercial AI products may be challenging in the face of proprietary software and data, but the opacity around these technologies makes this governance function that much more important.
- AI governance generates three novel challenges in the context of responsibility: (1) The open-source development of AI makes it harder to trace where in the AI supply chain harms occur, and to mitigate their downstream impacts; (2) Defining and prohibiting illegitimate uses is challenging in the context of AI because of the wide range of possible uses of AI systems in practice; (3) The

individualised nature of AI decisions generates challenges for contestability in both a legal and political sense.

- The challenges highlighted above may require the use of a combination of previously developed regulatory tools, as well as the development of some new, creative measures. Some potential approaches include: the use of algorithmic registers and licensing mechanisms at different levels of the supply chain; amendments to procurement law to ensure that there is a stronger scientific basis for government procurement of AI systems; and the establishment of robust processes to enable individuals to report algorithmic harms, which are then aggregated, publicised, and actioned into the policymaking process.

● ● ● Executive summary

Why This Matters

This report explains the historical role that governments have assumed in creating a more responsible ecosystem around new technologies, and protecting the safety and fundamental rights of citizens. This role is particularly crucial in the context of AI, given how ubiquitously it is used and how it is now almost impossible to opt out of its impacts. Analogies to past governance challenges, such as regulations targeting smog/air pollution, can help policymakers understand their role in the presence of externalised risks. Additionally, examining past governance *responses* to harms and risks associated with new technologies, enables policymakers to build AI governance mechanisms based on existing governance tools, rather than approaching AI governance as an unprecedented and daunting task. Moreover, governance remains an *iterative* exercise, in which initial governance efforts are typically crucial in enabling governments to revise and improve subsequent efforts.

Governance responses to AI are important because the failure to meet the concerns of affected parties risks creating responsibility gaps, or outcomes for which society bears the costs but no-one ultimately faces the consequences. It also inhibits the *relational*

practice of responsibility, in which affected parties can consistently feed into the process of improving AI systems. The ability to have concerns heard and responded to is the hallmark of any working democracy. Persistent responsibility gaps threaten social trust and solidarity, by diminishing the trust that individuals and communities have both in institutions and each other. Although the concept of trust is more often used to describe the loss of trust by entities looking to use or adopt AI, the impacts of losing *social* trust and solidarity present even greater challenges, by threatening broader social well-being and political stability.

Research Questions

1. How does the history of the field of AI factor into the increasingly ambiguous definition of AI? What considerations should policymakers take into account when defining AI in policy documents?
2. What role does governance play in ensuring a responsible AI ecosystem?
3. In what ways can policymakers enable effective AI governance, accountability, and compliance?
4. What useful lessons can AI policymakers draw from other regulatory/governance sectors and histories?
5. How does the ecosystem of AI development, deployment, and use pose novel challenges related to responsibility and accountability? How can these challenges be effectively met using existing tools? Are there newer tools of governance that should be considered/pursued to meet these challenges?

Methodology

This report uses an interdisciplinary set of methods. Concepts from philosophy and AI ethics are used to define the challenges to responsibility posed by the ecosystem of AI design, development, and use. Secondary historical research is used to collect evidence

of policy interventions in response to the harms and risks introduced by historically new technologies. Empirical methods in political science, such as structured, focused comparison and comparative process tracing, are used to compare different historical events and interpret their relevance to the policy process. Finally, policy analysis is used to discuss the ramifications of different potential policy choices.

Key Outputs

- Discussion of how AI has been defined in policy documents, and the impact this definition has on the remit of policymaking efforts.
- Conceptual framing of responsibility challenges posed by AI.
- Comparative analysis of past policy responses to historically new innovations.
- Review of existing/proposed EU, US, and UK uses of policy instruments, such as registration/adverse event reporting, licensing, inspection, risk monitoring, prohibition of false advertising, and standards setting, for the purposes of AI regulation.
- Cautioning against repeating past policy failures related to the use of insurance and audits/inspections for ensuring compliance and improving regulatory capacity.
- Discussion of novel challenges posed by AI, with possible solutions related to transparency, robust complaints processes, revised government procurement laws, and research on risks/harms emerging in the varied uses of AI systems in practice.
- Recommendations for increasing the statutory authority and power of governing bodies to enable more robust recording of adverse events, collecting of reported harms, and investigation/certification.

Limitations

While policymaking tends to build on previous efforts, there are limitations to taking this approach, particularly if doing so limits the ability of governing bodies to come up with creative ways of meeting specific AI-related governance challenges. The approach of building on previous policy efforts also risks limiting the types of harms that can be addressed to those which were previously addressable. This may not make sense if current attitudes have changed, and there is societal support for tackling a wider range of harms, such as those not previously studied or understood sufficiently in marginalised communities. An overreliance on past policy approaches may also create more opportunities for repeating past policy failures.

● ● ● Report

Introduction

Artificial intelligence (AI) has quickly come to underpin a wide range of systems in both the public and private sector. The expansion of AI has enabled exciting new advances such as the enhanced prediction and labelling of protein structures and quicker drug discovery¹ and disaster recovery.² In addition to these scientific advances, AI recommender systems are used in social media platforms and apps to recommend anything from online content, to potential dating partners. AI and its related tools are also used to allocate resources, such as public benefits, transplant organs, and insurance benefits, as well as for the automation of arrest decisions and fraud detection. However, along with this myriad of uses have come a number of well-

¹ Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. "Highly accurate protein structure prediction with AlphaFold." *Nature* 596 (7873): 583–89. <https://doi.org/10.1038/s41586-021-03819-2>

² Johns Hopkins University Applied Physics Laboratory. n.d. "APL shaping an intelligent approach to disaster response and relief." [press release] Accessed 28 March, 2023. <https://www.jhuapl.edu/news/news-releases/190926-apl-shaping-intelligent-approach-disaster-response-and-relief>

publicised harms, such as wrongful arrests,³ amplification of misinformation,⁴ and discrimination in hiring.⁵ As a result of these harms, a growing number of scholars, practitioners, and members of the public have called for the design and implementation of safeguards to protect those affected by algorithmic decision-making. This policy report explores the role of governance in reifying citizen rights and promoting the adoption of AI in socially beneficial ways. Contextualising AI governance challenges within other regulatory/governance sectors and histories, this report concludes with some potential ways of leveraging existing policy instruments to address some of the novel challenges posed by the rapid deployment of AI.

Defining AI: An Important Step for AI Governance

Since the term “artificial intelligence” was first coined in the 1950s by Stanford University professor John McCarthy as the “science and engineering of making intelligent machines”,⁶ the underlying approaches to and uses of AI have continued to evolve. Despite the exciting advances that followed, including well-publicised wins at chess tournaments, AI largely failed to meet the lofty goal of human intelligence envisioned by McCarthy and others.⁷ Consequently, funding for AI research waxed and waned until the late 1990s.⁸ The field then experienced a resurgence from the

³ Hill, Kashmir. 2020. “Another arrest, and jail time, due to a bad facial recognition match.” *New York Times*, 29 December, sec. Technology. <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html> Accessed July 24, 2024.

⁴ Donovan, Joan. 2020. “Social-media companies must flatten the curve of misinformation.” *Nature* [World View]. <https://doi.org/10.1038/d41586-020-01107-z>

⁵ Reuters. 2018. “Amazon scraps secret AI recruiting tool that showed bias against women,” 10 October, sec. Retail. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>, Accessed July 24, 2024.

⁶ McCarthy, J. n.d. “What Is AI?” Accessed March 28, 2023. <http://jmc.stanford.edu/articles/whatisai.html>

⁷ University of Washington. 2006. “The history of artificial intelligence.” [course website], December. <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf> Accessed July 24, 2024.

⁸ Agar, Jon. 2020. “What is science for? The Lighthill report on artificial intelligence reinterpreted.” *British Journal for the History of Science* 53 (3): 289–310. <https://doi.org/10.1017/S0007087420000230>

1990s to the early 2000s, as statistical methods (not always called AI) succeeded at specific business-related tasks, such as targeted advertising.

In the past 15 years or so, as a result of advances in the efficiency of statistical/machine learning algorithms, improved computing storage infrastructure, increased data processing speeds, and the availability of large amounts of data, AI has been used largely as a term to define the use of statistical/machine learning models for prediction-related tasks using a range of modelling techniques as simple as linear and logistic regression, to more complex models based on neural networks (deep learning). AI is also used to describe interactive systems that learn from environments (reinforcement learning), and knowledge-based systems that answer queries about images and documents (symbolic learning). Deep learning methods, in particular, have become commercially popular because they continually outperform other types of models and their performance can scale reliably with capital investment. Additionally, these models have become widely accessible, as the machine learning research community has developed an open-source marketplace in which pre-trained models (trained on large amounts of data) are made publicly available, and easily adaptable for specific uses.

In contrast to the design of AI for narrow domain-specific tasks, a subfield of researchers stayed focused on the field's traditional and more ambitious goal of developing what has been called "artificial general intelligence" (AGI), going on to create a number of successful companies and well-funded AGI research initiatives. However, as the most recent crop of generative models, such as GPT-4, now demonstrate a capacity for more flexible and general applications than previous commercial AI tools, at least one leading AI researcher⁹ has called for abandoning the

⁹ Arul, Akashdeep. 2022. "Yann LeCun sparks a debate on AGI vs human-level AI". *Analytics India Magazine*, 27, January. <https://analyticsindiamag.com/yann-lecun-sparks-a-debate-on-agi-vs-human-level-ai/> Accessed July 24, 2024.

“narrow AI/general AI” dichotomy for a more stable axis of comparison between machine-like and human-level intelligence.

The constant fluctuation of the definition of AI, particularly as the field continues to struggle to define itself, presents challenges for governments looking to understand what the focus of AI regulation and governance should be. As the hype around AI has grown, companies have seized on the marketing benefits of using the term to describe virtually any data-driven architecture or system. The myriad of diverse use cases makes it more challenging to disambiguate between lower- and higher-risk applications, and makes it near impossible to develop universal evaluation standards. The ambiguity around AI’s capabilities has also incentivised companies to overstate the ability of the systems they are selling,^{10,11} and to utilise pseudoscientific applications, such as emotion detection, to drive influential decisions, such as deciding who should be hired for a job.¹² Additionally, proponents of AGI tend to overstate the capabilities of current modelling techniques and platforms developed based on large language models.¹³ These overstated claims also drive a disproportionate focus on the long-term risks and threats of superintelligent machines, even though it is unlikely that human-level intelligence is achievable within the next 3–5 years, and unclear whether it is even possible in the longer term.¹⁴

¹⁰ Raji, Inioluwa Deborah, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst. 2022. “The fallacy of AI functionality.” In FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 959–72. <https://doi.org/10.1145/3531146.3533158>

¹¹ Narayanan, Arvind, and Sayash Kapoor. 2022. “Introducing the AI snake oil book project.” AI Snake Oil [blog]. 25 August. <https://aisnaeoil.substack.com/p/introducing-the-ai-snake-oil-book> Accessed July 24, 2024.

¹² Chen, Angela, and Karen Hao. n.d. “Emotion AI researchers say overblown claims give their work a bad name.” MIT Technology Review. Accessed January 31, 2023. <https://www.technologyreview.com/2020/02/14/844765/ai-emotion-recognition-affective-computing-hirevue-regulation-ethics/>

¹³ Narayanan, Arvind, and Sayash Kapoor. 2023. “GPT-4 and professional benchmarks: The wrong answer to the wrong question.” AI Snake Oil [blog]. 20 March. <https://aisnaeoil.substack.com/p/gpt-4-and-professional-benchmarks> Accessed July 24, 2024.

¹⁴ Fjelland, Ragnar. 2020. “Why general artificial intelligence will not be realized.” Humanities and Social Sciences Communications 7 (1): 1–9. <https://doi.org/10.1057/s41599-020-0494-4>

Against the backdrop of this complex information environment around the harms and risks of AI, governing bodies are presented with the tough challenge of continuing to protect the fundamental rights and safety of their citizens. To that end, policy documents such as the Draft EU AI Act¹⁵ and the US's Blueprint for an AI Bill of Rights,¹⁶ have sought to declare the rights citizens have in interacting with AI systems, and the role of governments in ensuring that these rights are protected and preserved. However, the EU and US have taken different approaches when it comes to defining the scope of technologies covered by their respective policy documents. The EU's Draft AI Act tends to focus mostly on public sector and law enforcement use of AI,¹⁷ and generates different risk tiers, including unacceptable risks, high risks, limited risks, and minimal risks, on the basis of categories that have been somewhat arbitrarily decided.¹⁸ In addition, the AI Act tends to focus on the duties of "providers" and "users" (influenced by previous product safety law), without consideration of the impacts on decision subjects, or those who are influenced by the decision made by a given system, but would not qualify as users of the system.¹⁹ In contrast, while the US's AI Bill of Rights is currently not legally binding, it more broadly describes the rights of *all* citizens as they apply to *all* automated decision-making systems.

The EU's approach of focusing on a narrow set of technologies and types of deployments presents a number of challenges to adequately address the harms emerging from the ubiquitous use of AI. For one, by targeting an arbitrary set of

¹⁵ European Commission. 2021. "Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts." COM(2021) 206 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206> Accessed July 24, 2024.

¹⁶ Office of Science and Technology Policy (OSTP). n.d. "Blueprint for an AI Bill of Rights." The White House. Accessed January 31, 2023. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> Accessed July 24, 2024.

¹⁷ Edwards, L. 2022. "Expert explainer: The EU AI Act Proposal." Accessed 30 March, 2023. <https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/>

¹⁸ Mahler, Tobias. 2021. "Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal." *Nordic Yearbook of Law and Informatics*. <https://papers.ssrn.com/abstract=4001444> Accessed July 24, 2024.

¹⁹ Edwards, L. 2022. "Expert opinion: Regulating AI in Europe." Accessed 30 March, 2023. <https://www.adalovelaceinstitute.org/report/regulating-ai-in-europe/>

technologies that are “high risk”, it fails to understand that risk can arise in any situation where there are complex interactions between the many humans involved in inputting, annotating, interpreting, and managing data and data-driven systems throughout the lifecycle of model design, development, and deployment. Thus, the concept of risk for AI-based systems cannot be neatly reduced to pre-determined risk tiers. Additionally, by privileging certain types of systems over others, the AI Act risks creating perverse incentives for regulatory avoidance, and potentially stifling innovation and the social and economic benefits that could be realised by using those systems in a responsible manner.²⁰ Finally, by stating that human oversight can be a means of risk mitigation, the draft AI Act risks generating even more algorithmic harms, particularly if the full responsibility for a system outcome is delegated to a human who lacks the adequate knowledge or control to act on that delegation.²¹

While the scope of technology considered by AI-related regulations poses some genuine tensions and challenges due to the far-reaching definition of AI, it is nevertheless important that regulatory efforts focus on targeting current practical use cases of AI and ensure that there is a robust process to assess risks and collect reported harms for all systems that utilise data-driven techniques. In addition, public sector bodies should consider conducting impact assessments before procuring systems falling within the wide range of AI-related products and services, and focus more on the manner of technological implementation, rather than the features of the underlying technology, when developing procurement-related laws and policies.

The Role of Governance in Addressing Responsibility Gaps

Researchers within the fields of engineering and AI ethics have long expressed the concern that, as a result of automated decision-making, it is increasingly difficult to

²⁰ Render, A and Engler, A. 2023. “What’s in a name?” CEPS [blog]. 22 February. <https://www.ceps.eu/ceps-publications/whats-in-a-name/> Accessed July 24, 2024.

²¹ Crootof, Rebecca, Margot E. Kaminski, and W. Nicholson Price II. 2022. “Humans in the loop”. 76 *Vanderbilt Law Review* 429 (2023). <https://doi.org/10.2139/ssrn.4066781>

hold anyone accountable for the outcomes of AI.^{22,23} Since Andreas Matthias coined the term “responsibility gap” to describe machine outcomes for which society bears the cost but for which no one is held responsible,²⁴ there have been a number of papers describing how these gaps arise in the context of AI.^{25,26,27,28} In particular, the literature on this topic highlights the challenge of assigning responsibility in the presence of two features: (1) the issue of “many hands” – or the wide range of distributed actors involved in designing, developing, and deploying AI systems; and (2) the diminished knowledge and control of individual actors involved in both designing and operating these systems.

At present, some of these gaps are being filled by members of the research community and civil society, who have worked to raise awareness of AI-related harms, such as wrongful arrests based on racial biases in facial recognition technology.²⁹ In response to these efforts, some technology companies have, at least publicly, stated that they

²² Nissenbaum, Helen. 1996. “Accountability in a computerized society.” *Science and Engineering Ethics* 2 (1): 25–42. <https://doi.org/10.1007/BF02639315>

²³ Cooper, A. Feder, Emanuel Moss, Benjamin Laufer, and Helen Nissenbaum. 2022. “Accountability in an algorithmic society: Relationality, responsibility, and robustness in machine learning.” In *FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 864–76. <https://doi.org/10.1145/3531146.3533150>

²⁴ Matthias, Andreas. 2004. “The responsibility gap: Ascribing responsibility for the actions of learning automata.” *Ethics and Information Technology* 6 (3): 175–83. <https://doi.org/10.1007/s10676-004-3422-1>

²⁵ Santoni de Sio, Filippo, and Giulio Mecacci. 2021. “Four responsibility gaps with artificial intelligence: Why they matter and how to address them.” *Philosophy & Technology* 34 (4): 1057–84. <https://doi.org/10.1007/s13347-021-00450-x>

²⁶ Nyholm, Sven. 2018. “Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci.” *Science and Engineering Ethics* 24 (4): 1201–19. <https://doi.org/10.1007/s11948-017-9943-x>

²⁷ Coeckelbergh, Mark. 2020. “Artificial intelligence, responsibility attribution, and a relational justification of explainability.” *Science and Engineering Ethics* 26 (4): 2051–68. <https://doi.org/10.1007/s11948-019-00146-8>

²⁸ Lima, Gabriel, Nina Grgić-Hlača, Jin Keun Jeong, and Meeyoung Cha. 2022. “The conflict between explainable and accountable decision-making algorithms.” In *FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2103–13. <https://doi.org/10.1145/3531146.3534628>

²⁹ Algorithmic Justice League. n.d. “Unmasking AI harms and biases.” Accessed March 28, 2023. <https://www.ajl.org/>

plan to limit their use of facial recognition,³⁰ and researchers within academia and the private sector have worked to create technical toolkits³¹ and research technical methods to mitigate these harmful outcomes.

While these design solutions are necessary and important, they typically only emerge in a narrow subset of cases when those affected by negative outcomes are in an influential enough position to be able to surface these harms, there are design fixes available to mitigate them, and the company/institution building the system views the issue to be important enough to merit design fixes. Thus, **the status quo does not enable the consistent *relational* practice of responsibility, in which the concerns of affected parties can effectively feed into the process of improving systems at scale, or issuing appropriate remedies.**³² Moreover, many algorithmic harms³³ cannot be effectively addressed through design changes alone, particularly if the entire product is built based on faulty assumptions or dubious causal links, or harms come about due to the unanticipated effects of interactions between different actors and systems within the AI supply chain.

Although some aspects of the current ecosystem of AI development, deployment, and use generate novel challenges (which will be described in later sections of this report), **in general, many of the challenges posed by AI are tractable and can be addressed through the amendment and effective use of existing policy instruments.** The narrative that AI is an entirely unprecedented governance challenge has long been used as an

³⁰ Hill, Kashmir. 2022. "Microsoft plans to eliminate face analysis tools in push for 'Responsible A.I.'" *New York Times*, 21 June. sec. Technology. <https://www.nytimes.com/2022/06/21/technology/microsoft-facial-recognition.html> Accessed July 24, 2024.

³¹ IBM Research. 2018. "Introducing AI fairness 360, a step towards trusted AI." *IBM Research Blog*. 19 September. <https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/> Accessed July 24, 2024.

³² Vargas, Manuel. 2013. *Building better beings: A theory of moral responsibility*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199697540.001.0001>

³³ Shelby, Renee, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla, et al. 2023. "Identifying sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction". *arXiv*. <https://doi.org/10.48550/arXiv.2210.05791>

excuse to delay needed governance action using known modes of responsible public management of technological risk. However, democratic governments have historically played a role in addressing both information deficits – a lack of understanding of just how new technologies contribute to harmful outcomes – and the issue of “many hands”, by mediating between different interests and prioritising the safety and fundamental rights of citizens. The section below compares different governance responses to new innovations, and discusses how AI governance can be informed by, and build upon, these previous efforts.

The Historical Role of Governance in Response to Technological Advances

The method of structured, focused comparison was developed by political scientists to study historical events and their implications for important policy problems.³⁴ This method involves asking the same questions of different case studies to determine the circumstances under which certain phenomena occur. Table 1 presents the outcome of this method when applied to the following question: *how does governance emerge in response to new innovations?* The case studies explored characterise regulatory efforts in the response to steamboat accidents, air pollution, airline safety concerns, the sale of harmful/misleading drugs, and financial crises. The questions presented below draw on a framework developed by historical scholar Peter Maust,³⁵ in which he classifies US government responses to steamboat accidents into four categories: *information collection, testing of inventions, legal penalties, and comprehensive regulatory efforts.*

³⁴ George, Alexander L., Paul I. Gordon Case, Ed Laurén, Timothy J. McKeown and Case Studies. “Chapter 3 The method of structured, focused comparison.” *International School for Advanced Studies*.

³⁵ Maust, Peter. 2012. “Preventing ‘those terrible disasters’: Steamboat accidents and congressional policy, 1824–1860,” August. <https://ecommons.cornell.edu/handle/1813/31121> Accessed July 24, 2024.

<i>Domain Area</i>	<i>Regulated UK Legislations</i>	<i>Intended outcome?</i>	<i>Arguments made in opposition to regulation?</i>	<i>Government role in collecting and disseminating information?</i>	<i>Government role in testing/evaluation?</i>	<i>What regulatory/legal tools were used?</i>
Steamboats ^{36, 37}	Steam Navigation Act(s) of 1846 and 1851	Reducing steamboat accidents	Steamboat accidents were isolated incidents; Budget deficits; Disagreement by engineers about optimal solutions; Employer might cease to	Record keeping of accidents; Dissemination of findings to general public; Expert testimony	Accident investigations; Providing funding to external research organisations / professional associations to test / develop standards	Government-backed inspections of steamboiler, navigation system and vessel; Professional certification of engineers and other personnel; Mandatory insurance; Tort/criminal liability to enforce regulatory statutes

³⁶ Bartrip, Peter W. J. 1980. "The state and the steam-boiler in nineteenth-century Britain." *International Review of Social History* 25 (1): 77–105. <https://doi.org/10.1017/S0020859000006222>

³⁷ Armstrong, J., and D. M. Williams. 2003. "The steamboat, safety and the state: Government reaction to new technology in a period of laissez-faire." *The Mariner's Mirror* 89 (2): 167–84. <https://doi.org/10.1080/00253359.2003.10659284>

			exercise care under assumption official body had taken responsibility; Threats to innovation			
Civil aviation ³⁸	British Air Navigations for Civil Flying (1919) and subsequent joint international efforts	Instilling public confidence in air travel	Threats to innovation/development of civil aircraft; Debates on what features constituted necessary safety features versus desirable commercial	Record keeping of accidents; Dissemination of findings to general public	Accident investigations	Licensing of personnel (pilots, especially); Registering of aircraft and nationality marks; Prohibited areas; Rules of the air/ground operations; Customs regulations (taxes, etc.);

³⁸ Chaplin, J.C. 2011. "Safety regulation- The first 100 years." <https://www.aerosociety.com/media/4858/safety-regulation-the-first-100-years.pdf> Accessed July 24, 2024.

			features; Industry complaints that standards on private pilots were too stringent			Certification of aircrafts; Mandatory Insurance; Tort/criminal liability to enforce regulatory statutes
Drug development and sale³⁹	Pharmacy and Medicines Act of 1941; Medicines Act of 1968 and subsequent regulations	Restricting advertising and illegitimate claims of therapeutic benefit	Proponents of alternative medicine opposed; Newspapers opposed as their advertising business was impacted	Information initially collected and disseminated by professional associations such as the Pharmaceutical Society and British Medical Association; Later legislations took	As legislation evolved, it became more and more stringent about the kinds of tests that needed to be carried out to obtain a license	Professional requirements for administering medicines; Labelling requirements on medicines; Licensing required to manufacture and sell medicines; Restrictions on advertisement;

³⁹ Ferner, Robin E., and Jeffrey K. Aronson. 2023. "Medicines legislation and regulation in the United Kingdom 1500-2020." *British Journal of Clinical Pharmacology* 89 (1): 80–92. <https://doi.org/10.1111/bcp.15497>

				over regulatory duties from Pharmaceutical Society and expanded capacity of government to collect and disseminate information		Tort/criminal liability for medical administration; Criminal sanctions for specific regulatory statutes, e.g. illegal advertising
Smog/pollution caused by coal-fired power stations and automobiles⁴⁰	Clean Air Act of 1956	Reducing air pollution/smog	Some Members of Parliament felt that the previous Public Health Act of 1936 had adequately regulated smoke; Worries about regulating the domestic	Collection of monitoring data and publishing of National Survey of Air Pollution	Funded research on air pollution	Continued earlier smoke abatement clauses; Specified chimney heights; Designated smoke control areas; Increased scope from previous industrial restrictions to restriction on domestic use

⁴⁰ Brimblecombe, P. 2006. "The clean air act after 50 years." *Weather*, 61 (11). Accessed March 29, 2023. <https://doi.org/10.1256/wea.127.06>

			usage of fuels; Concerns about adequate availability of smokeless fuels			
Financial crises related to complex financial instruments/innovations such as collateralised loan obligations (CLOs) and attendant complex derivative markets⁴¹	Financial Services Act of 2012; Makes changes to Bank of England Act of 1998, Banking Act of 2009, and Financial Services and Markets Act of 2000	Creating new regulatory framework for supervision and management of banking and financial services industry (in response to the 2008	The banking industry has tended to be very influential, and in general has argued that regulation is expensive, puts the UK at a competitive disadvantage, and restricts the ability of banks to lend. Arguments were	Under the previous regulatory regime, the Financial Services Authority (FSA) was tasked with promoting public awareness and promoting public understanding	In-depth investigations conducted to understand reasons for financial crisis and understand gaps in existing policymaking	Replacing FSA with three new regulators: Financial Policy Committee (FPC), which is responsible for macro-prudential regulation, the Prudential Regulatory Authority (PRA), which is responsible for micro-prudential regulation, and the Financial Conduct Authority (FCA), which is focused on consumer protection.

⁴¹ Rawlings, Phillip, Andromach Georgosouli, Costanza Russo. 2014. "Regulation of financial services: Aims and methods" *Centre for Commercial Law Studies*. Accessed 29 March, 2023.

		financial crisis)	made that banks would move out of London if laws were too strict. Criticism also targeted the complexity of having three regulators	of the financial system		<p>Licensing of firms done by both FCA and PRA (firms undertaking certain activities are licensed by PRA);</p> <p>FCA’s expanded authority on regulating business practices (regardless of which licensing regime a firm belongs to) enables more proactive regulation through banning of financial products;</p> <p>Investigation of business practices;</p> <p>Inspection of compliance; Publicising of deceptive practices</p>
--	--	-------------------	---	-------------------------	--	---

Table 1. Summary of Government Interventions Across Different Domain Areas



While the examples presented in Table 1 are not meant to be exhaustive, several observations emerge regarding both the regulatory process and the types of regulatory proficiencies that the government has built over the years. **For one, the history of technology governance shows that although governing bodies are typically reluctant to regulate technologies, the process of governance is necessarily iterative.** None of the regulatory efforts above was perfected in the first instance. In fact, drug regulations have been written and re-written numerous times over the past 100 years. However, initial governance efforts traditionally served as an important building block in establishing safer technologies over time.

Additionally, **the examples above demonstrate the role of governance in managing the distorted information ecosystem around new technologies.** When steam boilers first exploded, engineers, operators, and pilots would not always agree on the reasons for the adverse outcome^{42,43} because the science was at a nascent stage, and there were still many unsolved questions. Additionally, when a steamboat or plane crashed, it was not always possible to get eyewitness accounts of what happened, because accidents would destroy much of the evidence, and pilots/operators did not always survive. In the absence of consistent and reliable information, the media filled in the gaps with sensationalist accounts, often filled with misinformation and fuelled by common myths.⁴⁴ Similar challenges pervaded the development and sale of drugs, as it was not always easy to understand which drugs were effective and which drugs were mainly based on pseudoscience. **Governments have historically managed distorted information ecosystems by facilitating external information-gathering and independent testing of outcomes. They have also utilised tools, such as registering artefacts (i.e. steamboats/planes), accident reporting, investigations, and other risk-**

⁴² See footnote 35.

⁴³ Burke, John G. 1966. "Bursting boilers and the federal power." *Technology and Culture* 7 (1): 1-23. <https://doi.org/10.2307/3101598>

⁴⁴ McLachlan, Scott, Burkhard Schafer, Kudakwashe Dube, Evangelia Kyrimi, and Norman Fenton. 2022. "Tempting the fate of the furious: Cyber security and autonomous cars." *International Review of Law, Computers & Technology* 36 (2): 181-201. <https://doi.org/10.1080/13600869.2022.2060466>

monitoring techniques, to understand whether incidents were one-off events or posed broader systemic risks. The information gathered was made publicly available and communicated through efforts such as press releases as well.

In all the cases presented, **governments have also managed to resolve concerns and tensions around the distribution of responsibility and liability for adverse outcomes.** For example, in the case of aviation, there were concerns regarding the degree to which pilots should be held liable for outcomes, and in the case of drug development and sale, early efforts struggled to disambiguate between the responsibilities of those selling drugs and those administering them. Furthermore, in the wake of the financial crisis, it was clear that there were many factors responsible for the collapse of the housing market, including inappropriate business practices by banks, insurers, and rating agencies, complexity in the financial instruments themselves, and the failures of risk-prediction models throughout the system. **Governments have historically managed these challenges by utilising a combination of different policy tools, including licensing (both technological artefacts and professional personnel), inspection (for compliance), monitoring (proactive assessment of systemic risk), prohibiting false advertising, and restricting/prohibiting use of high-risk technologies.**

Many AI governance challenges reproduce the same questions and tensions as those presented in the cases above. AI governance has similarly been characterised by regulatory apprehension about constructing novel approaches to countering emerging risks. The information ecosystem around AI is also similarly distorted. Researchers within the broad field of AI disagree about the capabilities of AI-based systems, the reasons for adverse outcomes, the severity of these outcomes (one off events vs. systematic issues), and the appropriate responses to these outcomes. The media, while serving an important role in surfacing AI-related harms, has, in some cases, also

inadvertently played a role in sensationalising both the benefits and risks of AI.⁴⁵ Additionally, the issue of distributing responsibility and liability appropriately across a wide set of actors remains salient, as AI continues to be developed and deployed on a global scale and in a wide range of sectors. Below, we discuss some of the current proposed approaches for AI governance within the context of previous approaches.

AI Proposals Similar to Previous Approaches

This section places AI policy efforts undertaken in the US,⁴⁶ UK, and EU within the taxonomy of governance approaches developed in the previous section.

Independent Product Testing/Standards Setting

In the US, the Consumer Products Safety Commission (CPSC) is working alongside the National Institute of Standards and Technology (NIST), a subsidiary of the Department of Commerce in the US, to develop staff expertise for testing AI/ML products.⁴⁷ In the UK, the British Standards Institute (BSI) is in charge of testing, verification, and certification of products and services. NIST recently released a risk management framework⁴⁸ for voluntary use, and the BSI is actively developing AI-related standards.⁴⁹

⁴⁵ Kapoor, Sayash, and Arvind Narayanan. 2022. "Eighteen pitfalls to beware of in AI journalism." AI Snake Oil [blog]. 30 September. <https://aisnakeoil.substack.com/p/eighteen-pitfalls-to-beware-of-in> Accessed July 24, 2024.

⁴⁶ OSTP. 2022. "Biden-Harris administration announces key actions to advance tech accountability and protect the rights of the American public." The White House. 4 October. <https://www.whitehouse.gov/ostp/news-updates/2022/10/04/fact-sheet-biden-harris-administration-announces-key-actions-to-advance-tech-accountability-and-protect-the-rights-of-the-american-public/> Accessed July 24, 2024.

⁴⁷ U.S. Consumer Product Safety Commission. n.d. "Artificial intelligence and machine learning in consumer products". Accessed 30 March 2023. <https://www.cpsc.gov/About-CPSC/artificial-intelligence-and-machine-learning-in-consumer-products>

⁴⁸ Computer Security Division, Information Technology Laboratory. 2016. "About the RMF - NIST Risk Management Framework" *CSRC*. 30 November. <https://csrc.nist.gov/projects/risk-management/about-rmf> Accessed July 24, 2024.

⁴⁹ British Standards Institute. n.d. "Artificial intelligence." Accessed 30 March, 2023. <https://www.bsigroup.com/en-GB/industries-and-sectors/artificial-intelligence/>

Recording/Investigation of Adverse Events

In the US, some agencies, such as the Federal Trade Commission (FTC), the Department of Labor's Equal Employment Opportunity Commission (EEOC), the Consumer Financial Protection Bureau (CFPB), and the Federal Communications Commission (FCC), have a longstanding complaints process in place, through which anyone can submit a complaint about consumer products or business practices to the relevant authorities. Both the FTC and EEOC have explicitly stated their intention to investigate AI-related complaints, particularly as they relate to deceptive business practices (FTC),⁵⁰ and discriminatory hiring practices (EEOC).⁵¹

Requiring Registration/Licensing of Technological Artefacts

Because AI-based systems are already used extensively in the public sector, a number of European cities have created algorithmic registers for the purposes of making the public aware of the algorithms being used by governing bodies.⁵² In addition to these efforts, the Dutch government recently unveiled a register, which is currently contributed to on a voluntary basis, but will soon be legally required.⁵³

For private sector registration, the UK's Information Commissioner's Office (ICO) requires organisations processing personal data to register with the ICO and pay a data protection fee.⁵⁴ Additionally, pre-market approvals and licensing, which are currently required for new medical devices by the US Federal Drug Administration

⁵⁰ Federal Trade Commission. n.d. "Keep your AI claims in check." Accessed 29 March, 2023.

<https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check>

⁵¹ US EEOC. n.d. "Artificial intelligence and algorithmic fairness initiative." Accessed 29 March, 2023.

<https://www.eeoc.gov/ai>

⁵² Algorithm Register. n.d. "Algorithmic transparency standard." Accessed 30 March, 2023.

<https://www.algorithmregister.org/>

⁵³ "Het Algoritmeregister van de Nederlandse Overheid." n.d. Accessed March 29, 2023.

<https://algoritmes.overheid.nl/>

⁵⁴ ICO. "Data protection fee." 2022. 17 October. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-fee/> Accessed July 24, 2024.

(FDA)⁵⁵ and the UK Medicines and Healthcare products Regulatory Agency (MHRA),⁵⁶ will continue to be required for devices that utilise AI.

Defining/Prohibiting Illegitimate Uses

Although there is no US federal law related to facial recognition, some state-level advocacy efforts in the US led some states to ban the use of facial recognition in police departments. The situation continues to evolve, however, and some of the states that previously banned police use of facial recognition are subtly re-introducing these technologies.⁵⁷ The EU recently released a new draft of the EU AI Act, which bans biometric mass surveillance.⁵⁸

Conducting Inspections to Ensure Compliance with Standards

The FDA released a draft framework for the post-market surveillance of AI-related medical devices,⁵⁹ and the Medicines and Healthcare Products Regulatory Agency (MHRA)⁶⁰ has proposed a similar effort within their Algorithmic Change Programme. A discussion paper by the UK Digital Regulation Cooperation Forum, consisting of the Competition and Markets Authority (CMA), Ofcom, ICO, and FCA, suggests that algorithmic audits are a promising mechanism for assessing regulatory compliance.⁶¹

⁵⁵ FDA. 2022. “Artificial intelligence and machine learning in software as a medical device.” September. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> Accessed July 24, 2024.

⁵⁶ Medicines and Healthcare products Regulatory Agency. n.d. “Software and AI as a medical device change programme - roadmap.” *GOV.UK*. Accessed 29 March, 2023. <https://www.gov.uk/government/publications/software-and-ai-as-a-medical-device-change-programme/software-and-ai-as-a-medical-device-change-programme-roadmap>

⁵⁷ CNN Business. n.d. “First, they banned facial recognition. now they’re not so sure.” Accessed 29 March, 2023. <https://edition.cnn.com/2022/08/05/tech/facial-recognition-bans-reversed/index.html>

⁵⁸ News European Parliament. 2023. “AI Act: A step closer to the first rules on artificial intelligence.” 11 May. <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> Accessed July 24, 2024.

⁵⁹ See footnote 58.

⁶⁰ See footnote 58.

⁶¹ Digital Regulation Cooperation Forum. n.d. “Auditing algorithms: The existing landscape, role of regulators and future outlook.” *GOV.UK*. Accessed March 29, 2023. <https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing->

Curbing False Advertising About Efficacy

The FTC recently made clear its commitment to curbing false advertising related to the efficacy of AI-related products.⁶² The EU's Unfair Commercial Practices Directive also addresses unfair business practices, and the European Commission recently launched a "fitness check" or evaluation of existing EU consumer protection legislations (expected to be released in the second half of 2024) to ensure that they provide adequate protection in the digital environment.⁶³

Restricting Use of High-risk Technologies to Professionally Licensed Personnel

Existing regulations restricting the use of certain medical devices to licensed healthcare professionals will likely continue to apply in the case of AI-related devices.

Devolving Responsibility, Where Appropriate, to Insurance Markets

The use of insurance for underwriting AI-related risk has not yet been discussed in much detail. Insurance has historically been used as a means of improving the regulatory capacity of governments, particularly in the face of budget constraints.

workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook

⁶² See footnote 51.

⁶³ European Commission. 2022. "Digital fairness – fitness check on EU consumer law." 14 June. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13413-Digital-fairness-fitness-check-on-EU-consumer-law_en Accessed July 24, 2024.

Agency	Goals/type of harms targeted	Legal/policy framework
Consumer Financial Protection Bureau (CFPB) ⁶⁴	Discrimination/bias in lending (auto, student, mortgage)	Anti-discrimination law; Equal Credit Opportunity Act (ECOA)
Equal Employment Opportunity Commission (EEOC) (subsidiary of Department of Labor) ⁶⁵	Discrimination/bias in hiring algorithms	Anti-discrimination law; American Disabilities Act (ADA)
Federal Trade Commission (FTC) ⁶⁶	Discrimination/bias; Transparency; Commercial Surveillance; Data security/privacy	FTC Act's Unfair and Deceptive Practices (UDAP); Fair Credit Reporting Act (FCRA); Equal Credit Opportunity Act (ECOA)

⁶⁴ Consumer Financial Protection Bureau. 2022. "CFPB acts to protect the public from black-box credit models using complex algorithms". 26 May. <https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms/>

⁶⁵ U.S. Equal Employment Opportunity Commission. n.d. "The Americans with Disabilities Act and the use of software, algorithms, and artificial intelligence to assess job applicants and employees." Accessed 31 March, 2023. <https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence>

⁶⁶ Federal Trade Commission. n.d. "Using artificial intelligence and algorithms." Accessed 31 March, 2023. <https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms>

Consumer Products Safety Commission (CPSC) ⁶⁷	Discrimination/bias; Data privacy related to healthcare data; Transparency	Civil Rights Act; Rehabilitation Act; Education Amendments; Age Discrimination Act; Patient Protection and Affordable Care Act; Public Health Service Act; Federal, Food, Drug and Cosmetic Act; Safe Medical Devices Act; Mammography Quality Standards Act (MQSA)
Department of Labor ⁶⁸	Discrimination/bias; Threats to collective action/labor representation; Workplace health/safety	Labor-Management Reporting and Disclosure Act

Table 2: Overview of existing US sectoral approaches that apply to AI systems

⁶⁷ See footnote 48.

⁶⁸ U.S. Department of Labor. n.d. "What the blueprint for an AI bill of rights means for workers." [blog]. Accessed 31 March, 2023. <https://blog.dol.gov/2022/10/04/what-the-blueprint-for-an-ai-bill-of-rights-means-for-workers>

Recommendations for Avoiding Past Policy Missteps

- While the US and UK appear to have developed mechanisms for collecting AI-related harms in specific sectors (healthcare, transportation, etc.), **governments should also consider creating one unified body to record AI-related adverse events, so as to understand the systemic risks posed by AI.** This could be modelled on existing voluntary/nonprofit efforts documenting all AI-related adverse incidents and harms (in both the private and public sector), such as the Partnership on AI's AI Incident Database⁶⁹ and the AI, Algorithmic, and Automation Incidents and Controversies (AIAAIC).⁷⁰
- Audits/inspections remain an important possible avenue to pursue for AI governance, but **governments should take care to ensure that audits remain completely independent, so as to avoid issues such as regulatory capture, which were highlighted during both the Enron⁷¹ and Boeing/FAA scandals.⁷²**
- **Although curbing AI false advertising may be challenging in the face of proprietary software and data, the opacity around these technologies makes this governance function that much more important.**
- **Governments might consider expanding the use of licensing (of artefacts and personnel) in the case of high-risk applications of AI.**
- While insurance could potentially be utilised as a policy instrument for AI governance as well, **past policy failures such as the US financial crisis of 2008 suggest the importance of ensuring that issues like *moral hazard*,⁷³ where the**

⁶⁹ Partnership on AI. n.d. "AI incidents database." Accessed 29 March, 2023.

<https://partnershiponai.org/workstream/ai-incidents-database/>

⁷⁰ AIAAIC. n.d. "AIAAIC Repository." Accessed March 29, 2023. <https://www.aiaaic.org/aiaaic-repository>

⁷¹ Thomas, C. William. 2002. "The rise and fall of Enron." 2002. *Journal of Accountancy*. April 1, 2002. <https://www.journalofaccountancy.com/issues/2002/apr/theriseandfallofenron.html> Accessed July 24, 2024.

⁷² Reuters. 2020. "U.S. house report blasts failures of Boeing, FAA in 737 MAX Certification." September 16, , sec. Aerospace & Defense. <https://www.reuters.com/article/boeing-737max-congress-idUSL1N2GD046> Accessed July 24, 2024.

⁷³ Bhutta, N. and B.J. Keys. n.d. "Eyes wide shut? The moral hazard of mortgage insurers during the housing boom." *Zell and Lurie Real Estate Center*. Accessed 29 March, 2023.

presence of insurance encourages undue risk-taking, or *adverse selection*,⁷⁴ where information asymmetries distort market outcomes, are carefully studied, to avoid falling victim to the same policy failures of the past.

Novel and Specific AI Governance Challenges

Although AI reproduces many of the governance challenges presented by previous technological advances, AI does generate some novel challenges due to the sheer scale and scope of its impacts. Three major challenges are highlighted below, followed by some policy recommendations for addressing them.

Open-Source Development and “many hands”

The open-source nature of models and widespread availability of data has increased the average level of access to AI, making its use possible in any sector imaginable. Areas such as computer vision, initially consisted of a small group of researchers with specialised skills and training. However, as the commercial viability of using these techniques has become more apparent, an entire open-source marketplace has emerged, with large pretrained models available online and increasingly adapted to domain-specific applications.⁷⁵ While other global supply chains with many hands, such as food systems, have licensed producers, sellers, and inspectors, the same cannot be said for the AI supply chain.

<https://realestate.wharton.upenn.edu/working-papers/eyes-wide-shut-the-moral-hazard-of-mortgage-insurers-during-the-housing-boom/>

⁷⁴ Kahn, James A., and Benjamin S. Kay. 2020. “The impact of credit risk mispricing on mortgage lending during the subprime boom.” *BIS Working Papers* 875. Accessed July 24, 2024.

⁷⁵ Thomas, Suzanne L. 2019. “Migration versus management: The global distribution of computer vision engineering work.” In *2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE)*, 12–17. <https://doi.org/10.1109/ICGSE.2019.00017>

The relatively few barriers to uploading or using AI-based models has a number of implications for policymakers. For one, it is hard to surface the risks posed by AI if it is unclear who is using it and for what purpose. Highly regulated sectors such as healthcare have processes in place for putting commercial products on the market, and monitoring their impacts, but less regulated sectors do not. Additionally, although the legal liability tends to be placed on the end user of the system, that user is not always aware of the risks they are taking, or in a position to mitigate that risk. Moreover, if the system the end user is purchasing contains a combination of open-source models and proprietary data, infrastructure, or other elements, intellectual property laws may make it impossible for the end user to adequately inspect the system.⁷⁶ A number of uncertainties also remain around the appropriate ways to test and evaluate AI-based systems to prevent downstream harms, particularly given the limitations of modular design, which tends to push responsibility down the supply chain.⁷⁷

The challenges outlined above risk creating persistent responsibility gaps, particularly since end users have the power to make consequential decisions about decision subjects. While users can decide not to use a certain technology because they do not trust it, decision subjects often lack the ability to opt out of consequential decisions made about them. One potential way in which governments looking to protect the fundamental rights and safety of their citizens can mitigate algorithmic harms and risks is by making amendments to existing procurement law (as the US federal government has suggested it will do),⁷⁸ and ensuring that any public sector use of AI has adequately considered future impacts, with appropriate remedies in place in the case of adverse algorithmic outcomes.

⁷⁶ Wenn, Shelby. 2017. "Houston teachers to pursue lawsuit over secret evaluation system". 11 May. *Houston Chronicle*. Accessed 31 March, 2023. <https://www.houstonchronicle.com/news/houston-texas/houston/article/Houston-teachers-to-pursue-lawsuit-over-secret-11139692.php>

⁷⁷ Widder, David Gray, and Dawn Nafus. 2022. "Dislocated accountabilities in the AI supply chain: Modularity and developers' notions of responsibility." *arXiv*. <https://doi.org/10.48550/arXiv.2209.09780>

⁷⁸ See footnote 47.

When it comes to regulating private sector development of AI, in addition to recording/monitoring private sector use (through existing modes of licensing/registration), governments may also need to consider more heavily utilising novel licensing/certification mechanisms to distinguish between models that have adequately considered downstream impacts in their design and distribution, versus those which have not. Governments may look to licensing mechanisms such as the responsible AI licenses (RAIL) recently proposed by researchers⁷⁹ as one possible framework. **As past policy efforts demonstrate, regulatory efforts in the presence of “many hands” need to ensure that there are appropriate mechanisms for holding each of the hands accountable, so as to not pass on the risk to actors who are not able to adequately mitigate it.**

Defining Illegitimate Uses

One challenge generated by the current ecosystem around AI development is the issue of determining legitimate versus illegitimate uses of AI-based systems. In many ways, the benefits and risks of using AI are not necessarily immediately visible. However, the continued hype around these technologies has led to massive investment in their development and deployment. As a result, much like the case of drug development and sale, there have been many pseudoscientific applications of AI arising from the technological hype cycle, such as computer vision software measuring/evaluating worker productivity.⁸⁰

Even in more regulated sectors such as healthcare, because AI-based tools enable the decision-making function previously undertaken by licensed professionals to be

⁷⁹ Contractor, Danish, Daniel McDuff, Julia Katherine Haines, Jenny Lee, Christopher Hines, Brent Hecht, Nicholas Vincent, and Hanlin Li. 2022. “Behavioral use licensing for responsible AI.” In *FACCT ’22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 778–88. <https://doi.org/10.1145/3531146.3533143>

⁸⁰ Biddle, Sam. 2018. “Artificial intelligence experts issue urgent warning against facial scanning with a ‘dangerous history.’” *The Intercept*. Accessed 29 March, 2023. <https://theintercept.com/2018/12/06/artificial-intelligence-experts-issue-urgent-warning-against-facial-scanning-with-a-dangerous-history/>

partially undertaken by a tool, there are a wide range of ways in which these tools can be used, with understudied impacts. For example, a radiologist using an AI-based system to triage which scans should be prioritised (i.e. generate the most uncertainty) generates different risks than a radiologist using an AI-based system to sense check scans they have already read themselves. Until clear practice guidelines develop on the use of these technologies, there will continue to be risks that are not necessarily addressed through licensing/approval of the technologies alone. Moreover, the use of technologies to take over the decision-making of well-trained professionals risks deskilling, such that the workforce is less able to spot machine errors or remedy them over time.⁸¹

It is important that public sector bodies using AI ensure that there is a justified use for these technologies, based on scientific knowledge or understanding. When it comes to public sector use of AI systems, governments should consider the bureaucratic counterfactual⁸² to decide if the AI system offers any benefits over current approaches, or is merely a hyped-up technology with an uncertain risk profile. Impact assessments, such as the Canadian Impact Assessment,⁸³ offer a potential mechanism for proactively assessing potentially false claims made by AI vendors.

Sectoral bodies may also consider prohibiting certain private sector uses of AI, much like previous interventions in drug development prohibited the sale of drugs that had very high-risk side effects. Restrictions or prohibitions may be necessary if it appears

⁸¹ Aquino, Yves Saint James, Wendy Rogers, Annette Braunack-Mayer, Helen Frazer, Khin Win, Nehmat Houssami, Christopher Degeling, Christopher Semsarian, and Stacy M. Carter. 2022. "Professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills." SSRN Scholarly Paper. <https://doi.org/10.2139/ssrn.4129747>

⁸² Johnson, Rebecca Ann, and Simone Zhang. 2022. "What is the bureaucratic counterfactual? Categorical versus algorithmic prioritization in U.S. social policy." In *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1671–82. <https://doi.org/10.1145/3531146.3533223>

⁸³ Secretariat, Treasury Board of Canada. 2021. "Algorithmic impact assessment tool." Guidance. 22 March. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html> Accessed July 24, 2024.

as though AI is being used as a justification to engage in behaviour that is in violation of fundamental rights or based on pseudoscientific claims. Doing this effectively may require new statutory powers and more resources for sectoral bodies so that they can have the authority to intervene on fraudulent business practices, and the resources to hire staff with the appropriate skills to inspect and evaluate proprietary systems and data.

Individualised Outcomes

One of the outcomes of the wide-scale use of machine learning models is the production and commercial/public use of increasingly individualised products and services. Individualised products and services are seen as a way of enhancing the customer experience, and outcomes that are better tailored to individual circumstances. However, the individualised nature of AI decisions generates challenges for contestability in both a legal and political sense. Building a strong legal case typically requires a group of harmed claimants to jointly contest a given outcome. If an outcome is personalised, the person harmed by an outcome may not know that there are other claimants who experienced the outcome for the same reason. As a result, the harmed claimants may not have the political power to contest harmful systems through joint advocacy efforts, or the legal basis for contesting outcomes in court.

The issue of personalisation and its impacts on contestability suggests the need for governments to empower harmed claimants by (1) making them aware of the decision they have been subjected to, and the basis upon which it was made, and (2) providing them with the opportunity to report harmful outcomes to governing bodies. The individualised nature of algorithmic decisions suggests that it is more important than ever that governing bodies act as an aggregator of different types of reported harms, by collecting, pooling, and publicising algorithmic harms. Governments could use existing mechanisms, such as the complaints databases created and managed by the

US's CFPB⁸⁴ and Federal Communications Commission (FCC),⁸⁵ or the super-complaints⁸⁶ process part of the UK's Police Reform Act of 2002 to collect and analyse harms, and then determine whether they indicate larger-scale risks that need to be addressed through further regulatory action. In taking on this role, governments can also mitigate the loss of political power and agency caused by algorithmic decisions. The wide-ranging nature of AI applications suggests the importance of having a centralised authority collect AI-related complaints and issue appropriate remedies. Again, it will be important to ensure that the entity tasked with this role is provided with the adequate statutory authority to take on this function.

⁸⁴ Consumer Financial Protection Bureau. n.d. "Consumer complaint database." Accessed 29 March, 2023. <https://www.consumerfinance.gov/data-research/consumer-complaints/>

⁸⁵ Federal Communications Commission. n.d. "Consumer complaint data center." Accessed 29 March, 2023. <https://www.fcc.gov/consumer-help-center-data>

⁸⁶ Independent Office for Police Conduct, College of Policing, and HM Inspectorate of Constabulary and Fire & Rescue Services. 2022. "Police super-complaints: Guidance on submitting a super-complaint about policing." *GOV.UK*. 13 June. <https://www.gov.uk/guidance/police-super-complaints> Accessed July 24, 2024.

● ● ● Conclusion

This report uses the lens of responsibility gaps used in AI ethics literature to evaluate the role of governments in responding to the risks and harms posed by AI. First, the report briefly presented the challenges of defining AI, and the implications of using a narrow versus broad definition on the remit of policymaking efforts. Then, the report contextualised AI governance within other regulatory/governance sectors and histories. Additionally, the report reviewed governance efforts related to automated decision-making and AI currently being undertaken in the US and EU, and identified issues not currently encompassed by these approaches. Finally, the report made some suggestions for existing policy instruments that can be leveraged in addressing novel AI governance challenges.

● ● ● References

1. Agar, Jon. 2020. "What is science for? The Lighthill report on artificial intelligence reinterpreted." *British Journal for the History of Science* 53 (3): 289–310. <https://doi.org/10.1017/S0007087420000230>, accessed 20.05.2024.
2. Aquino, Yves Saint James, Wendy Rogers, Annette Braunack-Mayer, Helen Frazer, Khin Win, Nehmat Houssami, Christopher Degeling, Christopher Semsarian, and Stacy M. Carter. 2022. "Professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills." *SSRN Scholarly Paper*. <https://doi.org/10.2139/ssrn.4129747>, accessed 20.05.2024.
3. AIAAIC. n.d. "AIAAIC Repository." Accessed March 29, 2023. <https://www.aiaaic.org/aiaaic-repository>, accessed 20.05.2024.
4. Algorithmic Justice League. n.d. "Unmasking AI harms and biases." Accessed March 28, 2023. <https://www.ajl.org/>, accessed 20.05.2024.
5. Algorithm Register. n.d. "Algorithmic transparency standard." Accessed 30 March, 2023. <https://www.algorithmregister.org/>, accessed 20.05.2024.
6. Arul, Akashdeep. 2022. "Yann LeCun sparks a debate on AGI vs human-level AI". *Analytics India Magazine*, 27, January. <https://analyticsindiamag.com/yann-lecun-sparks-a-debate-on-agi-vs-human-level-ai/>, accessed 20.05.2024.
7. Armstrong, J., and D. M. Williams. 2003. "The steamboat, safety and the state: Government reaction to new technology in a period of laissez-faire." *The Mariner's Mirror* 89 (2): 167–84. <https://doi.org/10.1080/00253359.2003.10659284>, accessed 20.05.2024.
8. Bartrip, Peter W. J. 1980. "The state and the steam-boiler in nineteenth-century Britain." *International Review of Social History* 25 (1): 77–105. <https://doi.org/10.1017/S0020859000006222>, accessed 20.05.2024.
9. Bhutta, N. and B.J. Keys. n.d. "Eyes wide shut? The moral hazard of mortgage insurers during the housing boom." *Zell and Lurie Real Estate Center*. Accessed 29 March, 2023. <https://realestate.wharton.upenn.edu/working-papers/eyes-wide-shut-the-moral-hazard-of-mortgage-insurers-during-the-housing-boom/>, accessed 20.05.2024.
10. Biddle, Sam. 2018. "Artificial intelligence experts issue urgent warning against facial scanning with a 'dangerous history.'" *The Intercept*. Accessed 29 March, 2023. <https://theintercept.com/2018/12/06/artificial-intelligence-experts-issue-urgent-warning-against-facial-scanning-with-a-dangerous-history/>, accessed 20.05.2024.

11. Brimblecombe, P. 2006. "The clean air act after 50 years." *Weather*, 61 (11). Accessed March 29, 2023. <https://doi.org/10.1256/wea.127.06>, accessed 20.05.2024.
12. British Standards Institute. n.d. "Artificial intelligence." Accessed 30 March, 2023. <https://www.bsigroup.com/en-GB/industries-and-sectors/artificial-intelligence/>, accessed 20.05.2024.
13. Burke, John G. 1966. "Bursting boilers and the federal power." *Technology and Culture* 7 (1): 1–23. <https://doi.org/10.2307/3101598>, accessed 20.05.2024.
14. Chaplin, J.C. 2011. "Safety regulation- The first 100 years." <https://www.aerosociety.com/media/4858/safety-regulation-the-first-100-years.pdf>, accessed 20.05.2024.
15. Chen, Angela, and Karen Hao. n.d. "Emotion AI researchers say overblown claims give their work a bad name." *MIT Technology Review*. Accessed January 31, 2023. <https://www.technologyreview.com/2020/02/14/844765/ai-emotion-recognition-affective-computing-hirevue-regulation-ethics/>, accessed 20.05.2024.
16. CNN Business. n.d. "First, they banned facial recognition. now they're not so sure." Accessed 29 March, 2023. <https://edition.cnn.com/2022/08/05/tech/facial-recognition-bans-reversed/index.html>, accessed 20.05.2024.
17. Crootof, Rebecca, Margot E. Kaminski, and W. Nicholson Price II. 2022. "Humans in the loop". 76 *Vanderbilt Law Review* 429 (2023). <https://doi.org/10.2139/ssrn.4066781>, accessed 20.05.2024.
18. Coeckelbergh, Mark. 2020. "Artificial intelligence, responsibility attribution, and a relational justification of explainability." *Science and Engineering Ethics* 26 (4): 2051–68. <https://doi.org/10.1007/s11948-019-00146-8>, accessed 20.05.2024.
19. Computer Security Division, Information Technology Laboratory. 2016. "About the RMF - NIST Risk Management Framework" CSRC. 30 November. <https://csrc.nist.gov/projects/risk-management/about-rmf>, accessed 20.05.2024.
20. Consumer Financial Protection Bureau. n.d. "Consumer complaint database." Accessed 29 March, 2023. <https://www.consumerfinance.gov/data-research/consumer-complaints/>, accessed 20.05.2024.
21. Consumer Financial Protection Bureau. 2022. "CFPB acts to protect the public from black-box credit models using complex algorithms". 26 May.

- <https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms/>, accessed 20.05.2024.
22. Contractor, Danish, Daniel McDuff, Julia Katherine Haines, Jenny Lee, Christopher Hines, Brent Hecht, Nicholas Vincent, and Hanlin Li. 2022. "Behavioral use licensing for responsible AI." In *FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 778–88. <https://doi.org/10.1145/3531146.3533143>, accessed 20.05.2024.
23. Cooper, A. Feder, Emanuel Moss, Benjamin Laufer, and Helen Nissenbaum. 2022. "Accountability in an algorithmic society: Relationality, responsibility, and robustness in machine learning." In *FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 864–76. <https://doi.org/10.1145/3531146.3533150>, accessed 20.05.2024.
24. Donovan, Joan. 2020. "Social-media companies must flatten the curve of misinformation." *Nature [World View]*. <https://doi.org/10.1038/d41586-020-01107-z>, accessed 20.05.2024.
25. Digital Regulation Cooperation Forum. n.d. "Auditing algorithms: The existing landscape, role of regulators and future outlook." GOV.UK. Accessed March 29, 2023. <https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook>, accessed 20.05.2024.
26. Edwards, L. 2022. "Expert opinion: Regulating AI in Europe." Accessed 30 March, 2023. <https://www.adalovelaceinstitute.org/report/regulating-ai-in-europe>, accessed 20.05.2024.
27. European Commission. 2021. "Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts." COM(2021) 206 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>, accessed 20.05.2024.
28. European Commission. 2022. "Digital fairness – fitness check on EU consumer law." 14 June. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13413-Digital-fairness-fitness-check-on-EU-consumer-law_en, accessed 20.05.2024.
29. Federal Communications Commission. n.d. "Consumer complaint data center." Accessed 29 March, 2023. <https://www.fcc.gov/consumer-help-center-data>, accessed 20.05.2024.

30. FDA. 2022. "Artificial intelligence and machine learning in software as a medical device." September. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>, accessed 20.05.2024.
31. Federal Trade Commission. n.d. "Keep your AI claims in check." Accessed 29 March, 2023. <https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check>, accessed 20.05.2024.
32. Federal Trade Commission. n.d. "Using artificial intelligence and algorithms." Accessed 31 March, 2023. <https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms>, accessed 20.05.2024.
33. Ferner, Robin E., and Jeffrey K. Aronson. 2023. "Medicines legislation and regulation in the United Kingdom 1500-2020." *British Journal of Clinical Pharmacology* 89 (1): 80–92. <https://doi.org/10.1111/bcp.15497>, accessed 20.05.2024.
34. Fjelland, Ragnar. 2020. "Why general artificial intelligence will not be realized." *Humanities and Social Sciences Communications* 7 (1): 1–9.
35. George, Alexander L., Paul I. Gordon Case, Ed Laurén, Timothy J. McKeown and Case Studies. "Chapter 3 The method of structured, focused comparison." *International School for Advanced Studies*.
36. "Het Algoritmeregister van de Nederlandse Overheid." n.d. Accessed March 29, 2023. <https://algoritmes.overheid.nl/>, accessed 20.05.2024.
37. Hill, Kashmir. 2020. "Another arrest, and jail time, due to a bad facial recognition match." *New York Times*, 29 December, sec. Technology. <https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html>, accessed 20.05.2024.
38. Hill, Kashmir. 2022. "Microsoft plans to eliminate face analysis tools in push for 'Responsible A.I.'" *New York Times*, 21 June. sec. Technology. <https://www.nytimes.com/2022/06/21/technology/microsoft-facial-recognition.html>, accessed 20.05.2024.
39. IBM Research. 2018. "Introducing AI fairness 360, a step towards trusted AI." *IBM Research Blog*. 19 September. <https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/>, accessed 20.05.2024.
40. ICO. "Data protection fee." 2022. 17 October. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection->

[regulation-gdpr/accountability-and-governance/data-protection-fee/](#), accessed 20.05.2024.


41. Independent Office for Police Conduct, College of Policing, and HM Inspectorate of Constabulary and Fire & Rescue Services. 2022. "Police super-complaints: Guidance on submitting a super-complaint about policing." GOV.UK. 13 June. <https://www.gov.uk/guidance/police-super-complaints>, accessed 20.05.2024.
42. Johns Hopkins University Applied Physics Laboratory. n.d. "APL shaping an intelligent approach to disaster response and relief." [press release] Accessed 28 March, 2023. <https://www.jhuapl.edu/news/news-releases/190926-apl-shaping-intelligent-approach-disaster-response-and-relief>, accessed 20.05.2024.
43. Johnson, Rebecca Ann, and Simone Zhang. 2022. "What is the bureaucratic counterfactual? Categorical versus algorithmic prioritization in U.S. social policy." In *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1671–82. <https://doi.org/10.1145/3531146.3533223>, accessed 20.05.2024.
44. Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. "Highly accurate protein structure prediction with AlphaFold." *Nature* 596 (7873): 583–89. <https://doi.org/10.1038/s41586-021-03819-2>, accessed 20.05.2024.
45. Kahn, James A., and Benjamin S. Kay. 2020. "The impact of credit risk mispricing on mortgage lending during the subprime boom." *BIS Working Papers* 875. <https://www.bis.org/publ/work875.htm>, accessed 20.05.2024.
46. Kapoor, Sayash, and Arvind Narayanan. 2022. "Eighteen pitfalls to beware of in AI journalism." *AI Snake Oil* [blog]. 30 September. <https://www.aisnakeoil.com/p/eighteen-pitfalls-to-beware-of-in#:~:text=Pitfall%201.,nothing%20to%20do%20with%20robots>, accessed 20.05.2024.
47. Lima, Gabriel, Nina Grgić-Hlača, Jin Keun Jeong, and Meeyoung Cha. 2022. "The conflict between explainable and accountable decision-making algorithms." In *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2103–13. <https://doi.org/10.1145/3531146.3534628>, accessed 20.05.2024.
48. Mahler, Tobias. 2021. "Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal." *Nordic Yearbook of Law and Informatics*. <https://papers.ssrn.com/abstract=4001444>, accessed 20.05.2024.

49. Matthias, Andreas. 2004. "The responsibility gap: Ascribing responsibility for the actions of learning automata." *Ethics and Information Technology* 6 (3): 175–83. <https://doi.org/10.1007/s10676-004-3422-1>, accessed 20.05.2024.
50. Maust, Peter. 2012. "Preventing 'those terrible disasters': Steamboat accidents and congressional policy, 1824–1860," August. <https://ecommons.cornell.edu/handle/1813/31121>, accessed 20.05.2024.
51. McCarthy, J. n.d. "What Is AI?" Accessed March 28, 2023. <http://jmc.stanford.edu/articles/whatisai.html>, accessed 20.05.2024.
52. University of Washington. 2006. "The history of artificial intelligence." [course website], December. <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>, accessed 20.05.2024.
53. McLachlan, Scott, Burkhard Schafer, Kudakwashe Dube, Evangelia Kyrimi, and Norman Fenton. 2022. "Tempting the fate of the furious: Cyber security and autonomous cars." *International Review of Law, Computers & Technology* 36 (2): 181–201. <https://doi.org/10.1080/13600869.2022.2060466>, accessed 20.05.2024.
54. Medicines and Healthcare products Regulatory Agency. n.d. "Software and AI as a medical device change programme - roadmap." GOV.UK. Accessed 29 March, 2023. <https://www.gov.uk/government/publications/software-and-ai-as-a-medical-device-change-programme/software-and-ai-as-a-medical-device-change-programme-roadmap>, accessed 20.05.2024.
55. Narayanan, Arvind, and Sayash Kapoor. 2023. "GPT-4 and professional benchmarks: The wrong answer to the wrong question." *AI Snake Oil [blog]*. 20 March. <https://aisnakeoil.substack.com/p/gpt-4-and-professional-benchmarks>, accessed 20.05.2024.
56. Narayanan, Arvind, and Sayash Kapoor. 2022. "Introducing the AI snake oil book project." *AI Snake Oil [blog]*. 25 August. <https://aisnakeoil.substack.com/p/introducing-the-ai-snake-oil-book>, accessed 20.05.2024.
57. Nissenbaum, Helen. 1996. "Accountability in a computerized society." *Science and Engineering Ethics* 2 (1): 25–42. <https://doi.org/10.1007/BF02639315>, accessed 20.05.2024.
58. News European Parliament. 2023. "AI Act: A step closer to the first rules on artificial intelligence." 11 May. <https://www.europarl.europa.eu/news/en/press->

[room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence](#), accessed 20.05.2024.

59. Nyholm, Sven. 2018. "Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci." *Science and Engineering Ethics* 24 (4): 1201–19. <https://doi.org/10.1007/s11948-017-9943-x>, accessed 20.05.2024.
60. Partnership on AI. n.d. "AI incidents database." Accessed 29 March, 2023. <https://partnershiponai.org/workstream/ai-incidents-database/>, accessed 20.05.2024.
61. Raji, Inioluwa Deborah, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst. 2022. "The fallacy of AI functionality." In *FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 959–72. <https://doi.org/10.1145/3531146.3533158>, accessed 20.05.2024.
62. Rawlings, Phillip, Andromach Georgosouli, Costanza Russo. 2014. "Regulation of financial services: Aims and methods" *Centre for Commercial Law Studies*. Accessed 29 March, 2023.
63. Reuters. 2020. "U.S. house report blasts failures of Boeing, FAA in 737 MAX Certification." September 16, , sec. *Aerospace & Defense*. <https://www.reuters.com/article/boeing-737max-congress-idUSL1N2GD046>, accessed 20.05.2024.
64. Render, A and Engler, A. 2023. "What's in a name?" *CEPS [blog]*. 22 February. <https://www.ceps.eu/ceps-publications/whats-in-a-name/>, accessed 20.05.2024.
65. Reuters. 2018. "Amazon scraps secret AI recruiting tool that showed bias against women," 10 October, sec. *Retail*. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>, accessed 20.05.2024.
66. Santoni de Sio, Filippo, and Giulio Mecacci. 2021. "Four responsibility gaps with artificial intelligence: Why they matter and how to address them." *Philosophy & Technology* 34 (4): 1057–84. <https://doi.org/10.1007/s13347-021-00450-x>, accessed 20.05.2024.
67. Secretariat, Treasury Board of Canada. 2021. "Algorithmic impact assessment tool." *Guidance*. 22 March. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>, accessed 20.05.2024.

68. Shelby, Renee, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla, et al. 2023. "Identifying sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction". *arXiv*. <https://doi.org/10.48550/arXiv.2210.05791>, accessed 20.05.2024.
69. Thomas, C. William. 2002. "The rise and fall of Enron." 2002. *Journal of Accountancy*. April 1, 2002. <https://www.journalofaccountancy.com/issues/2002/apr/theriseandfallofenron.html>, accessed 20.05.2024.
70. Thomas, Suzanne L. 2019. "Migration versus management: The global distribution of computer vision engineering work." In *2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE)*, 12–17. <https://doi.org/10.1109/ICGSE.2019.00017>, accessed 20.05.2024.
71. Office of the Chief Information. 2021. "HHS Artificial Intelligence (AI) Strategy." U.S. Department of Health & Human Services. 22 December. 2021. <https://www.hhs.gov/about/agencies/asa/ocio/ai/strategy/index.html>, accessed 20.05.2024.
72. Office of Science and Technology Policy (OSTP). n.d. "Blueprint for an AI Bill of Rights." The White House. Accessed January 31, 2023. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>, accessed 20.05.2024.
73. OSTP. 2022. "Biden-Harris administration announces key actions to advance tech accountability and protect the rights of the American public." The White House. 4 October. <https://www.whitehouse.gov/ostp/news-updates/2022/10/04/fact-sheet-biden-harris-administration-announces-key-actions-to-advance-tech-accountability-and-protect-the-rights-of-the-american-public/>, accessed 20.05.2024.
74. Vargas, Manuel. 2013. *Building better beings: A theory of moral responsibility*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199697540.001.0001>, accessed 20.05.2024.
75. U.S. Consumer Product Safety Commission. n.d. "Artificial intelligence and machine learning in consumer products". Accessed 30 March 2023. <https://www.cpsc.gov/About-CPSC/artificial-intelligence-and-machine-learning-in-consumer-products>, accessed 20.05.2024.
76. U.S. Department of Labor. n.d. "What the blueprint for an AI bill of rights means for workers." [blog]. Accessed 31 March, 2023. <https://blog.dol.gov/2022/10/04/what-the-blueprint-for-an-ai-bill-of-rights-means-for-workers>, accessed 20.05.2024.

-
77. *USEEOC. n.d. "Artificial intelligence and algorithmic fairness initiative."* Accessed 29 March, 2023. <https://www.eeoc.gov/ai>, accessed 20.05.2024.
78. *U.S. Equal Employment Opportunity Commission. n.d. "The Americans with Disabilities Act and the use of software, algorithms, and artificial intelligence to assess job applicants and employees."* Accessed 31 March, 2023. <https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence>, accessed 20.05.2024.
79. *Wenn, Shelby. 2017. "Houston teachers to pursue lawsuit over secret evaluation system". 11 May. Houston Chronicle. Accessed 31 March, 2023.* <https://www.houstonchronicle.com/news/houston-texas/houston/article/Houston-teachers-to-pursue-lawsuit-over-secret-11139692.php>, accessed 20.05.2024.
80. *Widder, David Gray, and Dawn Nafus. 2022. "Dislocated accountabilities in the AI supply chain: Modularity and developers' notions of responsibility." arXiv.* <https://doi.org/10.48550/arXiv.2209.09780>, accessed 20.05.2024.
- 



This fellowship ran from January-May 2023 as part of BRAID.

BRAID is a UK-wide programme dedicated to integrating Arts and Humanities research more fully into the Responsible AI ecosystem, as well as bridging the divides between academic, industry, policy and regulatory work on responsible AI. Funded by the Arts and Humanities Research Council (AHRC), BRAID represents AHRC's major investment in enabling responsible AI in the UK. The Programme runs from 2022 to 2028. Working in partnership with the Ada Lovelace Institute and BBC, BRAID supports a network of interdisciplinary researchers and partnering organisations through the delivery of funding calls, community building events, and a series of programmed activities. Funding reference: Arts and Humanities Research Council grant number AH/X007146/1.

Learn more at www.braiduk.org

This research was supported via UK Research and Innovation by the R&D Science and Analysis Programme at the Department for Culture, Media & Sport. Any primary research, subsequent findings or recommendations do not represent Government views or policy and are produced according to research ethics, quality assurance, and academic independence.

To request an alternative format of this report please email braid@ed.ac.uk



Arts and
Humanities
Research Council

