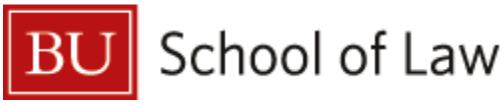

BEFORE THE NATIONAL TELECOMMUNICATION AND INFORMATION: COMMENTS FROM RESEARCHERS FROM BOSTON UNIVERSITY AND THE UNIVERSITY OF CHICAGO

Boston University School of Law
Research Paper Series No. 23-24

June 15, 2023

Ran Canetti
Boston University Graduate School of Arts and Sciences

Aloni Cohen
University of Chicago

Chris Conley
Boston University School of Law

Mark Crovella
Boston University Graduate School of Arts and Sciences

Stacey Dogan
Boston University School of Law

Marco Gaboardi
Boston University School Graduate School of Arts and Sciences

Woodrow Hartzog
Boston University School of Law

Christopher Robertson
Boston University School of Law

Katharine Silbaugh
Boston University School of Law

Rory Van Loo
Boston University School of Law

Before the

National Telecommunications and Information Administration

Washington, D.C. 20230

In the Matter of	)	
	)	
AI Accountability Policy	)	Docket No. 230407-0093
Request for Comment	)	

COMMENTS FROM RESEARCHERS AT BOSTON UNIVERSITY AND UNIVERSITY OF CHICAGO

These comments were composed by an interdisciplinary group of legal, computer science, and data science faculty and researchers at Boston University and the University of Chicago. This group collaborates on research projects that grapple with the legal, policy, and ethical implications of the use of algorithms and digital innovation in general, and more specifically regarding the use of online platforms, machine learning algorithms for classification, prediction, and decision making, and generative AI. Specific areas of expertise include the functionality and impact of recommendation systems; the development of Privacy Enhancing Technologies (PETs) and their relationship to privacy and data security laws; legal regulation of platforms under privacy, intellectual property, and antitrust laws; the science of monitoring and measuring the behavior of large deployed systems and networks; and programming languages and the science of rigorously specifying and verifying properties of algorithms and information systems.

This group has a wide range of views on AI governance but shares a set of core beliefs about the relationship between AI systems, legal accountability, technological transparency, and regulatory oversight. In particular, our research has led us to conclude:

(1) that mechanisms for AI monitoring and accountability must be implemented throughout the lifecycle of important AI systems, including requiring vendors to make concrete pre-release assertions regarding key properties and use cases of AI systems, as well as enabling ongoing monitoring of deployed systems as they evolve over time; and that the verification of vendor claims and the monitoring of AI system behavior are most effective when performed publicly by multiple independent parties, including experts, researchers, and the general public;

(2) that it is feasible and appropriate to demand that accountability mechanisms be robust and their outputs accessible to enable the development and evolution of technical, social, and legal safeguards for

users of and persons affected by AI systems, including (a) new or modified legal and regulatory regimes designed to take into account assertions, evidence and similar information provided by AI developers relevant to intended or known uses of their products, and (b) existing regimes such as product liability, consumer protection, and other laws designed to protect users and others against harm;

(3) that accountability and transparency requirements, including the need for public verification and access, are consistent with protecting privacy, trade secrets, and other intellectual property rights, and that vendors can and should be discreet but forthright, ensure downstream accountability, remain loyal to those who trust them, and use (or develop if necessary) adequate contractual obligations and technologies, including PETs and others, that will preserve the vendors' intellectual property rights and users' privacy;

(4) that accountability and transparency mechanisms are a necessary but not sufficient aspect of AI regulation. Procedural measures that place burdens on consumers are inadequate to prevent harmful uses of AI and should not be viewed as a complete solution. People should be protected from those designing and deploying AI systems regardless of how much information is disclosed or understood. To be effective, a regulatory approach for AI systems must go beyond procedural protections to include substantive, non-negotiable obligations that limit how AI systems can be built and deployed. For example, regulators should consider obligations to minimize data collection and use systems only for pre-specified and legitimate purposes. If the design and deployment of AI systems are unacceptably dangerous, they should be outright prohibited; and

(5) that effective AI regulation requires attention to both general AI-related risks and to specific risks associated with the use of AI in particular contexts. Existing regulatory agencies and enforcement authorities should be empowered to address AI-related risks within their subject matter domain. To enable these authorities to work most effectively and to ensure attention to generalizable risks of AI, we recommend establishment of a meta-agency with broad AI-related expertise (both technical and legal) which would develop baseline regulations regarding the general safety of AI systems, set standards, and enable review for compliance with substantive law, while collaborating with and lending its expertise to other agencies and lawmakers as they consider the impact of AI systems on their regulatory jurisdiction.

In the rest of this response we provide additional elaboration and motivation for the above points. To keep the flow of ideas more consistent and accessible, we are structuring the presentation around the above five points. At the beginning of each section we provide references to the questions in the RFC to which the discussion is responsive.

I. AI systems require multiple public-facing and testable accountability mechanisms throughout their lifecycle.

[Addressing Questions 1, 2, 3, 5, 15, 16, 18, 20, 24, 28]

No single accountability mechanism can suffice to keep AI systems in check. To be effective, an accountability standard must include multiple accountability mechanisms, deployed throughout an AI system's life cycle. Some of these requirements should be adapted to the character of the AI system and the scale of deployment. However, there are also requirements that should be imposed on all AI systems. Moreover, all assertions about AI systems must be public-facing and testable by third-party auditors.

Specifically, the development or deployment of AI systems, including any products that incorporate AI systems, should typically include:

- *Specifications:* AI systems, especially high-stakes ones, should be accompanied by documents that spell out the intended use cases and design goals and set clear and concrete specifications for the system and its claimed performance.
- *Vetting of components:* Any substantial components of an AI system that were developed by other vendors should be specified, along with the properties needed from that component for the overall system to function as specified pursuant to Section I(a) above and the evidence supporting the claim that the component possesses these properties.
- *Maintenance and update plans:* Developers of AI systems should commit to address any identified faults and deficiencies as soon as they become known, or alternatively to dismantle and take down the faulty AI system. In addition, AI products that are not static should include change control plans that indicate how the product will be modified throughout its lifecycle as well as how the developer will ensure continued validity of any assertions about the product. Because changes to AI systems may well have adverse and unexpected side effects, any changes–whether to address faults or to enhance or expand functionality–should require re-asserting that the system as a whole meets the specification as provided above.

AI system developers also should provide evidence that can be effectively and publicly verified that the developed AI system comports with its specifications at the time of the initial release and after any significant revision. This evidence can take multiple forms including, among others:

- mathematical analysis of the algorithms used;
- formal verification of relevant properties of the software (or key parts thereof);
- verifiable / repeatable results of tests of system behavior within the specified use cases;
- asserting the provenance and other relevant properties of the data used to train the AI.

Our research and that of many of our colleagues demonstrates the feasibility and effectiveness of validating assertions using these approaches.[1]

---

[1] See e.g.: Sanjit A. Seshia, Dorsa Sadigh, S. Shankar Sastry: Toward verified artificial intelligence. Commun. ACM 65(7): 46-55 (2022); Chao Huang, Jiameng Fan, Xin Chen, Wenchao Li, Qi Zhu: POLAR: A Polynomial Arithmetic Framework for Verifying Neural-Network Controlled Systems. ATVA 2022: 414-430; Mark Niklas Müller, Gleb Makarchuk, Gagandeep Singh, Markus Püschel, Martin T. Vechev: PRIMA: general and precise neural network certification via scalable convex hull approximations. Proc. ACM Program. Lang. 6(POPL): 1-33 (2022); Nina Narodytska, Shiva Prasad Kasiviswanathan, Leonid Ryzhyk, Mooly Sagiv, Toby Walsh: Verifying Properties of Binarized Deep Neural Networks. AAAI 2018: 6615-6624; Tommaso Dreossi, Daniel J. Fremont, Shromona Ghosh, Edward Kim, Hadi Ravanbakhsh, Marcell Vazquez-Chanlatte, Sanjit A. Seshia:VerifAI: A Toolkit for the Formal Design and Analysis of Artificial Intelligence-Based Systems. CAV (1) 2019: 432-442; Greg Anderson, Shankara Pailoor, Isil Dillig, Swarat Chaudhuri: Optimization and abstraction: a synergistic approach for analyzing neural network robustness. PLDI 2019: 731-744; Greg Anderson, Shankara Pailoor, Isil Dillig, Swarat Chaudhuri: Optimization and abstraction: a synergistic approach for analyzing neural network robustness. PLDI 2019: 731-744; Gilles Barthe, Gian Pietro Farina, Marco Gaboardi, Emilio Jesús Gallego Arias, Andy Gordon, Justin Hsu, Pierre-Yves Strub: Differentially Private Bayesian Programming. CCS 2016: 68-79; Tetsuya Sato, Alejandro Aguirre, Gilles Barthe, Marco Gaboardi, Deepak Garg, Justin Hsu: Formal verification of higher-order probabilistic programs: reasoning about approximation, convergence, Bayesian inference, and optimization. Proc. ACM Program. Lang. 3(POPL): 38:1-38:30 (2019).

AI vendors must also enable and facilitate the monitoring and auditing of deployed AI systems both by common users and by white-hat researchers, as long as this monitoring activity is not demonstrably intended to subvert the behavior of the system or harm users. Third-party monitoring provides an additional layer of protection against oversights in analysis, as well as a way to surface any unintended or unexpected adverse effects of the deployed system. Facilitation of third-party monitoring is especially important given the pervasive and dynamic nature of AI systems as well as the difficulties external researchers may experience in interrogating such systems without developer cooperation.[2]

II. Specification and verification of AI products is both feasible and necessary to develop a comprehensive accountability regime.

[Addressing Questions 1, 2, 7, 9, 11, 28]

The recommendations above contemplate that distributors of AI systems will have ongoing obligations to articulate the intended uses and expected behaviors of their systems, to update those claims throughout the life cycle of the systems, and to facilitate validation of those claims by third parties, regardless of whether the system is sold or offered free of charge. We acknowledge that this approach is a departure from existing practice in information systems in general and particularly in AI systems, and we expect there to be some reservation from the industry regarding such requirements. Yet other industries whose products present both individual and systemic risk, such as banking, energy, food, drugs, and transportation, are routinely required to disclose the attributes of their products both to ensure their safety and to enable evaluation by regulators, consumers, researchers, and watchdog groups.

At the outset, it is important to debunk the common misconception that AI systems are so complicated that they cannot be meaningfully described or understood. Claims that AI systems are too complex to be well understood and should be treated as “black boxes” are dangerous because they would significantly weaken the standards of accountability for these systems. To the contrary, the fields of computing and data sciences have developed a sophisticated set of tools designed to allow effective assessment of AI products. Our own research and that of other computer scientists demonstrates that it is feasible to require specification and assertion of salient properties of complex software.[1]

In addition, public-facing specifications and other documentation of AI systems are crucial to minimize inadvertent harm and to develop accountability mechanisms. While these requirements might indeed slow the release of new features, they are likely to enhance innovation in the important areas of verification, testing, and guaranteeing reliability of AI systems. Technical accountability mechanisms, including assertions and third-party audits, are also essential to both the development and the implementation of legal and social mechanisms to prevent harms arising from AI systems and to clarify the liability of

[2] See e.g. Spinelli, Crovella, How YouTube Leads Privacy-Seeking Users Away from Reliable Information, 28th ACM Conference on User Modeling, Adaptation and Personalization 244 (2020). Available at: https://doi.org/10.1145/3386392.3399566. This research investigates the nature of YouTube's recommendation system, showing that the platform typically directs users away from reliable sources and towards extreme or unreliable content, especially for privacy-seeking users. This "lead away" effect, exacerbated by privacy-preserving behavior, and its implications for user access to reliable information are critical concerns given the dominance of YouTube as an information source and its advertising-based business model. The researchers were frustrated in their efforts to extend this research, however, because the platform blocked them from using automated systems to conduct their experiments.

developers when their products are used in ways that cause such harms. We address such mechanisms in more detail below.

III. Accountability need not come at the expense of privacy and trade secrets.

[Addressing Questions 3, 21, 24, 25, 27]

The recommendations outlined above are compatible with vendor's trade secret and intellectual property rights, as well as with the privacy of individuals whose data is used to train the AI systems.[3] There exist a broad range of technologies that enable the generation and presentation of the information required to assess the performance of an AI system in ways that protect the privacy of individuals whose data was used in the training, as well as the trade secrets and other intellectual property rights of the vendor.[4] Furthermore, appropriate contractual and other legal mechanisms can be used to ensure that external research does not infringe upon the rights of the developer or third parties.[5]

Moreover, assertions of privacy or intellectual property rights have often been pretextual, used explicitly to minimize transparency and thus evade accountability.[6] We do not rule out the possibility that there may be circumstances where technical accountability mechanisms and other significant rights are in tension, but any claim of such tension must be substantiated, including by demonstrating the absence of any viable workaround. In general, we believe that vendors should be obligated to use—or develop, if needed—appropriate technologies that enable compliance with technical transparency obligations while protecting the privacy and intellectual property rights of all parties involved.

IV. Accountability requires substantive, non-negotiable obligations built upon but extending beyond technical transparency mechanisms.

[Addressing Questions 1, 3, 4, 11, 24, 30, 34]

---

[3] See e.g. Bamberger, Kenneth A. and Canetti, Ran and Goldwasser, Shafi and Wexler, Rebecca and Zimmerman, Evan, *Verification Dilemmas in Law and the Promise of Zero-Knowledge Proofs* (February 7, 2021), Berkeley Technology Law Journal, Vol. 37, No. 1, 2022, Available at SSRN: https://ssrn.com/abstract=3781082 or http://dx.doi.org/10.2139/ssrn.3781082; Rory Van Loo, *Privacy Pretexts,* 108 Cornell Law Review 1 (2022), https://cornelllawreview.org/wp-content/uploads/2023/03/2884.pdf.

[4] See more information on privacy enhancing technologies and their uses in Canetti, Ran; Kaptchuk, Gabe; Reyzin, Leonid; Smith, Adam; and Varia, Mayank: Request for Information (RFI) on Advancing Privacy Enhancing Technologies: https://www.nitrd.gov/rfi/2022/87-fr-35250/Canetti-Kaptchuk-Reyzin-Smith-Varia-PET-RFI-Response-2022.pdf. For a detailed example where PETs (in this case zero knowledge proofs) have been used to enable the assertion of salient properties regarding software, while making sure that the software itself remains private, see Bitan, Dor; Canetti, Ran; Goldwasser Shafi; Wexler, Rebecca: Using Zero-Knowledge to Reconcile Law Enforcement Secrecy and Fair Trial Rights in Criminal Cases, In Proceedings of the 2022 Symposium on Computer Science and Law (2022). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4074315

[5] See Woodrow Hartzog, *Chain-Link Confidentiality*, 46 Georgia Law Review 657 (2012), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2045818; Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. UNIV. L. REV. 961, 961 (2021), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3642217.

[6] See Rory Van Loo, *Privacy Pretexts*, 108 Cornell Law Review 1 (2022), https://cornelllawreview.org/wp-content/uploads/2023/03/2884.pdf.

Transparency and accountability are necessary for effective AI regulation, but without substantive and non-negotiable prohibitions, they will devolve into managerial box-checking exercises that are costly but fail to protect against the dangers of AI systems.[7] Decades of experience regulating information and technology demonstrate that without rules that explicitly limit how AI systems can be built or deployed, organizations will be able to comply superficially with transparency requirements while continuing to engage in the corrosive behavior that such rules were meant to discourage.[8] FTC commissioners have admitted that the agency's "notice and choice" approach, which relied heavily upon concepts of transparency and empowering consumers, doesn't work.[9] Lawmakers should not make the same mistake when confronting AI systems.

To be effective, a regulatory approach for AI systems must go beyond procedural protections to include substantive, non-negotiable obligations that limit how AI systems can be built and deployed. For example, regulators should consider obligations to minimize data collection and use systems only for pre-specified and legitimate purposes. Data minimization, the idea that companies should collect, keep, and use only information that is reasonably necessary and proportionate to what was requested and should use that data for a justified and specified purpose, is a core commitment of information governance and is critical for the regulation of AI systems.[10] Legislators have also recently proposed laws that include a duty of loyalty that would require companies to prioritize the best interests of people made vulnerable by their exposure to algorithms and data.[11]

---

[7] See Ari Waldman, Industry Unbound: The Inside Story of Privacy, Data, and Corporate Power (2021); Woodrow Hartzog, Privacy's Blueprint: The Battle to Control the Design of New Technologies (2018).

[8] See Woodrow Hartzog and Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection,* 61 Boston College Law Review 1687 (2020), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3441502; Richards, Hartzog, and Francis, *Comments of the Cordell Institute on the Prevalence of Commercial Surveillance and Data Security Practices that Harm Consumers*. Available at SSRN: https://ssrn.com/abstract=4284020.

[9] Remarks of Chair Lina M. Khan As Prepared for Delivery IAPP Global Privacy Summit 2022 Washington, D.C., available at https://www.ftc.gov/system/files/ftc_gov/pdf/Remarks%20of%20Chair%20Lina%20M.%20Khan%20at%20IAPP%20Global%20Privacy%20Summit%202022%20-%20Final%20Version.pdf; Preliminary FTC Staff Report, Protecting Consumer Privacy in an Age fo Rapid Change (December 2010), at https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-bureau-consumer-protection-preliminary-ftc-staff-report-protecting-consumer/101201privacyreport.pdf.

[10] Daniel J. Solove and Woodrow Hartzog, Breached! Why Data Security Law Fails and How to Improve It (Oxford University Press 2022).

[11] See, e.g., Data Care Act of 2019, S. 2961, 116th Cong. § 2 (2019) , https://www.congress.gov/bill/116th-congress/senate-bill/2961/text; Consumer Online Privacy Rights Act, S.2968, 116th Cong. § 101 , https://www.congress.gov/bill/116th-congress/senate-bill/2968/text#toc-idd95044fbea1d498f888e130c44e92067; New York Privacy Act, S. 5642 (2019), https://www.nysenate.gov/legislation/bills/2019/s5642; European Commission, Proposal for a Regulation on European data governance (Data Governance Act), https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-data-governance-data-governance-act (Nov. 25, 2020); Age Appropriate Design Code, U.K. Data Protection Act of 2018 § 123(1); An Act to provide facial recognition accountability and comprehensive enforcement, Mass. Bill H.117 (2021), https://malegislature.gov/Bills/192/H117 ("A covered entity shall be prohibited from taking any actions with respect to processing facial recognition data or designing facial recognition technologies that conflict with an end user's best interests."); see also Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 Wash. Univ. L. Rev. 961, 961 (2021); Woodrow Hartzog & Neil Richards, *The Surprising Virtues of Data Loyalty*, 71 Emory L. J. 985, 1033 (2022); Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 Notre Dame L. Rev. Reflection 356, 378–384 (2022).

If the design and deployment of specific AI systems are unacceptably dangerous, they should be outright prohibited. For example, multiple cities have recently prohibited government procurement and use of facial recognition systems, and the city of Portland prohibited its use in public places.[12] Lawmakers have proposed several new bills that include substantive prohibitions on actions like microtargeted political ads, the sale of geolocation data, and behavioral advertising to children.[13] Lawmakers have also recently passed specific design prohibitions such as rules against "dark patterns," which are manipulative and deceptive user interfaces meant to trick and persuade people into exposing themselves and making choices against their interest.[14] Lawmakers should consider a similar approach using specific substantive prohibitions to regulate AI systems to prevent the dilution and managerialization of otherwise robust legal frameworks.

V. AI accountability requires a dedicated oversight agency and public accountability mechanisms to foster the development of additional safeguards.

[Addressing Questions 2, 3, 11, 24, 26, 30, 31, 32, 34]

Finally, the government should develop a regulatory framework that enables effective oversight of AI systems, including both general AI systems and those intended for use in a particular regulated domain. In our view, the optimal structure would involve a meta-agency[15] that conducts its own rulemaking and enforcement actions applicable broadly to AI systems while supporting the development of robust public accountability mechanisms for AI systems in specific contexts (as described above) by other actors.

The meta-agency would develop rules addressed to the safety and accountability of AI systems generally. Rather than replicate AI expertise across every specialized agency, we encourage reliance on the meta-agency's experts in developing both substantive standards and accountability mechanisms. In addition, the meta-agency should develop rules that further facilitate the study of AI systems and efforts

---

[12] *See* Shannon Flynn, *13 Cities Where Police Are Banned from Using Facial Recognition Tech*, Innovation & Tech Today, https://innotechtoday.com/13-cities-where-police-are-banned-from-using-facial-recognition-tech/; Alred Ng, *Portland Passes the Toughest Ban on Facial Recognition in the US*, Cnet, Sept. 10, 2020, https://www.cnet.com/news/politics/portland-passes-the-toughest-ban-on-facial-recognition-in-the-us/.

[13] Press Release, Congresswoman Anna G. Eshoo, *Rep. Eshoo Reintroduces Legislation to Ban Microtargeted Political Ads*, Aug. 5, 2021, https://eshoo.house.gov/media/press-releases/rep-eshoo-reintroduces-legislation-ban-microtargeted-political-ads; Makena Kelly, *Warren Proposes Sweeping Ban on Location and Health Data Sales,* The Verge, June 15, 2022, https://www.theverge.com/2022/6/15/23169718/roe-wade-elizabeth-warren-location-data-tracking-ban-sale-brokers; Alfred Ng, *Biden Calls for Ban of Online Ads Targeting Children*, Politico, Feb. 7, 2023, https://www.politico.com/news/2023/02/07/biden-calls-for-ban-of-online-ads-targeting-children-00081731.

[14] Catherine Zhu, *Dark Patterns - A New Frontier in Privacy Regulation*, Reuters, June 29, 2021, https://www.reuters.com/legal/legalindustry/dark-patterns-new-frontier-privacy-regulation-2021-07-29/; Latham & Watkins LLP, *The Digital Services Act: Practical Implications for Online Services and Platforms*, March 2023, https://www.lw.com/admin/upload/SiteAttachments/Digital-Services-Act-Practical-Implications-for-Online-Services-and-Platforms.pdf; Felicity Slater, *The Future of Manipulative Design Regulation*, Future of Privacy Forum Blog, Jan. 19, 2023, https://fpf.org/blog/the-future-of-manipulative-design-regulation/.

[15] For a related area where the creation of a meta-agency has been proposed see Van Loo, *Rise of the Digital Regulator*, 66 Duke Law Journal 1267 (2017). Available https://ssrn.com/abstract=2902238 ("[T]he government should establish a meta-agency to create a comprehensive governance framework involving interdisciplinary experts and a centralized federal agency, which would enhance the efficacy of digital intermediaries while mitigating their potential shortcomings.")

to mitigate their risks. In particular, we encourage the enactment of legal protection for researchers seeking to study algorithms, as discussed above. The agency should also leverage the growing regulatory tool of pushing companies to help in the regulation of one another, in gatekeeper roles.[16] The meta-agency should also encourage the development and use of legal and technical tools, for example legal relationships of trust and confidentiality and public sector PETs, that protect businesses' and consumers' privacy interests without compromising accountability.

This meta-agency should support, rather than supplant, the roles of other expert agencies (such as the Federal Aviation Administration, the Federal Trade Commission, and the Food & Drug Administration) in regulating products relevant to their own expertise. Other agencies should receive the support of experts in the meta-agency to develop and execute rules directed to their statutory mandate, and provide input to the meta-agency as to the technical transparency mechanisms that best enable co-regulation of AI systems. Such an approach would facilitate the development of best practices across agencies while allowing each agency to tailor its practices based on its existing industry-specific expertise.

We believe that this approach will allow the development of technical, social, and legal approaches to assessing and regulating potentially-harmful uses of AI, including but not limited to additional regulation by other specialized federal agencies. By ensuring that AI vendors publicly disclose assertions about their product and proof of those assertions, and by enabling external researchers to validate those proofs, the meta-agency can contribute vital information to the development and evolution of legal and other safeguards concerning AI systems by state and federal courts and lawmakers. In particular, explicit and testable promises of suitability for a given purpose would enable various legal regimes, including state and federal consumer protection and products liability, to address both vendors and users of AI systems. More broadly, public-facing, robust technical accountability mechanisms allow businesses, consumers, courts and lawmakers alike to hold AI developers accountable for foreseeable harms while giving those developers a clear framework to evaluate their own systems and protect their interests.

VI. Conclusion

Public, robust technical accountability measures are an essential and entirely feasible tool to ensure that AI systems benefit the public and avoid foreseeable harms. Our research and those of our colleagues across the computer science, data science and legal fields demonstrates that AI systems are not "black boxes" that are impervious to oversight. Instead, well-studied approaches can be deployed to assess and verify the suitability of AI systems for their intended purposes, leveraging the expertise of a specialized AI oversight agency as well as independent researchers. These mechanisms need not violate interests in privacy or intellectual property, or interfere with the role of specialized agencies and other lawmakers in making and implementing policy. Instead, it serves as a foundation upon which legal and other regimes ensuring the responsible use of AI systems must be built.

Submitted on June 12, 2023 on behalf of the following signatories

[16] For more on this regulatory mechanism, see Rory Van Loo, *The New Gatekeepers: Private Firms as Public Enforcers*, 106 Va. L. Rev. 467 (2020), https://virginialawreview.org/articles/new-gatekeepers-private-firms-public-enforcers.

Ran Canetti
Wang Professor of Computer Science, Boston University
Director of Graduate Studies, Boston University Computer Science
Director of the Reliable Information Systems and Cyber Security Center, Boston University

Aloni Cohen
Assistant Professor of Computer Science and Data Science, University of Chicago

Chris Conley
Assistant Director, Technology Law Clinic at Boston University School of Law
Clinical Associate Professor of Law, Boston University School of Law

Mark Crovella
Professor of Computer Science and
Professor Computing & Data Sciences,
Boston University

Stacey Dogan
Professor of Law & Law Alumni Scholar, Boston University School of Law
Affiliated Faculty, Boston University Faculty of Computing and Data Sciences

Marco Gaboardi
Associate Professor of Computer Science, Boston University

Woodrow Hartzog
Professor of Law & Class of 1960 Scholar, Boston University School of Law

Christopher Robertson
N. Neal Pike Scholar and Professor, Boston University School of Law
Professor of Health Law, Policy, & Management (Secondary), Boston University School of Public Health

Katharine B. Silbaugh
Professor of Law & Law Alumni Scholar, Boston University School of Law

Rory Van Loo
Professor of Law & Michaels Faculty Research Scholar, Boston University School of Law

Appendix

Atleson, *Keep your AI claims in check*. Federal Trade Commission Division of Advertising Practices (2023). Available at: https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check

> Businesses often make false or unsubstantiated claims about AI products. This memo from the FTC provides guidance to firms engaged in marketing or advertising about AI features of their products. Key considerations for advertisers include avoiding exaggerations about product

capabilities, having sufficient evidence for comparative claims, understanding foreseeable risks of AI products, and ensuring that the product genuinely uses AI. The Federal Trade Commission (FTC) will investigate and enforce regulations against deceptive advertising practices in this sector.

Canetti et al., Using Zero-Knowledge to Reconcile Law Enforcement Secrecy and Fair Trial Rights in Criminal Cases, In Proceedings of the 2022 Symposium on Computer Science and Law (2022). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4074315

There is a recurring conflict in criminal cases between preserving the secrecy of law enforcement's investigative tools and upholding defendants' rights to a fair trial. One possible solution to this challenge comes in the form of Zero-Knowledge Proofs (ZKPs), which are cryptographic tools that allow independent verification of software's structure and operation without revealing sensitive details. This technique could be a general solution for verification dilemmas across AI; it could enable audit and verification of use claims, while protecting companies' private data during the process.

Hartzog, *Unfair and Deceptive Robots*. Maryland Law Review 74, 785 (2015). Available at SSRN: https://ssrn.com/abstract=2602452

Consumer robotics (*including AI*) presents a number of issues related to consumer protection, privacy, data security, and vulnerability exploitation. The Federal Trade Commission (FTC) must play a pivotal role in regulation, given its authority and history of consumer protection. However, there must be a collaborative approach, involving other agencies and possibly a Federal Robotics Commission (*or Federal AI Commission*), to safeguard consumers while ensuring the industry thrives in the face of unique challenges presented by the complexity and social significance of consumer robotics.

Richards, Hartzog, and Francis, *Comments of the Cordell Institute on the Prevalence of Commercial Surveillance and Data Security Practices that Harm Consumers*. Available at SSRN: https://ssrn.com/abstract=4284020

Data privacy regulation in the US is insufficient because extensive commercial surveillance disproportionately benefits data-intensive firms, undermines consumer choice, and causes harm, including threats to our privacy and mental health. This article advocates for a shift in the FTC's approach to data privacy rules, urging it to prioritize trust, loyalty, and the relational vulnerability of consumers, through measures such as data minimization, limits on targeted advertising, and comprehensive age protections, to foster a sustainable, trustworthy digital marketplace.

Smart, Grimm, and Hartzog, *An Education Theory of Fault For Autonomous Systems*, Notre Dame Journal on Emerging Technologies 2, 33 (2021). Available at SSRN: https://ssrn.com/abstract=3854927

This paper proposes a novel approach to determining fault in the context of autonomous systems such as AI, which emphasizes the crucial role of effective communication, testing, and education among stakeholders (developers, procurers, and end-users). The authors outline four specific "education-failure" points (Syntactic, Semantic, Testing, and Warning) where lapses could occur

in the creation, deployment, and use of these systems, leading to harm, thus providing a basis for attributing liability in a tort law fault-based legal framework that could apply to AI.

Smith, Using Artificial Intelligence and Algorithms. Federal Trade Commission Bureau of Consumer Protection (2020). Available at: https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms

Advertisements must be transparent, fair, and accurate regarding the use of AI and algorithms in products, especially when they are used for decision-making about consumers. The FTC offers guidance to companies utilizing AI, requiring them to be honest about their use of automated tools, disclose data collection methods, explain algorithmic decisions to consumers, ensure fair decisions that do not discriminate against protected classes, validate data and models, and uphold standards for compliance, ethics, and nondiscrimination.

Spinelli, Crovella, *How YouTube Leads Privacy-Seeking Users Away from Reliable Information*, 28th ACM Conference on User Modeling, Adaptation and Personalization 244 (2020). Available at: https://doi.org/10.1145/3386392.3399566

This research investigates the nature of YouTube's recommendation system, showing that the platform typically directs users away from reliable sources and towards extreme or unreliable content, especially for privacy-seeking users. This "lead away" effect, which is strongest for YouTube users who choose privacy-protecting options, and its implications for user access to reliable information are critical concerns given the dominance of YouTube as an information source and its advertising-based business model. This research demonstrates the need for independent third-party interrogation of AI systems and suggests the need for governmental oversight to limit their harms.

Van Loo, *Digital Market Perfection*, Michigan Law Review 117, 815 (2019), Available at SSRN: https://ssrn.com/abstract=3308524

AI is increasingly serving as a digital intermediary in the market, with large companies like Apple, Google, and Microsoft developing systems to help consumers make informed buying decisions. However, there must be a reexamination of the market framework to address potential challenges, such as “hyperswitching,” which is rapid and large-scale customer migration between services or products that could destabilize industries and the economy. A comprehensive regulatory framework is necessary to manage the implications of automated commerce, bridge the gap between microeconomic and macroeconomic concerns, and incorporate trade and financial regulations.

Van Loo, Rory, *The Missing Regulatory State: Monitoring Businesses in an Age of Surveillance*. 72 Vanderbilt Law Review 1563 (2019), Available at SSRN: https://ssrn.com/abstract=3444341

This article proposes regulatory monitoring of technology platforms, summarizes the approaches taken in other industries such as financial and environmental regulation, and argues that the FTC has existing authority to develop a more substantial technology monitoring program. “If implemented, a monitoring program would initially enable learning to develop new regulatory

> standards and later provide a mechanism for adapting regulations to a fast-changing industry. In terms of enforcement, monitoring would serve to help regulators identify platforms' violations of broad existing laws, such as general consumer protection and antitrust statutes, as well as violations of any future platform-specific regulation."

Viljoen, *A Relational Theory of Data Governance*, Yale Law Journal 131, 370 (2021). Available at: https://www.yalelawjournal.org/feature/a-relational-theory-of-data-governance

> The current and proposed data-governance laws fall short in addressing the socioeconomic impact and normative importance of data relations. These laws generally view the datafication process from an individual perspective and fail to recognize that data production is primarily aimed at deriving population-level insights rather than individual-specific ones. The article argues for a shift in the conceptualization of data-governance, from individual to collective, and proposes an approach called "data as a democratic medium," which calls for democratic, rather than personal, institutional governance to address the socioeconomic issues that result from data production.