R Street
Free markets. Real solutions.

R Street Policy Study
No. 283
April 2023

Flexible, Pro-Innovation Governance Strategies for Artificial Intelligence

By Adam Thierer

Getting governance balance right—and ensuring that it remains flexible, responsive and pragmatic—is essential if the United States hopes to remain at the forefront of global AI innovation and competitiveness.

Executive Summary

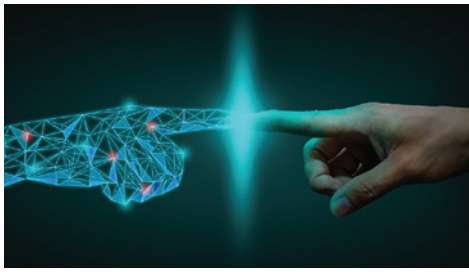
Policy interest in artificial intelligence (AI) and algorithmic systems continues to expand. Regulatory proposals are multiplying rapidly as academics and policymakers consider ways to achieve "AI alignment"—that is, to make sure that algorithmic systems promote human values and well-being. The process of embedding and aligning ethics in AI design is not static; it is an ongoing, iterative process influenced by many factors and values. It is therefore crucial that we build resiliency into algorithmic systems. The goal should be algorithmic risk mitigation—not elimination, which would be unrealistic. As we undertake this process, there will be much trial and error in creating ethical guidelines and finding better ways of keeping these systems aligned with human values. As a result, one-size-fits-all, top-down (i.e., regulatory-driven) mandates are unlikely to be workable or effective.

This article summarizes how more flexible, adaptive, bottom-up, less restrictive governance strategies can address algorithmic concerns and help ensure that AI innovation continues apace. Various organizations are already working to professionalize the process of AI ethics through sophisticated best-practice frameworks, algorithmic auditing and impact-assessment efforts. Multi-

Table of Contents

stakeholder efforts are helping to build consensus around these matters. These decentralized "soft-law" governance efforts build on existing hard law in many ways. Ex-post enforcement of existing laws and court-based remedies will provide an important backstop when AI developers fail to live up to their claims or promises about safe, effective and fair algorithms. Existing consumer protection laws and agency product recall authority will play a particularly important role in this regard.

Government can play an important role as a facilitator of ongoing dialogue and multi-stakeholder negotiations to address problems as they arise. The National Telecommunications and Information Administration (NTIA) and the National Institute of Standards and Technology (NIST), which have already done crucial work in this regard, can form a standing AI working group that brings parties together like this over time on an as-needed basis. Government actors can also facilitate digital literacy efforts and technology awareness-building, which can help lessen public fears about emerging algorithmic and robotic technologies.



Government actors can facilitate digital literacy efforts and technology awareness-building, which can help lessen public fears about emerging algorithmic and robotic technologies.

Introduction

AI and its governance have become topics of considerable public and political attention.[1] Regulatory proposals are multiplying rapidly with many media analysts, academics and politicians calling for interventions to address various algorithmic risks or potentially malicious uses.[2] Politicians have pitched the idea of robot taxes and a new federal agency—the Federal Automation and Worker Protection Agency—to "oversee automation and safeguard jobs and communities."[3] Several AI-related laws were introduced during the last session of Congress, including the Algorithmic Accountability Act, which would create a new federal office to oversee mandatory AI impact assessments.[4] Academics have also floated a variety of new laws like an Artificial Intelligence Development Act or a statute that would authorize the equivalent of "an FDA for algorithms."[5] Other proposals for a new oversight body include a Federal Robotics Commission, an AI Control Council, a National Algorithmic Technology Safety Administration, a National Technology Strategy Agency and even a new global regulatory body called the International Artificial Intelligence Organization.[6] Meanwhile, a variety of state and local measures are proposing different ways to regulate algorithmic systems.[7]

1. Henry A. Kissinger et al., *The Age of A.I.: And Our Human Future* (Little, Brown, 2021); Shira Ovide, "Why are we so afraid of AI?," *The Washington Post*, Feb. 24, 2023. https://www.washingtonpost.com/technology/2023/02/21/ai-polls-skeptics.
2. François Candelon et al., "AI Regulation Is Coming," *Harvard Business Review*, September–October 2021. https://hbr.org/2021/09/ai-regulation-is-coming.
3. Darren Orf, "Bernie Sanders Thinks Robots Should Pay Taxes. He's Right.," *Popular Mechanics*, Feb. 24, 2023. https://www.popularmechanics.com/technology/robots/a43046423/should-robots-pay-taxes-bernie-sanders; Bill de Blasio, "Why American Workers Need to Be Protected From Automation," *Wired*, Sept. 5, 2019. https://www.wired.com/story/why-american-workers-need-to-be-protected-from-automation.
4. H.R.6580, "Algorithmic Accountability Act of 2022," 117th Congress. https://www.congress.gov/bill/117th-congress/house-bill/6580.
5. Matthew U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies," *Harvard Journal of Law & Technology* 29:2 (Spring 2016), pp. 393-397. http://jolt.law.harvard.edu/articles/pdf/v29/29HarvJLTech353.pdf; Andrew Tutt, "An FDA for Algorithms," *Administrative Law Review* 69:1 (March 15, 2016). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994.
6. Ryan Calo, "The case for a federal robotics commission," Brookings, Sept. 15, 2014. https://www.brookings.edu/research/the-case-for-a-federal-robotics-commission; Anton Korinek, "Why we need a new agency to regulate advanced artificial intelligence: Lessons on AI control from the Facebook Files," Brookings, Dec. 8, 2021. https://www.brookings.edu/research/why-we-need-a-new-agency-to-regulate-advanced-artificial-intelligence-lessons-on-ai-control-from-the-facebook-files; Tutt. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994; Erica R.H. Fuchs, "What a National Technology Strategy Is—and Why the United States Needs One," *Issues in Science and Technology*, Sept. 9, 2021. https://issues.org/national-technology-strategy-agency-fuchs; Olivia J. Erdélyi and Judy Goldsmith, "Regulating Artificial Intelligence: Proposal for a Global Solution," *AIES '18: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (Dec. 27, 2018), pp. 95-101. https://dl.acm.org/doi/10.1145/3278721.3278731.
7. Neil Chilson and Adam Thierer, "The Coming Onslaught of 'Algorithmic Fairness' Regulations," Regulatory Transparency Project, Nov. 2, 2022. https://rtp.fedsoc.org/paper/the-coming-onslaught-of-algorithmic-fairness-regulations.

Earlier R Street Institute research identified some of the specific concerns driving these calls for algorithmic regulation.[8] Another R Street report contrasted different governance paradigms for technological systems and explained why highly precautionary and technocratic regulatory regimes for AI and machine learning (ML) are both unwise and impractical.[9]

Building on that research, this paper explains why more flexible governance strategies can address algorithmic concerns and help ensure that AI innovation continues apace. Although the precautionary principle is not the proper governance default for AI/ML, it can nonetheless help guide the governance of these technologies in a broader sense. Two general principles undergird many of the precautionary proposals around AI.[10] The first is the idea of "baking in" best practices and aligning AI design with widely shared goals and values. The second is the idea of keeping humans "in the loop" at critical stages of the algorithmic design process to ensure that they can continue to guide and occasionally realign those values and best practices as needed. These are wise principles, but they need not always be imposed in a highly regulatory, top-down fashion.

Nimble AI governance will be essential, as law lags behind the pace of technological change.

This paper also explains how it is possible to use flexible governance strategies to address various ethical concerns about AI to ensure that these technologies benefit humanity. Society can pursue this AI alignment without undermining advances in computational sciences or algorithmic innovation. The optimal governance approach for algorithmic systems should seek to establish certain best practices for development and use without foreclosing the important benefits associated with these technologies. Herein, we outline this type of agile and iterative approach to AI governance.

In addition, we describe how this flexible approach is already taking hold while more formal legislative and regulatory proposals continue to be stymied. Nimble AI governance will be essential, as law lags behind the pace of technological change. For example, government agencies are already behind in implementing the basic plans required by recent AI-related laws and presidential executive orders, and major technology legislative proposals have failed to pass in Congress—even when they enjoyed widespread support.[11] Experts note that "[f]ormal rulemaking is simply too time-consuming" for many emerging technology issues.[12] This inability to implement comprehensive technology legislation or regulation leads us to question whether we have strategies that can be put in place if more formal governance plans never get finalized.

8. Adam Thierer, "Mapping the AI Policy Landscape Circa 2023: Seven Major Fault Lines," R Street Institute, Feb. 9, 2023, https://www.rstreet.org/commentary/mapping-the-ai-policy-landscape-circa-2023-seven-major-fault-lines.
9. Adam Thierer, "Getting AI Innovation Culture Right," *R Street Policy Study* No. 281 (March 2023). https://www.rstreet.org/research/getting-ai-innovation-culture-right.
10. Benjamin Cedric Larsen, "Governing Artificial Intelligence: Lessons from the United States and China," Copenhagen Business School, 2022. https://research.cbs.dk/en/publications/governing-artificial-intelligence-lessons-from-the-united-states-.
11. Christie Lawrence et al., "Implementation Challenges to Three Pillars of America's AI Strategy," Stanford University Human-Centered Artificial Intelligence, December 2022. https://hai.stanford.edu/white-paper-implementation-challenges-three-pillars-americas-ai-strategy; Adam Thierer, "Governing Emerging Technology in an Age of Policy Fragmentation and Disequilibrium," American Enterprise Institute, April 2022. https://platforms.aei.org/can-the-knowledge-gap-between-regulators-and-innovators-be-narrowed.
12. Mark D. Fenwick et al., "Regulation Tomorrow: What Happens When Technology Is Faster than the Law?," *American University Business Law Review* 6:3 (2017), p. 572. https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1028&context=aublr.

This paper answers that question by identifying the decentralized soft-law governance techniques and existing regulatory authorities that are filling that governance gap. Although the decentralized governance techniques described herein can be amorphous, such iterative approaches are usually more in line with modern technological realities and policymaking needs; their application will contribute to the successful navigation of advances in AI. Algorithmic auditing and impact assessments are also emerging as leading governance mechanisms for AI. Although such assessments have a role, it is important that they not be imposed in a burdensome, inflexible fashion. Fortunately, there are ways to use those tools to help align values without disrupting important innovations.

Finally, this study explains what other steps governments can take to address algorithmic concerns. While some additional ex-ante regulatory constraints on algorithmic innovation may eventually become more necessary, it is sensible to use alternative legal and regulatory remedies that already exist before adding new rules and agencies. Many such solutions are available and can be adapted to algorithmic systems. One of the best roles for the government is to act as a facilitator of ongoing dialogue and a convener of multi-stakeholder discussions aimed at hammering out voluntary, consensus-driven best practices for algorithmic systems in an iterative fashion as problems develop. A case study is included to explore how these governance mechanisms are already being used for autonomous vehicles.

Policy interest in AI is multi-dimensional; lawmakers are interested in both controlling for risk and promoting the potential for algorithmic systems to advance global industrial competitiveness and geopolitical power.

Importantly, policy interest in AI is multi-dimensional; lawmakers are interested in both controlling for risk and promoting the potential for algorithmic systems to advance global industrial competitiveness and geopolitical power.[13] Policymakers also have a growing interest in countering China's expanding tech ambitions.[14] For example, a newly formed House Select Committee on the Strategic Competition Between the United States and the Chinese Communist Party is studying how the United States can better compete against China, especially on the high-tech front.[15] As policymakers examine these important issues, it is vital to consider how U.S. technology companies "currently face an erratic and often aggressive regulatory environment," due to both existing burdens and new legal threats.[16] Heavy-handed regulation of algorithmic systems would hurt the United States in terms of its global competitive standing relative to rivals like China and the many other countries vying to be the home of AI innovation.[17] The flexible, bottom-up governance strategy described in this paper can help the United States meet the challenge of global competition from China and other nations in cutting-edge emerging technology sectors while also addressing legitimate concerns about algorithmic systems.[18]

13. "Mid-Decade Challenges to National Competitiveness," Special Competitive Studies Project, September 2022. https://www.scsp.ai/reports/mid-decade-challenges-for-national-competitiveness.
14. Daitian Li et al., "Is China Emerging as the Global Leader in AI?," *Harvard Business Review*, Feb. 18, 2021. https://hbr.org/2021/02/is-china-emerging-as-the-global-leader-in-ai.
15. Deirdre Walsh and Barbara Sprunt, "Congress zeroes in on China — as economic and security threats loom," NPR, Feb. 28, 2023. https://www.npr.org/2023/02/28/1159132544/congress-zeroes-in-on-china-as-economic-and-security-threats-loom.
16. Adam J. White, "A Domestic Agenda for the House Select China Committee," *The Wall Street Journal*, Feb. 27, 2023. https://www.wsj.com/articles/a-domestic-agenda-for-the-china-committee-mike-gallagher-congress-strategic-competition-american-leadership-education-chips-semiconductors-rare-earth-minerals-692b421e.
17. Li et al. https://hbr.org/2021/02/is-china-emerging-as-the-global-leader-in-ai.
18. Adam Thierer, "A global clash of visions: The future of AI policy," *The Hill*, May 4, 2021. https://thehill.com/opinion/technology/551562-a-global-clash-of-visions-the-future-of-ai-policy.

Why Alternative Governance Approaches Are Needed for AI

The implicit premise of many academic papers and books about AI governance today is that the imposition of formal AI regulation is just a matter of time and political will. In reality, there are many practical reasons why AI governance will be much harder to implement than many advocates imagine.

To begin exploring this issue, it is important to recognize that the term technology governance can refer to more than just formal legislative and regulatory enactments. While such hard-law efforts are the leading form of governance for technology and many other things, they are not the only type. Many other forces and mechanisms beyond hard law can govern the development and use of emerging technologies. It is useful, therefore, to adopt a broader concept of governance in which the term includes an array of tools and solutions to address various ethical concerns and policy challenges.

Scholars refer to the governance issues surrounding emerging technologies as "wicked problems" for which "there is often no single, optimal solution." It is, therefore, important to consider "a collection of second-best strategies [that] intersect, coexist, and—in some ways—compete."

When considering governance approaches for emerging technologies, one scholar notes, "it is useful to speak not about a 'policy' but about the 'policy space.' Otherwise, there is a risk that the basket of policy alternatives and tools is conceived too narrowly."[19] This concept of a policy space "recognizes that oversight power and regulatory authority are not held within a single formal body, but may be dispersed—or shared—between any number of entities, both private and public, within the relevant space."[20] These other entities can include media entities, professional associations, standards bodies, activist watchdog groups, civil society organizations and various other stakeholders.

This broadened perspective on the policy space surrounding technological governance is particularly relevant when considering the challenges posed by highly disruptive technologies today.[21] Scholars refer to the governance issues surrounding emerging technologies as "wicked problems" for which "there is often no single, optimal solution [...] but rather a mix of substandard solutions that must 'satisfice.'"[22] It is, therefore, important to consider "a collection of second-best strategies [that] intersect, coexist, and—in some ways—compete."[23]

The relentless pace of technological change demands this sort of reconceptualization. Almost every discussion of technological governance today alludes to the challenge posed by the so-called pacing problem, which refers to the quickening pace of technological developments and the inability of governments to keep up with those changes.[24] Another name for the pacing problem is the

19. Richard D. Taylor, "Quantum Artificial Intelligence: A 'precautionary' U.S. approach?," *Telecommunications Policy* 44:6 (July 2020), p. 10. https://www.sciencedirect.com/science/article/abs/pii/S030859612030001X.
20. Ibid.
21. Araz Taeihagh et al., "Assessing the regulatory challenges of emerging disruptive technologies," *Regulation & Governance* 15:4 (October 2021), pp. 1009-1019. https://onlinelibrary.wiley.com/doi/full/10.1111/rego.12392.
22. Gary E. Marchant, "Governance of Emerging Technologies as a Wicked Problem," *Vanderbilt Law Review* 73:6 (Dec. 22, 2020), p. 1862. https://vanderbiltlawreview.org/lawreview/2020/12/governance-of-emerging-technologies-as-a-wicked-problem.
23. Ibid.
24. Adam Thierer, "The Pacing Problem and the Future of Technology Regulation," Mercatus Center at George Mason University, Aug. 8, 2018. https://www.mercatus.org/bridge/commentary/pacing-problem-and-future-technology-regulation.

law of disruption, which describes how "technology changes exponentially, but social, economic, and legal systems change incrementally."[25] Whatever one calls this problem, there is no denying that the phenomenon presents a fundamental challenge to the regulation of many modern technological systems—most especially digital and algorithmic systems where pure computer code lies at the heart of innovation.

There is no denying that the pacing problem presents a fundamental challenge to the regulation of many modern technological systems—most especially digital and algorithmic systems where pure computer code lies at the heart of innovation.

Pacing-problem scholars explain the concept in more detail:

> In contrast to this accelerating pace of technology, the legal frameworks that society relies on to regulate and manage emerging technologies have not evolved as rapidly, fueling concerns about a growing gap between the rate of technological change and management of that change through legal mechanisms.[26]

Even advocates of AI regulation admit that the pacing problem creates significant challenges for traditional regulatory regimes. A major AI study group organized by Stanford University concluded that "[c]urrent regulatory systems are already struggling to keep up with the demands of technological evolution, and AI will continue to strain existing processes and structures."[27]

Other scholars have identified how the pacing problem gives rise to an exponential gap or competency trap for policymakers because, just as quickly as they are coming to grips with new technological developments, other technologies are emerging.[28] "Formal rulemaking is simply too time-consuming," another expert observes, adding that "[t]he speed of product innovation makes it possible to bring a new product to market while formal rulemaking in the existing regulatory infrastructure, taking months and often years of regulatory procedure, is still dealing with the last product launch."[29] Thus, regulations designed to apply to a specific innovation could be outdated before they are even finalized.[30]

All these factors are particularly relevant when considering the fast-moving and global nature of AI markets. As two prominent AI scholars summarize:

> Regulatory strategies developed in the public sector operate on a time scale that is much slower than AI progress, and governments have limited public funds for investing in the regulatory innovation to keep up with the complexity of AI's evolution. AI also operates on a global scale that is misaligned with regulatory regimes organized on the basis of the nation state.[31]

AI is also becoming the "most important *general-purpose technology* of our era."[32] General-purpose technologies are intertwined with almost every other sector of

25. Larry Downes, *The Laws of Disruption: Harnessing the New Forces That Govern Life and Business in the Digital Age* (Basic Books, 2009), p. 2.
26. Gary E. Marchant, "The Growing Gap Between Emerging Technologies and the Law," in Gary E. Marchant et al., eds., *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Springer, 2011), p. 19.
27. "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report," Stanford University, September 2021, p. 42. http://ai100.stanford.edu/2021-report.
28. Azeem Azhar, *The Exponential Age: How Accelerating Technology is Transforming Business, Politics and Society* (Diversion Books, 2021); David Rejeski, "Public Policy on the Technological Frontier," in Gary E. Marchant et al., eds., *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Springer, 2011), p. 57.
29. Fenwick et al. https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1028&context=aublr.
30. Ibid.
31. Jack Clark and Gillian K. Hadfield, "Regulatory Markets for AI Safety," *Computers and Society* (Dec. 11, 2019). https://arxiv.org/abs/2001.00078.
32. Erik Brynjolfsson and Andrew McAfee, "The Business of Artificial Intelligence," *Harvard Business Review*, July 18, 2017. https://hbr.org/2017/07/the-business-of-artificial-intelligence.

the economy and used ubiquitously throughout society.[33] For example, almost all organizations will use AI to help improve analytics and marketing, enhance customer service and boost sales or performance in various new ways. AI will completely upend the way production and work is done in countless fields and professions. This is both what makes AI so important for future innovation and growth and what complicates its governance.[34]

Moreover, AI's definitional boundaries are amorphous and constantly expanding, and many technologies today build on top of one another in a symbiotic fashion (i.e., combinatorial innovation), further blurring the lines between formerly distinct technologies and sectors.[35] Consider how these definitional challenges are relevant to the governance of autonomous vehicle systems. On one hand, a driverless car is something quite new—essentially an AI-powered computer on wheels with many sophisticated technological sub-components, including powerful sensors and wireless communications capabilities. On the other hand, an autonomous vehicle is still an automobile—and automobiles already face many legacy regulations.[36] Thus, as vehicles become more sophisticated and incorporate a broader range of technologies, these advances will place enormous pressure on the hard-law regulatory scheme developed for the driving machines of an earlier era.

There is another driver of the pacing problem: public demand. Once the public gains access to new technological capabilities, they expect that more and better tools will follow.

There is another driver of the pacing problem: public demand. Once the public gains access to new technological capabilities, they expect that more and better tools will follow. Product development lifecycles are shrinking not only because innovators supply new and better goods and services, but also because the public expects them to be forthcoming. As experts explain, "Regulators cannot unwind the widespread commercial adoption of AI techniques," and "tools powered by [ML] are […] unlikely to be abandoned given consumer demand and the real welfare gains derived from them."[37] Even if one government seeks to clamp down on innovation, others will welcome it.[38] This is known as innovation arbitrage, a term that refers to the fact that innovators and their innovations often move to wherever they receive the most hospitable treatment.[39] "When the results come back and show that the economic and health benefits are tremendous," experts have argued, "the floodgates will open everywhere."[40]

This is another reason decentralized governance approaches are needed to ensure that the public can enjoy the life-enriching and even life-saving AI applications they will increasingly desire, while also working to ensure that those applications

33. Timothy F. Bresnahan and M. Trajtenberg, "General purpose technologies 'Engines of growth'?," *Journal of Econometrics* 65:1 (January 1995), pp. 83-108. https://www.sciencedirect.com/science/article/abs/pii/030440769401598T.
34. Nicholas Crafts, "Artificial intelligence as a general-purpose technology: an historical perspective," *Oxford Review of Economic Policy* 37:3 (Autumn 2021), pp. 521-536. https://academic.oup.com/oxrep/article/37/3/521/6374675.
35. Hal R. Varian, "Computer Mediated Transactions," *American Economic Review* 100:2 (May 2010), pp. 1-10. https://www.aeaweb.org/articles?id=10.1257/aer.100.2.1.
36. Rebecca Bellan, "Buckle up, autonomous vehicles finally get federal safety standards," TechCrunch, March 10, 2022. https://techcrunch.com/2022/03/10/nhtsa-first-autonomous-vehicle-occupant-safety-standards.
37. Mariano-Florentino Cuéllar and Aziz Z. Huq, "Artificially Intelligent Regulation," *Dædalus* 151:2 (May 1, 2022), p. 339. https://direct.mit.edu/daed/article/151/2/335/110625/Artificially-Intelligent-Regulation.
38. Garry Kasparov, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins* (PublicAffairs, 2017), p. 118.
39. James Pethokoukis, "Global Innovation Arbitrage and Driverless Cars," AEI, Aug. 23, 2016. https://www.aei.org/economics/global-innovation-arbitrage-and-driverless-cars.
40. Kasparov, p. 118.

are safe. Flexible, soft-law governance tools can also operate at the global scale required for innovation today.

Finally, traditional, hard-law mechanisms are also under strain because of a variety of other political realities.[41] Hyper-partisanship and general legislative dysfunction seem to be the new norm in Congress, frustrating efforts to advance broad-based legislation on many issues.[42] When combined with the pacing problem, this makes the prospect of hard-law enactments for AI issues even less likely.[43] Decentralized governance mechanisms and soft-law approaches will need to fill the vacuum out of necessity.

Decentralized Governance and Soft Law: Conceptions and Characteristics

Some scholars worry about the prospect of "self-regulation in a vacuum of government input" and wonder whether it "usurps the traditional role of public regulators."[44] While such concerns are understandable, the definitional issues and pacing problem challenges described above are driving the development of new governance mechanisms for many modern technology sectors. Traditional hard-law regulatory approaches tend to be more top-down driven and often lack flexibility. These older mechanisms focus on control and compliance with a strictly defined set of policies. Unfortunately, as a scholar on this topic explained, "the control paradigm is too limited to address all the issues that arise in the context of emerging technologies."[45] The problems with top-down, command-and-control regulation are well documented, and the World Economic Forum (WEF) argues that as new ideas, products and business models develop, prescriptive regulation can become obsolete quickly.[46]

The WEF has called upon governments to adopt more flexible and agile approaches to regulation that are better suited to an era of fast-paced technological change, noting that "[t]he 'regulate-and-forget' era has passed."

This is why the WEF has called upon governments to adopt more flexible and agile approaches to regulation that are better suited to an era of fast-paced technological change, noting that "[t]he 'regulate-and-forget' era has passed."[47] The WEF explains that "to grasp the opportunities and mitigate the risks from innovation and disruption, governments need to adopt an 'adapt-and-learn' approach instead."[48] They call this "agile regulation" and suggest that the goal should be to reconceptualize technological governance "as a cycle of continuous learning and adaptation as the technology develops."[49]

41. Cecilia Kang and Adam Satariano, "As A.I. Booms, Lawmakers Struggle to Understand the Technology," *The New York Times*, March 3, 2023. https://www.nytimes.com/2023/03/03/technology/artificial-intelligence-regulation-congress.html.
42. Drew Desilver, "The polarization in today's Congress has roots that go back decades," Pew Research Center, March 10, 2022. https://www.pewresearch.org/fact-tank/2022/03/10/the-polarization-in-todays-congress-has-roots-that-go-back-decades.
43. Thierer, "Governing Emerging Technology in an Age of Policy Fragmentation and Disequilibrium." https://platforms.aei.org/can-the-knowledge-gap-between-regulators-and-innovators-be-narrowed.
44. Michael Guihot et al., "Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence," *Vanderbilt Journal of Entertainment & Technology Law* 2:2 (Winter 2017), pp. 432, 434-435.
45. Marc A. Saner, "The Role of Adaptation in the Governance of Emerging Technologies," in Gary E. Marchant et al., eds., *Innovative Governance Models for Emerging Technologies* (Edward Elgar, 2014), p. 106.
46. "Agile Regulation for the Fourth Industrial Revolution: A Toolkit for Regulators," World Economic Forum, December 2020, p. 16. https://www.weforum.org/pages/agile-regulation-for-the-fourth-industrial-revolution-a-toolkit-for-regulators.
47. Ibid., p. 4.
48. Ibid., p. 11.
49. Ibid.

The touchstones of the new governance approaches tend to include flexibility, agility, adaptability, experimentation and decentralization. Governance experts at Deloitte have listed some of the many names these new approaches go by, including adaptive regulation, outcome-based regulation and sandboxing.[50] Others use terms like co-regulation, flexible regulation, policy prototyping and entrepreneurial administration.[51] There are subtle differences among these concepts, but they all share an approach to technological governance made up of many different elements and possible solutions—not all of which are regulatory or highly formal.

Even governance scholars who work within the growing intellectual movement known as responsible research and innovation (RRI) advocate for new decentralized governance approaches.[52] While many RRI scholars favor precautionary, hard-law solutions, there is a growing recognition among these scholars that decentralized and experimental governance approaches will need to be on the table when hard law fails, for whatever reason. Leading RRI scholars have documented "the emergence of new, more hybrid styles of governance" for a wide variety of tech sectors.[53] They highlight how, within these new schemes, "governance is considered […] as a learning process, less directed to direct intervention and 'decision-making', and more towards experimentation."[54] These authors identify a shift away from applying governance as a quick fix because clear and anticipated solutions no longer exist.[55]

RRI scholars highlight how "governance is considered […] as a learning process, less directed to direct intervention and 'decision-making', and more towards experimentation." They identify a shift away from applying governance as a quick fix because clear and anticipated solutions no longer exist.

This is why soft law is ascendant in emerging-technology policy circles today. While hard law includes formal statutory enactments and administrative promulgations, soft law is "a shorthand term to cover a variety of nonbinding norms and techniques for implementing them."[56] Scholars at the Arizona State University (ASU) School of Law have tracked and coordinated much of the cross-disciplinary research around soft-law governance. They explain in more detail what soft law entails and why it has quickly become a major trend in the field of emerging technology governance, especially for AI:

> Soft law is defined as a program that sets substantive expectations, but is not directly enforceable by government. Because soft law is not bound by a geographic jurisdiction and can be developed, amended, and adopted by any entity, it will be the dominant form of [AI] governance for the foreseeable future. […] Soft law is not a panacea or silver bullet. By itself, it is unable to solve all of the governance issues experienced by society due to AI. Nevertheless, whether by choice or necessity, soft law is and will continue to play a central role in the governance of AI for some time.[57]

50. William D. Eggers et al., "The future of regulation: Principles for regulating emerging technologies," Deloitte, June 19, 2018. https://www2.deloitte.com/insights/us/en/industry/public-sector/future-of-regulation/regulating-emerging-technology.html; Matt Perault and Andrew Keane Woods, "A Road Map for Tech Policy Experimentation," *Lawfare*, Aug. 12, 2022. https://www.lawfareblog.com/road-map-tech-policy-experimentation.
51. Philip J. Weiser, "Entrepreneurial Administration," *Boston University Law Review* 97 (2017), pp. 2011-2081. http://scholar.law.colorado.edu/articles/838.
52. Laurens Landeweerd et al., "Reflections on different governance styles in regulating science: a contribution to 'Responsible Research and Innovation,'" *Life Sciences, Society and Policy* 11:8 (August 2015). https://lsspjournal.biomedcentral.com/articles/10.1186/s40504-015-0026-y.
53. Ibid., p. 17.
54. Ibid.
55. Ibid.
56. Kenneth W. Abbott et al., "Soft Law Oversight Mechanisms for Nanotechnology," *Jurimetrics* 52:3 (Spring 2012), p. 285. https://www.jstor.org/stable/23240003.
57. Ibid.

It is easiest to think of soft law as a type of pragmatic governance rooted in incremental learning and ongoing improvement. Flexibility and adaptability are its core virtues. In this sense, soft law embodies what has been famously referred to as "the science of muddling through."[58] In 1959, this scholar observed that policymaking is a rough process and that policy "is not made once and for all; it is re-made endlessly."[59] He argued that policymakers should appreciate the benefits of incremental change and understand that policies will often only be partially successful while also producing some unintended consequences.[60]

This more incrementalist approach to governance has many benefits, allowing policymakers, firms and society to:

- Gain knowledge by testing predictions and policies before advancing to other steps
- Limit the damage that more sweeping policies might entail
- More easily remedy past errors once discovered[61]

It is easiest to think of soft law as a type of pragmatic governance rooted in incremental learning and ongoing improvement. Flexibility and adaptability are its core virtues.

Soft law embodies this mindset by encouraging even more outside-the-box and on-the-fly approaches to technology policy, including governance mechanisms of a non-regulatory and voluntary manner. It is an approach rooted in humility about the challenges surrounding emerging technologies and their governance. Technology scholars argue that, for these reasons, "we should not expect perfection, only partial success" when devising governance solutions.[62]

Compared with hard law, soft law has some obvious advantages that make it better suited for fast-moving technologies like AI. Soft-law scholars stress how it can be more rapidly and flexibly adapted to suit new circumstances, allowing for the level of agility necessary to address complex technological governance challenges.[63] Moreover, according to the ASU scholars, "unlike hard regulation adopted by regulatory authorities that are legally restricted to specific geographical jurisdictions, soft-law measures have no similar restrictions, and thus tend to be inherently international in scope," which is important when a technology is being developed and used globally, as is the case with AI.[64]

Finally, soft-law mechanisms can fill the gap while other more formal hard-law policies are being formulated and can help policymakers determine which types of hard law might work best when addressing specific concerns around emerging technologies like AI.

58. Charles E. Lindblom, "The Science of 'Muddling Through,'" *Public Administration Review* 19:2 (Spring 1959), pp. 79-88. https://www.jstor.org/stable/973677.
59. Ibid., p. 86.
60. Ibid.
61. Ibid.
62. Marchant, "Governance of Emerging Technologies as a Wicked Problem." https://vanderbiltlawreview.org/lawreview/2020/12/governance-of-emerging-technologies-as-a-wicked-problem.
63. Ryan Hagemann, "New Rules for New Frontiers: Regulating Emerging Technologies in an Era of Soft Law," *Washburn Law Journal* 57:2 (Spring 2018), p. 249. https://contentdm.washburnlaw.edu/digital/collection/wlj/id/7163.
64. Marchant et al., "Governing Emerging Technologies Through Soft Law: Lessons for Artificial Intelligence," *Jurimetrics* 61:1 (2020), p. 8. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3761871.

Soft-Law Methods and Current Applications

A diverse array of soft-law strategies exist, and the universe of soft-law tools and methods is constantly evolving. To reiterate, we need best practices for AI development free of the regulatory baggage that accompanies precautionary, principle-oriented efforts. More specifically, AI development needs to be guided by the principles of "ethics by design" and the concept of keeping "humans in the loop" to ensure that important values are protected. Luckily, many decentralized governance techniques already build upon the same set of principles that some want enshrined into hard-law edicts.

Scholars have noted that soft law is an amorphous term and that it is helpful to view it "as part of a continuum" of ever-changing governance options.

Scholars have noted that soft law is an amorphous term and that it is helpful to view it "as part of a continuum" of ever-changing governance options.[65] Some of the leading types of soft-law governance mechanisms include:

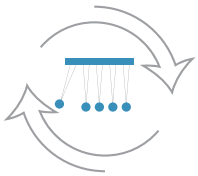
- **Multi-stakeholder processes**, in which various stakeholders are assembled (often by government bodies) to devise governance guidelines for a particular sector or technology
- **Agency guidance documents**, often developed through agency workshops and workshop reports
- **Informal consultations** between government and nongovernmental actors
- **"Sandboxes,"** or special trial-run approaches to alternative regulatory arrangements (which can also include geographically defined innovation zones)
- **Best practices and voluntary codes of conduct** (either for organizations or individual practitioners), often crafted through multi-stakeholder processes
- **Education and awareness-building efforts**, by both government and nongovernmental actors

Soft-Law Mechanisms: Governance Examples

Soft law can also include more market-driven activities or private-sector-led steps such as:

- **Insurance markets**, which serve as risk calibrators and correctional mechanisms
- **Third-party accreditation and standards-setting bodies**
- **Social norms and reputational effects**, especially the growing importance of reputational feedback mechanisms[66]
- **Societal pressure and advocacy** from media, academic institutions, nonprofit advocacy groups and the general public, all of which can put pressure on technology developers
- **Ongoing innovation and competition** within markets

Soft-Law Mechanisms: Market-Driven Activities or Private-Sector-Led Examples



Many federal agencies in the United States have been tapping new governance approaches to address novel questions raised by new technologies. The Federal Trade Commission (FTC), the NTIA, the Food and Drug Administration (FDA), the Department of Transportation (DOT) and the Federal Communications Commission

65. Kenneth W. Abbott et al., "Soft Law Oversight Mechanisms for Nanotechnology," *Jurimetrics* 52 (Fall 2012), p. 286. https://www.jstor.org/stable/23240003.
66. Adam Thierer et al., "How the Internet, the Sharing Economy, and Reputational Feedback Mechanisms Solve the 'Lemons Problem,'" *University of Miami Law Review* 70:3 (2016). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2610255.

(FCC) have all utilized soft-law mechanisms to address new technical challenges, including:

- "Big data" machine-learning[67]
- The "Internet of Things" (i.e., internet-enabled devices and applications)[68]
- Online advertising practices[69]
- Autonomous-vehicle (i.e., driverless car) policy[70]
- Motor vehicle cybersecurity[71]
- Cybersecurity of advanced medical devices[72]
- Facial recognition technologies[73]
- Health and medical smartphone applications[74]
- Medical advertising on social media platforms[75]
- Mobile phone privacy disclosures and mobile applications for children[76]
- 3D-printed medical devices[77]
- Small, unmanned aircraft systems (i.e., drones)[78]

Soft-Law Mechanisms: Federal Agency Examples

Soft-law approaches are often tailored to specific issues and risks that are evolving constantly, so the governance recommendations flowing out of these efforts can be quite detailed and context-specific. One common best practice recommended in many soft-law efforts involves devising appropriate data collection and storage procedures. Innovators are typically encouraged to use commonly accepted encryption techniques and ensure that data is handled properly; only used for clearly specified and sensible purposes; and deleted after a certain amount of time. For example, in the NHTSA's 2016 workshop and corresponding report on "Cybersecurity Best Practices for Modern Vehicles," the agency said, "[w]idely

67. "Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues (FTC Report)," Federal Trade Commission, January 2016. https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report; "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," Executive Office of the President, May 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.
68. "internet of things: Privacy & Security in a Connected World," Federal Trade Commission, January 2015. https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf; "Careful Connections: Keeping the Internet of Things Secure," Federal Trade Commission, January 2015. https://www.ftc.gov/system/files/documents/plain-language/913a_careful_connections.pdf.
69. "Native Advertising: A Guide for Businesses," Federal Trade Commission, last accessed March 3, 2023. https://www.ftc.gov/tips-advice/business-center/guidance/native-advertising-guide-businesses.
70. "Federal Automated Vehicles Policy: Accelerating the Next Revolution In Roadway Safety," U.S. Department of Transportation, September 2016. https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf.
71. "Cybersecurity Best Practices for Modern Vehicles," U.S. Department of Transportation, October 2016. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/812333_cybersecurityformodernvehicles.pdf.
72. "Postmarket Management of Cybersecurity in Medical Devices," U.S. Food & Drug Administration, December 2016. https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm482022.pdf.
73. "Privacy Best Practice Recommendations For Commercial Facial Recognition Use," National Telecommunications and Information Administration, last accessed March 3, 2023. https://www.ntia.doc.gov/files/ntia/publications/privacy_best_practices_recommendations_for_commercial_use_of_facial_recogntion.pdf.
74. "Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff," U.S. Food & Drug Administration, Feb. 9, 2015. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf.
75. "Internet/Social Media Platforms with Character Space Limitations— Presenting Risk and Benefit Information for Prescription Drugs and Medical Devices," U.S. Food & Drug Administration, June 2014. https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-drugs-gen/documents/document/ucm401087.pdf.
76. "Mobile Privacy Disclosures: Building Trust Through Transparency: A Federal Trade Commission Staff Report," Federal Trade Commission (February 2013). https://www.ftc.gov/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission; "Mobile Apps for Kids: Disclosures Still Not Making the Grade," Federal Trade Commission, December 2012. https://www.ftc.gov/sites/default/files/documents/reports/mobile-apps-kids-disclosures-still-not-making-grade/121210mobilekidsappreport.pdf.
77. "Technical Considerations for Additive Manufactured Medical Devices: Guidance for Industry and Food and Drug Administration Staff," U.S. Food & Drug Administration, December 2017. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM499809.pdf.
78. Federal Aviation Administration, "Small Unmanned Aircraft Systems (sUAS)," U.S. Department of Transportation, June 21, 2016. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_107-2.pdf; Federal Aviation Administration, "UAS Integration Pilot Program," U.S. Department of Transportation, June 3, 2022. https://www.faa.gov/uas/programs_partnerships/completed/integration_pilot_program.

accepted encryption methods should be employed in any IP-based operational communication between external servers and the vehicle."[79] In some cases, technical specifications and procedures are worked out during multi-stakeholder negotiations, often assisted by governmental bodies. For example, with mobile phone privacy disclosures and mobile applications for children, the NTIA and FTC used multi-stakeholder processes to push for stronger developer privacy codes of conduct. Other times, the process of hammering out best practices is left to industry bodies or third-party accreditors to address and enforce.

Although some consider soft law's informality and amorphous nature to be a weakness, that is also its primary strength. Soft law is particularly well suited to address governance issues in fast-evolving sectors like AI in which "there is a growing consensus that traditional government regulation is not sufficient for the oversight of emerging technologies" because hard-law mechanisms either cannot keep pace with technological developments or are simply too inflexible to accommodate new realities.[80]

Although some consider soft law's informality and amorphous nature to be a weakness, that is also its primary strength. Soft law is particularly well suited to address governance issues in fast-evolving sectors like AI.

Much of the academic scholarship surrounding AI governance either ignores soft-law efforts or belittles their importance, typically due to a preference for more aggressive, hard-law proposals of a precautionary, principle-based orientation. For many of these scholars and various AI critics, nothing short of a comprehensive federal (or even international) law and corresponding regulatory regime will suffice.[81]

Excessive preemptive regulation would greatly limit beneficial AI innovations.[82] It is also shortsighted because it ignores the practical challenges associated with attempts to slow rapidly evolving and fully global technologies like AI and ML.

The Growth of AI Ethical Codes and Best-Practice Frameworks

A recent AI report from a top university noted that one of the most important trends in the field of algorithmic governance was "the rise of AI ethics everywhere."[83] The report summarized the explosive growth of ethical frameworks and guidelines for AI that has been occurring throughout academia and industry:

> Research on fairness and transparency in AI has exploded since 2014, with a fivefold increase in related publications at ethics-related conferences. Algorithmic fairness and bias has shifted from being primarily an academic pursuit to becoming firmly embedded as a mainstream research topic with wide-ranging implications. Researchers with industry affiliations contributed 71% more publications year over year at ethics-focused conferences in recent years.[84]

79. "Cybersecurity Best Practices for Modern Vehicles," p. 20. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/812333_cybersecurityformodernvehicles.pdf.
80. Wendell Wallach and Gary Marchant, "Toward the Agile and Comprehensive International Governance of AI and Robotics," *Proceedings of the IEEE* 107:3 (March 2019), p. 506. https://ieeexplore.ieee.org/document/8662741.
81. John Frank Weaver, "We Need to Pass Legislation on Artificial Intelligence Early and Often," *Slate*, Sept. 12, 2014. https://slate.com/technology/2014/09/we-need-to-pass-artificial-intelligence-laws-early-and-often.html.
82. Thierer, "Getting AI Innovation Culture Right." https://www.rstreet.org/research/getting-ai-innovation-culture-right.
83. "Measuring trends in Artificial Intelligence," Stanford University Human-Centered Artificial Intelligence, 2022, p. 105. https://aiindex.stanford.edu/report.
84. Ibid.

Academic researchers who aim to analyze and classify the resulting ethical recommendations are closely studying this "avalanche of initiatives and policy documents" around AI ethics.[85] A 2019 survey by a group of researchers based in Switzerland analyzed 84 AI ethical frameworks and found "a global convergence emerging around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy)," noting that there were differences in which of these values were most important and how each of them should be interpreted and implemented.[86] The authors explained that, even with those limitations, these ethical frameworks and soft-law governance approaches "are aimed at assisting with—and have been observed to have significant practical influence on—decision making in certain fields, comparable to that of legislative norms."[87]

In 2021, a team of ASU legal scholars published the most comprehensive survey of soft-law efforts for AI to date.[88] They analyzed 634 soft-law AI programs that were formulated between 2016 and 2019. More than one-third of these efforts were initiated by governments, with the others being led by nonprofits or private-sector bodies. Echoing the findings from the Swiss researchers, the ASU report found widespread consensus among these soft-law frameworks on values such as transparency and explainability, ethics/rights, security and bias. This makes it clear that considerable consistency exists among ethical soft-law frameworks in that most of them focus on a core set of values to embed within AI design. The Alan Turing Institute boils their list down to four "FAST Track Principles": fairness, accountability, sustainability and transparency.[89]

An ASU report found widespread consensus among soft-law frameworks on values such as transparency and explainability, ethics/rights, security and bias. This makes it clear that considerable consistency exists among ethical soft-law frameworks in that most of them focus on a core set of values to embed within AI design.

The scholars noted how ethical best practices for product design already influence developers by creating powerful norms and expectations about responsible product design, noting that "[o]nce a soft law program is created, organizations may seek to enforce it by altering how their employees or representatives perform their duties through the creation and implementation of internal procedures."[90] They point out that "[p]ublicly committing to a course of action is a signal to society that generates expectations about an organization's future actions."[91]

This is important because many major trade associations and individual companies have been formulating governance frameworks and ethical guidelines for AI development and use. For example, among large trade associations, the U.S. Chamber of Commerce, the Business Roundtable, the BSA | The Software Alliance and ACT | The App Association have all recently released major AI best practice

85. Mark Coeckelbergh, *AI Ethics* (MIT Press, 2020), p. 148.
86. Anna Jobin et al., "The global landscape of AI ethics guidelines," *Nature Machine Intelligence* 1 (Sept. 2, 2019), pp. 389-399. https://www.nature.com/articles/s42256-019-0088-2.
87. Ibid., p. 389.
88. Gutierrez and Marchant. https://lsi.asulaw.org/softlaw/wp-content/uploads/sites/7/2022/08/final-database-report-002-compressed.pdf.
89. David Leslie, "Understanding Artificial Intelligence Ethics and Safety," The Alan Turing Institute, 2019. https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety.
90. Ibid., p. 17.
91. Ibid., p. 18.

guidelines.[92] Notable corporate efforts to adopt guidelines for ethical AI practices include statements or frameworks by Amazon, IBM, Intel, Google, Microsoft, Salesforce, SAP and Sony.[93] There is remarkable consistency across these corporate statements in terms of the best practices and ethical guidelines they endorse. The guidelines from these trade associations or corporations align closely with the core values identified in the hundreds of other soft-law frameworks that ASU scholars surveyed. These efforts go a long way toward helping to promote a culture of responsibility among leading AI innovators.[94]

Of course, more work remains to be done, especially by smaller developers. A 2022 survey of 225 AI startups found that 58 percent of them have established a set of AI principles.[95] The authors of the report argue that "it is apparent that many AI startups are aware of possible ethical issues" and that many are taking steps to address them proactively.[96] Yet more efforts are needed to ensure that other AI providers are adopting ethical guidelines and best practices, especially as calls for formal regulation grow louder.

Notable corporate efforts to adopt guidelines for ethical AI practices exist, but more work is needed.

A 2022 survey of

225

AI startups found that 58 percent of them have established a set of AI principles.

With the ethical frameworks coalescing around a core set of widely accepted principles, the next stage of AI soft-law governance will involve efforts to formalize their implementation. As the Swiss team of AI researchers noted, "[a]t the policy level, greater interstakeholder cooperation is needed to mutually align different AI ethics agendas and to seek procedural convergence not only on ethical principles but also their implementation."[97] (The mechanics of implementation will be discussed later in this paper.)

The best hope for scaling up ethical principles on a more widespread basis lies in the crucial work done by professional organizations and standards bodies such as the Association of Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), the International Organization for Standardization (ISO) and UL (previously known as Underwriters Laboratories).[98] Such organizations serve

92. "U.S. Chamber Releases Artificial Intelligence Principles," U.S. Chamber of Commerce, Sept. 23, 2019. https://www.uschamber.com/technology/us-chamber-releases-artificial-intelligence-principles; "Artificial Intelligence," Business Roundtable, last accessed March 3, 2022. https://www.businessroundtable.org/policy-perspectives/technology/ai; "BSA Releases Framework to Confront Bias in Artificial Intelligence and Calls for Legislation," BSA | The Software Alliance, June 8, 2021. https://www.bsa.org/news-events/news/bsa-releases-framework-to-confront-bias-in-artificial-intelligence-and-calls-for-legislation; "ACT | The App Association's Policy Principles for Artificial Intelligence," ACT | The App Association, last accessed March 3, 2023. https://www.nist.gov/document/act-app-associations-policy-principles-artificial-intelligence-online-submission.
93. "Responsible use of artificial intelligence and machine learning," AWS, last accessed March 3, 2023. https://aws.amazon.com/machine-learning/responsible-machine-learning; "Precision Regulation for Artificial Intelligence," IBM, Jan. 21, 2020. https://www.ibm.com/policy/ai-precision-regulation; David Hoffman and Riccardo Masucci, "Intel's AI Privacy Policy White Paper: Protecting individuals' privacy and data in the artificial intelligence world," Intel, 2018. https://www.intel.com/content/dam/www/public/us/en/ai/documents/Intels-AI-Privacy-Policy-White-Paper-2018.pdf; "Responsible AI practices," Google AI, last accessed March 3, 2023. https://ai.google/responsibilities/responsible-ai-practices; "Introducing the Model Card Toolkit for Easier Model Transparency Reporting," Google Research, July 29, 2020. https://ai.googleblog.com/2020/07/introducing-model-card-toolkit-for.html; "Putting principles into practice at Microsoft," Microsoft, last accessed March 3, 2023. https://www.microsoft.com/en-us/ai/our-approach; "Salesforce Debuts AI Ethics Model: How Ethical Practices Further Responsible Artificial Intelligence," Salesforce, Sept. 2, 2021. https://www.salesforce.com/news/stories/salesforce-debuts-ai-ethics-model-how-ethical-practices-further-responsible-artificial-intelligence; Kathy Baxter, "AI Ethics Maturity Model," Salesforce, last accessed March 3, 2023. https://www.salesforceairesearch.com/static/ethics/EthicalAIMaturityModel.pdf; "SAP's Guiding Principles for Artificial Intelligence," SAP, Sept. 18, 2018. https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence; "AI Engagement within Sony Group," Sony Group, Sept. 25, 2018. https://www.sony.net/SonyInfo/csr_report/humanrights/AI_Engagement_within_Sony_Group.pdf.
94. Miles Brundage et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," Future of Humanity Institute, February 2018, p. 56. https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.
95. James Bessen et al., "Ethical AI development: Evidence from AI startups," Brookings, March 29, 2022. https://www.brookings.edu/research/ethical-ai-development-evidence-from-ai-startups.
96. Ibid.
97. Jobin et al. https://www.nature.com/articles/s42256-019-0088-2.
98. "Request for Information to the Update of the National Artificial Intelligence Research and Development Strategic Plan: Responses," Association of Computing Machinery, March 4, 2022. https://www.ai.gov/rfi/2022/87-FR-5876/NAIRDSP-RFI-2022-Eisgrau-ACM.pdf; "Presenting the Standard for Safety for the Evaluation of Autonomous Vehicles and Other Products," UL Standards & Engagement, last accessed March 3, 2023. https://ul.org/UL4600.

as independent standards-creation bodies and help hold innovators accountable by designing guidelines and best practices that have been established through soft-law processes. Industry trade associations, such as the Consumer Technology Association, also develop industry-wide standards for AI technologies.[99] Analysts note that the general U.S. system of voluntary consensus standards "has been exceptionally successful in generating technological innovation in the United States."[100]

The work of the ISO, IEEE and ACM deserves greater attention because these three organizations have labored to create detailed international standards for AI and ML development. These organizations possess enormous sway in professional circles, and the employees of most major technology companies have some sort of membership in them—or at least work closely with them to create international standards in various technology fields.

ISO

"[A]n independent, non-governmental international organization with a membership of **163 national standards bodies**"

The ISO is one of the oldest global standard-making bodies. Formed in 1946, the ISO "is an independent, non-governmental international organization with a membership of 163 national standards bodies" that seeks to build global consensus through multi-stakeholder efforts.[101] Through this work, the ISO plays an important role in establishing international norms for emerging technologies. The organization convenes dozens of technical committees that include global experts in diverse fields, such as industry, consumer associations, academia, nongovernmental organizations and governments.[102] It has already played an important role in formulating global best practices for robotics and AI-based applications. In 2014, for example, the ISO crafted requirements and guidelines "for the inherently safe design, protective measures, and information for use of personal care robots."[103] That standard is just one of dozens of robotics-related guides that the ISO has published.[104] The organization also has a suite of standards governing a wide variety of AI, including a particularly detailed set of guidelines for AI risk management.[105] The ISO has also issued other guidance standards for information data security that are relevant to AI systems development.[106]

IEEE

420,000 members in 160+ countries

"[T]he world's largest technical professional organization dedicated to advancing technology for the benefit of humanity."

With more than 420,000 members in more than 160 countries, the IEEE boasts that it is "the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity."[107] Over the past several years, the IEEE worked to finalize a massive *Ethically Aligned Design* project is an effort to craft "A

99. "Artificial Intelligence," Consumer Technology Association, last accessed March 3, 2023. https://www.cta.tech/Topics/Artificial-Intelligence.
100. Hodan Omaar, "U.S. AI Policy Report Card," Center for Data Innovation, July 27, 2022. https://datainnovation.org/2022/07/ai-policy-report-card.
101. "About us," ISO, last accessed March 3, 2023. http://www.iso.org/iso/home/about.htm.
102. "Developing standards," ISO, last accessed March 3, 2023. http://www.iso.org/iso/home/standards_development.htm.
103. "ISO 13482:2014: Robots and robotic devices—Safety requirements for personal care robots," ISO, February 2014. https://www.iso.org/standard/53820.html.
104. "Standards by ISO/TC 299: Robotics," ISO, last accessed March 3, 2023. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_browse.htm?commid=5915511.
105. "ISO/IEC JTC 1/SC 42 Artificial intelligence," ISO, 2017. https://www.iso.org/committee/6794475.html; "ISO/IEC DIS 23894:2023: Information technology — Artificial intelligence — Guidance on risk management," ISO, February 2023. https://www.iso.org/standard/77304.html.
106. "ISO 27001 – Information Security," IMSM, last accessed March 3, 2023. https://www.imsm.com/us/iso-27001.
107. Ibid., p. 5.

Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems."[108] The IEEE's new effort seeks to incorporate five key principles into AI design that involve the protection of human rights, better wellbeing metrics, designer accountability, systems transparency and efforts to minimize the misuse of these technologies. The second iteration of the group's report was 263 pages and contained a suite of standards to satisfy each of those objectives.[109] The IEEE also continues to oversee an Organizational Governance of Artificial Intelligence working group to formulate standards and best practices for the development or use of AI within global organizations.

ACM

The ACM developed a *Code of Ethics and Professional Conduct* in the early 1970s, refined it in the early 1990s and then updated it again in 2018.[110] Each iteration of the ACM *Code* has reflected ongoing technological developments from the mainframe era to the PC and internet revolution and on through today's ML and AI era. The latest version of the ACM *Code* "affirms an obligation of computing professionals, both individually and collectively, to use their skills for the benefit of society, its members, and the environment surrounding them," and insists that computing professionals "should consider whether the results of their efforts will respect diversity, will be used in socially responsible ways, will meet social needs, and will be broadly accessible."[111] The *Code* also stresses how "[a]n essential aim of computing professionals is to minimize negative consequences of computing, including threats to health, safety, personal security and privacy. When the interests of multiple groups conflict, the needs of those less advantaged should be given increased attention and priority."[112]

Code of Ethics

ACM develops a Code of Ethics and Professional Conduct that "affirms an obligation of computing professionals, both individually and collectively, to use their skills for the benefit of society, its members, and the environment surrounding them."

Others

Many other academic institutions and international organizations play an important watchdog role by formulating AI ethical development guidelines and holding private developers accountable for the commitments they make through various soft-law frameworks. Some of the more notable efforts include:

- The Markkula Center for Applied Ethics at Santa Clara University produces "An Ethical Toolkit for Engineering/Design Practice," with a seven-step process for tech developers to follow when considering how to mitigate risks associated with new products.[113] The Markkula Center also partnered with the WEF and Deloitte to produce a white paper titled "Ethics by Design."[114]
- To focus on ethical AI in the fintech sector, experts at The Wharton School at The University of Pennsylvania created an Artificial Intelligence/Machine Learning

108. "Ethically Aligned Design," IEEE, last accessed March 3, 2023. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf.
109. "Autonomous and Intelligent Systems (AIS)," IEEE, last accessed March 3, 2023. https://ethicsinaction.ieee.org/p7000.
110. "ACM Code of Ethics and Professional Conduct," Association for Computing Machinery, 2018. https://www.acm.org/code-of-ethics.
111. Ibid.
112. Ibid.
113. Shannon Vallor et al., "An Ethical Toolkit for Engineering/Design Practice," Markkula Center for Applied Ethics at Santa Clara University, June 22, 2018. https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit.
114. "Ethics by Design: An organizational approach to responsible use of technology," World Economic Forum, Dec. 10, 2020. https://www.weforum.org/whitepapers/ethics-by-design-an-organizational-approach-to-responsible-use-of-technology.

Risk & Security Working Group, "to promote, educate, and advance AI/ML governance for the financial services industry by focusing on risk identification, categorization, and mitigation."[115]

- The Partnership on AI began as an industry-led effort formed by Apple, Amazon, Google, Facebook, IBM and Microsoft, but it has grown to include more than 100 members, including the American Civil Liberties Union and Human Rights Watch. The Partnership is billed as a multi-stakeholder organization that brings those diverse groups together "to study and formulate best practices on AI, to advance the public's understanding of AI, and to provide a platform for open collaboration between all those involved in, and affected by, the development and deployment of AI technologies."[116]
- OpenAI is a nonprofit research organization created in 2015 with seed money from notable tech innovators and investors like Elon Musk of Tesla, Sam Altman of Y Combinator, venture capitalist Peter Thiel, Reid Hoffman of LinkedIn and others. In addition to developing important algorithmic applications such as ChatGPT, OpenAI publishes research reports discussing how to make sure AI development "is used for the benefit of all, and to avoid enabling uses of AI or (artificial general intelligence) that harm humanity" and to ensure that it does not become "a competitive race without time for adequate safety precautions."[117] OpenAI is also a member of the Partnership on AI.
- The UL has produced many different standards in the area of AI, including its ANSI/UL 4600 "Standard for Safety for the Evaluation of Autonomous Products."[118] Similarly, in the United Kingdom, the British Standards Institution published a "Guide to the Ethical Design and Application of Robots and Robotic Systems" in 2016.[119] Developed by a committee of scientists, academics, ethicists and philosophers, the guide "recognizes that potential ethical hazards arise from the growing number of robots and autonomous systems being used in everyday life" and aims to "eliminate or reduce the risks associated with these ethical hazards to an acceptable level."[120] Specifically, protective measures create best practices for the safe design and use of robotic applications in a wide range of fields, from industrial services to personal care to medical services.[121]
- Additional noteworthy AI ethics groups, programs and efforts include: AI Now, Anthropic, Future of Life Institute, Future of Humanity, Center for Human-Compatible AI at UC Berkeley, the Centre for the Governance of AI at Oxford, and the Leverhulme Centre for the Future of Intelligence.

Many other academic institutions and international organizations play an important watchdog role by both formulating AI ethical development guidelines and holding private developers accountable.

115. Artificial Intelligence/Machine Learning Risk & Security Working Group (AIRS), "Artificial Intelligence Risk & Governance," University of Pennsylvania, last accessed March 3, 2023. https://ai.wharton.upenn.edu/artificial-intelligence-risk-governance.
116. "Building a Community of Practice: Reflections from our 2nd All Partners Meeting," Partnership on AI, Nov. 21, 2018. https://partnershiponai.org/building-a-community-of-practice-reflections-from-our-2nd-all-partners-meeting.
117. "OpenAI Charter," OpenAI, last accessed Feb. 4, 2019. https://openai.com/charter.
118. "Artificial Intelligence Risk Management Framework [Docket Number: 210726-0151]," U.S. National Institute of Standards and Technology, Aug. 19, 2021. https://www.nist.gov/document/ai-rmf-rfi-comments-underwriters-laboratories.
119. Hannah Devlin, "Do no harm, don't discriminate: official guidance issued on robot ethics," *The Guardian*, Sept. 18, 2016. https://www.theguardian.com/technology/2016/sep/18/official-guidance-robot-ethics-british-standards-institute.
120. "BS 8611:2016: Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems," European Standards, April 30, 2016. https://www.en-standard.eu/bs-8611-2016-robots-and-robotic-devices-guide-to-the-ethical-design-and-application-of-robots-and-robotic-systems.
121. Ibid.

How the Embedding of AI Ethics Works in Practice, and How It Could Be Improved

Efforts such as these can go a long way toward improving accountability and responsibility among various emerging technology companies and individual innovators. Standards, codes, ethical guidelines and multi-stakeholder collaborations create powerful social norms and expectations that are often equal to or even more important than what laws and regulations might accomplish.[122] Powerful reputational factors are at work in every sector that—when combined with efforts such as these—create a baseline of accepted practice. These efforts are also likely to get more initial buy-in among private innovators, at least compared to heavy-handed regulatory proposals, which could undermine new business models. Finally, these efforts deserve more attention if for no other reason than the continuing reality of the pacing problem. Soft-law mechanisms will always be easier to adopt and adapt as new circumstances demand.

Powerful reputational factors are at work in every sector that create a baseline of accepted practice. These efforts are also likely to get more initial buy-in among private innovators, at least compared to heavy-handed regulatory proposals, which could undermine new business models.

For codes of conduct, voluntary standards and professional ethical codes to have a lasting impact, however, additional steps are needed. The ASU scholars mentioned earlier argue that "[i]t is not enough to just have AI companies sign onto a list of ethical principles […] Rather, these principles must be operationalized into effective practices and credible assurances."[123] This need for "transitioning from ideas to action" represents the major challenge for soft law and decentralized governance efforts going forward.[124]

The first phase of AI soft-law development has been aspirational and focused on the formulation of values and best practices by soft-law scholars, government officials, industry professionals and various other stakeholder groups. Currently and in years to come, the focus will increasingly shift to the implementation and enforcement of these values and best practices. The ultimate success of soft-law mechanisms as a governance tool for AI will come down to how well aspirational goals like "baking in" certain key values and keeping humans "in the loop" are translated into concrete development practices.

There are other ways to conceptualize this process of AI alignment. AI experts increasingly talk about the importance of transfer learning when thinking about how to improve ML techniques and develop more sophisticated AI systems.[125] Transfer learning refers to "the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned."[126] Through transfer-learning techniques, algorithms are trained to reference and learn from related datasets and processes to achieve superior outcomes in a different domain. Human programmers oversee the process and constantly look to refine and improve those systems.

122. Gregory N. Mandel, "Regulating Emerging Technologies," *Law, Innovation and Technology* 1:1 (May 1, 2015), pp. 75-92. https://www.tandfonline.com/doi/abs/10.1080/17579961.2009.11428365.
123. Marchant et al. "Governing Emerging Technologies through Soft Law: Lessons for Artificial Intelligence—An Introduction." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3761871.
124. Carlos Ignacio Gutierrez, "Transitioning From Ideas to Action: Trends in the Enforcement of Soft Law for the Governance of Artificial Intelligence," *IEEE Transactions on Technology and Society* 2:4 (December 2021). https://ieeexplore.ieee.org/abstract/document/9548780.
125. Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (Farrar, Straus and Giroux, 2019), p. 166.
126. Emilio Soria Olivas et al., *Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques* (Information Science Reference, 2009), p. 242.

This is also a useful way to think about how to embed and align ethics. We essentially need the equivalent of transfer learning for ethical principles within AI systems as they evolve such that important values and principles are embedded at each step of the process. Optimally, as algorithms and AI systems learn and develop new capabilities, the goal should be to ensure that the same guiding principles we have attempted to "bake in" remain and are extended. If AI systems can gain greater capacity to transfer and use the knowledge they have learned from one task or application to another, by extension, they should be able to transfer and apply ethical principles and guidelines they have learned from one task or application to another. Of course, human operators still need to be "in the loop" to correct for inevitable errors along the way. This does not mean the process is foolproof; both machines and humans will err.[127] Moreover, as already noted, sometimes important values and best practices will conflict with other values and will need to be balanced in ways that will upset some policymakers or stakeholders. Nonetheless, the general framework of trained learning for AI ethics remains valuable.

As algorithms and AI systems learn and develop new capabilities, the goal should be to ensure that the same guiding principles we have attempted to "bake in" remain and are extended. Of course, human operators still need to be "in the loop" to correct for inevitable errors along the way.

Iterative amplification is another way of thinking about how to improve AI systems over time. The leader of the Alignment Research Center, a nonprofit research organization whose mission is to align future algorithmic systems with human interests, frames iterative amplification as:

> The idea in iterative amplification is to start from a weak AI. At the beginning of training you can use a human. A human is smarter than your AI, so they can train the system. As the AI acquires capabilities that are comparable to those of a human, then the human can use the AI that they're currently training as an assistant, to help them act as a more competent overseer.
>
> Over the course of training, you have this AI that's getting more and more competent, the human at every point in time uses several copies of the current AI as assistants, to help them make smarter decisions. And the hope is that that process both preserves alignment and allows this overseer to always be smarter than the AI they're trying to train.[128]

The hope here is that, "as you move along the training, by the end of training, the human's role becomes kind of minimal" and "at each step it remains aligned. You put together a few copies of the AI to act as an overseer for itself."[129] When we think about iterative amplification as a governance strategy, the general goal is the same one stressed repeatedly above: baking important values into AI development and keeping humans in the loop along the way to refine and improve the alignment process until it becomes safer and more useful.

Taken together, transfer learning and iterative amplification are essentially forms of learning by doing. It is a mistake to think of AI safety or algorithmic ethics as a static phenomenon that has a single solution or final destination. Incessant and unexpected change is the new normal. That means that many different strategies and much ongoing experimentation will be needed to address the many

127. Lorrie Faith Cranor, "A Framework for Reasoning About the Human in the Loop," Carnegie Mellon University, 2008, pp. 1-15. https://perma.cc/JA53-8AL8.
128. Robert Wiblin and Keiran Harris, "Dr. Paul Christiano on how OpenAI is developing real solutions to the 'AI alignment problem', and his vision of how humanity will progressively hand over decision-making to AI systems," 80,000 Hours, Oct. 2, 2018. https://80000hours.org/podcast/episodes/paul-christiano-ai-alignment-solutions.
129. Ibid.

challenges we must confront today and in the future. The goal is to assess and prioritize risks continuously and then formulate and reformulate our response toolkit to those risks using the most practical and effective solutions available.

Red teaming is an example of one strategy that AI firms already use to accomplish this. It involves testing algorithmic systems in a closed or highly controlled setting to determine how things could go wrong. Anthropic is an AI safety and research company that has done important red-teaming research, and their researchers have documented how "using manual or automated methods to adversarially probe a language model for harmful outputs, and then updating the model to avoid such outputs" is a useful tool for addressing potential harms.[130] By intentionally eliciting problematic results from generative AI models and then taking steps to counter those results, red teaming represents the idea of ethical transfer learning and iterative amplification in action. However, Anthropic researchers correctly note that "[t]he research community lacks shared norms and best practices for how to release findings from red teaming," and that "it would be better to have a neutral forum in which to discuss these issues."[131]

KEY TAKEAWAY

Many different strategies and much ongoing experimentation will be needed to address the many challenges we must confront today and in the future. The goal is to assess and prioritize risks continuously and then formulate and reformulate our response toolkit to those risks using the most practical and effective solutions available.

Luckily, there are many useful soft-law mechanisms—some old, some new—that can address that problem and facilitate collaborative efforts. As noted earlier, many broad-based ethical guidelines already exist for AI development, and they are organized increasingly around a common set of values and best practices such as transparency, privacy, security and nondiscrimination. Again, professional associations like IEEE, ACM, ISO and others are particularly important coordinators in this regard. Industry trade associations and other nongovernmental organizations (NGOs) also play a crucial role. These organizations and bodies need to work together to align alignment efforts. That should include finding ways to better publicize red-team research methods and results while identifying useful collective solutions to other identified vulnerabilities.

Once that is underway, we must ensure that such values get translated into concrete guidelines and guardrails at the developer level. ASU scholars have highlighted the growth of important internal measures that can help AI developers prioritize the embedding of ethics by design and ensure that humans remain "in the loop" along the way.[132] In addition to professional bodies and trade associations, they identify many other important strategies to give shared norms and best practices real meaning, including:

- **Corporate boards**: Building on widespread corporate social responsibility themes and efforts, corporate boards can act to align business practices and decision-making by encouraging firms to adopt widely held values or guidelines.[133] These

130. Deep Ganguli et al., "Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned," Cornell University, Aug. 23, 2022. https://arxiv.org/abs/2209.07858.
131. Ibid., p. 15.
132. Marchant et al., "Governing Emerging Technologies through Soft Law: Lessons for Artificial Intelligence—An Introduction. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3761871.
133. Jamie Baker, "Ethics and Artificial Intelligence: A Policymaker's Introduction," Center for Security and Emerging Technology, April 2021, pp. 13-16. https://cset.

efforts can help ensure that the firms guard against misuses of their technologies, which could have negative reputational effects and financial ramifications for the company and its shareholders.

- **Ethics committees**: Firms can establish and empower internal bodies or technology review boards to help embed and enforce ethics by design.[134] Microsoft established an Office of Responsible AI to help establish and enforce "company-wide rules for responsible AI through the implementation of our governance and public policy work."[135] Microsoft has also developed a robust harms-modeling framework to build on their ethical best practices. This framework includes what they refer to as a "community juries" process to bring together groups affected by various technologies.[136] Likewise, IBM created an internal AI Ethics Board that built on its preexisting Privacy Advisory Committee to consider how to educate employees about embedding ethics when designing new services.[137]

- **Ethics officers**: Another type of internal champion is a Chief Ethical Officer (or ethical champion) who plays a role similar to that of a Chief Privacy Officer.[138] These professionals have a formal responsibility to help establish best practices for technological developments and then ensure that organizations live up to their commitments.

- **Ombudsmen or whistleblower mechanism**: AI developers can enlist the support of internal and external individuals and experts to help monitor these efforts and evaluate ethical development and use on an ongoing basis. Some firms have already formed external ethics boards or watchdog bodies, but not always without controversy. A notable effort by Google to form an Advanced Technology External Advisory Council in 2019 shut down less than a week after its launch due to protests about certain members of the council.[139] Meanwhile, in mid-2022, Axon, a firm involved in law enforcement contracting, announced a plan to move forward with an effort to develop Taser-equipped drones to address mass shootings and school shootings, even though an AI Ethics Board recommended against it. In response, nine members of that body resigned in protest over the company's decision to ignore their advice.[140] But then Axon announced it was halting the development of the Taser drones in response to the resignations.[141] Other firms have developed similar external ethics boards, and whistleblowers have made news in recent years for outing algorithmic practices at Facebook and

georgetown.edu/publication/ethics-and-artificial-intelligence.

134. Wallach and Marchant. https://ieeexplore.ieee.org/document/8662741.
135. "Putting principles into practice at Microsoft." https://www.microsoft.com/en-us/ai/our-approach.
136. "Responsible innovation: a best practices toolkit," Microsoft, Jan. 24, 2023. https://docs.microsoft.com/en-us/azure/architecture/guide/responsible-innovation.
137. "Responsible Use of Technology: The IBM Case Study," World Economic Forum, Sept. 28, 2021. https://www.weforum.org/whitepapers/responsible-use-of-technology-the-ibm-case-study.
138. "Chief Privacy Officers: Who Are They and Why Education Leaders Need Them," Center for Democracy & Technology, January 2019. https://cdt.org/wp-content/uploads/2019/01/Student-Privacy-Chief-Privacy-Officer-Issue-Brief.pdf.
139. Kelsey Piper, "Google cancels AI ethics board in response to outcry," *Vox*, April 4, 2019. https://www.vox.com/future-perfect/2019/4/4/18295933/google-cancels-ai-ethics-board.
140. Drew Harwell, "Taser maker proposed shock drones for schools. What could go wrong?," *The Washington Post*, June 6, 2022. https://www.washingtonpost.com/technology/2022/06/06/taser-drone-school-shootings-clash.
141. Michael Balsamo, "Axon halts plans for Taser drone as 9 on ethics board resign," AP News, June 6, 2022. https://apnews.com/article/technology-government-and-politics-shootings-655fc0df3588e3e6afcd2a81b9619724.

> Twitter, among other tech companies.[142] That will likely continue and influence the creation of more internal and external oversight mechanisms to avoid liability or unwanted public relations.

The good news is that many developers are getting more serious about embedding ethics in the AI design process using such approaches. As a *Vox* reporter summarized, "we can build AI systems that are aligned with human values, or at least that humans can safely work with. That is ultimately what almost every organization with an artificial general intelligence division is trying to do."[143]

Tensions sometimes exist between AI-related ethical values and best practices. This is a continuing challenge for both hard- and soft-law efforts.

Balancing Ethical Values: Complications and Tradeoffs

Importantly, the many reports and efforts cited here typically also acknowledge that defining and categorizing these ethical values can be complicated, and tensions may exist between some of these ethical values and best practices. This is a continuing challenge for both hard- and soft-law efforts.

Consider values like transparency and explainability. Transparency is a value that can be tricky to define, and, as the author of *AI Ethics* notes, "it is questionable if it is possible to always have transparent AI."[144] If transparency requirements are applied aggressively, they could conflict with corporate confidentiality and user privacy. For example, developers who were forced to be completely transparent about how their algorithms work could essentially be forced to divulge their core intellectual property. User privacy could also be compromised if transparency requirements resulted in security vulnerabilities that made it easier for others to access the data that powered certain AI applications.

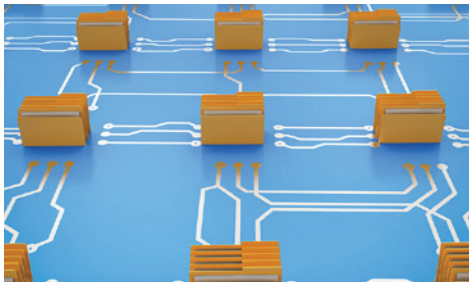
Likewise, some critics argue that AI systems be made more "explainable" to avoid the so-called "black-box" problem (i.e., algorithms being opaque and mysterious).[145] It seems like a reasonable governance requirement, but the problem is that "AI's outputs remain difficult to explain."[146] A leading AI expert has identified the challenges associated with explainability as a general governance concept:

> While it would be easy to program the computer to print out a list of all the additions and multiplications performed by a network for a given input, such a list would give us humans *zero* insight into how the network arrived at its answer. A list of a billion operations is not an explanation that a human can understand. Even the humans who train deep networks generally cannot look under the hood and provide explanations for the decision their networks make.[147]

142. Billy Perrigo, "Inside Frances Haugen's Decision to Take on Facebook," *Time*, Nov. 22, 2021. https://time.com/6121931/frances-haugen-facebook-whistleblower-profile; John D. McKinnon and Dave Michaels, "Twitter Comes Under Washington Spotlight With Whistleblower Complaint," *The Wall Street Journal*, Aug. 24, 2022. https://www.wsj.com/articles/twitter-comes-under-washington-spotlight-with-whistleblower-complaint-11661291987.
143. Kelsey Piper, "The case for taking AI seriously as a threat to humanity," *Vox*, Oct. 15, 2020. https://www.vox.com/future-perfect/2018/12/21/18126576/ai-artificial-intelligence-machine-learning-safety-alignment.
144. Coeckelbergh, p. 120.
145. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2016).
146. Henry Kissinger et al., "ChatGPT Heralds an Intellectual Revolution," *The Wall Street Journal*, Feb. 24, 2023. https://www.wsj.com/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homo-technicus-technology-cognition-morality-philosophy-774331c6.
147. Mitchell, p. 108.

Many other scholars have documented the challenges associated with trying to explain exactly how algorithmic systems arrive at certain answers or solutions.[148]

There is also a tradeoff between data minimization and the overall quality or effectiveness of algorithmic systems. Most data minimization proposals are premised on fears about data privacy or abuse. Too much information, some worry, could give rise to new types of discrimination.[149] The best way to improve datasets and eliminate bias, however, is through more—and better—data, not less. Better data requires constant refinement and improvement of existing datasets and the collection of more accurate data going forward. "The capacity to sort and mine through immense amounts of data enables algorithms to educate us about inequality," notes the author of *The Equality Machine: Harnessing Digital Technology for a Brighter, More Inclusive Future*.[150] She argues that calls for mandatory data minimization undermine that process because "addressing inequality starts with better data."[151] She believes that "data done right is the best of disinfectants, and digital illumination the most powerful social equalizer."[152]



Most data minimization proposals are premised on fears about data privacy or abuse. Too much information, some worry, could give rise to new types of discrimination. The best way to improve datasets and eliminate bias, however, is through more—and better—data, not less.

Such tensions and trade-offs will continue to complicate AI governance efforts going forward, especially for matters involving bias and "fairness."[153] No rigid formula can provide a simple answer to how to strike this balance. "There's no perfect consensus" about what constitutes discrimination and fairness and, therefore, "AI models will never be completely free from bias," says the author of the *AI Ethics* handbook.[154] Likewise, the authors of *The Ethical Algorithm: The Science of Socially Aware Algorithm* correctly observe that "the tension between fairness and accuracy will always remain" because "such trade-offs have always been implicitly present in human decision making."[155] Moreover, the root of the AI bias problem is often the underlying biases of humans who provided or interpreted bad data from the past. This is the so-called "garbage in, garbage out" problem, or the reality that "the model will be only as good as the data training it."[156] Again, the solution to this problem is improved data collection techniques.

Consequently, the quest for algorithmic fairness and AI alignment will be a process of ongoing trial and error; values will be calibrated and recalibrated depending on the specific use case being considered. Context is everything, and datasets and models will need to undergo constant refinement to address bad prior inputs or new social realities.

148. Chloe Xiang, "Scientists Increasingly Can't Explain How AI Works," *Vice*, Nov. 1, 2022. https://www.vice.com/en/article/y3pezm/scientists-increasingly-cant-explain-how-ai-works.
149. Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown, 2016).
150. Orly Lobel, *The Equality Machine: Harnessing Digital Technology for a Brighter, More Inclusive Future* (PublicAffairs, 2022), p. 8.
151. Ibid.
152. Ibid.
153. Cem Dilmegani, "Bias in AI: What it is, Types, Examples & 6 Ways to Fix it in 2022," *AI Multiple*, Sept. 12, 2020. https://research.aimultiple.com/ai-bias.
154. Coeckelbergh, p. 131.
155. Michael Kearns and Aaron Roth, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (Oxford University Press, 2020), p. 72.
156. Bronwyn Howell, "AI Algorithm Bias: What Can Be Done About It?," AEI, Oct. 31, 2022. https://www.aei.org/technology-and-innovation/ai-algorithms-bias-what-can-be-done-about-it.

Once again, soft-law mechanisms at least offer a more flexible way to address these tensions than slow-moving, binary hard-law regulatory approaches. "Regardless of its use," notes the recent ASU study, "soft law's flexibility has made it the dominant form of AI governance," and its ability to be nimbler in responding to such trade-offs is part of the reason why that is the case.[157]

"Professionalizing" AI Ethical Oversight

What AI governance needs now is an even more unified effort to formalize AI ethics and to make this "baking in" process routine for AI developers of all sizes and in all sectors. For soft law to make a lasting difference, the aspirational values found in the many ethical frameworks outlined above need to be translated into more concrete deliverables that hold innovators to certain standards. We might think of this as the "professionalization" of AI ethical oversight, in that the goal is to make the embedding of ethical best practices a more routine part of AI development.

For soft law to make a lasting difference, the aspirational values found in proposed ethical frameworks need to be translated into more concrete deliverables that hold innovators to certain standards.

One model for how to do so might mimic the role played by the International Association of Privacy Professionals (IAPP) for privacy best practices. Founded in 2000, the IAPP trains and certifies privacy professionals through formal credentialing programs, supplemented by regular meetings, annual awards, and a variety of outreach and educational initiatives.[158] The IAPP offers credentialing programs for the roles of Certified Information Privacy Professional (CIPP), the Certified Information Privacy Manager (CIPM), Certified Information Privacy Technologist (CIPT) and others. We can think of this as the professionalization of privacy practices, and it has become a robust and widely accepted system within data-driven industries, even in the absence of any overarching federal privacy law in the United States.

Of course, it is somewhat easier to create a professional credentialing system for a narrower category of concern like privacy. Broad-based credentialing for AI ethics will prove more challenging and may need to build on more narrowly drawn efforts by organizations working to address privacy, safety and security.

Some groups are already looking to fill this gap. The Trust and Safety Professional Association (TSPA) seeks to "support the global community of professionals who develop and enforce principles and policies that define acceptable behavior and content online."[159] The TSPA creates and circulates resources and tools to digital-safety professionals, including best practices and a formal Code of Conduct to enable the creation of safer online spaces and experiences that are free from bias and harassment and that protect privacy.[160] Likewise, the Digital Trust & Safety Partnership (DTSP) is an effort "to promote a safer and more trustworthy internet" through the application of various industry best practices, backed up by

157. Gutierrez and Marchant, p. 3. https://lsi.asulaw.org/softlaw/wp-content/uploads/sites/7/2022/08/final-database-report-002-compressed.pdf.
158. "IAPP Mission and Background," IAPP, last accessed March 3, 2023. https://iapp.org/about/mission-and-background.
159. "What We Do," Trust & Safety Professional Association, last accessed March 3, 2023. https://www.tspa.org/what-we-do.
160. "Code of Conduct," Trust & Safety Professional Association, last accessed March 3, 2023. https://www.tspa.org/code-of-conduct.

assessments and audits.[161] The DTSP looks to create a process for training people who will carry out such responsibilities in a professional context for major data-handling operators.[162]

Even better might be an effort to combine this professionalization approach with some sort of formal seal of approval for AI products deemed compliant with the ethical frameworks and best practices outlined above. To the extent that there is a problem in the field of AI soft law and AI ethics today, it could be that there are too many efforts currently underway. Some degree of consolidation is needed in terms of the major efforts by IEEE, ACM, ISO and other organizations. We do not have four different movie- or video-game-rating systems, for example. If multiple rating bodies existed for movies and games, they would likely create considerable confusion among content creators and the public. Standardized rating systems have been quite effective in informing the public of what they can expect to see and hear in movies and video games because they are applied in a fairly consistent, comprehensive and understandable fashion.[163]

To the extent that there is a problem in the field of AI soft law and AI ethics today, it could be that there are too many efforts currently underway. Some degree of consolidation is needed.

While a formal rating system is likely unworkable for AI ethics, it might be possible to have certification efforts for general compliance with ethical best practices. In the United Kingdom, the BSI has issued "Kitemark" seals of approval since 1903, which are quality certification awards for products or services that pass a rigorous assessment for safety and reliability.[164] As noted, the UL offers similar seals and certifications here in the United States. Perhaps it would be possible to certify Chief Ethical Officers in a similar way to Chief Privacy Officers, and then those Chief Ethical Officers could work to ensure that their companies satisfy various best-practice guidelines to receive seals of approval or certifications from leading bodies. The details need to be worked out, but the general framework already exists in other fields. This approach has the added benefit of relieving some of the pressure involved with more formal regulation of AI systems, so it is in the best interest of developers to work diligently to create such governance systems.

The government's role in this process could be to again play the role of convener and advisor, helping to bring various stakeholders together regularly to formulate and reformulate ethical best practices as needed to address various AI use cases. Policymakers can also help advise parties and remind them about existing hard- or soft-law governance frameworks that can guide the formulation and enforcement of best practices. Finally, government can play the backstop role described in detail below, using tools such as consumer protection rules or product recall authority to supplement soft-law frameworks when things go wrong. The courts will also continue to play an important role as cases come before them involving more serious and unforeseen harms.

161. David Morar, "Tech Firms Take First Step Toward Self-Regulation on Trust & Safety," *Tech Policy Press*, Sept. 25, 2022. https://techpolicy.press/tech-firms-take-first-step-toward-self-regulation-on-trust-safety.
162. "The Safe Assessments: An Inaugural Evaluation of Trust & Safety Best Practices," Digital Trust & Safety Partnership, July 2022. https://dtspartnership.org/wp-content/uploads/2022/07/DTSP_Report_Safe_Assessments.pdf.
163. Adam Thierer, "Soft Law in ICT Sectors: Four Case Studies," *Jurimetrics* 61:1 (April 2021), pp. 94-100. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3777490.
164. "The BSI Kitemark™ – trust and confidence," BSI, last accessed March 3, 2023. https://www.bsigroup.com/en-GB/kitemark.

The Ins and Outs of Algorithmic Auditing and AI Impact Assessments

The professionalization of AI ethics could be further formalized through algorithmic auditing and AI impact assessments.[165] Other business sectors use audits and impact assessments to address safety practices, financial accountability, labor practices, human rights issues, supply chain practices and various environmental concerns. AI audits and impact assessments would require those who develop or deploy algorithmic systems to conduct reviews to evaluate how well aligned the systems were with various ethical values or other commitments.[166] These evaluations could be conducted before or after a system launch, or both. Governments, private companies and any other institution developing or deploying algorithmic systems could employ such audits or assessments.[167]

Algorithmic audits and impact assessments face the same sort of definitional challenges that pervade AI more generally. For example, what constitutes a risk or harm in any given context will often be a complicated and contentious matter.

However, many complexities exist. Algorithmic audits and impact assessments face the same sort of definitional challenges that pervade AI more generally. For example, what constitutes a risk or harm in any given context will often be a complicated and contentious matter. In some cases, the potential harm or impact on a group might be easier to assess, such as when so-called predictive policing algorithms are used by law enforcement officials or the courts to judge or sentence individuals from marginalized groups.[168] Governmental uses of algorithmic processes will always raise greater concern and require greater oversight because governments possess coercive powers that private actors do not.

The focus here, however, will be on how audits or assessments might be used to address private-sector uses of AI and ML that give rise to concerns about privacy, safety, security or bias. Many current academic proposals for algorithmic auditing regimes imagine that this must be a formal regulatory certification process, modeled after other existing regulatory regimes.[169] For example, some of the scholars advocating for these ideas want to use the National Environmental Policy Act (NEPA) as a model.[170] Passed in 1969, NEPA requires formal environmental impact statements for major federal actions "significantly affecting the quality of the human environment."[171] Many states have adopted similar requirements.

U.S. policymakers are already floating bills that would mandate algorithmic auditing and impact assessments. Once such measure, the Algorithmic Accountability Act of 2022, proposed that developers perform impact assessments and file them

165. Rich Ehisen, "Could Algorithm Audits Curb AI Bias?," *State Net Insights*, Feb. 18, 2022. https://www.lexisnexis.com/community/insights/legal/capitol-journal/b/state-net/posts/could-algorithm-audits-curb-ai-bias; Ilana Golbin, "Algorithmic impact assessments: What are they and why do you need them?," pwc, Oct. 28, 2021. https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-impact-assessments.html.
166. Jacob Metcalf et al., "Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts," FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (March 2021), pp. 735-746. https://doi.org/10.1145/3442188.3445935.
167. Dillon Reisman et al., "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability," AI Now, April 2018. https://ainowinstitute.org/aiareport2018.pdf.
168. Jamie Grierson, "Predictive policing poses discrimination risk, thinktank warns," *The Guardian*, Sept. 15, 2019. https://www.theguardian.com/uk-news/2019/sep/16/predictive-policing-poses-discrimination-risk-thinktank-warns.
169. Andrew D. Selbst, "An Institutional View Of Algorithmic Impact Assessments," *Harvard Journal of Law & Technology* 35 (Fall 2021), pp. 117-191. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3867634.
170. Emanuel Moss et al., "Assembling Accountability: Algorithmic Impact Assessment for the Public Interest," Data & Society (June 29, 2021). https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest.
171. "National Environmental Policy Act," United States Environmental Protection Agency, July 6, 2022. https://www.epa.gov/nepa.

with the FTC. The Act creates a new Bureau of Technology inside the FTC to oversee the process. The law would also "require each covered entity to attempt to eliminate or mitigate, in a timely manner, any impact made by an augmented critical decision process that demonstrates a likely material negative impact that has legal or similarly significant effects on a consumer's life."[172] Similar algorithmic auditing requirements are also included in the American Data Protection and Privacy Act of 2022, a comprehensive federal privacy proposal that attracted widespread bipartisan support.[173] The proposed law would require large data handlers to perform an annual algorithm impact assessment that includes a "detailed description" of both "the design process and methodologies of the covered algorithm," as well as a "steps the large data holder has taken or will take to mitigate potential harms from the covered algorithm."[174]

Advocates of auditing mandates insist that "increasingly robust regulatory requirements" will mean that "the public will have greater confidence in using highly automated systems," but they typically fail to consider whether those systems will even be developed if they are preemptively suffocated by layers of red tape and lengthy approval timetables.

The full scope of this sort of mandate remains to be seen. If enforced through a rigid regulatory regime, compliance with algorithmic auditing mandates would likely become a time-consuming, convoluted, bureaucratic process that could significantly slow the pace of AI development. Unfortunately, most of the academic literature surrounding algorithmic auditing fails to discuss the potential costs associated with the paperwork burdens and compliance delays that would likely be associated with such a regulatory regime. Advocates of auditing mandates insist that "increasingly robust regulatory requirements" will mean that "the public will have greater confidence in using highly automated systems," but they typically fail to consider whether those systems will even be developed if they are preemptively suffocated by layers of red tape and lengthy approval timetables.[175]

Consider the complexities of NEPA. Although well intentioned, NEPA environmental impact statements create significant compliance costs and project delays.[176] NEPA assessments were initially quite short (sometimes less than 10 pages), but the average length of these statements now exceeds 600 pages and can include appendices that push the total to more than 1,000 pages.[177] Moreover, these assessments take an average of 4.5 years to complete; some have taken 17 years or longer.[178] What this means in practice is that many important public projects are not completed, or they take much longer to complete at considerably higher expenditure than originally predicted. For example, NEPA has slowed many infrastructure projects and clean energy initiatives, and even Democratic presidential administrations have suggested the need to reform the assessment process due to its rising costs.[179]

172. H.R.6580, "Algorithmic Accountability Act of 2022," 117th Congress. https://www.congress.gov/bill/117th-congress/house-bill/6580.
173. H.R.8152, "American Data Privacy and Protection Act," 117th Congress. https://www.congress.gov/bill/117th-congress/house-bill/8152.
174. American Data Protection and Privacy Act, § 207(c)(1).
175. Gregory Falco et al., "Governing AI safety through independent audits," *Nature Machine Intelligence* 3 (2021), p. 570. https://www.nature.com/articles/s42256-021-00370-7.
176. Eli Dourado, "Why are we so slow today?," The Center for Growth and Opportunity, March 12, 2020. https://www.thecgo.org/benchmark/why-are-we-so-slow-today.
177. Ibid.
178. Ibid.
179. Ibid.

The author of *Construction Physics* referred to NEPA as an "anti-law" in the sense that it largely accomplishes the exact opposite of what the underlying statute intended.[180] Instead of creating predictability, the law "greatly reduces predictability and increases coordination cost and risk, because it's so unclear what's needed to meet NEPA requirements," he says.[181] Politicization is also a serious problem because NEPA "seems easily captured by small groups with strongly held opinions" who stand ready to block almost all progress on important projects and, therefore, "is effectively a bias towards the status quo."[182] Sadly, it is not clear that the law does anything to improve environmental outcomes because it makes it so difficult for many important initiatives to be completed in a timely or effective manner—assuming they are allowed to move forward at all. "The NEPA process is effectively a tax on any major government action, and like any tax, we'd expect it to result in less of what it taxes."[183] NEPA's laboriously complicated and slow permitting processes—and the failure of policymakers to address them—have led to questions about whether some in the environmental movement are concerned more about the process itself rather than concrete results. An *Atlantic* reporter suggested that "many people within the environmentalist movement are undermining the nation's emissions goals in the name of localism and community input."[184]

Applying the NEPA model to algorithmic systems would likely grind AI innovation to a halt in the face of lengthy delays, paperwork burdens and significant compliance costs.

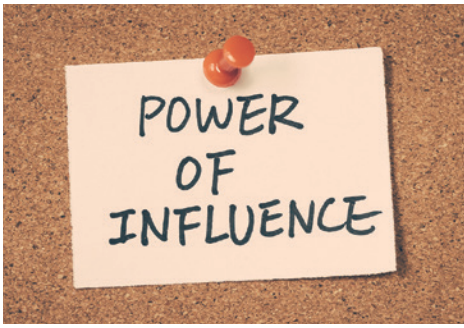
For similar reasons, applying the NEPA model to algorithmic systems would likely grind AI innovation to a halt in the face of lengthy delays, paperwork burdens and significant compliance costs.[185] Converting audits into a formal regulatory process would also create several veto points that opponents of AI could use to slow progress in the field. Many scholars today decry the United States' growing culture of "vetocracy," which describes the many veto points within modern political systems that hold back innovation, development and economic opportunity.[186] This endless accumulation of potential veto points in the policy process in the form of mandates and restrictions can greatly curtail innovation opportunities. NEPA-like algorithmic auditing mandates would create many such veto points within the product development process.

Algorithmic systems evolve at an incredibly rapid pace and undergo constant iteration, with some systems being updated on a weekly or even daily basis. One AI analyst observed that "algorithms can be fearsomely complex entities to audit" because of the combination of their daunting size, complexity and obscurity.[187] Society cannot wait years or even months for bureaucracies to get around to formally signing off on audits or assessments, many of which would be obsolete

180. Brian Potter, "How NEPA works," *Construction Physics*, Aug. 19, 2022. https://constructionphysics.substack.com/p/how-nepa-works.
181. Ibid.
182. Ibid.
183. Ibid.
184. Jerusalem Demsas, "Not Everyone Should Have a Say," *The Atlantic*, Oct. 19, 2022. https://www.theatlantic.com/ideas/archive/2022/10/environmentalists-nimby-permitting-reform-nepa/671775.
185. Philip Rossetti, "Addressing NEPA-Related Infrastructure Delays," *R Street Policy Study* No. 234 (July 2021). https://www.rstreet.org/research/addressing-nepa-related-infrastructure-delays; Jeremiah Johnson, "The Case for Abolishing the National Environmental Policy Act," *Liberal Currents*, Sept. 6, 2022. https://www.liberalcurrents.com/the-case-for-abolishing-the-national-environmental-policy-act.
186. William Rinehart, "Vetocracy, the costs of vetos and inaction," The Center for Growth and Opportunity at Utah State University, March 24, 2022. https://www.thecgo.org/benchmark/vetocracy-the-costs-of-vetos-and-inaction; Adam Thierer, "Red tape reform is the key to building again," *The Hill*, April 28, 2022. https://thehill.com/opinion/finance/3470334-red-tape-reform-is-the-key-to-building-again.
187. James Kobielus, "How We'll Conduct Algorithmic Audits in the New Economy," *InformationWeek*, March 4, 2021. https://www.informationweek.com/ai-or-machine-learning/how-we-ll-conduct-algorithmic-audits-in-the-new-economy.

before they were completed. Many AI developers would likely look to innovate elsewhere if auditing or impact assessments became a bureaucratic and highly convoluted compliance nightmare.

Additionally, algorithmic auditing will always be an inexact science because of the inherent subjectivity of the values being considered. Auditing algorithms is not like auditing an accounting ledger, where the numbers either do or do not add up. When evaluating algorithms, there are no binary metrics that can quantify the scientifically correct amount of privacy, safety or security in a given system.



Legislatively mandated algorithmic auditing could give rise to the problem of significant political meddling in speech platforms powered by algorithms.

Legislatively mandated algorithmic auditing could give rise to the problem of significant political meddling in speech platforms powered by algorithms. In recent years, both Republican and Democratic lawmakers have accused digital technology companies of manipulating algorithms to censor their views. For example, during a heated 2022 debate over a bill to regulate algorithmic content moderation, lawmakers from both parties accused social media companies of censoring them or their favored content.[188] Aside from the fact that both sides cannot be right, the fact that they all want to use government leverage to influence private content management decisions illustrates the danger of mandatory algorithmic auditing. Whichever party is in power at any given time could use the auditing process to politicize terms like "safety," "security" and "nondiscrimination" to nudge or even force private AI developers to alter their algorithms to satisfy political desires.

Political issues like this arose at the FCC when the agency abused its ambiguous authority to regulate "in the public interest" and indirectly censored broadcasters through intimidation.[189] The agency would send radio and television broadcasters letters of inquiry (LOIs) asking about programming decisions and not-so-subtly suggesting how the stations might reconsider what they put on the air. This tactic was used frequently enough that it came to be known in policy circles as "regulation by raised eyebrow," or "regulatory threats that cajole industry members into slight modifications" of their programming content.[190] This became an effective way for the FCC to avoid First Amendment battles that would ensue in the courts if the agency had taken formal steps to revoke the license of a broadcaster. The agency used the LOIs in combination with jawboning tactics and other threats in speeches and public statements to shape industry speech decisions. Congressional lawmakers also used these same jawboning tactics in hearings and public statements to influence private content choices.[191] These tactics were used in other ways during merger reviews or other regulatory processes when policymakers realized that they possessed leverage to extract demands from private parties.[192]

188. Adam Thierer, "Left and right take aim at Big Tech—and the First Amendment," *The Hill*, Dec. 8, 2021. https://thehill.com/opinion/technology/584874-left-and-right-take-aim-at-big-tech-and-the-first-amendment.
189. Randolph J. May, "The Public Interest Standard: Is It Too Indeterminate to Be Constitutional?," *Federal Communications Law Journal* 53:3 (May 2011), pp. 427-468. https://www.repository.law.indiana.edu/fclj/vol53/iss3/3.
190. Thomas Streeter, *Selling the Air: A Critique of the Policy of Commercial Broadcasting in the United States* (The University of Chicago Press, 1996), p. 189.
191. Jerry Brito, "'Agency Threats' and the Rule of Law: An Offer You Can't Refuse," *Harvard Journal of Law & Public Policy* 37:2 (2014), p. 553. https://www.harvard-jlpp.com/wp-content/uploads/sites/21/2014/05/37_2_553_Brito-1.pdf.
192. Thierer, "Soft Law in ICT Sectors: Four Case Studies," pp. 94-96. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3777490.

It is not a stretch to imagine how regulators or lawmakers could use mandated algorithmic audits or impact statements to unduly influence AI decision-making in similar ways. We have already witnessed intense debates over what constitutes online "disinformation" following a short-lived Biden administration effort to create a Disinformation Governance Board within the Department of Homeland Security.[193] If a new algorithmic oversight law or agency were created, similar fights would ensue. While not explored here, there are potentially profound First Amendment issues at play with the regulation of algorithms. These considerations could become a major part of AI regulatory efforts going forward if the AI auditing process were mandated and then became politicized in this fashion.[194]

Algorithmic Auditing Done Right

Despite these problems, algorithmic auditing and AI impact assessments can still be a part of a more decentralized, polycentric governance framework and can help innovations by "ensuring that programs are not inadvertently 'learning' the wrong lessons from the information entered into the systems."[195] Algorithmic audits can help developers constantly improve their systems and avoid damaging market losses or liability threats.

Algorithmic auditing and AI impact assessments can still be a part of a more decentralized, polycentric governance framework and can help developers constantly improve their systems and avoid damaging market losses or liability threats.

Even in the absence of any sort of hard-law mandates, algorithmic auditing and impact reviews represent a sensible way to help formalize the ethical frameworks and best practices already formulated by professional associations such as the IEEE, ISO, ACM and others. Once again, the focus of those efforts is to get developers to think more seriously about how to bake in widely shared goals and values and consider how to keep humans in the loop at critical stages of this process to ensure that they can continue to guide and occasionally realign those values as needed.

Such an auditing and impact assessment process can be rooted in the voluntary risk assessment frameworks that the OECD and the NIST have been formulating. The OECD has developed a *Framework for the Classification of AI Systems* with the goals of helping "to develop a common framework for reporting about AI incidents that facilitates global consistency and interoperability in incident reporting," and advancing "related work on mitigation, compliance and enforcement along the AI system lifecycle, including as it pertains to corporate governance."[196]

NIST also recently released a comprehensive *Artificial Intelligence Risk Management Framework*, which is a voluntary, consensus-driven guidance document intended "to offer a resource to the organizations designing, developing, deploying, or using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems."[197] The *Framework* builds on

193. Adam Thierer and Patricia Patnode, "Disinformation About the Real Source of the Problem," *Real Clear Policy*, May 23, 2022. https://www.realclearpolicy.com/articles/2022/05/23/disinformation_about_the_real_source_of_the_problem_833681.html.
194. Stuart Minor Benjamin, "The First Amendment and Algorithms," in Woodrow Barfield, ed, *The Cambridge Handbook of the Law of Algorithms* (Cambridge University Press, 2021), pp. 606-631.
195. Keith E. Sonderling et al., "The Promise and The Peril: Artificial Intelligence and Employment Discrimination," *University of Miami Law Review* 77:1 (2022), p. 80. https://repository.law.miami.edu/umlr/vol77/iss1/3.
196. "OECD AI Principles overview," OECD.AI, last accessed March 3, 2023. https://oecd.ai/en/ai-principles; "OECD Framework for the Classification of AI Systems," OECD, Feb. 22, 2022, p. 6. https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm.
197. "NIST Risk Management Framework Aims to Improve Trustworthiness of Artificial Intelligence," NIST, Jan. 26, 2023, p. 2. https://www.nist.gov/news-events/news/2023/01/nist-risk-management-framework-aims-improve-trustworthiness-artificial.

the ethical frameworks developed by the many different organizations mentioned earlier, such as the IEEE, ISO and ACM.

Many AI developers and business groups have endorsed the use of such audits and assessments. BSA|The Software Alliance has said that "[b]y establishing a process for personnel to document key design choices and their underlying rationale, impact assessments enable organizations that develop or deploy high-risk AI to identify and mitigate risks that can emerge throughout a system's lifecycle."[198] As noted below, developers can still be held accountable for violations of certain ethical norms and best practices through both private and formal sanctions by consumer protection agencies (like the FTC or comparable state offices) or by state attorneys general.

Independent AI auditing bodies are already developing and could play an important role in helping to professionalize AI ethics going forward. EqualAI is a group that works with lawyers, businesses, and policy leaders to create and monitor ethical AI best practices. In collaboration with the WEF, EqualAI is creating a "Responsible AI Badge Certification" program.[199] The WEF has recently produced two major reports that can guide such efforts: "Empowering AI Leadership: AI C-Suite Toolkit" and "A Blueprint for Equity and Inclusion in Artificial Intelligence."[200] Meanwhile, the WEF is also involved in a partnership with AI Global, a nonprofit organization focused on advancing the responsible and ethical adoption of AI, and the Institute for Technology and Society at the University of Toronto to "create a globally recognized certification mark for the responsible and trusted use of AI systems."[201]

"To win customer, regulator, and investor trust," a journalist explained, "AI companies need to address [internal auditing] proactively, rather than waiting for regulations."

According to The Institute of Internal Auditors (IIA), a widespread internal auditing profession already exists, with professional auditors "identifying the risks that could keep an organization from achieving its goals, making sure the organization's leaders know about these risks, and proactively recommending improvements to help reduce the risks." The IIA collectively represents these auditors, helps establish standards for the profession and awards a Certified Internal Auditor designation through rigorous examinations.[202] Eventually, more and more organizations will expand their internal auditing efforts to incorporate AI risks because it makes good business sense to stay on top of these issues to help avoid liability, negative publicity or other customer backlash.[203] "To win customer, regulator, and investor trust," a journalist explained, "AI companies need to address these concerns proactively, rather than waiting for regulations."[204]

198. "Enhancing Innovation and Promoting Trust: BSA's Artificial Intelligence Policy Agenda," BSA | The Software Alliance, 2022, p. 2. https://www.bsa.org/files/policy-filings/03222022bsausaiagenda.pdf.
199. Kay Firth-Butterfield and Miriam Vogel, "5 ways to avoid artificial intelligence bias with 'responsible AI,'" World Economic Forum, July 5, 2022. https://www.weforum.org/agenda/2022/07/5-governance-tips-for-responsible-ai.
200. "Empowering AI Leadership: AI C-Suite Toolkit," World Economic Forum, Jan. 12, 2022. https://www.weforum.org/reports/empowering-ai-leadership-ai-c-suite-toolkit; "A Blueprint for Equity and Inclusion in Artificial Intelligence," World Economic Forum, June 29, 2022. https://www.weforum.org/whitepapers/a-blueprint-for-equity-and-inclusion-in-artificial-intelligence.
201. Jovana Jankovic, "U of T's Schwartz Reisman Institute and AI Global to develop global certification mark for trustworthy AI," Dec. 1, 2020. https://www.utoronto.ca/news/u-t-s-schwartz-reisman-institute-and-ai-global-develop-global-certification-mark-trustworthy-ai.
202. "All in a Day's Work: A Look at the Varied Responsibilities of Internal Auditors," The Institute of Internal Auditors, last accessed March 3, 2023. https://www.theiia.org/globalassets/documents/about-us/promote-the-profession/informational-resources/all-in-a-days-work-brochure.pdf.
203. Jeff Bleich and Bradley J. Strawser, "Tool or Trouble: Aligning Artificial Intelligence with Human Rights," Harvard Advanced Leadership Initiative, April 25, 2022. https://www.sir.advancedleadership.harvard.edu/articles/tool-or-trouble-aligning-artificial-intelligence-with-human-rights.
204. Karen Hao, "Worried about your firm's AI ethics? These startups are here to help," MIT Technology Review, Jan. 15, 2021. https://www.technologyreview.com/2021/01/15/1016183/ai-ethics-startups.

Meanwhile, the field of algorithmic consulting continues to expand and will supplement these efforts with tailored expert oversight on technical, ethical and legal matters. For example, a leading AI social scientist created O'Neil Risk Consulting and Algorithmic Auditing to help organizations manage and audit algorithmic risks—specifically those pertaining to fairness, bias and discrimination.[205] The legal profession will also expand its focus to assist potential clients on these matters. For example, BNI.ai launched in 2020 and describes itself as a "boutique law firm that leverages world-class legal and technical expertise to help our clients avoid, detect, and respond to the liabilities of AI and analytics."[206] Other specialized AI law firms like this are sure to develop in coming years.

Another benefit of voluntary AI auditing and impact assessments is that these efforts can have a global reach when companies and trade associations adopt principles and frameworks like those described earlier. Finally, the governance mechanisms discussed herein will continue to be supplemented by various hard-law legal remedies to hold developers to the promises they make to the public while also addressing more serious AI harms that emerge or prove too challenging for soft law to address.

Highly precautionary and technocratic regulatory regimes for AI are both unwise and impractical. Although some ex-ante constraints may eventually become more necessary, it is more sensible to tap alternative legal and regulatory remedies that are already available.

How Ex-Post Hard Law Complements Soft Law

Much of the literature surrounding AI governance ignores the many existing ex-post legal mechanisms that can complement various AI soft-law governance approaches. This may be because many advocates of more precautionary regulatory regimes insist that ex-ante anticipatory regulation must lie at the heart of AI governance efforts.

Highly precautionary and technocratic regulatory regimes for AI are both unwise and impractical, however. Although some ex-ante constraints may eventually become more necessary and perhaps workable, it is more sensible to tap alternative legal and regulatory remedies that are already available. New ethical frameworks and soft-law governance mechanisms can build on these existing legal solutions and remedies.[207] "Voluntary codes as soft-law interventions do not exist in isolation from hard law, as codes and laws can interact to support or dampen the efficacy or creation of each other," observes one technological governance scholar.[208] It is also the case that "entities generally seek to comply with adopted codes because noncompliance may compel those entities to publicly explain their departure from the code."

In this way, soft law is buttressed by hard law, much as is already the case in other technology sectors, such as consumer electronics and computing. The United States does not have a Federal Computer Commission or Bureau of Consumer Electronics, for example, but when things go wrong, many legal remedies are available to address problems in those fields. In these and many other industries, innovators are generally free to develop new products. When harms develop, they are addressed

205. "It's the Age of the Algorithm and We Have Arrived Unprepared," ORCAA, last accessed March 3, 2023. https://orcaarisk.com.
206. "Why BNH," BNH.AI, last accessed March 3, 2023. https://www.bnh.ai/why-bnh; Seth Colaner, "Bnh.ai is a new law firm focused only on AI," Venture Beat, March 19, 2020. https://venturebeat.com/2020/03/19/bnh-ai-is-a-new-law-firm-focusing-only-on-ai.
207. John Villasenor, "Soft law as a complement to AI regulation," Brookings, July 31, 2020. https://www.brookings.edu/research/soft-law-as-a-complement-to-ai-regulation.
208. Walter G. Johnson, "Governance Tools for the Second Quantum Revolution," *Jurimetrics* 59:4 (April 27, 2019), p. 511. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3350830.

in a remedial fashion. In a similar way, existing legal remedies can help address risks associated with algorithmic and robotic systems. Some of these solutions include:

- **Federal and state consumer protection statutes and agencies:** The FTC possesses broad consumer protection powers to police "unfair or deceptive acts or practices in or affecting commerce."[209] Over the past decade, the agency has used this authority to address many data security matters and, in 2022, issued a major report highlighting its concerns with various AI risks.[210] Thus, when defective or deceptive algorithmic technologies create substantial harm to consumers, the FTC can intervene.[211] An attorney with the FTC's Division of Advertising Practices was even more hard-nosed about this in a February 2023 blog post, asserting, "[i]f you think you can get away with baseless claims that your product is AI-enabled, think again […] In an investigation, FTC technologists and others can look under the hood and analyze other materials to see if what's inside matches up with your claims."[212] Meanwhile, state Attorneys General and state consumer protection agencies also routinely address unfair practices and continue to advance their own privacy and data security policies, some of which are more stringent than federal law.

- **Product recall authority:** Several regulatory agencies in the United States possess recall authority that allows them to remove products from the market when certain unforeseen problems manifest. For example, the National Highway Traffic Safety Administration (NHTSA), FDA and Consumer Product Safety Commission (CPSC) all possess broad recall authority that can address risks that develop from algorithmic or robotic systems.[213] In February 2023, for instance, the NHTSA mandated a recall of Tesla's full self-driving autonomous driving system, and the agency required an over-the-air software update to over 300,000 vehicles that had the software package.[214] While the NHTSA's and FDA's recall authority is more targeted to vehicle and medical technologies, respectively, the CPSC can recall any consumer product that contains a defect if it poses "a substantial risk of injury to the public to warrant such remedial action."[215] A July 2022 poll commissioned by the CPSC revealed that 80 percent of consumers do everything that a recall notice encourages them to do to address a safety lapse.[216] While encouraging, that result could be further improved using education and awareness efforts. The CPSC has already issued staff reports highlighting how the agency has many policy tools to address emerging technology risks.[217]

209. 15 U.S.C. § 45(a).
210. "FTC Report Warns About Using Artificial Intelligence to Combat Online Problems," Federal Trade Commission, June 16, 2022. https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-report-warns-about-using-artificial-intelligence-combat-online-problems.
211. Inioluwa Deborah Raji et al., "The Fallacy of AI Functionality," Cornell University, June 20, 2022. https://arxiv.org/abs/2206.09511.
212. Michael Atleson, "Keep your AI claims in check," Federal Trade Commission, Feb. 27, 2023. www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check.
213. "Recalls, Corrections and Removals (Devices)," U.S. Food & Drug Administration, Sept. 29, 2020. https://www.fda.gov/medical-devices/postmarket-requirements-devices/recalls-corrections-and-removals-devices.
214. David Shepardson, "Tesla recalls 362,000 U.S. vehicles over Full Self-Driving software," *Reuters*, Feb. 16, 2023. https://www.reuters.com/business/autos-transportation/tesla-recalls-362000-us-vehicles-over-full-self-driving-software-2023-02-16.
215. United States Consumer Product Safety Commission, *Recall Handbook* (March 2012), pp. 2, 12.
216. "Qualtrics Final Report on Consumer Attitudes and Behaviors Regarding Product Safety," United States Consumer Product Safety Commission, July 26, 2022. https://www.cpsc.gov/content/Qualtrics-Final-Report-on-Consumer-Attitudes-and-Behaviors-Regarding-Product-Safety.
217. "Artificial Intelligence and Machine Learning In Consumer Products," United States Consumer Product Safety Commission, May 19, 2021. https://www.cpsc.gov/About-CPSC/artificial-intelligence-and-machine-learning-in-consumer-products; "Potential Hazards Associated with Emerging and Future Technologies," United States Consumer Product Safety Commission, Jan. 18, 2017. https://www.cpsc.gov/content/Potential-Hazards-Associated-with-Emerging-and-Future-Technologies.

- **Common law remedies:** Various court-enforced common law remedies exist that can address AI risks. These include product liability; negligence; design defects law; failure to warn; breach of warranty; property law and contract law; and other torts.[218] Common law evolves to meet new technological concerns and incentivizes innovators to make their products safer over time to avoid lawsuits and negative publicity.[219] It also evolves to incorporate new social and ethical norms. "[W]hen confronted with new, often complex, questions involving products liability, courts have generally gotten things right," notes a Brookings Institution scholar. He goes on to explain that "[p]roducts liability law has been highly adaptive to the many new technologies that have emerged in recent decades" and, by extension, it will adapt to other technologies and developments as cases and controversies come before the courts.[220] This also creates powerful incentives for developers to improve the safety and security of their systems and avoid liability, unwanted press attention and lost customers. The question is not whether common law liability will come to cover AI and robotics; it is whether it will impose too great a burden because the United States tends to have a highly litigious legal system.[221]

- **Property and contract law:** Federal and state laws covering contractual rights and property rights can address many perceived harms associated with algorithmic technologies. Property law already governs trespass claims, for example, which will come in handy as drones and other autonomous robotic systems proliferate. Contract law can also help developers live up to the promises they make to the public, including other business customers. Of note, class-action lawsuits will become more common if firms fail to honor their contractual terms.

- **Insurance and other accident-compensation mechanisms:** Many organizations have improved their digital cybersecurity practices "driven by demands from insurance underwriters and a better understanding of the risks of ransomware following high-profile attacks."[222] The market for highly tailored algorithmic insurance instruments is growing—and not just to address cybersecurity risks.[223] New insurance instruments will likely cover even more broad-based, amorphous algorithmic concerns ranging from physical safety risks to various other risks. Although broad-based algorithmic regulation is unlikely in the short term, lawsuits alleging algorithmic harm are likely going to proliferate in the future. As that occurs, insurance markets are going to continue to evolve and respond, especially for industrial robotics.[224]

218. "Torts of the Future II: Addressing the Liability and Regulatory Implications of Emerging Technologies," U.S. Chamber Institute for Legal Reform, April 2018. https://instituteforlegalreform.com/wp-content/uploads/2020/10/tortsofthefuturepaperweb.pdf; Richard A. Epstein, "Liability Rules in the Internet of Things: Why Traditional Legal Relations Encourage Modern Technological Innovation," Hoover Institution, Jan. 8, 2019. https://www.hoover.org/research/liability-rules-internet-things-why-traditional-legal-relations-encourage-modern.
219. Donald G. Gifford, "Technological Triggers to Tort Revolutions: Steam Locomotives, Autonomous Vehicles, and Accident Compensation," *Journal of Tort Law* 11:1 (Sept. 5, 2018), pp. 71-143. https://doi.org/10.1515/jtl-2017-0029.
220. John Villasenor, "Who is at fault when a driverless car gets in an accident?," UCLA Newsroom, May 2, 2014. https://newsroom.ucla.edu/stories/who-is-at-fault-when-a-driverless-car-gets-in-an-accident.
221. Adam Thierer, "When the Trial Lawyers Come for the Robot Cars," *Slate*, June 10, 2016. https://slate.com/technology/2016/06/if-a-driverless-car-crashes-who-is-liable.html.
222. Robert McMillan et al., "Hackers Extort Less Money, Are Laid Off as New Tactics Thwart More Ransomware Attacks," *The Wall Street Journal*, Feb. 22, 2023. https://www.wsj.com/articles/ransomware-attacks-decline-as-new-defenses-countermeasures-thwart-hackers-23b918a3.
223. Jeff Qiu, "Improving U.S. Cybersecurity by Solving Issues in the Cyber Insurance Market Part One: Current State and Challenges," R Street Institute, Aug. 8, 2022. https://www.rstreet.org/commentary/improving-u-s-cybersecurity-by-solving-issues-in-the-cyber-insurance-market-part-one-current-state-and-challenges; Josephine Wolff, "A Brief History of Cyberinsurance," *Slate*, Aug. 30, 2022. https://slate.com/technology/2022/08/cyberinsurance-history-regulation.html.
224. Andrea Bertolini et al., "On Robots and Insurance," *International Journal of Social Robotics* 8 (March 3, 2016), pp. 381-391. https://link.springer.com/article/10.1007/s12369-016-0345-z.

- **Existing statutes and agencies:** Many long-standing statutes and agency rules exist that can address concerns about algorithmic bias, privacy or security. Regarding the accusations of potential algorithmic bias and discrimination, the United States has a wide array of broad-based civil rights statutes that apply, including the Civil Rights Act, the Age Discrimination in Employment Act and the Americans with Disabilities Act.[225] Targeted financial laws could address discrimination in the allocation of credit, including the Fair Credit Reporting Act and Equal Credit Opportunity Act. The Fair Housing Act already addresses discrimination for real estate.[226] On the privacy front, laws such as the Health Insurance Portability and Accountability Act, the Gramm-Leach-Bliley Act and the Children's Online Privacy Protection Act already govern data flows.[227] Moreover, the United States already has a veritable alphabet soup of regulatory agencies that oversee technological developments in various sectors touched by algorithmic and robotic developments. These laws, regulations and agencies can provide a backstop when AI developers fail to live up to any claims they make about safe, effective and fair algorithmic systems.[228] If needed, Congress could tweak existing laws and regulations should novel or persistent problems develop. Many states also have laws that could apply to algorithmic or robotic systems. For example, "Peeping Tom" laws and antiharassment statutes exist that prohibit spying into homes and other private spaces.[229] Before enacting new laws, policymakers should consider how such existing policies might already cover new technological developments.

Case Study: Bottom-Up Governance of Autonomous Vehicles

All the flexible governance strategies mentioned throughout this report have already been leveraged in one particularly important AI sector: autonomous vehicles. As noted, there are many academic proposals to have government impose preemptive certification regimes on new AI systems. The U.S. DOT briefly considered such a precautionary regulatory regime for autonomous vehicles late in the Obama administration. In September 2016, the NHTSA published the government's first report on Federal Automated Vehicles Policy and said that the agency was considering "a pre-market approval approach" for highly automated vehicles (HAVs).[230] This regulatory approach, the agency said, "would prohibit the manufacture, introduction into commerce, offer for sale and sale of HAVs unless, prior to such actions, NHTSA has assessed the safety of the vehicle's performance

225. "Civil Rights Act (1964)," National Archives, last accessed March 3, 2023. https://www.archives.gov/milestone-documents/civil-rights-act; Keith E. Sonderling et al., "The Promise and The Peril: Artificial Intelligence and Employment Discrimination," *University of Miami Law Review* 77:1 (2022), p. 6. https://repository.law.miami.edu/umlr/vol77/iss1/3; "The Americans with Disabilities Act (ADA) protects people with disabilities from discrimination," U.S. Department of Justice, last accessed March 3, 2023. https://www.ada.gov.
226. "The Fair Housing Act," U.S. Department of Justice, last accessed March 3, 2023. https://www.justice.gov/crt/fair-housing-act-1.
227. "Health Insurance Portability and Accountability Act of 1996 (HIPAA)," Centers for Disease Control and Prevention, last accessed March 3, 2023. https://www.cdc.gov/phlp/publications/topic/hipaa.html; "Gramm-Leach-Bliley Act," Federal Trade Commission, last accessed March 3, 2023. https://www.ftc.gov/business-guidance/privacy-security/gramm-leach-bliley-act; "Children's Online Privacy Protection Rule ("COPPA")," Federal Trade Commission, last accessed March 3, 2023. https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa.
228. Joshua New and Daniel Castro, "How Policymakers Can Foster Algorithmic Accountability," Center for Data Innovation, May 21, 2018. https://datainnovation.org/2018/05/how-policymakers-can-foster-algorithmic-accountability.
229. See, e.g., Va. Code Ann. § 18.2-130 Peeping or spying into dwelling or enclosure.
230. "Federal Automated Vehicles Policy," U.S. Department of Transportation, Sept. 20, 2016, p. 72. https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016.

and approved the vehicle."[231] The agency suggested that the Federal Aviation Administration's (FAA) might provide a model for how such premarket approval could work for autonomous vehicles.

The NHTSA was surprisingly forthcoming about the potential negative tradeoffs associated with a pre-market approval regulatory regime for autonomous vehicles. At a minimum, the agency admitted, this "would be a wholesale structural change in the way NHTSA regulates motor vehicle safety and would require both fundamental statutory changes and a large increase in Agency resources."[232] There would be other costs, too. In a short appendix to the report, the agency noted that "the duration of the [FAA] certification processes varies. Typically, they last three to five years."[233] Of note, however, the FAA's certification the Boeing 787 Dreamliner took much longer; the agency estimated it took 200,000 hours of FAA staff time over an eight-year period.[234]

Imposing pre-market approval on driverless cars would likely result in long delays in product approval, which could have significant costs—not just for product developers but also for the public.

Thus, imposing the same sort of pre-market approval on driverless cars would likely result in long delays in product approval, which could have significant costs—not just for product developers but also for the public.[235] The death and injury toll associated with human-driven vehicles continues to be a public health catastrophe, and improved roadway safety remains a top priority for transportation regulators.[236] Most experts agree that HAVs could help reduce these road risks, meaning that significant regulatory delays would have harmful real-world consequences.

Perhaps for that reason, the DOT quietly moved away from its initial consideration of pre-market approval regime for autonomous vehicles. Instead, the agency released a series of guidance documents that mimic the way software upgrades are "versioned" in the tech sector. The DOT's second autonomous vehicle report, released in September 2017, was titled "Automated Driving Systems: A Vision for Safety 2.0," and the third, released in October 2018, was referred to as "Automated Vehicles 3.0."[237] In them, the DOT turned away from preemptive regulatory efforts and toward more flexible, soft-law approaches. This included an array of recommended—but not required—industry best practices. Whereas the old regulatory playbooks were filled with "shall" and "must" requirements, the language of the new soft-law guidance focused more on "should consider" suggestions.

The DOT's reliance on a soft-law approach expanded in 2019 when the agency created the Non-Traditional and Emerging Transportation Technology (NETT) Council.[238] The fact that the agency described the effort as "non-traditional" signaled its continuing departure from past regulatory practices. In 2020, the NETT Council published

231. Ibid.
232. Ibid.
233. Ibid., 95.
234. Ibid., 95-96.
235. Adam Thierer and Caleb Watney, "Comment on the Federal Automated Vehicles Policy," Mercatus Center at George Mason University, Dec. 5, 2016. https://www.mercatus.org/publications/technology-and-innovation/comment-federal-automated-vehicles-policy.
236. U.S. Department of Transportation, "As Part of Major Push to Bring Down Traffic Deaths, USDOT Launches Roadway Safety Call to Action," Feb. 3, 2023. https://www.transportation.gov/briefing-room/part-major-push-bring-down-traffic-deaths-usdot-launches-roadway-safety-call-action.
237. Jennifer Huddleston Skees et al., "'Soft Law' Is Eating the World: Driverless Car Edition," Mercatus Center at George Mason University, Oct. 11, 2018. https://www.mercatus.org/bridge/commentary/soft-law-eating-world-driverless-car.
238. "U.S. Department of Transportation's NETT Council," U.S. Department of Transportation, April 17, 2019. https://www.transportation.gov/policy-initiatives/nett/us-department-transportations-nett-council.

"Pathways to the Future of Transportation"—a guidance document aiming to provide "a clear path for innovators of new, cross-modal technologies to engage with the Department."[239] The report stressed that the new NETT Council "will engage with innovators and entrepreneurs" to strike the balance between continued safety and increased innovation, and, while acknowledging existing agency regulatory authority, it placed a premium on expanding dialogue among affected stakeholders when addressing policy on an ongoing basis. This model relied on ongoing consultation and collaboration with various stakeholders in an attempt to build a rough consensus around a variety of best practices for driverless vehicles.

Thus far, the Biden administration mostly continues to use this soft-law framework, and those guidelines constitute the rough "rules of the road" for autonomous vehicles at the federal level in the absence of any formal legislative action. It remains to be seen whether federal regulators will continue to build on this more agile governance model or instead take a turn toward hard-law-oriented mandates.[240] Major safety or security lapses could change this equation. But even amid some recent autonomous vehicle incidents and investigations, soft-law mechanisms continue to be the norm. Meanwhile, as mentioned above, the NHTSA has used its investigatory power and recall authority to look into Tesla's full self-driving autonomous driving system and has required an over-the-air software update to vehicles with deficiencies.[241]

With the prospects of legislation remaining quite dim on the autonomous vehicle front, a flexible, bottom-up approach will likely continue to be dominant and can be a model for other algorithmic sectors.

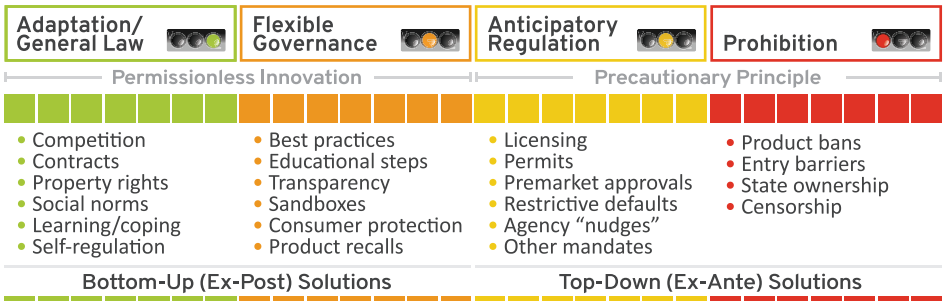
Thus, the United States' current rules of the road for autonomous vehicles are driven by soft law, multi-stakeholder negotiations, industry best practices, agency guidance, existing agency regulatory authority and other agile governance mechanisms. With the prospects of legislation remaining quite dim on this front, this flexible, bottom-up approach will likely continue to be dominant and can be a model for other algorithmic sectors.

What Should Government Do?

This paper has surveyed a broad spectrum of possible responses to AI risk and discussed how more flexible, adaptive and bottom-up governance approaches are often better suited to address rapidly evolving algorithmic concerns. As NIST notes, "flexibility is particularly important where impacts are not easily foreseeable and applications are evolving."[242] Figure 1 attempts to identify the range of governance options along this spectrum. To maximize the potential for algorithmic innovation, the governance default for AI policy should be set closer to the green light of permissionless innovation—a general freedom to innovate—before moving down the spectrum toward more restrictive measures.[243]

239. "Pathways to the Future of Transportation," U.S. Department of Transportation, July 2020, p. iii. https://www.transportation.gov/policy-initiatives/nett/pathways-future-transportation.
240. Adam Thierer and John Croxton, "Elon Musk and the Coming Federal Showdown Over Driverless Vehicles," Discourse, Nov. 22, 2021. https://www.discoursemagazine.com/economics/2021/11/22/elon-musk-and-the-coming-federal-showdown-over-driverless-vehicles.
241. David Shepardson, "Tesla recalls 362,000 U.S. vehicles over Full Self-Driving software," *Reuters*, Feb. 16, 2023. https://www.reuters.com/business/autos-transportation/tesla-recalls-362000-us-vehicles-over-full-self-driving-software-2023-02-16.
242. "Artificial Intelligence Risk Management Framework [Docket Number: 210726-0151]," p. 4. https://www.nist.gov/document/ai-rmf-rfi-comments-underwriters-laboratories.
243. Thierer, "Getting AI Innovation Culture Right." https://www.rstreet.org/research/getting-ai-innovation-culture-right.

Figure 1: Spectrum of Technological Governance Options



The goal of AI policy should be risk mitigation—not a completely unrealistic pursuit to preemptively eliminate all hypothetical risks which could be accomplished only by stopping progress altogether. The sensible governance of AI systems can foster both a culture of innovation as well as a culture of responsibility and resiliency. Iteration and fine-tuning over time will be crucial to build public understanding and acceptance. "Understanding and managing the risks of AI systems will help enhance trustworthiness, and, in turn, cultivate public trust," NIST noted.[244]

Government policy for algorithmic systems should be rooted in humility about the limits of our knowledge of future developments and should appreciate that not every problem can be addressed preemptively. A former acting chair of the FTC put it best when she argued that:

> It is […] vital that government officials, like myself, approach new technologies with a dose of regulatory humility, by working hard to educate ourselves and others about the innovation, understand its effects on consumers and the marketplace, identify benefits and likely harms, and, if harms do arise, consider whether existing laws and regulations are sufficient to address them, before assuming that new rules are required.[245]

Government policy for algorithmic systems should be rooted in humility about the limits of our knowledge of future developments and should appreciate that not every problem can be addressed preemptively.

As a result, forbearance will often be the best first option for AI policy, but regulation will still play an important role, and a wide diversity of remedies already exist that should be tapped before rushing to impose costly new ex-ante regulations.[246]

The other smart role for government would be to act as a facilitator of ongoing dialogue and multi-stakeholder negotiations to solve thorny problems on the fly. This paper identified how government agencies such as the NTIA and NIST have played a crucial role in recent years as conveners of working groups, workshops, roundtables and other discussion fora. Under this approach, government officials can set the stage for discussions and then let various stakeholders develop best practices and solutions as problems arise.[247] Instead of trying to create an expensive and cumbersome new regulatory bureaucracy for AI, the easier approach is to have the NTIA and NIST form a standing committee that brings parties together as needed. These efforts will be informed by the extensive work already done by professional associations, academics, activists and other stakeholders.

244. "Artificial Intelligence Risk Management Framework [Docket Number: 210726-0151]," p. 1. https://www.nist.gov/document/ai-rmf-rfi-comments-underwriters-laboratories.
245. Ibid.
246. Dan Castro, "Ten Principles for Regulation That Does Not Harm AI Innovation," Center for Data Innovation, Feb. 8, 2023. https://itif.org/publications/2023/02/08/ten-principles-for-regulation-that-does-not-harm-ai-innovation.
247. Thierer, "Soft Law in ICT Sectors: Four Case Studies," pp. 79-119. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3777490.

Finally, government actors can also facilitate technology education and awareness-building—sometimes referred to as digital literacy—to help lessen public fears about emerging algorithmic and robotic technologies.[248] "Digital literacy—and improving digital rationality—should be a national strategy," argues one scholar.[249] The goal of such an approach is to foster a healthy balance of trust and skepticism that identifies the trade-offs associated with new technologies and considers sensible responses.[250]

This framework can then be supplemented on an as-needed basis to address more complicated challenges or serious harms as they are identified.[251] Getting this governance balance right—and ensuring that it remains flexible, responsive and pragmatic—is essential if the United States hopes to remain at the forefront of global AI innovation and competitiveness.

Summary of Key Points

- The process of embedding ethics in AI design is not set in stone. Aligning ethics is an ongoing, iterative process influenced by many forces and factors. We should expect much trial and error when devising ethical guidelines for AI and hammering out better ways of keeping these systems aligned with human values.
- Building redundancy and resiliency into AI/ML systems is crucial. The goal is risk mitigation, not the completely unrealistic elimination of all risks.
- A top-down regulatory framework is unwise. It would be folly to imagine that a one-size-fits-all governance solution exists for all AI challenges. A more decentralized, polycentric governance approach is needed—nationally and globally.
- Various organizations are already working together to professionalize the process of AI ethics through sophisticated best-practice frameworks as well as through algorithmic auditing and impact-assessment efforts.
- Decentralized governance efforts build on hard law in many ways. Ex-post enforcement of existing laws and court-based remedies will provide an important backstop when AI developers fail to live up to their claims or promises about safe, effective and fair algorithms.
- Government's best role will be to act as a facilitator of ongoing dialogue and multi-stakeholder negotiations to solve problems as they arise. The NTIA and NIST could form a standing AI working group that brings parties together as needed. Government actors can also help facilitate digital literacy efforts and technology awareness-building to help lessen public fears about emerging algorithmic and robotic technologies.

248. Liana Loewus, "What Is Digital Literacy?," *EducationWeek*, Nov. 8, 2016. https://www.edweek.org/teaching-learning/what-is-digital-literacy/2016/11.
249. Orly Lobel, "The Law of AI for Good," San Diego Legal Studies Paper No. 23-001 (Jan. 26, 2023), p. 68. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4338862.
250. Ibid., p. 69.
251. Adam Thierer, "U.S. Chamber AI Commission Report Offers Constructive Path Forward," R Street Institute, March 9, 2023. https://www.rstreet.org/commentary/u-s-chamber-ai-commission-report-offers-constructive-path-forward.

About the Author

Adam Thierer is a senior fellow in the Technology and Innovation Policy program at the R Street Institute in Washington, D.C.