ADMINISTERING ARTIFICIAL INTELLIGENCE

ALICIA SOLOW-NIEDERMAN*

As AI increasingly features in everyday life, it is not surprising to hear calls to step up regulation of the technology. In particular, a turn to administrative law to grapple with the consequences of AI is understandable because the technology's regulatory challenges appear facially similar to those in other technocratic domains, such as the pharmaceutical industry or environmental law. But AI is unique, even if it is not different in kind. AI's distinctiveness comes from technical attributes—namely, speed, complexity, and unpredictability—that strain administrative law tactics, in conjunction with the institutional settings and incentives, or strategic context, that affect its development path. And this distinctiveness means both that traditional, sectoral approaches hit their limits, and that turns to a new agency like an "FDA for algorithms" or a "federal robotics commission" are of limited utility in constructing enduring governance solutions.

This Article assesses algorithmic governance strategies in light of the attributes and institutional factors that make AI unique. In addition to technical attributes and the contemporary imbalance of public and private resources and expertise, AI governance must contend with a fundamental conceptual challenge: algorithmic applications permit seemingly technical decisions to de facto regulate human behavior, with a greater potential for physical and social impact than ever before. This Article warns that the current trajectory of AI development, which is dominated by large private firms, augurs an era of private governance. To maintain the public voice, it suggests an approach rooted in governance of data—a

*. 2020–2022 Climenko Fellow and Lecturer on Law, Harvard Law School; 2017–2019 PULSE Fellow, UCLA School of Law and 2019-2020 Law Clerk, U.S. District Court for the District of Columbia. Alicia Solow-Niederman drafted this work during her tenure as a PULSE Fellow, and the arguments advanced here are made in her personal capacity. This Article reflects the regulatory and statutory state of play as of early March 2020. Thank you to Jon Michaels, Ted Parson, and Richard Re for substantive engagement and tireless support; to Jennifer Chacon, Ignacio Cofone, Rebecca Crootof, Ingrid Eagly, Joanna Schwartz, Vivek Krishnamurthy, Guy Van den Broeck, Morgan Weiland, Josephine Wolff, Jonathan Zittrain, participants at We Robot 2019, and the UCLAI working group for invaluable comments and encouragement; to Urs Gasser for conversations that inspired this research project; and to the editors of the *Southern California Law Review* for their hard work in preparing this Article for publication. Thanks also to the Solow-Niederman family and especially to Nancy Solow for her patience and kindness, and to the Tower 26 team for helping me to maintain a sound mind in a sound body. Any errors are my own.

fundamental AI input—rather than only contending with the consequences of algorithmic outputs. Without rethinking regulatory strategies to ensure that public values inform AI research, development, and deployment, we risk losing the democratic accountability that is at the heart of public law.

TABLE OF CONTENTS

INTRODUCTION

Popular media coverage about artificial intelligence ("AI")[1] often makes it sound as though the technology itself is an autonomous actor. It's easy to understand the urge to anthropomorphize AI: sometimes, the results of algorithmic research are altogether different from what the data scientists who created the algorithm expected in ways that suggest algorithmic autonomy.[2] Consider, for instance, an algorithm that was told to sort data. Like Amelia Bedelia, the software took this directive a bit too literally and deleted all the data fed to it, apparently on the theory that data that did not exist could not be considered unsorted.[3] But thinking about the algorithm as the actor here is dangerous because it elides the role that humans, and the institutions within which they sit, are in fact responsible in the first instance for the data selection and programmatic choices that create what we call AI today.[4] In the case of the data-sorting algorithm, the algorithm was not thinking independently about how to playfully evade a directive; rather, it was following instructions given to it by data scientists, who had not realized that their algorithmic agent could comply with the spirit but not the intent behind the directive in an unexpected way.

Such algorithmic creativity might, in the right setting, be exciting and generative insofar as it catalyzes new research approaches. As anyone who has grown frustrated with a toddler can attest, though, childlike evasion of a directive is not always funny. Take the researchers who programmed an

1. This Article defines artificial intelligence as a class of technologies that rely on some form of automated decisionmaking executed by a computer. AI, as used in this Article, includes both AI-driven robots and AI algorithms that lack a bodied form, whether they employ machine learning or another method. *See* Jack M. Balkin, *The Path of Robotics Law*, 6 CALIF. L. REV. CIR. 45, 45–46 (2015) ("I do not distinguish sharply between robots and artificial intelligence (AI) agents. As innovation proceeds, the distinction between these two kinds of technologies may be far less important to the law . . . ."); Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN L. REV. 83, 85 n.2 (2017) (noting a "terminological divide in legal scholarship" wherein "[s]ome of the most prominent authors in the field prefer to conceive of algorithmic regulation as the problem of regulating robots" and asserting that "algorithms are the appropriate unit of regulation"). The understanding used in this Article thus applies "regardless of whether [the particular technologies under consideration] are physical robots, such as care robots or driverless cars, or software systems, such as medical diagnosis systems, intelligent personal assistants, or algorithmic chat bots, in real, virtual, contextual, and mixed-reality environments." INST. OF ELEC. & ELEC. ENG'RS, ETHICALLY ALIGNED DESIGN: A VISION FOR PRIORITIZING HUMAN WELL-BEING WITH AUTONOMOUS AND INTELLIGENT SYSTEMS 17, (1st ed. 2019), https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf [https://perma.cc/MC7P-KZJX].

2. *See* Joel Lehman et al., *The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities*, ARXIV 7 (Nov. 21, 2019), https://arxiv.org/pdf/1803.03453.pdf [https://perma.cc/2U68-SCX9] (documenting research results that differed from what data scientists expected).

3. *Id.* at 8; *see also Meet Amelia Bedelia*, AMELIA BEDILIA, https://www.ameliabedeliabooks.com/meet-amelia-bedelia [https://perma.cc/FX8U-YQPN] ("She does exactly what you tell her to do – so be careful what you say.").

4. This Article limits itself to present and near-term AI capabilities, in which humans supervise and are involved with the technology to at least some extent.

aircraft landing simulation to identify alternative ways to decelerate planes. Rather than actually slow down the planes, the algorithm found a loophole: by generating extremely large force calculations during landing, it was possible to overflow the measurement and cause the force to read out as nearly zero.[5] The algorithm thus appeared to be rapidly decelerating when, in reality, it was speeding into a crash landing.[6] Fortunately, the researchers caught the error, and the harm remained virtual.

But imagine that the algorithm is not a simulation. Rather, it is part of the operating code for an autonomous vehicle ("AV") truck that a private company is using to deliver packages or groceries.[7] And that code is created and overseen by the private employees of that company. How can we trust that any errors or unpredictable algorithmic steps will have been tested and detected—before that AV crashes into a human being? There is presently Department of Transportation ("DOT") guidance for safety testing and private self-certification.[8] Yet it is nonbinding. All that really protects the public is faith in the technical abilities of the private firm's employees, along with confidence that the firm will adopt best practices when it comes to safety protocols.

Reliance on self-governance by market players, however, is problematic. In 2019, Boeing's all-too-real 737 Max crashes vividly illustrated the perils of this tactic. By delegating so much of the safety

5. Lehman et al., *supra* note 2, at 11.

6. *Id.* This example is not sui generis. *See, e.g.*, Devin Coldewey, *This Clever AI Hid Data from Its Creators to Cheat at Its Appointed Task*, TECHCRUNCH (Dec. 31, 2018, 3:14 PM), https://techcrunch.com/2018/12/31/this-clever-ai-hid-data-from-its-creators-to-cheat-at-its-appointed-task [https://perma.cc/JKY2-6AZD] ("A machine learning agent intended to transform aerial images into street maps and back was found to be cheating by hiding information it would need later in 'a nearly imperceptible, high-frequency signal.' " (quoting Casey Chu et al., *CycleGAN, a Master of Steganography*, ARXIV (Dec. 16, 2017), https://arxiv.org/pdf/1712.02950.pdf [https://perma.cc/TMU9-GFGB])).

7. *See, e.g.*, Jennings Brown, *UPS Has Been Delivering Cargo in Self-Driving Trucks for Months and No One Knew*, GIZMODO (Aug. 15, 2019, 12:50 PM), https://gizmodo.com/ups-has-been-delivering-cargo-in-self-driving-trucks-fo-1837272680 [https://perma.cc/2TR7-6VQD]; Megan Rose Dickey, *Nuro and Kroger Are Deploying Self-Driving Cars for Grocery Delivery in Arizona Today*, TECHCRUNCH (Aug. 16, 2018, 5:00 AM), https://techcrunch.com/2018/08/16/nuro-and-kroger-are-deploying-self-driving-cars-for-grocery-delivery-in-arizona-today [https://perma.cc/PP5E-WTMG].

8. *See* U.S. DEP'T OF TRANSP., PREPARING FOR THE FUTURE OF TRANSPORTATION: AUTOMATED VEHICLES 3.0 (2018) ("AV 3.0 [p]rovides [n]ew [m]ultimodal [s]afety [g]uidance [that] . . . [a]ffirms the approach outlined in *A Vision for Safety 2.0* and encourages automated driving system developers to make their Voluntary Safety Self-Assessments public to increase transparency and confidence in the technology."). In January 2020, the DOT issued a request for public comments on a draft version of AV 4.0. Notice of Request for Comments: Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0 (AV 4.0), 85 Fed. Reg. 7011 (Feb. 6, 2020). This draft, in keeping with earlier versions of these guidance documents, indicates that the government will "promote voluntary consensus standards as a mechanism to encourage increased investment and bring cost-effective innovation to the market more quickly." *See* NAT'L SCI. & TECH. COUNCIL & U.S. DEP'T OF TRANSP., ENSURING AMERICAN LEADERSHIP IN AUTOMATED VEHICLE TECHNOLOGIES 29 (2020).

certification process to the manufacturer of the plane, the Federal Aviation Administration ("FAA") failed to take a hard look at how changing one automation software component affected the safety of the overall system, including both other aspects of the plane's hardware and software and the training and reactions of its pilots.[9] The Boeing catastrophe should ring alarm bells for those considering the present and near future of AI research and development ("R&D").

When it comes to AI R&D, the contemporary leaders are the same firms that have historically embraced an ethos of "mov[ing] fast and break[ing] things"[10] (Facebook) and asked consumers to trust them not to "be evil"[11] (Google). For instance, at a leading AI conference in 2018, Google published more papers than any other institution, including the Massachusetts Institute of Technology and Stanford University.[12] In the midst of the "techlash"[13] against private social media platforms' poor privacy and data security practices, it is not clear why the public should trust that technology

9. For a detailed exposé, see Darryl Campbell, *Redline: The Many Human Errors that Brought Down the Boeing 737 Max*, VERGE (May 2, 2019, 8:03 AM), https://www.theverge.com/2019/5/2/18518176/boeing-737-max-crash-problems-human-error-mcas-faa [https://perma.cc/EG2D-2557]; Jack Nicas et al., *Changes to Flight Software on 737 Max Escaped F.A.A. Scrutiny*, N.Y. TIMES (Apr. 11, 2019), https://www.nytimes.com/2019/04/11/business/boeing-faa-mcas.html [https://perma.cc/GN52-8XXB].

10. Facebook's internal motto in its early days was "move fast and break things." *See* Seth Fiegerman, *Are Facebook's 'Move Fast and Break Things' Days Over?*, MASHABLE (Mar. 13, 2014), https://mashable.com/2014/03/13/facebook-move-fast-break-things [https://perma.cc/G3XR-KWK4]. Facebook CEO Mark Zuckerberg continued with a less well-known second part of the motto: "[I]f you never break anything, you're probably not moving fast enough." *See id.*

11. Google's code of conduct included variations of this phrase from 2000 to 2018. *See* Kate Conger, *Google Removes 'Don't Be Evil' Clause from Its Code of Conduct*, GIZMODO (May 18, 2018, 5:31 PM), https://gizmodo.com/google-removes-nearly-all-mentions-of-dont-be-evil-from-1826153393 [https://perma.cc/H5H6-UK2R].

12. NATHAN BENAICH & IAN HOGARTH, STATE OF AI REPORT 42 (2019), https://www.slideshare.net/StateofAIReport/state-of-ai-report-2019-151804430 [https://perma.cc/Q3ZQ-4TYK] (reporting fifty-seven first author papers by Google-affiliated authors at NeurIPS 2018, as compared to forty-four by MIT and thirty-eight by Stanford).

13. This term refers to public coverage of and reaction to large companies such as Amazon, Facebook, and Google in the wake of findings that they breached consumers' trust with regard to data privacy and security practices and/or facilitated foreign interference with electoral politics. *See generally* Eve Smith, *A Memo to Big Tech: The Techlash Against Amazon, Facebook and Google—and What They Can Do*, ECONOMIST (Jan. 20, 2018), https://www.economist.com/briefing/2018/01/20/the-techlash-against-amazon-facebook-and-google-and-what-they-can-do [https://perma.cc/7RBD-8RAQ]. For a report on social media's role in permitting foreign interference in elections, see *Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisement*, U.S. HOUSE OF REPRESENTATIVES PERMANENT SELECT COMMITTEE ON INTELLIGENCE, https://democrats-intelligence.house.gov/social-media-content [https://perma.cc/JE2Z-RQV6]. For coverage of privacy and data security controversies and reported losses of user trust, see, e.g., Akiko Fujita et al., *Tech World Experiencing a Major 'Trust Crisis,' Futurist Warns*, CNBC (Mar. 20, 2018, 3:16 AM), https://www.cnbc.com/2018/03/20/tech-world-experiencing-a-major-trust-crisis-futurist-warns.html [https://perma.cc/P98T-8XNC]; Timothy B. Lee, *Zuckerberg: Cambridge Analytica Leak a "Breach of Trust" with Users*, ARS TECHNICA (Mar. 21, 2018, 2:24 PM), https://arstechnica.com/tech-policy/2018/03/zuckerberg-cambridge-analytica-leak-a-breach-of-trust-with-users [https://perma.cc/8G2Q-XAKF].

companies can or will "do the right thing." Nor is it clear that, standing alone, the threat of ex post sanctions through tort law and/or criminal law will influence firms or entire industries,[14] at least outside of egregious cases.

So, given the lack of any real algorithmic governance strategy, if human safety is at stake, then why not step in and regulate by insisting on government oversight with far greater rigor than the FAA's approach with respect to Boeing? Such interventions might take the form of sector-specific regulation within the existing administrative state,[15] a new agency or bureau dedicated to AI, robotics, and/or algorithms,[16] a process-driven agency that focuses exclusively on protecting consumers from specified algorithmic harms,[17] or some combination of these approaches.

The catch is that a traditional public regulatory response, such as an "FDA for algorithms" or a similar agency, is no panacea for AI pathologies. First, generalist policymakers are likely to lack the expertise or resources to make informed governance choices about highly technical digital code.[18] Second, even for specialists or informed generalists with the requisite expertise, AI algorithms are a poor conceptual fit for top-down, prescriptive

14. For an insightful analysis of how tort law might play a role, see generally Rebecca Crootof, *The Internet of Torts: Expanding Civil Liability Standards to Address Corporate Remote Interference*, 69 DUKE L.J. 583 (2019) (proposing how tort law might evolve to address new harms enabled by "Internet of Things" devices).

15. For example, the first recommendation in a 2018 AI Now report urges, "[g]overnments need to regulate AI by expanding the powers of sector-specific agencies to oversee, audit, and monitor these technologies by domain." MEREDITH WHITTAKER ET AL., AI NOW REPORT 2018 4 (2018), https://ainowinstitute.org/AI_Now_2018_Report.pdf [https://perma.cc/K93Z-MXLK].

16. *See* 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 [https://perma.cc/D2Y9-HLXR].

17. *See generally* Tutt, *supra* note 1 (making the case for an "FDA for algorithms"); Ben Schneiderman, Professor, Univ. of Md., Lecture at The Alan Turing Institute Turing Lecture Series: Algorithmic Accountability (May 30, 2017); Michael Segal, *We Need an FDA for Algorithms*, NAUTILUS (Nov. 1, 2018), http://nautil.us/issue/66/clockwork/we-need-an-fda-for-algorithms [https://perma.cc/29L5-LP4U]; *cf.* Calo, *supra* note 16 (advocating creation of commission to "advise on issues at all levels . . . that touch upon the unique aspects of robotics and artificial intelligence and the novel human experiences these technologies generate").

18. *See* Kim Zetter, *Of Course Congress Is Clueless About Tech—It Killed Its Tutor*, WIRED (Apr. 21, 2016, 7:00 AM), https://www.wired.com/2016/04/office-technology-assessment-congress-clueless-tech-killed-tutor [https://perma.cc/B4PJ-EPTM] (lamenting defunding of the Office of Technology Assessment, which previously advised Congress). Technological hearings in Congress from the past few years illustrate the average congressmember's lack of technological understanding. *See, e.g.*, Lorelei Kelly & Robert Bjarnason, *Our 'Modern' Congress Doesn't Understand 21st Century Technology*, TECHCRUNCH (May 6, 2018, 9:44 AM), https://techcrunch.com/2018/05/06/our-modern-congress-doesnt-understand-21st-century-technology [https://perma.cc/M9UB-RS77] (discussing 2018 Facebook hearings); Issie Lapowsky, *The Sundar Pichai Hearing Was a Major Missed Opportunity*, WIRED (Dec. 11, 2018, 4:29 PM), https://www.wired.com/story/congress-sundar-pichai-google-ceo-hearing [https://perma.cc/4UHA-8EWN] (arguing that the House Judiciary Committee "blew it" in its hearing with Google executive).

regulation.[19] Imagine, for instance, a popular consensus that there should be strict premarket clearance of particular AI applications. What, precisely, would such a regime clear for the market? The dominant AI method, machine learning ("ML"),[20] is not fixed in the same way as, for instance, the molecules in a pharmaceutical compound. In ML, a statistical model "learns" to identify a pattern by analyzing training data. This observed pattern is then deployed in a "working algorithm" that applies the predictive model to new data.[21] Significantly, to ensure that the model does not become stale, the working algorithm might be designed to incorporate new data, resulting in updating of the algorithm itself.[22] This dynamism contrasts markedly with static regulatory objects that may be more amenable to prescriptive controls.[23]

19. *See* discussion *infra* Section II.A.2.

20. Machine learning is a subset of artificial intelligence. In ML, a system learns without ex ante, explicit programming. *See* A. L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, 3 IBM J. RES. & DEV. 211, 211 (1959). For more discussion of ML in general, see generally IAN GOODFELLOW ET AL., DEEP LEARNING (2016), http://www.deeplearningbook.org [https://perma.cc/YPP6-RSHK]. For more detail on ML in the law, see generally David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653 (2017) (providing accessible overview of stages of machine learning for legal scholars, and Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043 (2019) (assessing constitutional implications of machine learning in criminal justice contexts); Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87 (2014) (analyzing the potential impact of machine learning on legal practice).

21. For an accessible overview of ML, see Karen Hao, *What Is Machine Learning?*, MIT TECH. REV. (Nov. 17, 2018), https://www.technologyreview.com/s/612437/what-is-machine-learning-we-drew-you-another-flowchart [https://perma.cc/HB9N-HLXY]; *see also* VISHAL MAINI & SAMER SABRI, MACHINE LEARNING FOR HUMANS (Sachin Maini ed., 2017), https://www.dropbox.com/s/e38nil1dnl7481q/machine_learning.pdf?dl=0 [https://perma.cc/5AJ3-RCJ8].

22. *See* Lehr & Ohm, *supra* note 20, at 701–02 (noting that, once they are trained and deployed, many ML algorithms "are not run merely occasionally, but continuously," which requires "continuously feed[ing] new data into the trained algorithms" and explaining that, once scaled-up, running ML algorithms "may also be turned into *online* learning systems . . . in which the algorithms are regularly and automatically re-trained upon the collection of new data"). This approach is a promising way for systems to adapt to changing real-world settings without requiring a data scientist to re-train the entire model. It has already been deployed by companies in commercial applications. *See, e.g.*, Karen Hao, *Car-Hailing Firm Didi Has a New Dispatching Algorithm that Adapts to Rider Demand*, MIT TECH. REV. (Dec. 12, 2018), https://www.technologyreview.com/the-download/612568/car-hailing-firm-didi-has-a-new-dispatching-algorithm-that-adapts-to-rider [https://perma.cc/58S2-RXMQ] (discussing major Chinese AV firm's adaptive dispatching algorithm); Braden Hancock et al., *Learning from Dialogue after Deployment: Feed Yourself, Chatbot!*, ARXIV 1 (June 13, 2019), https://arxiv.org/pdf/1901.05415.pdf [https://perma.cc/YA2Q-VHWE] (discussing "a dialogue agent with the ability to extract new training examples from the conversations it participates in" that Facebook has adopted).

23. To its credit, the Food and Drug Administration has announced "steps to consider a new regulatory framework specifically tailored to promote the development of safe and effective medical devices that use" what it terms "adaptive" or "continuously learning," as opposed to "fixed" or static, AI algorithms. Press Release, U.S. Food & Drug Admin., Statement from FDA Commissioner Scott Gottlieb, M.D. on Steps Toward a New, Tailored Review Framework for Artificial Intelligence-Based Medical Devices (Apr. 2, 2019), https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-steps-toward-new-tailored-review-framework-artificial [https://perma.cc/F22L-8CZ4]. This release accentuates the point, addressed in greater depth *infra* Part II: traditional prescriptive regulatory models do not map neatly onto this emerging technology.

What about, as an alternative, a solution that entails collaborative governance in the form of shared public-private efforts over time?[24] The necessary conditions for this form of governance do not exist today and are unrealistic future targets.[25] Public-private partnerships by definition require an equal public partner. But this prerequisite is missing when it comes to AI. For one, contemporary AI research and development is centered outside of the state, particularly in America.[26] True, since 2018, there have been new national security investments in AI research and development.[27] And on the civilian side, the legislative and executive branches have signaled a renewed interest in AI policy, including the 2019 publication of an American AI R&D national strategy.[28] The stark reality, though, is that the private sector

24. Technology law scholars have not yet robustly explored collaborative governance for AI. As Margot Kaminski observes, the literature on collaborative governance and algorithms tends to be limited to a specific context, such as health law, *see, e.g.*, W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421 (2017) [hereinafter Price, *Regulating Black-Box Medicine*], or copyright law, *see, e.g.*, Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016). *See* Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1535 n.8 (2019). Relatively few works specifically discuss collaborative governance and algorithms. *See* Michael Guihot et al., *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence*, 20 VAND. J. ENT. & TECH. L. 385 (2017); Kaminski, *supra*. This Article goes beyond an initial survey of collaborative governance as a regulatory option and focuses in greater depth on systemic AI development and deployment choices with both "virtual" and "real" consequences, including the prospect of physical harm. Blending analytic tools from administrative law, collaborative governance, and cyberlaw, it is the first account to not only assess the available regulatory toolkit, but also take a hard look at the public-private institutional dynamics and incentives that might make collaborative governance more or less feasible in both the immediate and longer term.

25. For explanation and definition of "governance," see discussion *infra* Section I.A.

26. *See* discussion *infra* Section II.B.2. The analysis in this Article focuses on the United States, recognizing that conditions in other jurisdictions, particularly the European Union and China, differ tremendously with respect to both governing laws and government investments in AI.

27. *See* discussion *infra* Section II.B.2.

28. *See* NAT'L SCI. & TECH. COUNCIL, THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN: 2019 UPDATE (June 2019) [hereinafter 2019 NATIONAL AI R&D STRATEGIC PLAN], https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf. For more discussion of this plan, see *infra* Part III.B.2. Scholars and policymakers have also advocated for a set of consumer protection, public utility, and antitrust approaches to contend with the power of Big Tech firms. *See, e.g.*, TIM WU, THE CURSE OF BIGNESS: ANTITRUST IN THE NEW GILDED AGE (2018); Frank Pasquale, *Tech Platforms and the Knowledge Problem*, 2 AM. AFF. 3 (2018); Lina Khan, Note, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710 (2017); K. Sabeel Rahman, *The New Octopus*, LOGIC MAG., Apr. 1, 2018, https://logicmag.io/04-the-new-octopus [https://perma.cc/44UW-Q4TN]. There has already been policy uptake of such scholarship; for instance, in July 2018, U.S. Senator Mark Warner published a white paper on ways to regulate Big Tech "platforms," including steps to combat disinformation, tactics to protect consumer privacy, and interventions to constrain or control corporate structures through competition law. *See* MARK R. WARNER, POTENTIAL POLICY PROPOSALS FOR REGULATION OF SOCIAL MEDIA AND TECHNOLOGY FIRMS (drft. 2018), https://graphics.axios.com/pdf/PlatformPolicyPaper.pdf [https://perma.cc/PX64-RGJM].

This Article proceeds in parallel to such work. Though the "bigness" of firms or degree of market concentration may be part of the problem insofar as it affects the balance of public and private decisionmaking authority, this Article focuses on the broader governance challenges that AI reveals—challenges that are present to at least some degree regardless of the market power of the firm developing or deploying AI, even if they may be most acute when decision-making authority is concentrated in a outclasses the public sector in terms of both expertise and resources.[29] Even assuming massive public resource investments that do not presently exist, both the speed of development and the highly specialized nature of the technology[30] will make it challenging if not outright Sisyphean to shift the center of gravity away from the private sector.[31] To account for this public-private imbalance, a functional theory of AI governance requires us to pay more attention to the actions of private entities and individuals.

The reason that a private center of gravity matters is because the development path of a particular AI technology is unavoidably bound up in the strategic context, or institutional settings and incentive structures, that inform its creation and deployment.[32] Take, for example, the finding that IBM's AI-powered Watson supercomputer recommended "unsafe and incorrect" cancer treatments.[33] A report on these failures concluded that the problem did not arise from the tool in isolation. Rather, IBM engineers and doctors interacted with and trained the tool in a way that created the problems.[34] And the same reporting suggests that IBM executives pushed forward to market it, despite awareness of its flaws.[35] These errors, then, did not occur merely because of some flaw in the technology, in isolation. Nor did they occur merely because of a single regulatory gap that the administrative state might fill. They reflect a dynamic interaction among technological attributes,[36] regulatory constraints, and choices about how to train the tool that were made by individuals operating within particular

small number of technology companies. For a discussion of code as policy, see discussion *infra* Part III.

29. *See infra* text accompanying notes 197–220.

30. For a discussion of speed and the regulatory challenges of complexity and unpredictability, see discussion *infra* Section II.A.2.

31. *See* discussion *infra* Section II.B.2.

32. This framing is adapted from a May 2018 conference hosted by the Program on Understanding Law, Science, and Evidence at UCLA School of Law and inspired by conversations with Ted Parson and Richard Re. *See AI in Strategic Context: Development Paths, Impacts, and Governance*, UCLA LAW: PULSE (May 7, 2018), https://law.ucla.edu/centers/interdisciplinary-studies/pulse/news/2018/05/ai-in-strategic-context-development-paths-impacts-and-governance [https://perma.cc/4UCA-7HU3].

33. Casey Ross & Ike Swetlitz, *IBM's Watson Supercomputer Recommended 'Unsafe and Incorrect' Cancer Treatments, Internal Documents Show*, STAT (July 25, 2018), https://www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf [https://perma.cc/X6Z3-UPJM].

34. The errors are reportedly due to the use of hypothetical patient data, not real data, in training the algorithm. *See id*.

35. Doctors interviewed for the report did not mince words: "[IBM] should be called out on this . . . . I would bet this is a calculated risk they took. . . . They're kind of messing with people, but it's within the marketing spin that is increasingly allowed these days . . . ." *Id*.

36. The use of the term "technical attributes" does not imply a techno-deterministic lens that assumes technologies possess characteristics apart from their social and political contexts. Quite the opposite: this Article intends to pinpoint what might be technically distinct about AI because of these complex social and political interactions.

business settings.[37]

Attention to the strategic context from which AI applications emerge is especially critical, moreover, given a rapidly-growing body of literature that recognizes how the impact of algorithms can be a significant problem from a fairness, accountability, and transparency perspective.[38] Consider Google's "Smart Compose" email feature, which relies on ML. This "smart" product would not stop associating words such as *investor* or *engineer* with men, to the point that all gender-specific pronouns were removed from the tool in late 2018.[39] This same kind of demographic bias is likely to emerge

37. These points build from Meg Leta Jones' work on the social construction of technology, see Meg Leta Jones, *Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw*, 2018 U. ILL. J.L. TECH & POL'Y 249, to consider these principles in the AI context.

38. *See* Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389, 1394–1406 (2019); Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 774, 780–82 (2019) [hereinafter Mulligan & Bamberger, *Procurement as Policy*]; Kaminski, *supra* note 24 at 1532–37 (providing an overview of algorithmic accountability scholarship) (citing Mireille Hildebrandt, *The Dawn of a Critical Transparency Right for the Profiling Era*, *in* DIGITAL ENLIGHTENMENT YEARBOOK 2012 41 (Jacques Bus et al. eds., 2012); Isak Mendoza & Lee A. Bygrave, *The Right Not to Be Subject to Automated Decisions Based on Profiling*, *in* EU INTERNET LAW 77, 96–97 (Tatiani-Eleni Synodinou et al. eds., 2017); Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973, 983–84 (2016); Kenneth A. Bamberger, *Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State*, 56 DUKE L.J. 377, 385–92 (2006); Lee A. Bygrave, *Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, 17 COMPUTER L. & SECURITY REP. 17 (2001); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93 (2014); Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1 (2017); James Grimmelmann & Daniel Westreich, *Incomprehensible Discrimination*, 7 CALIF. L. REV. ONLINE 164 (2017); Guihot et al., *supra* note 24; Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 61 (2019); Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189 (2017); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633 (2017); Perel & Elkin-Koren, *supra* note 24; Price, *supra* note 24; Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393 (2014); Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 373–76 (2016); Paul Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Computer*, 43 HASTINGS L.J. 1321 (1992); Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1, 64–78 (2005); Tutt, *supra* note 1; Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503)).

39. Paresh Dave, *Fearful of Bias, Google Blocks Gender-Based Pronouns from New AI Tool*, REUTERS (Nov. 26, 2018, 10:06 PM), https://www.reuters.com/article/us-alphabet-google-ai-gender/fearful-of-bias-google-blocks-gender-based-pronouns-from-new-ai-tool-idUSKCN1NW0EF [https://perma.cc/N5HT-6J9F]. According to publicly-available blog posts, this tool works by parsing an extremely large corpus of past emails from many users, identifying trends in those email responses, and then suggesting in real-time how the current emailer might wish to complete a sentence or phrase. It appears to make this recommendation based on its analysis of how people tend to email in combination with signals such as what words the user has typed in the email, the email subject, and any text in the body of prior emails in the chain. *See* Paul Lambert, *Subject: Write Emails Faster with Smart Compose in Gmail*, GOOGLE: THE KEYWORD (May 8, 2018), https://www.blog.google/products/gmail/subject-write-emails-faster-smart-compose-gmail [https://perma.cc/V6BF-DXEH]; Yonghui Wu, *Smart Compose:*

whenever an AI system relies on data sets that draw connections from past human behavior.[40] But all machine learning relies on precisely these kinds of correlations—raising the risk of even more troubling outcomes, such as racial disparities in criminal justice algorithms that are already being used to identify recidivism risks and provide data for pretrial detention and sentencing decisions.[41] Though the code is digital, any harms caused by such systems are not limited to a computer screen. These algorithms are interacting with real-world lived experiences, with harms disproportionately borne by minority or vulnerable populations.[42]

In short, technical decisions about algorithms are not only mediating public safety, but also encoding values, without any uniform oversight, normative requirements, or public accountability.[43] As AI technologies are embedded in more and more applications, given the leading role of the private sector in AI development and deployment, assessing AI in strategic context reveals the extent to which we increasingly live in an age of private governance.

Building from the "governance-by-design" literature, which focuses on how technical decisions by public actors are implementing particular directives, this Article suggests that a fundamental reorientation is required. Functionally, private actors making technical decisions about AI are also making policy decisions. As choices about how to design and deploy an algorithm are regulating[44] human behavior, it is time to recognize what this

Using Neural Networks to Help Write Emails, GOOGLE AI BLOG (May 16, 2018), https://ai.googleblog.com/2018/05/smart-compose-using-neural-networks-to.html [https://perma.cc/S53H-QVXB].

40. *See, e.g.*, Aylin Caliskan et. al, *Semantics Derived Automatically from Language Corpora Contain Human-Like Biases*, 356 SCIENCE 183, 183 (2017) (describing the risk of bias in natural language processing, given human biases); Sandra G. Mayson, *Bias in, Bias out*, 128 YALE L.J. 2218, 2218 (2019) ("In a racially stratified world, any method of prediction will project the inequalities of the past into the future. This is as true of the subjective prediction that has long pervaded criminal justice as it is of the algorithmic tools now replacing it.").

41. *See Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools*, ELECTRONIC PRIVACY INFO. CTR., https://epic.org/algorithmic-transparency/crim-justice [https://perma.cc/K9XH-GMV4] (surveying states that require or recommend risk assessment algorithms).

42. *See generally, e.g.*, VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2017) (underscoring inequities in algorithmic tools); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018) (revealing bias embedded in search engines).

43. *See* discussion *infra* Part III.

44. "Regulation," as used in this Article, does not refer only to formal, top-down regulation promulgated by a policymaker, but rather invokes a broader understanding of regulation in the sense that Lawrence Lessig describes it: "the constraining effect of some action, or policy, whether intended by anyone or not." Lawrence Lessig, *The New Chicago School*, 27 J. LEGAL STUD. 661, 662 n.1 (1998). This definition spans "hard" and "soft" law approaches and also accounts for the built environment, or "architecture," and the potential role of social norms. *See* discussion *infra* Section I.B. This Article uses the term "public regulation" to refer to top-down regulation by the state.

Article terms *code as policy*. When commercial actors outpace public sector resources and expertise in an algorithmic domain, the outcome is de facto private governance. Such private governance, like "governance-by-design" in the public sector,[45] may evade democratic checks at the same time that it fails to provide a clear regulatory rule. Within the terms of traditional governance and administrative law models, we thus face an increasingly stark dilemma: regulate and constrain AI firms far more, ex ante—even if there are major costs to private efficiency or innovative potential—or step away and accept that AI augurs a new order of governance by commercial entities and not by the state.[46] Rather than accept the constraining terms of this public-private dilemma, this Article seeks a third way in the form of regulation of AI through the public governance of data—the resource that powers leading AI methods.

The following analysis of governance options in the age of AI takes a hard look at traditional models and their limits before reframing the solution space to suggest data-centered interventions. Part I offers a brief survey of governance theory in general and theories of regulation of digital technologies in particular. Part II considers the fit between existing public regulatory approaches and AI. Section II.A discusses prescriptive[47] regulation, such as the premarket drug clearance regime of the Food and Drug Administration ("FDA"), and rebuts the case for an "FDA for algorithms." Section II.B considers governance as a public regulatory alternative, looking to environmental regulation as an example. It concludes that AI's highly dynamic, complex, and interdisciplinary challenges are indeed similar to ecosystem management, yet warns that the necessary starting conditions for accountable public-private negotiation are missing because of the commercial sector's lead in AI research and development.

Part III builds from this analysis and argues that, even if there were presently less of a public-private imbalance, fundamental challenges for meaningful public regulation of AI will always remain, given the extent to which algorithmic decisions are de facto policy choices. Invoking lessons from governance and cyberlaw scholarship, Part III calls for recognition of the power of code as policy and closes with a series of suggestions that leverage markets and norms to govern data and thereby account for public

45. *See* sources cited *infra* note 238 and accompanying text.

46. This point extends upon Margaret Radin's leading work on the law of the firm. *See, e.g.*, Margaret Jane Radin, *Regulation by Contract, Regulation by Machine*, 160 J. INSTITUTIONAL & THEORETICAL ECON. 142, 143 (2004); Margaret Jane Radin, *Regime Change in Intellectual Property: Superseding the Law of the State with the "Law" of the Firm*, 1 U. OTTAWA L. & TECH. J. 173, 178–88 (2004).

47. This Article uses the terms "prescriptive regulation" and "command-and-control" interchangeably. Both of these terms refer to public regulatory responses by the state.

input in AI research and development. Only by updating the ways that we are responding to code-based innovation that touches lives, in virtual and physical realms, can we harness the full power of emerging technologies today and administer AI in a way that advances public and private interests alike.

I. BEYOND FORMAL REGULATION

Governance of AI requires strategies to contend with the impact of algorithmic technologies on human lives in the physical and digital world. To situate this Article's forthcoming assessment of the regulatory toolbox for AI, the following Sections provide an overview of administrative theory and cyberlaw theory.

A. FROM REGULATION TO COLLABORATION

Since the late twentieth century, administrative law scholars have grown skeptical of traditional paradigms of regulatory administration.[48] From this point of view, administrative law should move away from a top-down model of public agencies staffed by specialized experts who issue rules or adjudicatory orders to bind private actors.[49] These scholars contend that public regulation in the modern state requires active participation by both governmental and non-governmental actors.[50]

48. *See* JON D. MICHAELS, CONSTITUTIONAL COUP: PRIVATIZATION'S THREAT TO THE AMERICAN REPUBLIC 39–50 (2017) (describing twentieth century "*pax administrativa*" and detailing how far we have deviated from the "administrative era" of the 1930s–1970s). For a classic account of the twentieth century model, see generally JAMES M. LANDIS, THE ADMINISTRATIVE PROCESS (1938).

49. *See* Cynthia Estlund, *Rebuilding the Law of the Workplace in an Era of Self-Regulation*, 105 COLUM. L. REV. 319, 341 n.94 (2005) ("The alternatives to command and control have many variations and varied names . . . . [A]ll of them involve some devolution of regulatory activity to the regulated entities themselves, all aim for greater flexibility, and all struggle with the tension between flexibility and accountability."). As Estlund summarizes, this school of approaches has been called "responsive regulation," *see* IAN AYRES & JOHN BRAITHWAITE, RESPONSIVE REGULATION: TRANSCENDING THE DEREGULATION DEBATE 4–6 (1992); "democratic experimentalism," *see* Michael C. Dorf & Charles F. Sabel, *A Constitution of Democratic Experimentalism*, 98 COLUM. L. REV. 267, 267-70 (1998); " 'contractarian regulation,' *see* David A. Dana, *The New 'Contractarian' Paradigm in Environmental Regulation*, 2000 U. ILL. L. REV. 35, 36; 'collaborative governance,' *see* Jody Freeman, *Collaborative Governance in the Administrative State*, 45 UCLA L. REV. 1, 22 (1997); 'regulatory flexibility,' *see* Marshall J. Breger, *Regulatory Flexibility and the Administrative State*, 32 TULSA L.J. 325, 328 (1996); 'cooperative implementation,' *see* Douglas C. Michael, *Cooperative Implementation of Federal Regulations*, 13 YALE J. ON REG. 535, 540–41 (1996), and 'reconstitutive law,' *see* Richard B. Stewart, *Reconstitutive Law*, 46 MD. L. REV. 86, 108–09 (1986)." Estlund*, supra*.

50. *Cf.* Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342, 344 (2004) ("The new governance model . . . challeng[es] the traditional focus on formal regulation as the dominant locus of change. The model enables practices that dislocate traditional state-produced regulation from its privileged place, while at the same time maintaining the cohesion and large-scale goals of an integrated legal system."). Whether this shift reflects partisan moves to deregulate or principled theoretical evolution is irrelevant to this Article's more practical point: a bevy of new scholarship around governance has emerged since the

Though precise formulation of this alternative governance model varies, the unifying thread in these accounts is the need for an updated public regulatory script that casts the regulator and the regulated in a less adversarial light. For instance, Richard Stewart discusses twenty-first century administrative law as evolving into "government-stakeholder network structures," wherein "regulatory agencies have developed a number of strategies to enlist a variety of governmental and nongovernmental actors, including business firms and nonprofit organizations, in the formulation and implementation of regulatory policy."[51] Jody Freeman uses the term "collaborative governance"[52] to describe contemporary administrative efforts and invokes the metaphor of contracts to situate governance as "a set of negotiated relationships."[53] And Orly Lobel chronicles the fall of "New Deal" regulation and the rise of "Renew Deal" governance, which features a "range of activities, functions, and exercise of control by both public and private actors in the promotion of social, political, and economic ends."[54] This literature thus proposes public regulatory theories and tactics that consist of ongoing public-private collaboration in lieu of top-down commands dictated by a public regulator to constrain a private regulated entity.

B. CODE, LAW, AND REGULATION

Less formal regulatory models have also emerged in contexts outside of administrative law. In the late 1990s and early 2000s, legal scholars confronted with the rise of the Internet as a mass medium developed their own distinct theory of regulation. Rather than assess how the state interacted with private entities, they articulated how digital programming strings—"code"—functions as a regulatory modality that both creates and controls the online world.[55] These scholars, particularly Lawrence Lessig, established a theory of regulation that flows through code, both directly and indirectly.[56] In Lessig's words: "[i]n real space, we recognize how laws regulate—through constitutions, statutes, and other legal codes. In cyberspace we must

pax administrativa era.

51. Richard B. Stewart, *Administrative Law in the Twenty-First Century*, 78 N.Y.U. L. REV. 437, 448 (2003).

52. Freeman, *supra* note 49.

53. Jody Freeman, *The Private Role in Public Governance*, 75 N.Y.U. L. REV. 543, 571 (2000).

54. Lobel *supra* note 50, at 344.

55. *See* LAWRENCE LESSIG, CODE: VERSION 2.0, at 5–7 (2006) [hereinafter LESSIG, CODE V2.0]; *see also* James Grimmelmann, Note, *Regulation by Software*, 114 YALE L.J. 1719 (2005); Lawrence Lessig, Cyberspace's Constitution, Lecture at the American Academy, Berlin, Germany (Feb. 10, 2000), https://cyber.harvard.edu/works/lessig/AmAcd1.pdf [https://perma.cc/RA9Q-MJ2B].

56. Lessig's development of this framework builds from Joel Reidenberg's important earlier contributions to the field. *See* Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553 (1998).

understand how a different 'code' regulates—how the software and hardware (i.e., the 'code' of cyberspace) that make cyberspace what it is also regulate cyberspace as it is."[57] In cyberspace, "code is law."[58]

This thesis draws from an earlier piece, *The New Chicago School*, in which Lessig introduced a four-part model that positions law, norms, markets, and architecture[59] as forces that affect the human beings at the center of the regulatory equation.[60] Lessig explains that each of these modalities is regulatory, individually and cumulatively, not in the more traditional understanding of the term "regulation," but rather because it exercises a "constraining effect . . . [on] some action, or policy, whether intended by anyone or not."[61] Within this model, in other words, each of the regulatory modalities constrains in the sense that it simultaneously creates and limits the possibilities for the entity at the center of the model. Regulation, then, consists of more than top-down administrative and legislative constraints and sanctions. It is the net product of all the forces that act on human beings.

In this model, code operates as architecture as well as law.[62] Code constructs digital realms in a literal sense: the digital bits that make up strings of code create online environments.[63] Code also determines the affordances and limitations of those realms by dictating what an Internet user can or cannot do in a particular online setting.[64] As Lessig explains, there are no laws of physics in cyberspace.[65] And because there are no laws of nature in virtual space, code also creates all of the fundamental parameters that apply

57. LESSIG, CODE V2.0, *supra* note 55, at 5.

58. *Id*. (citing WILLIAM J. MITCHELL, CITY OF BITS 111 (1995); Reidenberg, *supra* note 56).

59. By "architecture," Lessig referenced the natural and built environment that constrains and/or enables human behavior, such that this category encompasses the cumulative effect of found environments, past planning, design, and investment decisions that create the world around us. Lessig, *supra* note 44, at 663 ("I mean by 'architecture' the world as I find it, understanding that as I find it, much of this world has been made. . . . [F]eatures of the world—whether made, or found—restrict and enable in a way that directs or affects behavior. They are features of this world's architecture . . . .").

60. *Id*. at 664; *see also* LESSIG, CODE V2.0, *supra* note 55, at 122–24. For an overview of each of the modalities, see Lessig, *supra* note 44, at 662–63.

61. Lessig, *supra* note 44, at 662 n.1.

62. *See* LESSIG, CODE V2.0, *supra* note 55, at 4–6.

63. As Lessig explains, "Code is a regulator in cyberspace because it defines the terms upon which cyberspace is offered." *See id*. at 84. Operating as architecture, code determines the values of the space and can "change[] the mix of benefits and burdens that people face[]" when they interact within the space. *Id.* at 87.

64. *See id.* at 114 ("[C]ode embeds values. It enables, or not, certain control. And . . . it is also a tool of control . . . .").

65. *Cf. id.* at 15 (" '[O]n the Internet' and 'in cyberspace,' technology constitutes the environment of the space, and it will give us a much wider range of control over how interactions work in that space than in real space.").

to life within that environment.[66] Code, then, could be said both to constitute the online world and to represent the constraints and controls that govern cyberspace.[67]

* * *

More than ever before, individuals experience the tangible and intangible effects of algorithms on the world—online and offline.[68] Yet cyberlaw theory and technology law have not adequately accounted for the ways that digital technologies themselves shape and regulate physical space, for better and for worse. There have been some important partial steps, to be sure. Significantly, scholars such as Ryan Calo have asserted that robotics has such an effect because it combines "the promiscuity of information" with a "physical" impact.[69] This point, however, does not go far enough in recognizing the role that code plays in contemporary lived experiences.

Code can touch physical life, whether or not an algorithm takes an embodied form.[70] This is a lived reality today because contemporary life is not virtual or physical. When we stroll down the street and text someone across the country, while ignoring someone right next to us, which world do

66. *See id.*

67. Lessig's account is both descriptive and normative: he is concerned with the way that the state can use code to exercise indirect regulatory control when it might otherwise have to regulate directly through law. For a visual depiction of this model and further elaboration, see *id.* at 130.

68. Within the original cyberlaw framework, "code" is an online, or cyberspace, modality of regulation, with incidental impacts in offline, or physical space. Lessig's original account does contemplate how the outcomes of virtual interactions may affect individuals in the physical world. *See, e.g.*, *id.* at 114–15 (discussing "the code of digital technologies" programmed into video cassettes, and thereby affecting consumer behavior in the physical world). Others have also addressed interactions between virtual and physical realities. For instance, David C. Clark, who served as the chief protocol officer for the Internet's development in the 1980s, maintains that it may be erroneous to consider cyberspace wholly separate from "real" space. *See* DAVID CLARK, CHARACTERIZING CYBERSPACE: PAST, PRESENT, AND FUTURE 5 (2010), http://docshare01.docshare.tips/files/9608/96080638.pdf [https://perma.cc/R2HM-RHJF] ("[T]he right image for cyberspace may be a thin veneer that is drawn over 'real' space, rather than a separate space that one 'goes to.' "). But in the original theoretical model, architecture is "real-space code," Lessig, CODE V2.0, *supra* note 55, at 342, and code is digital-space architecture, with the two realms pitched as descriptively distinct.

In addition, though beyond the scope of this Article, the idea that "cyber" and "real" are distinct zones has long elided sociological nuance about the ways in which individuals' interactions in cyberspace entail complex negotiations with previous, "real" world understandings of cultural identity, socioeconomic status, and political empowerment. *See, e.g.*, CYBERGHETTO OR CYBERTOPIA?: RACE, CLASS, AND GENDER ON THE INTERNET (Bosah Ebo ed., 1998); EUBANKS, *supra* note 42; RACE AFTER THE INTERNET (Lisa Nakamura & Peter A. Chow-White eds., 2012); RACE IN CYBERSPACE (Beth E. Kolko et al. eds., 2000).

69. *See, e.g.*, Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 515, 549–62 (2015) ("Robotics combines, arguably for the first time, the promiscuity of information with the capacity to do physical harm.").

70. *See* discussion *supra* note 1 and sources cited *infra* note 77; *see generally* KAI-FU LEE, AI SUPERPOWERS (2018) (discussing what he terms "online-merge-of-offline" as the digital and physical worlds combine).

we inhabit?[71] Algorithmic technology's potency is particularly stark when it comes to artificial intelligence. AI by definition places code-driven autonomous and intelligent systems—from moderation of social media content to consumer applications like AVs to medical applications like algorithmic diagnostics—directly into social and political interactions in the physical world.[72]

Perhaps because digital technology questions have seemed "virtual," legal scholarship has not focused sustained attention on how emerging digital technologies do or do not fit within either prescriptive administrative law or newer governance paradigms. There are connections to a number of existing literatures, including normative work regarding democratic governance and technology, sociological work interrogating the interplay between social values and technology, interdisciplinary work on governance of emerging technology and science more generally,[73] and prior legal scholarship assessing how existing law applies to the Internet[74] and cyber-physical systems,[75] as well as how cyberspace might affect traditional notions of state sovereignty and jurisdiction.[76] And legal scholars have begun an important conversation that interrogates the ways in which contemporary technological developments challenge longstanding assumptions. For instance, Mireille Hildebrandt has assessed the ways in which technology may demand fundamental changes in our conception of the law itself.[77] Others, such as

71. *See generally* Alicia Solow-Niederman, *Emerging Digital Technology and the "Law of the Horse*," UCLA L. REV. (Feb. 19, 2019), https://www.uclalawreview.org/emerging-digital-technology-and-the-law-of-the-horse [https://perma.cc/W47V-7RDX] (assessing status of technology law in a digitally-connected world).

72. For more detail, see discussion *supra* note 1 and sources cited therein.

73. *See, e.g.*, *Conference on Governance of Emerging Technologies & Science*, ARIZ. STATE U. SANDRA DAY O'CONNOR C. OF L. [hereinafter *GETS Conference*], http://events.asucollegeoflaw.com/gets [https://perma.cc/778Y-ZGR8].

74. For the classic debate, compare Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207 (suggesting that cyberlaw is about as real, and about as useful, as a "law of the horse"), *with* Lawrence Lessig, Commentary, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501 (1999) (suggesting that a close study of cyberspace yields distinct lessons for law and regulation).

75. *See* Crootof, *supra* note 14, at 587–88 nn.13–15 (surveying "Internet of Things" scholarship).

76. *See generally, e.g.*, Jack L. Goldsmith, *Against Cyberanarchy*, 65 U. CHI. L. REV. 1199 (1998); David G. Post, *Against "Against Cyberanarchy*," 17 BERKELEY TECH. L.J. 1365 (2002); Timothy S. Wu, Note, *Cyberspace Sovereignty?—The Internet and the International System*, 10 HARV. J.L. & TECH. 647 (1997).

77. *See generally* MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW (2015). Hildebrandt uses the term "onli*fe*" to refer to "a transformative life world, situated beyond the increasingly artificial distinction between online and offline." *Id.* at 8. Whereas Hildebrandt focuses on how these developments affect the rule of law, this Article emphasizes the related but distinct question of how governance models can contend with the probable development path and risk trajectory of AI, in light of both the regulatory and market status quo in the United States and the policy challenges that the technology presents. *See also* JULIE E. COHEN, CONFIGURING THE NETWORKED SELF (2012) (analyzing laws and technologies that control the flow of information about individuals and considering how the

Deirdre Mulligan and Kenneth Bamberger, have exposed ways in which using technology as a regulatory tool may compromise democratic accountability and rule of law norms, particularly when the technology is deployed by the state itself.[78] But in general, governance and cyberlaw analyses tend to occur in parallel streams of legal scholarship.[79] In particular, the broader collaborative governance discussion has not sufficiently engaged with cyberlaw theory or its implications for emerging digital technologies, especially when it comes to algorithmic interventions.[80] These dialogues can and should intersect, and this Article's analysis of AI governance options aims to kickstart this conversation.

II. ADMINISTRATIVE PARADIGMS

Before turning to less traditional governance approaches, it makes sense to start with the tried and true. How might administrative law tackle AI

"networked self" functions in contemporary society); Julie E. Cohen, *Cyberspace as/and Space*, 107 COLUM. L. REV. 210, 212–213 (2007) (underscoring the constructed social processes through which cyberspace users situated in the physical world experience the virtual world); Orin S. Kerr, *The Problem of Perspective in Internet Law*, 91 GEO. L.J. 357, 357 (2003) (dissecting legal issues posed by the Internet's simultaneous presentation of an "internal" perspective taken from the "viewpoint of virtual reality" and an "external perspective" taken from the "viewpoint of physical reality").

78. Mulligan & Bamberger, *Procurement as Policy*, *supra* note 38, at 781 (recognizing how "decisions about goals, values, risk, and certainty, along with the elimination of case-by-case discretion, inherent in machine-learning system design create policies . . . [with] no public participation, no reasoned deliberation, and no factual record, which abdicates Government responsibility for policymaking"); Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-By-Design*, 106 CALIF. L. REV. 697, 705–14 (2018) [hereinafter Mulligan & Bamberger, *Saving Governance*] (offering detailed literature review of science and technology studies ("STS") theory and its salience for public regulation); Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 708–11, 711 n.231 (2010) (raising normative questions about the accountability and political legitimacy of governing through code-based interventions and considering the ways in which technology can shape a "world view" or "frame"); *see also* THE OXFORD HANDBOOK OF LAW, REGULATION, AND TECHNOLOGY 7–14 (Roger Brownsword et al. eds., 2017) (compiling a cross-disciplinary set of essays centered on "[t]he 'disruptive' potential of technological innovation"); REGULATING TECHNOLOGIES 1–218 (Roger Brownsword & Karen Yeung eds., 2008) (compiling essays on the use of technology as a regulatory tool); Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121, 122–23 (1980) (situating technology as socially mediated and discussing the ways in which "artifacts can contain political properties"). This piece builds from Mulligan and Bamberger's incisive work to suggest that questions concerning democratic accountability for technological decisions emerge not only in the context of administrative decisions, *see, e.g.*, Mulligan & Bamberger, *Procurement as Policy*, *supra* note 38, at 812–22, but also in the context of seemingly quotidian design choices in AI products offered on the private market, *see* discussion *infra* Part III.

79. There are conversations about governance *by* technology, *see* sources cited *supra* notes 38, 78 (assessing need for accountability in algorithmic decisions and analyzing use of technology to regulate, respectively), and the Internet itself represents a decentralized form of governance, *cf.* Jonathan Masters, *What Is Internet Governance?*, COUNCIL ON FOREIGN REL. (Apr. 23, 2014), https://www.cfr.org/backgrounder/what-internet-governance [https://perma.cc/8T88-RKRJ]. And there are some moves to assess challenges of emerging technologies. *See, e.g.*, *GETS Conference*, *supra* note 73. Missing, though, is a robust body of legal scholarship that synthesizes collaborative governance theory and foundational cyberlaw theory to distill lessons about governance of emerging digital technologies like AI.

80. *See* discussion *supra* note 24 and sources cited therein.

governance? This Part considers two alternative administrative law approaches that public actors could apply to an emerging technology like AI: prescriptive regulation and collaborative governance. It first provides a stylized summary of each model and then explores each model's limits for AI.

A. PRESCRIPTIVE REGULATION

1. Pharmaceutical Clearance by FDA

The Food and Drug Administration, as the name suggests, regulates food and drugs.[81] Federal regulation of the American food and drug industry predates FDA's creation and is traceable to the Pure Food and Drug Act of 1906, inspired in part by muckraking journalism that exposed unsavory factory conditions and the threats that questionable products posed to the public.[82] By the 1930s, a new wave of concerns and controversies[83] led Congress to pass the Food, Drug, and Cosmetics Act of 1938, which established FDA as a federal "citizen-protection agency."[84]

In regulating drugs, FDA's ambit has been to permit commercial development of life-changing medical offerings while simultaneously intervening to protect citizens when the products brought to market threatened health and safety.[85] FDA's drug regulation involves a "command-and-control" tactic that requires industry players to meet certain safety and efficacy thresholds before permitting them to market drugs to the general

81. This Article considers drug regulation and leaves FDA's regulation of food and medical devices for separate study. *See, e.g.*, Margaret Gilhooley, *FDA and the Adaptation of Regulatory Models*, 49 ST. LOUIS U. L.J. 131, 132 (2004). Section II.A's stylized summary focuses primarily on policy challenges and patterns of administrative law responses.

82. *See Part I: The 1906 Food and Drugs Act and Its Enforcement*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/AboutFDA/History/FOrgsHistory/EvolvingPowers/ucm054819.htm [https://perma.cc/WCY7-WDGX] (last updated Apr. 24, 2019) ("This act . . . prohibited the interstate transport of unlawful food and drugs under penalty of seizure of the questionable products and/or prosecution of the responsible parties. The basis of the law rested on the regulation of product labeling rather than pre-market approval.").

83. An especially prominent and tragic case involved the marketing of a sulfa drug, Elixir Sulfanilamide, that contained a toxic chemical similar to antifreeze. The untested drug claimed over one hundred victims, including many children. *See* Rebecca S. Eisenberg, *The Role of the FDA in Innovation Policy*, 13 MICH. TELECOMM. & TECH. L. REV. 345, 345–46 nn.1–2 (2007) (citing PHILIP J. HILTS, PROTECTING AMERICA'S HEALTH (2003)); *Part II: 1938, Food, Drug, Cosmetic Act*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/AboutFDA/History/FOrgsHistory/EvolvingPowers/ucm054826.htm [https://perma.cc/8TA4-9VCE].

84. HILTS, *supra* note 83, at xi.

85. *See id.* at xii (describing FDA as "the people's investigator" and suggesting that the agency stood for "the principle that it was now the job of government not just to champion commerce but also to intervene when it got out of hand"). Early on, these efforts focused specifically on misbranding and adulteration of drugs in interstate commerce. *How Did the Federal Food, Drug, and Cosmetic Act Come About?*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/AboutFDA/Transparency/Basics/ucm214416.htm [https://perma.cc/Q384-PQXU].

public.[86] The platonic vision is an agency that can rely on empirical evidence and rigorous scientific testing to ensure that privately-created products neither defraud the public nor threaten individuals' physical wellbeing.[87] This vision requires FDA officials to possess both expertise and access to proprietary information in order to parse empirical evidence and operate as a meaningful check on industry claims.

These basic tenets have remained intact since the New Deal, with statutory expansions of FDA's authority over time.[88] Notably, responding in part to a spate of deaths and birth defects linked to a particular drug,[89] the Kefauver-Harris Amendments of 1962 increased FDA's pre-market authority.[90] Rather than putting the initial burden on FDA to screen submissions and adopting the default of market entry unless FDA actively barred it, these amendments required companies to provide "substantial" evidence of a drug's safety before FDA would clear the drug for market. In other words, this intervention changed the market for drugs from a premarket notification system to a premarket approval system.[91] Approval, moreover, required "adequate" and "well-controlled" scientific investigations carried out by "experts qualified by scientific training."[92] The responsibility of establishing a drug's safety thus shifted to the private manufacturer, with the company required to affirmatively demonstrate a product's safety and efficacy for its stated purpose in order to obtain market clearance.[93] Assuming that there is no agency capture or undue special interest influence, the operational premise is that FDA can use scientifically-verified, empirical testing as a way to protect the public's safety—without outright stopping innovative drugs from reaching the market if proper evidence is provided.

86. *See* Eric R. Claeys, *The Food and Drug Administration and the Command-and-Control Model of Regulation*, 49 ST. LOUIS U. L.J. 105, 107–111 (2004); Gilhooley, *supra* note 81, at 131; Price, *supra* note 24, at 424.

87. *See* HILTS, *supra* note 83, at xii, 93, 104–07. In reality, this vision must contend with the significant threat of agency capture by the regulated firms. *See generally* PREVENTING REGULATORY CAPTURE: SPECIAL INTEREST INFLUENCE AND HOW TO LIMIT IT (Daniel Carpenter & David A. Moss, eds., 2014) (addressing risk of regulatory capture in administrative bodies).

88. *See* Claeys, *supra* note 86, at 105–06.

89. *See* HILTS, *supra* note 83, at 154–58 (describing thalidomide controversy and reporting evidence that drug caused birth defects in up to eight thousand babies worldwide, with an estimated five thousand to seven thousand additional pre-birth deaths).

90. *See id*. at 161–65.

91. *See* Richard A. Merrill, *The Architecture of Government Regulation of Medical Products*, 82 VA. L. REV. 1753, 1764–67 (1996).

92. HILTS, *supra* note 83, at 160, 164–65.

93. *Id.* at 164.

2. Against "Command-and-Control" for AI

Why not deploy a similar model for AI? On the surface, FDA demonstrates that other areas of the administrative state have needed to contend with complex blends of technocratic topics, cross-cutting incentives, dynamic information, and commercial actors. This Section assesses how AI's technical attributes and associated policy challenges[94] make this approach problematic, above and beyond the political economy and agency capture questions that plague even comparatively simple regulatory contexts. First, the potential speed of algorithmic development leads to an especially stark instance of the classic "pacing problem" in regulation. Second, any policy intervention must contend with the technology's complexity, including both the need for domain expertise and barriers to interpretability. Third, any policy intervention must grapple with AI's unpredictability, which raises questions of both uncertainty and emergence.[95] Given these factors, even recognizing that similar challenges arise to some extent in the context of pharmaceutical regulation, this Section concludes that AI is meaningfully unique in ways that counsel against command-and-control approaches as a universal tactic for AI-based applications.

a. Speed

AI development and deployment might implicate speed in at least three ways: (1) initial creation of the end product or service; (2) subsequent adjustment of the algorithm in ways that affect the overall product or service; and (3) the time and resources required for algorithmic computation to occur. The relationship between AI and computing power warrants separate analysis,[96] and the focus of this Article is on the first two dimensions, with

94. This Article uses the term "policy challenge" or "regulatory challenge" to underscore that any technology emerges in social context. *See* Balkin, *supra* note 1, at 45 ("I do not think it is helpful to speak in terms of 'essential qualities' of a new technology that we can then apply to law. . . . [W]e should try not to think about characteristics of technology as if these features were independent of how people use technology in their lives and in their social relations with others.").

95. *Cf.* Calo, *supra* note 69, at 538–45 (discussing emergence in the context of robotics).

96. There are significant questions about how particular hardware advances, notably quantum computing, might revolutionize the path of technical development. *See, e.g.*, Tim Hwang, *Computational Power and the Social Impact of Artificial Intelligence*, ARXIV (Mar. 23, 2018), https://arxiv.org/pdf/1803.08971.pdf [https://perma.cc/Q7WP-FS5U]; *cf. Speed Conference*, CORNELL TECH, https://www.dli.tech.cornell.edu/sp0eed [https://perma.cc/9QY5-GFS8] (focusing on computing speed in particular applications and domains). Recent research underscores the link between hardware configuration and AI advances. *See* Yanping Huang et al., *GPipe: Efficient Training of Giant Neural Networks Using Pipeline Parallelism*, ARXIV (July 25, 2019), https://arxiv.org/abs/1811.06965 [https://perma.cc/Z55B-YYBA] ("Scaling up deep neural network capacity has been known as an effective approach to improving model quality for several different machine learning tasks."); *see also Google Obtains New ImageNet State-Of-The-Art Accuracy with Mammoth Networks Trained Via 'Gpipe' Infrastructure,* IMPORT AI (Nov. 26, 2018), https://us13.campaign-archive.com/?u=67bd06787e84d73db24fb0aa5&id=6d61d65ae0&e=[UNIQID] [https://perma.cc/N6GH-TBWX] (observing "trend of research bifurcating into large-compute and small-compute domains" and suggesting possible consequences).

an emphasis on settings in which AI is applied in a tangible product.[97]

Algorithmic development might occur more rapidly than past innovation cycles in part because the software on which AI runs can be developed, erased, and re-created with relatively fewer investments in physical infrastructure and resources.[98] True, considerable capital investment and research is required to build an AI algorithm, and the system requires a computer on which to run. Nonetheless, for technical innovation in general—and thus for AI in particular—producing software requires fewer upfront physical resources than those demanded by other sectors. And once developed, algorithms can be deployed an infinite number of times without further resource investments in a way that is not possible for non-digital innovative products.

Compare the high costs of pharmaceutical research and development and drug manufacturing. Any would-be drug manufacturer must make sizeable resource investments to create and test a drug,[99] including paying for laboratories and scientists' time. Then, once a drug is approved for the market, creating the infrastructure for mass-market drug development requires sizeable upfront physical investments and resource deployment[100] as well as ongoing investments to continue production at scale. In contrast, the core resources required to implement an algorithm are digital, not physical, and—assuming access to adequate processing power and electricity—it can be run an infinite number of times, in an infinite number of locations.

Even where physical hardware is required to run the algorithm, moreover, software updates can often change the manner in which it functions without needing to start from scratch.[101] With access to adequate data on which to run an algorithm, it is possible to create and change the

97. This is a pragmatic move with two advantages. One, it permits a more granular and concrete analysis of the relationship among AI technologies, market pressures, and social norms. Two, it fosters more direct dialogue between this Article and legal scholarship that has focused on embodied algorithms, or robots. *See supra* note 1.

98. The word "code" is sometimes used to refer to these strings of programming text; however, this Section uses a different term to denote software because Lessig's original account defines "code" to refer to both software and hardware. *See supra* note 57 and accompanying text. The category "code" thus includes the algorithms discussed here.

99. *See* Joseph A. DiMasi et al., *The Price of Innovation: New Estimates of Drug Development Costs*, 22 J. HEALTH ECON. 151, 166, 180–83 (2003) (finding a total preapproval cost estimate of US $802 million (in year 2000 dollars) to develop a new drug). *But see* MERRILL GOOZNER, THE $800 MILLION PILL 237–39 (2004) (contesting the cost of drug development).

100. *See* W. Nicholson Price II, *Making Do in Making Drugs: Innovation Policy and Pharmaceutical Manufacturing*, 55 B.C. L. REV. 491, 498–500 (2014) ("Overall, drug manufacturing makes up a very large portion of industry expenses across the different types of pharmaceutical firms.").

101. *See* Paul Ohm, Commentary, *We Couldn't Kill the Internet if We Tried*, 130 HARV. L. REV. F. 79, 81 (2016).

algorithm far more quickly as compared to other sectors, which require large investments of capital and sustained periods of development to implement a similar scale of change.[102] As Paul Ohm explains, the "change-per-effort ratio" is far lower for software, or algorithmic, construction as compared to past industrial processes.[103]

Particularly in the AI context, such algorithmic characteristics can also affect the nature of the consumer-facing product. Seeing how requires focusing on the nature of AI development and application. AI algorithms are not products on their own terms; rather, they consist of building blocks for other goods or services. Consider, for instance, an algorithm that gathers data from onboard sensors to determine how to steer an AV. This algorithm is a building block that becomes useful when it is applied in the context of a vehicle. It is akin to electricity, not a lamp.

Because of this utility-like nature of AI, the algorithm is typically embedded in another end product. And because the software is not the whole product, a manufacturer can change the way the product functions by adjusting a software parameter. An AV manufacturer, for instance, could adjust the software and change whether the emergency brake engages while the vehicle is operating in computer-assist mode.[104] This algorithmic alteration would fundamentally alter the product's functionality. Even if competitive considerations weigh heavily before a company chooses to implement such a change, the company's developers can adjust the underlying technical programming and thereby alter the final product.[105] This adjustment is far faster and easier than, say, attempting to modify the active molecules in a drug.

102. Because data is so critical for machine learning in particular, it can be an important stopgap where it is a scarce resource. This Article thus focuses on data governance policies as a particularly fruitful source of initial intervention points. *See* discussion *infra* Part III.B.2.

103. Ohm, *supra* note 101 ("To effect massive, structural, fundamental change to an operating code base, software developers need not erect new scaffolding, dismantle old structures, or create new blueprints . . . . [C]ompared to the industrial processes it tends to replace, coding is far more efficient and far less onerous, in a strict change-per-effort ratio.").

104. *Cf.* Colin Dwyer, *NTSB: Uber Self-Driving Car Had Disabled Emergency Brake System Before Fatal Crash*, NPR: THE TWO-WAY (May 24, 2018, 8:48 PM), https://www.npr.org/sections/thetwo-way/2018/05/24/614200117/ntsb-uber-self-driving-car-had-disabled-emergency-brake-system-before-fatal-cras [https://perma.cc/3YQ7-ZDTU] (discussing a National Transportation Safety Board ("NTSB") report that software in autonomous vehicle "did not engage the brakes on its own" because the company, Uber, had changed the factory-equipped settings to disable the automatic emergency braking function when the vehicle was in use). For further discussion of this accident and the NTSB's investigation, see *infra* text accompanying notes 155–61 and sources cited therein.

105. The change may not always be implemented; to the contrary, what Rebecca Crootof calls "technological–legal lock-in" may well thwart change. *See* Rebecca Crootof, *"Cyborg Justice" and the Risk of Technological–Legal Lock-In*, 119 COLUM. L. REV. F. 233, 235, 246–50 (2019). The point here is to suggest that more rapid change is possible, compared to other domains, because the relationship between inputs and outputs is distinct for algorithms.

Given the potential for such adjustments to proceed with fewer physical limitations, it might seem that prescriptive regulation is exactly what is needed to slow down any feverish development. Sometimes friction can enhance safety. And if it is cheaper and easier to change an algorithm than a physical product, then what excuse is there for a company that fails to get it just right?

The snag is that speed also amplifies the classic "pacing problem"[106] for regulation of emerging technologies. Even where it is productive to introduce friction points to enhance safety, there is a tradeoff between slowing down technical development to meet a regulatory prescription and allowing unthrottled innovation to occur. Furthermore, particularly because other algorithmic interventions are likely to speed ahead even if one category is more rigorously regulated, the more rapid pace of technical change risks compounding any gap between law on the books and the state of technology in the world.[107] There might well be narrow domains where we decide that the risks are so great that it is worth curtailing development in that zone.[108] But across the board, barring far more extensive state control of market development, speed will naturally accelerate product cycles and act as a force multiplier of any AI challenges. And as the next Sections explore, in light of other factors that make AI unique, it makes sense to pause before relying on premarket clearance measures as an overarching regulatory strategy.

b. Complexity

The technical complexity of AI introduces a number of additional potential regulatory challenges. Two dimensions are especially salient: (1) interpretability, including both inexplicability and incommensurability with traditional ways of understanding the world, and (2) domain expertise,

106. *See* Gary E. Marchant, *The Growing Gap Between Emerging Technologies and the Law*, *in* THE GROWING GAP BETWEEN EMERGING TECHNOLOGIES AND LEGAL-ETHICAL OVERSIGHT: THE PACING PROBLEM 19, 19 (Gary E. Marchant et al. eds., 2011) ("[T]he legal frameworks that society relies on to regulate and manage emerging technologies have not evolved as rapidly [as the technologies themselves] . . . . The consequence of this growing gap between the pace of technology and law is increasingly outdated and ineffective legal structures, institutions[,] and [regulatory] processes . . . ." (citations omitted)).

107. Though international considerations are reserved for future work, this issue becomes starker on a global scale. Even if the United States imposed strict regulatory controls, so long as there are open borders, other countries would ostensibly speed ahead, and the technology would reach U.S. markets. The two interventions that might change this dynamic—global government or strict import controls for all algorithmic technologies, globally—seem unlikely.

108. Delineating such zones is beyond the scope of this Article, which reserves for future work further consideration of whether there are subzones of AI development in which more prescriptive regulation remains the best method of governance. *Cf.* Finale Doshi-Velez et al., *Accountability of AI Under the Law: The Role of Explanation*, ARXIV 2–4, 19–21 (Dec. 20, 2019), https://arxiv.org/pdf/1711.01134.pdf [https://perma.cc/7N5E-42X6] (arguing that different levels of AI explanation may be appropriate in different contexts).

given the specialized nature of the technology itself and the knowledge required to conduct AI research or create applications.

i. Interpretability

ML has an interpretability problem. At a high level of abstraction, ML algorithms operate by identifying patterns in massive data sets. Specifically, a statistical model selected by a data scientist parses a large set of training data and identifies correlations in order to group together data points that possess similar attributes. This initial training data can either be partially labelled by a data scientist, in which case the system will aim to learn to identify similar cases, or the system can run unsupervised analysis. The result of this ML training is a running model with a decisional rule, sometimes called a "working algorithm," that can be applied to other data sets to which the model has not previously been exposed.

The trouble, however, is that human beings may not be able to comprehend what the "black box" working algorithm is doing.[109] As Andrew Selbst and Solon Barocas explore, there are two potential layers of incomprehensibility: First, a model might be inscrutable in the sense that, particularly in the "deep-learning" or complex "neural network" ML configurations often used for more complex tasks, it is not possible to observe and explain exactly how the model is interpreting the data.[110] Second, the manner in which the model is connecting the data to identify a pattern might be nonintuitive, particularly when compared to the cause-and-effect reasoning that drives the scientific method.[111] It is possible that advances in interpretable machine learning might decrease inscrutability,[112] and a researcher may be able to increase interpretability or implement auditing procedures that make a particular outcome less opaque.[113] Nonetheless, these are coping strategies, not cures.

The issue of nonintuitive connections, moreover, may be harder to resolve. For instance, in one well-noted example, a neural network built to

109. This Article does not claim that every ML algorithm is properly categorized as an opaque black box, but rather uses this term to refer to the common conception of such models.

110. *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1094–96 (2018).

111. *Id.* at 1096–99.

112. There has been a flurry of work in the past several years to develop "interpretable" and "explainable" AI models. *See, e.g.*, Finale Doshi-Velez & Been Kim, *Towards a Rigorous Science of Interpretable Machine Learning*, ARXIV (Mar. 2, 2017), https://arxiv.org/pdf/1702.08608.pdf [https://perma.cc/397A-GGSA]; Chris Olah et al., *The Building Blocks of Interpretability*, DISTILL (Mar. 6, 2018), https://distill.pub/2018/building-blocks [https://perma.cc/27U4-N5X6]; Matt Turek, *Explainable Artificial Intelligence (XAI)*, DEF. ADVANCED RES. PROJECTS AGENCY, https://www.darpa.mil/program/explainable-artificial-intelligence [https://perma.cc/LQ68-SDUE].

113. *See* Lehr & Ohm, *supra* note 20, at 656–58.

separate images of wolves from images of dogs did not develop an understanding of biological differences between the canines, but instead recognized that all of the wolves were standing on snow and the dogs on grass.[114] Along similar lines, in a possibly apocryphal story, the U.S. Army used neural networks to distinguish forests from camouflaged tanks—but did not realize that the algorithm was really identifying sunshine versus shade because all of the forest photographs were taken on sunny days and the tanks on cloudy days.[115] This sort of nonintuitive, yet rationalizable correlation abounds in ML; indeed, the technique's appeal comes from its capacity to identify patterns that humans would not necessarily discern or consider significant. But ML is consequently not susceptible to the kind of interpretation and careful testing upon which the scientific method has traditionally relied.

An administrative agency might try to contend with these sorts of interpretability issues using tactics that have worked in the past. The catch is that AI is still unique, even if it shares certain similarities with prior policy challenges. Compare ML's inscrutability to drugs that have been tested and cleared for a particular use, without the ability to explain why they are effective treatments. Consider, for instance, selective serotonin reuptake inhibitors ("SSRIs"), which are often used to treat depression or generalized anxiety disorders. Though these drugs are thought to alleviate symptoms by increasing the level of particular neurotransmitters,[116] the specific "mechanism of action" through which they operate remains unknown.[117] In other words, the precise cause-and-effect process that makes these drugs efficacious remains inscrutable, even if it is effective for many patients.

The problem, however, is that FDA's tactics may be nearing their limits when cutting-edge medical interventions come into the equation. Take "black-box medicine," which Nicholson Price defines as "the use of opaque computational models to make decisions related to health care."[118] For

114. Marco Tulio Ribeiro et al., *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, ARXIV (Aug. 9, 2016), https://arxiv.org/pdf/1602.04938.pdf [https://perma.cc/QHU9-F25G]; *see also* Selbst & Barocas, *supra* note 110, at 1123–24 (discussing the wolf-dog example).

115. Eliezer Yudkowsky, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, *in* GLOBAL CATASTROPHIC RISKS 308, 321 (Nick Bostrom & Milan M. Ćirković eds., 2008).

116. *Selective Serotonin Reuptake Inhibitors (SSRIs)*, NHS INFORM, https://www.nhsinform.scot/tests-and-treatments/medicines-and-medical-aids/types-of-medicine/selective-serotonin-reuptake-inhibitors-ssris [https://perma.cc/LV6Y-WYLE].

117. *See* ELI LILLY & CO., FULL PRESCRIBING INFORMATION 20 (2020), http://pi.lilly.com/us/prozac.pdf [https://perma.cc/QZ9G-DBUN] ("Although the exact mechanism of PROZAC is unknown, it is presumed to be linked to its inhibition of CNS neuronal uptake of serotonin.").

118. Price, *Regulating Black-Box Medicine*, *supra* note 24, at 429 (quoting W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 421, 429–34 (2015) [hereinafter Price, *Black-Box Medicine*]).

instance, an algorithm might use patient data to discover a new use of an existing drug, perhaps by mining medical records, parsing a patient's symptoms and their pharmaceutical regimen, and then identifying previously unrecognized patterns between ailments and potential remedies.[119] It might not be possible to explain the causal relationship between the algorithm's choice and the outcome, thereby straining the core premises that undergird FDA's preclearance authority. To date, the attitude has been that FDA can respond by adjusting its regulatory practices. But this conclusion seems to stem from a sense that there is no other agency that could step in, nor any other actor with legal authority to do so.[120] It is far from a first-best solution, particularly given the increased speed of algorithmic development, and—as the next Section explores—especially in an emerging field that is as cross-cutting as AI algorithms.

ii. Domain Expertise

AI's interpretability problem is also a practical problem in another sense: the technology's complexity puts further demands on the people developing the algorithms. Particularly when it comes to ML, general computer science knowledge is not enough. ML requires mastery of data science to "tune" the algorithmic levers that will allow a model to identify patterns. It also demands more specialized expertise to understand any idiosyncrasies of the issue area and sensitivity to potential bias or discrimination in the data set.[121] And even before seeking this more specialized expertise, individuals trained to tackle AI in general or ML in particular are in short supply. Although market demand may incentivize more individuals to enter the field, a 2018 study found just over "22,000 PhD-educated researchers in the entire world who are capable of working in AI research and applications. . . . [and, in an advanced subset, there are only] 5,400 AI experts in the world who are publishing and presenting at leading AI conferences."[122] Moreover, formal training is inadequate because "developing successful machine learning applications requires a substantial amount of 'black art' that is hard to find in textbook."[123]

119. *See* Price, *Black-Box*, *supra* note 118, at 436–37.

120. *See id*. at 452 (expressing concern that FDA lacks expertise to contend with complex black-box algorithms, yet doubting that another government agency is better positioned).

121. For a description of ML stages targeted at legal scholars, see Lehr & Ohm, *supra* note 20, at 669–702.

122. Jean-François Gagne et al., *Global AI Talent Pool Report 2018*, JFGAGNE, http://www.jfgagne.ai/talent [https://perma.cc/4KK9-LHN8]; *see also* Jean-François Gagne et al., *Global AI Talent Report 2019*, JFGAGNE, https://jfgagne.ai/talent-2019 [https://perma.cc/7H2A-ERZL] (reporting "22,400 unique individuals who published at one or more" top AI conferences in 2018 and about 4,000 individuals "doing research that is having a notable impact on the overall field as measured by citations received in the last two years").

123. Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, 55 COMM. ACM,

True, the "black art" of functional domain expertise is a prerequisite in many technocratic disciplines. As Arthur C. Clarke's third law holds, "[a]ny sufficiently advanced technology is indistinguishable from magic."[124] The uninitiated cannot understand the trick. What nonetheless makes AI uniquely challenging is the difficulty of delineating precisely what sort of expertise is required. The general utility of AI as a tool stands in contrast to a field like pharmaceutical regulation. Drugs, to be sure, have both social and economic ramifications. Drug development is a market in which the products can—depending on whether FDA can strike the right level of innovation and consumer protection—either save lives or cause tremendous pain and suffering.

But FDA's regulatory task differs from the algorithmic context because it can point to a specific regulatory object, the drug, and craft a regime around that object. Though these delineations are not perfect,[125] FDA can nonetheless confine itself to a relatively well-specified regulatory goal: confirming and communicating that a drug is safe and effective for the market.[126] Development of AI, however, is not like development of a drug, for which the molecules are ostensibly stable once taken out of the lab. Again, at a high level, ML works when a data scientist selects a statistical model, which processes training data to identify correlations between different attributes in the data. The goal is to pinpoint correlational patterns and thereby produce a running model with a decisional rule that can accurately make predictions about other data points. For this process to work, data scientists must constantly make choices about how to "play with the data" in order to run a particular model.[127] Moreover, it is imperative to conduct ongoing validation to confirm that the running model is not based on outdated data or beliefs about the world that lead to stale or biased predictions.[128] This requirement means that properly-validated running

Oct. 2012, at 78, 78.

124. *Clarke's Three Laws*, WIKIPEDIA, https://en.wikipedia.org/wiki/Clarke%27s_three_laws [https://perma.cc/D2R8-TE49].

125. Indeed, some of the challenges in pharmaceutical regulation may arise in contexts where it is harder to specify a clear regulatory target. "Black-box medicine," for instance, may be especially challenging in part because its algorithmic aspects seem an imperfect fit for the traditional command-and-control paradigm of drug regulation by testing the molecular compounds. *See* Price, *supra* note 24, at 424. This Article reserves further analysis of what factors may make an emerging technology especially ill-suited for traditional administrative law paradigms for future work.

126. *See* Amy Kapczynski, *Dangerous Times: The FDA's Role in Information Production, Past and Future*, 102 MINN. L. REV. 2357, 2358 (2018) ("The core function of the FDA as a drug regulator . . . is not to make choices for the public, or to certify the truth, but to generate and validate information about medicines.").

127. *See* Lehr & Ohm, *supra* note 20, at 655 (offering "that legal scholars should think of machine learning as consisting of two distinct workflows: 'playing with the data,' . . . and 'the running model' . . . .").

128. *See* Brandon L. Garrett & John Monahan, *Judging Risk*, 108 CALIF. L. REV. 439, 490 (2020)

algorithms are far less fixed than molecules. And even if a particular algorithm is formally approved for a particular use, any outputs would change with the introduction of new data, as the machine "learns"—requiring a new round of approval. It is not clear that any agency could administer such a far-reaching preclearance regime.

Moreover, the scale and complexity of the problem compounds if the institutional intervention is an overarching agency to oversee algorithms across different sectors.[129] Even if the goal shifts from prescriptive regulation to simply providing advice and oversight, such as, for instance, through a Federal Robotics Commission[130] or a National Algorithm Safety Board,[131] the sheer number of domains affected is daunting. For instance, the 2019 National AI Research and Development Strategic Plan, published by the National Science and Technology Council's Select Committee on AI, identified AI applications in fifteen domains as diverse as agriculture, defense, law, personal services, and transportation.[132]

Furthermore, the broad sweep of potential applications for AI becomes even more complex because the technology often implicates a range of public interests, such as cross-cutting privacy and safety issues, within a single application. For instance, picture a partially-automated vehicle that incorporates ML software. Imagine that this vehicle contains sensors and cameras to monitor its driver's alertness[133] and emotional state,[134] compiles

("Instruments should be re-validated over time at reasonable intervals and with attention to local variation in populations, resources, and crime patterns."); John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV.1725, 1757 (2018) ("For tools to make well-calibrated predictions from the start, they need to be trained on data that matches the conditions about which they are making predictions."); Alicia Solow-Niederman et al., *The Institutional Life of Algorithmic Risk Assessment*, 34 BERKELEY TECH. L.J. 705, 709 (2019) (discussing the importance of validation in the risk assessment context).

129. *See* WHITTAKER ET AL., *supra* note 15, at 4 ("[A] national AI safety body or general AI standards and certification model will struggle to meet the sectoral expertise requirements needed for nuanced regulation.").

130. *See, e.g.*, Calo, *supra* note 16.

131. *See, e.g.*, Schneiderman, *supra* note 17.

132. 2019 NATIONAL AI R&D STRATEGIC PLAN, *supra* note 28, at 6; *see also id.* at 8 ("Recent Federal investments have prioritized . . . fundamental ML and AI research . . . as well as the use of ML and AI across numerous application sectors, including defense, security, energy, transportation, health, agriculture, and telecommunications. Ultimately, AI technologies are critical for addressing a range of long-term challenges . . . .").

133. Vehicles are already being equipped with such systems. *See, e.g.*, Joann Muller, *Driver Monitoring Systems Are Here — And So Are Privacy Concerns*, AXIOS (Oct. 27, 2018), https://www.axios.com/driver-cameras-bring-privacy-concerns-873804d2-8897-468b-82f4-b3586bdfea31.html [https://perma.cc/DTA7-HWQG] (discussing "Super Cruise" driver monitoring system now available in GM Cadillac CT6).

134. Again, manufacturers are already developing this capability. *See* Viknesh Vijayenthiran, *Kia Concept for 2019 CES Previews a World with Self-Driving Cars*, MOTOR AUTHORITY (Jan. 3, 2019), https://www.motorauthority.com/news/1120570_kia-concept-for-2019-ces-previews-a-world-with-self-driving-cars [https://perma.cc/R84U-NKZQ] (discussing Kia's READ ("Real-time Emotion Adaptive

this information, and sends the data along with the car's location to a centralized database, aiming to create a transport ecosystem that is safest for the public in the aggregate. This vehicle presents issues of physical safety (for the driver, others on the road, and any pedestrians it might encounter) that interact with existing federal, state, and local regulations. These issues arise in tandem with civil liberty questions about individual privacy rights and the level of surveillance that society is willing to accept[135] (for the monitored individual and any others whose data is collected and analyzed) and normative questions about how to value human life (for cases in which the wellbeing of the driver and that of other individuals might be in conflict).[136] Other cross-cutting questions, such as cybersecurity[137] and how to define and operationalize fairness in a data set,[138] also arise in multiple sectors.

Cabining these and similar issues in the context of questions about a single domain risks obscuring or wholly eliding these cross-cutting considerations. Trying to solve these issues within a single sector, moreover, stymies the development of cross-sectoral principles and can make it harder to develop a publicly-shared consensus on the values that are not to be compromised in any setting.[139] If we are contending with a brave new

Driving") system, a not-yet-deployed feature that monitors and responds to the driver's emotional state).

135. *See* Saheli Roy Choudhury, *Driverless Cars Will Need Cities Covered in Sensors, China's Didi Chuxing Says*, CNBC (Dec. 4, 2018, 7:38 AM), https://www.cnbc.com/2018/11/27/east-tech-west-chinas-didi-chuxing-on-future-of-self-driving-cars.html [https://perma.cc/K72W-A454] (discussing need to embed sensors city-wide for AVs to function); Keith Naughton, *Ford Breaks with GM, Toyota on Future of Talking-Car Technology*, BLOOMBERG (Jan. 7, 2019, 5:00 AM), https://www.bloomberg.com/news/articles/2019-01-07/ford-breaks-with-gm-toyota-on-future-of-talking-car-technology [https://perma.cc/5KXG-VVBJ] (discussing Ford's plans, beginning in 2022, to outfit all new U.S. models with "cellular vehicle-to-everything technology," or "C-V2X," that will permit cars to "communicate with one another about road hazards, talk to stop lights to smooth traffic flow and pay the bill automatically while picking up fast food").

136. *Cf.* Maggie Miller, *Consumer Groups Say Senate's Revamped Self-Driving Car Bill Fails to Resolve Cyber, Safety Concerns*, INSIDE CYBERSECURITY (Dec. 6, 2018), https://insidecybersecurity.com/daily-news/consumer-groups-say-senates-revamped-self-driving-car-bill-fails-resolve-cyber-safety (discussing consumer groups' choice not to support the 2018 AV START Act on grounds it did not amply address cybersecurity and privacy issues).

137. Cybersecurity challenges are already arising with software updates for non-automated vehicles. *See*, *e.g.*, Joann Muller, *The Hidden Risks of Remote Software Updates*, AXIOS (Dec. 14, 2018), https://www.axios.com/risks-of-over-the-air-software-update-vehicles-117d0b7d-cb13-4b63-aa0f-0dae365f97dc.html [https://perma.cc/4YKP-6JY7] (discussing cybersecurity risks when "over-the-air" software updates are sent to vehicles).

138. The choice of a training data set for a machine-learning algorithm can interact with and reify problematic social stereotypes that discriminate against particular categories of individuals in making decisions about who to hire or where police should devote more aggressive resources. A growing body of work on fairness, accountability, transparency, and, increasingly, ethics, addresses these questions. *See ACM Conference on Fairness, Accountability, and Transparency*, ACM FACCT CONFERENCE, https://fatconference.org/index.html [https://perma.cc/B2L7-UBZM]; *cf.* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016) (discussing how big data increases inequality and threatens democracy).

139. Privacy law scholars may be reminded of the distinction between the U.S. regime of protecting

algorithmic world, then we should have first principles that reflect the actual ground conditions.

c. Unpredictability

In addition to these foundational normative challenges, AI governance becomes even more complicated as an algorithmic model interacts with the real world in unpredictable ways. Unpredictability encompasses two broad categories of problems: (1) uncertainty and (2) emergence, each of which applies for both algorithms and embodied AI, or robots.

i. Uncertainty

The inability to predict AI outcomes with certainty may produce accidents, particularly when a real-world AI system is poorly designed in ways that lead to unintended and harmful behavior.[140] Following research on "concrete problems in AI safety" by a team at Google Brain, Stanford University, UC Berkeley, and OpenAI, this Article defines an accident as "a situation where a human designer had in mind a certain (perhaps informally specified) objective or task, but the system that was designed and deployed for that task produced harmful and unexpected results."[141] The specific details of such an accident may not be foreseeable, but the high probability of an accident, absent some form of ex ante intervention, might be. Consider, for instance, the issue of "reward hacking," which occurs if an AI algorithm is told to optimize a particular task but instead "games" its reward function.[142]

AI researchers have begun to explore how careful study of categories of common issues such as reward hacking might minimize such accidents. If certain categories of accidents tend to recur in AI development, then it may be possible to implement mechanisms or strategies to decrease their likelihood or mitigate their impact. Proposed technical solutions to reward

informational privacy rights sector-by-sector, *see* Alicia Solow-Niederman, *Beyond the Privacy Torts: Reinvigorating a Common Law Approach for Data Breaches*, 127 YALE L.J. F. 617 n.13 (2018), as compared to the European Union's protection of privacy as a fundamental human right across all sectors.

140. *See* Dario Amodei et al., *Concrete Problems in AI Safety*, ARXIV 1 (July 25, 2016), https://arxiv.org/pdf/1606.06565.pdf [https://perma.cc/6EAB-3T76] (identifying and discussing "the problem of *accidents* in machine learning systems, defined as unintended and harmful behavior that may emerge from poor design of real-world AI systems").

141. *Id.* at 2. As the researchers acknowledge, this issue extends across many classes of engineering, yet may be uniquely pressing in the case of AI. *Id.* (citing Jacob Steinhardt, *Long-Term and Short-Term Challenges to Ensuring the Safety of AI Systems*, ACADEMICALLY INTERESTING (June 24, 2015), https://jsteinhardt.wordpress.com/2015/06/24/long-term-and-short-term-challenges-to-ensuring-the-safety-of-ai-systems [https://perma.cc/553T-7NCE]).

142. *See supra* text accompanying notes 1–6; *see also* Amodei et al., *supra* note 140, at 3, 7–11 (describing a "cleaning robot" that, if rewarded "for achieving an environment free of messes . . . might disable its vision so that it won't find any messes, or cover over messes with materials it can't see through, or simply hide when humans are around so they can't tell it about new types of messes").

hacking, for instance, include careful engineering through "formal verification or practical testing of parts of the system;" "reward capping," or placing a ceiling on the maximum possible reward; and "trip wires" that "deliberately introduce some plausible vulnerabilities (that an agent has the ability to exploit but should not exploit if its value function is correct)," thereby providing a clear signal in the event that something does go awry when the model runs.[143]

These specific technical problems and solutions may be new, yet the underlying challenge of contending with accidents is an old engineering problem. And accident management is itself one facet of a broader field of risk management: any system, built responsibly, must account for the risk that there will be errors downstream. Beyond engineering or computer science, there is not only a robust literature on risk analysis as "a systematic approach to science-based decision making" in general,[144] but also an important substrate of risk management of emerging technologies in particular,[145] including legal scholarship on point.[146] In fact, the principles of this growing field undergird pharmaceutical regulation.

Consider FDA's premarket clearance requirements for drugs.[147] As mentioned previously, this legislative amendment to the administrative regime was prompted, in large part, by the thalidomide disaster, in which a drug introduced to treat sleeping disorders produced severe birth defects when ingested by pregnant women.[148] A key problem that Congress sought to redress was a lack of adequate testing to establish safety and efficacy for a specified use before a drug like thalidomide could enter the market.[149] In other words, there were not adequate procedural requirements in place to decrease the risk of accidents. The 1962 amendment thus increased drug producers' burden to establish that the purported benefits of their products exceeded the risks. By forcing private firms to provide FDA with empirical evidence required to make this assessment before FDA would clear the drug

143. *See* Amodei et al., *supra* note 140, at 7–11.

144. *See, e.g.*, DANIEL M. BYRD III & C. RICHARD COTHERN, INTRODUCTION TO RISK ANALYSIS (2005).

145. *GETS Conference, supra* note 73.

146. *See, e.g.*, Gary E. Marchant, *Advancing Resilience Through Law*, *in* INT'L RISK GOVERNANCE CTR., 1 RESOURCE GUIDE ON RESILIENCE 158, 159 (M.V. Florin & I. Linkov eds., 2016) (citing S.A. Shapiro & R.L. Glicksman, *Improving Regulation Through Incremental Adjustment*, 52 U. KAN. L. REV. 1179 (2004)); Gary E. Marchant & Yvonne A. Stevens, *Resilience: A New Tool in the Risk Governance Toolbox for Emerging Technologies*, 51 U.C. DAVIS L. REV. 233, 241–47 (2017).

147. *See supra* text accompanying notes 91–93 and sources cited therein.

148. Bara Fintel et al., *The Thalidomide Tragedy: Lessons for Drug Safety and Regulation*, HELIX MAG. (July 28, 2009), https://helix.northwestern.edu/article/thalidomide-tragedy-lessons-drug-safety-and-regulation [https://perma.cc/R67M-LDLN].

149. *See supra* Section II.A.1.

for the market, Congress expanded FDA's authority to manage risk.

So, what is unique about AI, if anything? It comes down once more to the nature of the regulatory object and how to fit it within administrative law institutions. Again, attempting to equate FDA's approach to regulation of AI falters because of a unit of analysis problem. It is one thing for the engineers creating a product to develop systematic approaches to risk. It is another to task a single agency with doing so, given the dynamic and cross-cutting nature of AI as it is applied.

Algorithmic technologies thus present a catch-22 for the would-be regulator seeking to mitigate risk. Without much finer-grained specification of a particular regulatory object, AI's technical attributes do not fit neatly within risk management paradigms. But finer-grained specification to fit into this paradigm comes at a significant cost. Narrowing the regulatory scope to a single sector again compromises a critical opportunity to develop grounding principles that would apply across sectoral applications.[150] Without such grounding principles, we risk waking up tomorrow and being asked to consent to whatever the market delivers—without ever pausing to consider what we wanted as citizens, and not merely as consumers.

ii. Emergence

An even more intractable dimension of unpredictability is emergence, or the manner in which complex systems can interact in ways that would not be predicted by looking at any one of its subparts in isolation.[151] Emergence is, on one hand, a desirable property insofar as it catalyzes creative outcomes that human programmers would not necessarily have considered.[152] Indeed, in some contexts, it may be a new form of intelligence.[153]

But this intelligence has two faces. In an algorithm, each line of programming code operates as a low-level element of an emergent system.[154] Each individual line of programming will combine with other steps of the code and also with external actors and inputs to produce an outcome. For

150. *See supra* text accompanying notes 132–39.

151. *See* STEVEN JOHNSON, EMERGENCE 18 (2001); *see also* Calo, *supra* note 69, at 538–45.

152. *See* Lehman et al., *supra* note 2, at 5; *see also* Calo, *supra* note 69, at 539–40 (noting potential for "useful but unexpected problem solving by machines").

153. *See* JOHNSON, *supra* note 151 (discussing the collective intelligence of ants, brains, cities, and software); Steven Johnson, *Only Connect*, GUARDIAN (Oct. 15, 2001), https://www.theguardian.com/books/2001/oct/15/society [https://perma.cc/247F-WYWA] (describing Amazon product recommendations as an emergent system that "has got smart by looking for patterns in users' purchasing behaviour, and in their limited feedback about the items they've read" to create "a kind of collective wisdom . . . [that is] much more fluid and nuanced than the logic we traditionally expect from our computers").

154. Again, this Article breaks from much of past legal scholarship in focusing on algorithms in both their embodied, or robot, and intangible forms to underscore that the problem is not limited to regulation of robots. *See supra* note 1.

instance, in an application such as an AV, each line of code will come together to form a working algorithm that interacts with physical sensor data to make decisions about how to proceed on the road, forming an emergent intelligence to steer the vehicle.

The problem is that the same complexity that permits emergence as a desirable property of machine-based intelligence can also be dangerous. For instance, in March 2018, a car operating under computer control hit and killed a pedestrian who was walking a bicycle across a dark street. According to an investigation by the National Transportation Safety Board (NTSB), because the pedestrian was not near a crosswalk, the automated driving system ("ADS") "never accurately classified" the jaywalking individual "as a pedestrian or predicted her path."[155] Because the vehicle's software could not determine what it was driving toward, its waffling ate up precious response time.[156] Once it made a final determination, the ADS itself had inadequate time to avert the collision. The company's design choices relied on the human driver to remain attentive and operate as a safety backstop—but in this case, the driver was distracted by her cell phone.[157]

In theory, there might still have been other ways to avert the collision. In practice, these solutions were not possible because of the company's "inadequate safety culture,"[158] which prioritized comfort above safety and had turned off the emergency braking function "to reduce the potential for erratic vehicle behavior" and make the ride less turbulent.[159] Additionally, the software system had been designed so that it did not provide a warning to alert the human safety driver of a potential issue.[160] This tragedy emerged not from any single point of control, but rather from a combination of complex code, unexpected inputs, human design choices, and the manner in

155. NAT'L TRANSP. SAFETY BD., HWY18MH010, PUBLIC MEETING OF NOVEMBER 19, 2019: COLLISION BETWEEN VEHICLE CONTROLLED BY DEVELOPMENTAL AUTOMATED DRIVING SYSTEM AND PEDESTRIAN 1 (2019).

156. *Id.* ("By the time the ADS determined that a collision was imminent, the situation exceeded the response specifications of the ADS braking system.").

157. *Id.* at 1–3.

158. *Id.* at 2.

159. NAT'L TRANSP. SAFETY BD., HWY18MH010, PRELIMINARY REPORT: HIGHWAY 2 (2018); *see also* NAT'L TRANSP. SAFETY BD., *supra* note 155, at 3 ("The Uber Advanced Technologies Group's inadequate safety culture created conditions—including inadequate oversight of vehicle operators—that contributed to the circumstances of the crash and specifically to the vehicle operator's extended distraction during the crash trip."); Timothy B. Lee, *NTSB: Uber's Sensors Worked; Its Software Utterly Failed in Fatal Crash*, ARS TECHNICA (May 24, 2018, 8:10 AM), https://arstechnica.com/cars/2018/05/emergency-brakes-were-disabled-by-ubers-self-driving-software-ntsb-says [https://perma.cc/PSB8-R9Z4].

160. NAT'L TRANSP. SAFETY BD., *supra* note 155, at 2–3 ("The Uber Advanced Technologies Group's deactivation of the Volvo forward collision warning and automatic emergency braking systems without replacing their full capabilities removed a layer of safety redundancy and increased the risks associated with testing automated driving systems on public roads.").

which the human safety driver operated on the road. AV-human interactions are complex systems whose interactions cannot necessarily be predicted by focusing on any single subpart.

Legal interventions might seem to provide ways to avoid or redress the prospect for harm in such complex systems. An ex ante legislative or regulatory intervention could require manufacturers to optimize safety above comfort, for instance,[161] or ramp up the level of federal oversight of private choices before vehicles hit the road.[162] Or a system of stricter ex post sanctions in tort and/or criminal law could change the cost-benefit analysis of safety tradeoffs and thereby incentivize manufacturers to proceed more cautiously. But such uses of law to target organizational protocols do not address the underlying technical limitations. At the level of the code itself, better programming and good coding practices will not necessarily correct the liabilities of emergence—and even a safety-oriented corporate culture cannot redress this underlying reality. Corporate practices alone cannot contend with the ways that the complexity of the algorithm, as applied, strains our very understanding of the universe.[163]

Loss of agency notwithstanding, it might still be tempting to develop a public regulatory response to try to mitigate the most critical policy concerns. Logically, it would seem that FDA's pharmaceutical regulatory framework would also need to contend with emergent properties of drugs to ensure their safe usage. But it is a mistake to extrapolate from FDA's approach to emergence to AI.

To see why, consider SSRI medications once more. These pharmaceuticals have been rigorously tested, long marketed, and "are usually the first choice medication for depression because they generally have fewer side effects than most other types of antidepressant[s]."[164] But they can also be life-threatening if a patient develops serotonin syndrome, which occurs if the level of the neurotransmitter in the patient's body is too high. This rare condition is not because of the drug per se, in the sense that

161. *See* David Weinberger, *Optimization Over Explanation*, MEDIUM: BERKMAN KLEIN CTR. (Jan. 28, 2018), https://medium.com/berkman-klein-center/optimization-over-explanation-41ecb135763d [https://perma.cc/2DC6-7WEQ] (advocating a turn to existing policy processes to "to decide what we want [AI] systems optimized for").

162. *See* NAT'L TRANSP. SAFETY BD., *supra* note 155, at 3–4 (endorsing "[m]andatory submission of safety self-assessment reports" and National Highway Traffic Safety Administration evaluation of these reports).

163. *See* Jonathan Zittrain, *Intellectual Debt: With Great Power Comes Great Ignorance*, MEDIUM: BERKMAN KLEIN CTR. (July 24, 2019), https://medium.com/berkman-klein-center/from-technical-debt-to-intellectual-debt-in-ai-e05ac56a502c [https://perma.cc/5KGL-MPHS] (discussing the "intellectual debt" problem of machine learning, which "stands to produce answers that work, without offering any underlying theory").

164. *Selective Serotonin*, *supra* note 116.

consumption of an SSRI directly causes too much serotonin. Rather, according to the Mayo Clinic, it most frequently occurs from the ingestion of two medications that raise the level of serotonin, in combination. So if a patient is taking an SSRI like Prozac in combination with, for example, the herbal supplement St. John's Wort to treat their irritable bowel syndrome or insomnia, then they might be at risk of developing this dangerous syndrome.[165] Or if a patient is taking an acid reflux drug that contains the chemical compound omeprazole, then that compound can raise the individual's serum exposure to the SSRI.[166] At a constant dosage of some SSRIs, in other words, a patient who is also taking the acid reflux drug would likely exhibit a higher blood concentration of the SSRI than one who is not, such that it is as if they are taking a higher dosage—which could, in turn, heighten the risk of serotonin syndrome.

FDA does not in fact attempt to directly command and control these emergent interactions. For one, it does not regulate the safety or efficacy of herbal or botanical remedies that are used as dietary supplements at all.[167] Instead, it relies on prominent medical clinics to publicize information about dangerous interactions.[168] More generally, FDA contends with drug interaction challenges through alternative methods of control, in the form of a combination of labeling requirements that mandate disclosure of known interactions[169] and reliance on doctors' counsel in prescribing medical interventions in a way that accounts for the patient's full medical history.[170]

165. *Selective Serotonin Reuptake Inhibitors (SSRIs)*, MAYO CLINIC, https://www.mayoclinic.org/diseases-conditions/depression/in-depth/ssris/art-20044825 [https://perma.cc/9K57-WXH]; *see also St. John's Wort*, NIH: NAT'L CTR. FOR COMPLEMENTARY & INTEGRATIVE HEALTH, https://nccih.nih.gov/health/stjohnswort/ataglance.htm [https://perma.cc/4ATT-PHZQ] (describing uses of St. John's Wort and potentially dangerous interaction between herbal remedy and SSRIs).

166. *See, e.g.*, Caroline Gjestad et al., *Effect of Proton Pump Inhibitors on the Serum Concentrations of the Selective Serotonin Reuptake Inhibitors Citalopram, Escitalopram, and Sertraline*, 37 THERAPEUTIC DRUG MONITORING 90, 90 (2015).

167. *See* Dietary Supplement Health and Education Act of 1994, Pub. L. No. 103-417, 108 Stat. 4325 (codified as amended in scattered provisions of 21 U.S.C.).

168. *See Selective Serotonin Reuptake Inhibitors (SSRIs)*, *supra* note 165.

169. FDA market clearance of a drug comes with labelling requirements. As Hilts explains, "by the mid-1990s, the FDA, under pressure [to avoid mistakenly approving drugs with safety risks], was pinning its hopes on warning labels and doctors' care in prescribing." HILTS, *supra* note 83, at 234. The disclaimer is to include facts such as appropriate uses and dosage information, though the specific details vary by category of drug. *See Guidances (Drugs)*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs [https://perma.cc/D9HK-LBVC]; *see also* Eisenberg, *supra* note 83 at 382–83 (describing differences in disclosure requirements for prescription versus over-the-counter drugs).

170. The FDA-approved label for Prozac, for instance, states:

> Patients should be advised to inform their physician if they are taking, or plan to take, any prescription medication, including Symbyax, Sarafem, or over-the-counter drugs, including herbal supplements or alcohol. Patients should also be advised to inform their physicians if they plan to discontinue any medications they are taking while on PROZAC.

ELI LILLY & CO., *supra* note 117, at 26.

Thus, FDA's regulatory preclearance regime does not grapple with emergence head-on. The question of whether FDA should clear a particular drug as safe and effective based on clinical data addresses a narrowly circumscribed use of the drug. Yet it may not account for interactions between the drug and exogenous factors that are apparent after the drug is brought to market. These postmarket emergence questions are considered outside the scope of its market clearance regulatory mission. Instead, FDA relies on other tactics, in the form of labelling and disclosure requirements.

It is a mistake to extrapolate these lessons to the AI context because algorithmic applications are markedly different in at least two crucial ways. At a practical level, there is no natural learned intermediary to guide the would-be algorithmic consumer, nor is there an objective list of what a disclosure label should provide a warning about to inform and empower the end user.[171] Even more fundamentally, at a conceptual level, rigorously controlling emergent properties is a poor fit for algorithmic applications. In the context of drug development, emergent properties are failures of control caused by properties outside of the authorized, market-cleared use of the drug. For AI, however, emergent properties are not necessarily a bad thing. To the contrary, much of the creative promise of ML algorithms in particular comes from the ability to adapt to inputs in ways that humans would never have foreseen.[172] At times, these solutions can be better than the human would have predicted.[173] Accordingly, control in the sense of specifying particular use of a drug and providing warnings about its use, as FDA testing attempts to do, is inadvisable if the goal is to create an algorithm that arrives at the best possible outcome. In contrast to regulation of a single regulatory object defined by a clear-cut objective, such as an approved usage of the active molecules of a drug, the question of what to control when we talk about commanding-and-controlling AI is not so clear. AI's emergent properties are Janus-faced, making it hard to determine ex ante when a lack

171. Though distinct from labels to inform the end user, there are nascent efforts to label the data that goes into algorithms. *See* Hilary Ross & Nicole West Bassoff, *The "Dataset Nutrition Label Project" Tackles Dataset Health and Standards*, MEDIUM: BERKMAN KLEIN CTR. (Jan. 29, 2019), https://medium.com/berkman-klein-center/the-dataset-nutrition-label-project-tackles-dataset-health-and-standards-658dc162dfbb [https://perma.cc/JV9F-P9B3] (discussing effort to "make it easier to quickly assess the viability and fitness of a dataset, before it is used to train a model, by giving it a 'nutrition' label").

172. Lehman et al., *supra* note 2, at 5–24 (discussing "[s]urprise from [a]lgorithms and [s]imulations," noting that "the field of complex systems" is well aware "that simple programs can yield complex and surprising results when executed," and offering that "digital evolution" can produce surprising, and at times creative, outcomes); *cf.* David Weinberger, *Our Machines Now Have Knowledge We'll Never Understand*, WIRED (Apr. 18, 2017, 8:22 PM), https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand [https://perma.cc/95ZE-96WX] ("The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.").

173. *See* Lehman et al., *supra* note 2, at 13–17.

of a top-down control is in fact a desirable property of the system.

* * *

AI thus seems a less than ideal fit for prescriptive structures across the board. First, its status as software permits more rapid creation and adjustment, such that the implementation of such a regime would slow down development relatively more as compared to industries that rely on physical capital investments to bring products or services to market. This speed consideration, standing alone, might make AI truly different—yet it might not defeat the case for command-and-control regulatory intervention, particularly if policymakers could point to evidence of clear market failures. But two issues remain. As a practical matter, when speed is coupled with the complexity and unpredictability of the technology, AI's combined policy challenges strain the capacity of the traditional administrative framework. And as a normative matter, given AI's status as a general utility technology, cabining regulation too firmly within particular sectors splinters imperative conversations about cross-cutting values and norms for all domains. If the goal is a publicly accountable governance strategy for AI, then we need other options.

B. COLLABORATION AND NEGOTIATION

Given the limits of a prescriptive approach, the following Sections explore governance as a contrasting public regulatory strategy before exposing its contemporary limits for AI. Section II.B.1 presents why governance might seem like a good fit for AI by surveying how "collaborative governance" models have emerged as an alternative to either state-driven prescriptive frameworks or wholly market-driven, deregulatory paradigms[174] in environmental law,[175] another complex and dynamic

174. *See generally, e.g.*, Daniel A. Farber, *Triangulating the Future of Reinvention: Three Emerging Models of Environmental Protection*, 2000 U. ILL. L. REV. 61 (describing three models for "reinvention" of environmental protection); Daniel J. Fiorino, *Rethinking Environmental Regulation: Perspectives on Law and Governance*, 23 HARV. ENVTL. L. REV. 441 (1999) (discussing efforts to "reinvent" environmental regulation); Jody Freeman & Daniel A. Farber, *Modular Environmental Regulation*, 54 DUKE L.J. 795 (2005) (advocating a new, modular approach to environmental regulation); Bradley C. Karkkainen, *Collaborative Ecosystem Governance: Scale, Complexity, and Dynamism*, 21 VA. ENVTL. L.J. 189 (2002) (presenting "collaborative ecosystem governance" as an alternative to traditional regulatory and legal arrangements); Orly Lobel, *New Governance as Regulatory Governance*, *in* OXFORD HANDBOOK OF GOVERNANCE 65, 65 (David Levi-Faur ed., 2012) (positioning new governance as a "third-way vision between unregulated markets and top-down government controls").

175. Environmental law traditionally addresses pollution control, whereas "natural resources management" is used to refer to a distinct set of resource management challenges. Building from the work of Jody Freeman and Daniel Farber, this Article uses the term "environmental law" to refer to ecosystem-wide challenges, including "situations in which pollution issues (e.g., water quality) and traditional resource management issues (e.g., water allocation) arise together." Freeman & Farber, *supra* note 174, at 800 n.4.

domain. Section II.B.2 then addresses why the public-private balance of power in AI R&D makes such an approach a poor present-day fit for American AI governance.

1. From Regulation to Governance in Environmental Law

Environmental law emerged as a distinct field in the 1970s as the state expanded the set of legally protected rights.[176] This era witnessed a number of new statutes and associated regulatory regimes to protect natural resources such as air and water, including the National Environmental Protection Act ("NEPA"), the Clean Air Act of 1970, and the Clean Water Act of 1972, as well as a new agency, the Environmental Protection Agency ("EPA").[177] This first generation of federal laws emphasized aspirational goals above economic analysis, focusing on how "to force industry to develop new technology capable of substantially more reductions in existing levels of pollution."[178] The unprecedented substantive reach of these statutes reflected the vast number of implicated domains[179] and included many congressional mandates with a command-and-control flavor. For instance, the Clean Air Act of 1970 "mandated the achievement by 1975 of national ambient air quality standards necessary for the protection of public health (primary standard) and public welfare (secondary standard)"; "instructed the EPA to publish an initial listing of 'hazardous' air pollutants within ninety days and then, within a year of its listing, to publish final emissions standard regulations;" imposed similarly strict requirements "for the EPA's listing of categories of stationary sources that 'may contribute significantly to air pollution which causes or contributes to the endangerment of public health or welfare' and called for an even tighter schedule for promulgating regulations for new sources"; and "mandated that the administrator achieve a 90 percent reduction in existing levels of automotive emissions of hydrocarbons and carbon monoxide by 1975 and nitrogen oxides by 1976."[180] The Clean Air Act is, moreover, just one of eighteen major federal environmental protection statutes enacted in the 1970s.[181]

---

176. *See* CASS R. SUNSTEIN, AFTER THE RIGHTS REVOLUTION: RECONCEIVING THE REGULATORY STATE 1 (1990).

177. *See id.* at 25–27; *see also* Richard J. Lazarus, *The Greening of America and the Graying of United States Environmental Law: Reflections on Environmental Law's First Three Decades in the United States*, 20 VA. ENVTL. L.J. 75, 76–78 (2001).

178. Richard J. Lazarus, *supra* note 177, at 78.

179. *See* ROGER W. FINDLEY ET AL., CASES AND MATERIALS ON ENVIRONMENTAL LAW 1–2 (6th ed. 2003) (describing the complexity of the field, including myriad federal statutes and regulatory schemes, overlap with other areas of law, and interdisciplinary considerations from economics and the sciences).

180. RICHARD J. LAZARUS, THE MAKING OF ENVIRONMENTAL LAW 70 (2004).

181. *Id.*

Yet since the early 1980s,[182] there has been a shift away from such stringent, top-down statutory command and toward "reform,"[183] "rethinking,"[184] or "reinvention"[185] of traditional regulatory approaches for environmental law.[186] This shift might be a post hoc rationalization of deregulatory political forces, an organic rethinking of theory that informed policy interventions, or some combination of the two. No matter the root cause, the bottom line is that a changing relationship among the state, industry players, and local citizen and non-profit representatives has led to revision of environmental law theory and practice. As Jody Freeman and Daniel Farber explain, a growing chorus has urged that "success with every environmental problem . . . requires not only a suite of complementary regulatory tools and the coordination of multiple levels of government [or strategy], but also a wide variety of informal implementation mechanisms and the ongoing participation of key stakeholders [or tactics]."[187]

This collaborative approach has been adopted as a policy strategy to address complex environmental systems. Though the wide variety of informal mechanisms and dynamic nature of such programs makes it hard to characterize them definitively, by way of example, consider Freeman and Farber's detailed case study of the CalFed Bay-Delta Program in California.[188] This program required attention to two ostensibly competing goals: first, protection of the habitat, and second, water provision for the state. Each of these goals relied on myriad federal, state, and local public and private stakeholders with different sources of expertise, access to different data sets, and different incentives and goals. As Freeman and Farber explain, this project broke from the traditional approach, wherein "the EPA set[] water quality standards (either on its own or by approving state standards), whereas wildlife agencies independently list[ed] endangered species and designate[d] their critical habitat."[189] They describe how such a "divided approach" may not be able to contend with the interactions among discrete interventions. It likely cannot account, for instance, for the ways in which "species survival and recovery can depend on water quality, including not only pollutants discharged from point sources but also salinity and flow

182. Detailed documentation of the phases of this revolution is outside the scope of this Article. For exposition, see *id.* at 85–99.

183. *See, e.g.*, Bruce A. Ackerman & Richard B. Stewart, Comment, *Reforming Environmental Law*, 37 STAN. L. REV. 1333, 1340 (1985).

184. *See, e.g.*, Fiorino, *supra* note 174.

185. *See, e.g.*, *id.*; Farber, *supra* note 174.

186. *See* sources cited *supra* note 174.

187. Freeman & Farber, *supra* note 174, at 797–98.

188. *See id.* at 837–76.

189. *Id.* at 842.

criteria."[190] By rethinking the traditional regulatory paradigm and permitting a more collaborative approach that involved public and private actors, as CalFed did, Freeman and Farber offer that at least some environmental stakeholders were better able to consider the interactions among discrete interventions.

2. Governance Challenges for AI

Given that environmental law governance has been presented by scholars and employed by at least some policymakers as a strategy to contend with complex ecosystems and dynamic challenges, it might appear to be a natural playbook for team AI to adopt.[191] Ecosystem management in particular may seem a rich source of AI lessons. In the case of joint tributary management, for example, there is a similar need for domain expertise. Vital inputs include attention to the conditions of the local watershed, the surrounding tributaries, and the manner in which different nutrient inflows might differently affect each tributary in dynamic fashion.[192] As with ML, knowledge of the relevant details may demand considerable formal training, technical data, and skills gained on the job, in a particular context.

In addition, given its regulation of complex ecosystems, environmental law, like AI, has needed to contend with uncertainty and emergence. Both fields must address "complex dynamic systems" that consist of "many mutually interdependent parts operating in dynamic, co-evolutionary trajectories."[193] Accordingly, environmental law scholars have invoked systems theory to describe how the "non-linear" nature of complex systems may limit "our understanding of the . . . ultimate effect that particular inputs will have . . . ."[194] Faced with such dynamic and uncertain conditions, it can be incredibly challenging to predict how a particular intervention will

190. *Id.*

191. Again, there is limited scholarship to date on AI and collaborative governance. *See* Kaminski, *supra* note 24, at 1535 n.8 (offering that "[o]nly a handful of scholars have explicitly considered using collaborative governance, in sector-specific contexts, to govern algorithmic decision-making or AI" and discussing limited scholarship directly on point). Kaminski's insightful analysis turns to collaborative governance to grapple with multifaceted algorithmic governance challenges, proposing a two-pronged system of individual due process rights alongside the systemic regulation provided by collaborative governance. *See generally id*. As discussed in this Part and *infra* Part III, this Article takes a step back to argue that—notwithstanding the potential benefits of collaborative algorithmic governance, in theory—the conditions for it to succeed neither exist today nor are likely to exist anytime soon, as both a practical and a theoretical matter.

192. Karkkainen, *supra* note 174, at 207–08 and sources cited therein (describing Chesapeake Bay Program).

193. *Id.* at 194–95 (citing C.S. Holling et al., *Science, Sustainability and Resource Management*, *in* LINKING SOCIAL AND ECOLOGICAL SYSTEMS 342, 346–47 (Fikret Berkes & Carl Folke eds., 1998)) (emphasis omitted).

194. *Id* at 194–96.

unfold—such that in at least some instances, the result may well be surprising or downright nonintuitive, and in many instances, we will be unable to specify a cause-and-effect relationship.[195] The unintended consequences at the level of the system, moreover, may not be evident by assessing individual inputs. Indeed, the governance approach is offered as a solution in part for this very reason. If we cannot understand enough about the cause-and-effect relationships in a complex system, then we cannot effectively prescribe top-down interventions for it. Ongoing public-private information sharing, collaboration, and (re-)negotiation thus serve as alternatives.

Given these apparent systemic similarities, why not endorse a governance-influenced regulatory solution as a natural fit for AI, spearheaded by the algorithmic equivalent of the EPA? This approach is misguided because executing the moves in a playbook demands the right combination of players who can coordinate in the right way, and we presently lack the requisite preconditions for collaborative AI governance.

The AI field is missing the active public voice necessary for a democratically accountable governance model. Environmental law's governance model calls for increasing the role of nonstate actors alongside state actors, as compared to a more adversarial prescriptive model of stringent, top-down regulation by the state. Yet the very idea of delegating some authority away from the state, moving away from "command-and-control," and substituting public-private negotiation makes little sense unless there is both a strong cohort of public representatives and informed private stakeholders. If there is no robust partner in the state itself, then collaboration is an oxymoron.[196] Any negotiation will occur in an unregulated market, without democratically accountable coordination or enforceable checks on commercial profit motives.

195. *See* Zittrain, *supra* note 163 ("While machine learning systems can surpass humans at pattern recognition and predictions, they generally cannot explain their answers in human-comprehensible terms. They are statistical correlation engines — they traffic in byzantine patterns with predictive utility, not neat articulations of relationships between cause and effect.").

196. The Internet might seem to challenge this claim because Internet governance has emerged without state-centered leadership. Though full treatment awaits another paper, the history of the Internet's emergence reveals a subtler story. Internet governance is "bottom-up," yet is nonetheless contingent on a shared TCP/IP protocol as a central technical organizing principle. This protocol was developed by public officials in the Department of Defense and implemented with their backing. In other words, the network began with a strong, state-backed public voice that was able to implement a shared technical standard. Non-state-driven governance then developed atop this common central infrastructure. For a summary of the history of the Internet's development, see generally BARRY M. LEINER ET AL., BRIEF HISTORY OF THE INTERNET (1997), https://www.internetsociety.org/wp-content/uploads/2017/09/ISOC-History-of-the-Internet_1997.pdf [https://perma.cc/92R2-H38V]. *See also* David D. Clark, *The Design Philosophy of the DARPA Internet Protocols*, 18 ACM SIGCOMM COMPUTER COMM. REV. 106, 107 (1988) (describing DARPA's role in early Internet architecture).

For AI, the reality is that there is no robust, democratically accountable state body with which to forge public-private partnerships. Following the money underscores the extent to which the public sector has lagged the private sector in AI research and development—and why it will not be so simple to change this contemporary balance.[197] The February 2019 "American AI Initiative" did not include any lump sum funding for AI, instead directing federal funding agencies to prioritize AI investments at their own discretion.[198] The subsequent 2019 AI R&D National Strategy takes the same line, providing "an expectation for the overall portfolio for Federal AI R&D investments" by each agency without any material backing.[199] Rhetoric about priorities notwithstanding, the punchline is that public expenditures do not come close to the scale and scope of private-side R&D expenditures. According to a report by Stanford University's AI Index, two firms alone (Amazon and Alphabet) invested a combined $30 billion in R&D in 2017.[200] By comparison, federal government agencies requested nearly $1 billion for non-defense AI research for the 2020 fiscal year,[201] and

197. Because they are not subject to democratic checks through the political process, this Article places non-profit and academic actors on the private side of the ledger. It nonetheless recognizes that their incentives are ostensibly distinct from commercial profit motives.

198. *See* Will Knight, *Trump Has a Plan to Keep America First in Artificial Intelligence*, MIT TECH. REV. (Feb. 10, 2019), https://www.technologyreview.com/s/612926/trump-will-sign-an-executive-order-to-put-america-first-in-artificial-intelligence [https://perma.cc/BAW9-Q5WD] (quoting a former Obama Administration official's analysis: "The plan is aspirational with no details and is not self-executing"); *see also* Matthew Hutson, *Trump to Launch Artificial Intelligence Initiative, But Many Details Lacking*, SCI. MAG. (Feb. 11, 2019, 12:01 AM), https://www.sciencemag.org/news/2019/02/trump-launch-artificial-intelligence-initiative-many-details-lacking [https://perma.cc/9YTM-HYWK]. To avoid speculation about future budgetary developments or the not-yet fully realized impact of 2020 budgetary allocations, this Article focuses on data from 2019 and earlier. It is nonetheless worth noting that the President's 2021 budget includes proposed increases in federal budgetary allocations for AI R&D. *See* Office of Sci. & Tech. Policy, *President Trump's FY 2021 Budget Commits to Double Investments in Key Industries of the Future*, WHITE HOUSE (Feb. 11, 2020), https://www.whitehouse.gov/briefings-statements/president-trumps-fy-2021-budget-commits-double-investments-key-industries-future [https://perma.cc/T5ZS-UVA4] (indicating "significant increase in nondefense AI R&D compared to the FY 2020 Budget," with plans to double spending by 2022). That said, any such increase not only remains subject to congressional approval, but also appears targeted at economic growth and national security—while simultaneously cutting overall federal R&D investments for basic research. *See* David Shepardson, *Trump Administration to Propose Big Jump in Funding For AI, Quantum R&D: Sources*, REUTERS (Feb. 7, 2020, 2:22 PM), https://www.reuters.com/article/us-usa-trump-budget-research/trump-administration-to-propose-big-jump-in-funding-for-ai-quantum-rd-sources-idUSKBN2012OK [https://perma.cc/D7LP-6XGB] (reporting potential "big jump" in nonclassified AI R&D funding along with rumored cuts in overall federal R&D expenditures and quoting White House advisor's statement that "America's economic strength and national security depend on" the proposed budgetary increases). And as discussed above, historic and ongoing private sector investments outmatch even the doubled federal figure.

199. *See* 2019 NATIONAL AI R&D STRATEGIC PLAN, *supra* note 28, at i.

200. YOAV SHOHAM ET AL., AI INDEX 2018 ANNUAL REPORT 58 (2018) ("In 2017, private technology companies like Amazon and Alphabet invested $16.1 billion and $13.9 billion, respectively, in R&D. To put this in perspective, the total budget for the NSF, together with DARPA and DOT's investment in autonomous and unmanned systems totals $5.3 billion in the 2019 budget."). 2017 is the most recent year for which this data is available.

201. Sara Castellanos, *Executives Say $1 Billion for AI Research Isn't Enough*, WALL ST. J. (Sept.

the federal government's projected investment in AI R&D across federal civilian agencies and the Department of Defense ("DOD") is estimated at $4.98 billion for fiscal year 2020.[202] This federal investment amounts to a small fraction of what private firms have invested—and are continuing to—invest.[203] Nor can potential future investment change the contemporary dynamic: the sums spent by private firms are more than five times the combined 2019 budgets for the National Science Foundation ("NSF"), the Defense Advanced Research Projects Agency ("DARPA"),[204] and the Department of Transportation ("DOT") investments in autonomous and unmanned systems.[205] And in addition to corporate spending, consider, for example, that a single individual pledged $125 million over three years for a "common sense AI" initiative,[206] and that there are scores of such initiatives[207] further shifting the epicenter of AI R&D outside of the government sector.

Federal AI investments in basic research, moreover, have thus far tilted

10, 2019), https://www.wsj.com/articles/executives-say-1-billion-for-ai-research-isnt-enough-11568153863 [https://perma.cc/2Q5S-VKZV] (chronicling comments by executives at Intel and Nvidia that these investments are "not nearly enough"); *see also* Brandi Vincent, *Administration Projects Agencies Will Spend $1 Billion on Artificial Intelligence Next Year*, NEXTGOV (Sept. 10, 2019), https://www.nextgov.com/emerging-tech/2019/09/administration-projects-agencies-will-spend-1-billion-artificial-intelligence-next-year/159781 [https://perma.cc/22JA-T4CU] (same). For a detailed summary of the proposed investments, see NAT'L SCI. & TECH. COUNCIL, THE NETWORKING & INFORMATION TECHNOLOGY RESEARCH & DEVELOPMENT PROGRAM SUPPLEMENT TO THE PRESIDENT'S FY2020 BUDGET (Sept. 2019), https://www.whitehouse.gov/wp-content/uploads/2019/09/FY2020-NITRD-AI-RD-Budget-September-2019.pdf [https://perma.cc/5MNR-3RFG].

202. YOAV SHOHAM ET AL., AI INDEX 2019 ANNUAL REPORT 95 (2019).

203. *See* Steve Lohr, *At Tech's Leading Edge, Worry About a Concentration of Power*, N.Y. TIMES (Sept. 26, 2019), https://www.nytimes.com/2019/09/26/technology/ai-computer-expense.html [https://perma.cc/8HPX-KG9Y] (reporting that "increasingly expensive" AI research demands "complex calculations done by giant data centers," noting computer scientists' concerns that "pioneering artificial intelligence research will be a field of haves and have-nots," and predicting that "the haves will be mainly a few big tech companies like Google, Microsoft, Amazon and Facebook, which each spend billions a year building out their data centers").

204. DARPA sits within the DOD and also supports and funds AI research. *See Our Research*, DEF. ADVANCED RES. PROJECTS AGENCY, https://www.darpa.mil/our-research [https://perma.cc/EG9S-HG5T].

205. *See supra* note 200.

206. The individual is Paul Allen, who launched the Allen Institute for AI, or AI2, in 2014. *See About*, ALLEN INST. FOR AI, https://allenai.org/about.html [https://perma.cc/UPX2-LXUD]. In early 2018, he pledged another $125 million over three years for a new "common sense AI" initiative, Project Alexandria. Paul G. Allen, The Allen Institute for Artificial Intelligence to Pursue Common Sense for AI, PR NEWSWIRE (Feb. 28, 2018), https://www.prnewswire.com/news-releases/the-allen-institute-for-artificial-intelligence-to-pursue-common-sense-for-ai-300605609.html [https://perma.cc/LE4F-EMDV].

207. *See* Tate Williams, *As Concern Grows, Another Philanthropy-Backed AI Watchdog Launches*, INSIDE PHILANTHROPY (Jan. 13, 2017), https://www.insidephilanthropy.com/home/2017/1/13/as-concern-grows-another-philanthropy-backed-ai-watchdog-launches [https://perma.cc/L75X-RC99] (discussing a number of privately funded organizations, including Elon Musk's $10 million contribution to the Future of Life Institute and the industry-backed Partnership on AI).

toward the military and intelligence sectors.[208] For instance, the Department of Defense spent an unspecified $7.4 billion on AI in 2017[209] and has presumably made additional classified expenditures. The DOD made two further notable investments in 2018: First, in July 2018, it established a Joint Artificial Intelligence Center ("JAIC") dedicated to AI production and prototyping.[210] Second, in September 2018, it announced a two-billion-dollar campaign to develop the "next wave" of AI technologies.[211] Furthermore, in addition to AI research by DARPA in the DOD, the Office of the Director of National Intelligence's Intelligence Advanced Research Projects Activity ("IARPA") has several AI research projects.[212] Any military research might be dual-use in the sense that it could crossover to civilian applications.[213] Indeed, we owe transformative technologies such as the Internet to DARPA.[214] Yet a much longer timeline is forecast for such R&D efforts,[215] delaying any such cross-pollination. And in the meantime, private sector investments remain far greater than public sector R&D. Such crossover, moreover, is distinct from sustained programmatic support that comes from an institution that is not motivated primarily by military or intelligence concerns.[216]

208. There are limited exceptions, such as agricultural science research supported by NSF and biomedical research supported by the National Institute of Health. *See* 2019 NATIONAL AI R&D STRATEGIC PLAN, *supra* note 28, at 7 (discussing recent agency research and development programs). That said, as discussed above, private investments surpass government investments by a large margin.

209. GOVINI, DEPARTMENT OF DEFENSE ARTIFICIAL INTELLIGENCE, BIG DATA AND CLOUD TAXONOMY 7 (2018).

210. *See About the JAIC*, JOINT ARTIFICIAL INTELLIGENCE CTR., https://www.ai.mil/about.html [https://perma.cc/9PDC-9LBN].

211. *See AI Next Campaign*, DEF. ADVANCED RES. PROJECTS AGENCY, https://www.darpa.mil/work-with-us/ai-next-campaign [https://perma.cc/M5EG-GPWE].

212. *See, e.g.*, *Deep Intermodal Video Analytics (DIVA)*, INTELLIGENCE ADVANCED RES. PROJECTS AGENCY, https://www.iarpa.gov/index.php/research-programs/diva [https://perma.cc/5QHS-KMKX]; *Integrated Cognitive-Neuroscience Architectures for Understanding Sensemaking (ICArUS)*, INTELLIGENCE ADVANCED RES. PROJECTS AGENCY, https://www.iarpa.gov/index.php/research-programs/icarus [https://perma.cc/WRC7-R4SX]; *Knowledge Representation in Neural Systems (KRNS)*, INTELLIGENCE ADVANCED RES. PROJECTS AGENCY, https://www.iarpa.gov/index.php/research-programs/krns [https://perma.cc/J3LW-B65S].

213. DARPA's explainable AI work in particular, which could assist in making AI more interpretable, might have significant civilian applications. *See* Turek, *supra* note 112.

214. *See Paving the Way to the Modern Internet*, DEF. ADVANCED RES. PROJECTS AGENCY, https://www.darpa.mil/about-us/timeline/modern-internet [https://perma.cc/QVD4-ASLF].

215. *See The Public Policy Challenges of Artificial Intelligence*, HARV. KENNEDY SCH.: INST. POL. (Feb. 15, 2018), http://iop.harvard.edu/forum/public-policy-challenges-artificial-intelligence [https://perma.cc/3QCL-CTEV] (interviewing IARPA Director Jason Matheny, who predicts a ten-year timeline for IARPA's AI R&D).

216. Some state-level innovation is promising, such as efforts in New York to pass an "algorithmic accountability" statute. *See* Lauren Kirchner, *New York City Moves to Create Accountability for Algorithms*, PROPUBLICA (Dec. 18, 2017, 12:08 PM), https://www.propublica.org/article/new-york-city-moves-to-create-accountability-for-algorithms [https://perma.cc/9SQX-GKVG]. However, this Article focuses on federal action, both to situate AI within federal administrative paradigms and because state-level policy seems more likely to regulate effects of the technology than to foster basic R&D in a way

Nor do these trends seem likely to change anytime soon. Consider a 2019 government investment in AI: the 2019 National Defense Authorization Act's allocation of up to ten million dollars in support of an independent executive body, the National Security Commission on Artificial Intelligence ("NSCAI"). NSCAI is expected to issue recommendations on "action by the executive branch and Congress related to artificial intelligence, machine learning, and associated technologies, including recommendations to more effectively organize the Federal Government."[217] Given the inclusion of this funding in a defense authorization bill, it is likely that the focus will be on national security implications of AI—particularly because two-thirds of its members were chosen by congressmembers who sit on armed services and intelligence committees,[218] and who ostensibly selected individuals whom they believe will advance their institutional objectives as representatives of those governing bodies.[219]

If money talks, then the takeaway from this conversation is that the federal government in general and its non-military officials and entities in particular have not been sufficiently forceful in their AI speech.[220] The

that would be a direct counterpoint to actions by private firms.

217. Tajha Chappellet-Lanier, *Alphabet, Microsoft Leaders Named to National Security Commission on Artificial Intelligence*, FEDSCOOP (Nov. 14, 2018), https://www.fedscoop.com/alphabet-microsoft-leaders-named-national-security-commission-artificial-intelligence [https://perma.cc/8R5M-UQQY].

218. John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, § 1051, 132 Stat. 1636 (2018) (describing appointment process).

219. To NSCAI's credit, two of its members indicated an interest in broader public input with a call for "original, creative ideas that challenge the status quo, shake our assumptions, and will cause us to reconsider the arguments we've already heard and hear new arguments in a different light." Robert Work & Eric Schmidt, *In Search of Ideas: The National Security Commission on Artificial Intelligence Wants You*, WAR ON THE ROCKS (July 18, 2019), https://warontherocks.com/2019/07/in-search-of-ideas-the-national-security-commission-on-artificial-intelligence-wants-you [https://perma.cc/9BSB-F3AT]. That said, a call for ideas for NSCAI to consider is not synonymous with public accountability for the Commission's mission and processes. The initial report to Congress is light on substance and primarily describes preliminary organizational steps and plans to engage in the "assessment phase" of the work in 2020. *See generally* NAT'L SEC. COMM'N ON ARTIFICIAL INTELLIGENCE, INITIAL REPORT (July 31, 2019), https://www.nscai.gov/about/reports-to-congress#h.p_via-X-XQewCS [https://perma.cc/TK8Z-5GKT]. The November 2019 interim report is heartening insofar as it indicates consultation with over one hundred organizations. *See generally* NAT'L SEC. COMM'N ON ARTIFICIAL INTELLIGENCE, INTERIM REPORT 66–67 (Nov. 2019), https://www.nscai.gov/about/reports-to-congress#h.p_via-X-XQewCS [https://perma.cc/W7C6-ALLR]. It remains to be seen whether these consultation efforts permit robust public engagement with the NSCAI's plans or process in ways that meaningfully affect outputs. *Cf.* Erica Pandey, *The Unanswered Questions in America's AI Strategy*, AXIOS FUTURE (Nov. 6, 2019), https://www.axios.com/newsletters/axios-future-ea38a66a-7c97-4cea-86ea-532ea7dcd908.html?chunk=1&utm_term=emshare#story1 [https://perma.cc/6XK4-NMJ3] ("If you look at what is being said, we continue to hear the right things . . . . But if you look at what is being done, it's not a hell of a lot.").

220. *See* Fei-Fei Li & John Etchemendy, *We Need a National Vision for AI*, HUM.-CENTERED ARTIFICAL INTELLIGENCE STAN. UNIV. (Oct. 22, 2019), https://hai.stanford.edu/news/we-need-national-vision-ai [https://perma.cc/XYB7-3MQY] (characterizing $1 billion federal commitment to AI as "encouraging, but not nearly enough" and calling for $12 billion annual investment in AI R&D, education, and support for entrepreneurs).

strategic context for AI development thus features a public sector that lags woefully behind private players in terms of both resource allocations and policy development. Academia might in theory counterbalance some of these trends; however, recent history suggests that private companies are hiring a large proportion of the extremely limited supply of global AI talent out of academic labs, thereby taking would-be contributors away from public sector R&D efforts or basic research.[221] Furthermore, even where ostensibly public AI entities do exist, they continue to rely on private support both for technical assistance[222] and for policy counsel.[223] There is nothing inherently wrong with such private-side involvement. To the contrary, private expertise may be needed,[224] and it may be a wise move to bring in external counsel and include industry perspectives. Without a countervailing voice from inside the state that the public can hold directly accountable, however, such representation may be little more than private voices cloaked in public garb. And without a public lead in the first instance, this combination of forces is

221. For instance, in one well-publicized incident, Uber hired a large proportion of Carnegie Mellon's National Robotics Engineering Center ("NREC"), including its director. *See* Josh Lowensohn, *Uber Gutted Carnegie Mellon's Top Robotics Lab to Build Self-Driving Cars*, VERGE (May 19, 2015, 4:07 PM), https://www.theverge.com/transportation/2015/5/19/8622831/uber-self-driving-cars-carnegie-mellon-poached [https://perma.cc/6TSR-48J5]; *see also* Kaveh Waddell, *1 Big Thing: A Feud atop AI's Commanding Heights*, AXIOS FUTURE (Sept. 6, 2018), https://www.axios.com/newsletters/axios-future-f9f455cd-d89c-4406-a4b4-acca49281d12.html?chunk=0#story0 [https://perma.cc/BMR3-PV8M] (reporting on Big Tech's luring away of top AI academic talent).

222. *See* Patrick Tucker, *The Pentagon Is Building an AI Product Factory*, DEFENSE ONE (Apr. 19, 2018), https://www.defenseone.com/technology/2018/04/pentagon-building-ai-product-factory/147594 [https://perma.cc/7RKK-HWRC] ("Following the [Project] Maven example, the military will rely mostly on contractors and third parties for its AI, and the [JAIC] could help."). Though Project Maven's contract with Google was cancelled after employees protested the company's military work, the overall dynamic has not changed. *See* Lara Seligman, *Pentagon's AI Surge on Track, Despite Google Protest*, FOREIGN POL'Y (June 29, 2018, 4:11 PM), https://foreignpolicy.com/2018/06/29/google-protest-wont-stop-pentagons-a-i-revolution [https://perma.cc/B78D-8V8T] ("Google is not the only company that can do [the Project Maven] work. Its decision to pull out of Project Maven creates a market opening for other companies such as Amazon, Microsoft, and IBM . . . .").

223. For instance, corporate representatives form the largest bloc of the fifteen seats on the NSCAI. Specifically, six seats went to individuals affiliated with commercial firms, including Amazon, Google, Oracle, and Microsoft. The remaining nine seats are split among scholars and researchers (three seats), former government employees from the FCC and Department of Defense (three seats), and current government employees from IARPA, NASA, and the U.S. Senate (three seats). *See* Justin Doubleday, *Top Tech Execs Named to New National Security Commission on Artificial Intelligence*, INSIDE DEFENSE (Jan. 10, 2019, 12:44 PM), https://insidedefense.com/insider/top-tech-execs-named-new-national-security-commission-artificial-intelligence [https://perma.cc/DE4L-B3DG].

224. *See* Ash Carter, *Shaping Disruptive Technological Change for Public Good*, HARV. KENNEDY SCH.: BELFER CTR. (Aug. 2018), https://www.belfercenter.org/publication/shaping-disruptive-technological-change-public-good [https://perma.cc/FZ7A-4N4Z] (critiquing Project Maven protests at Google and asking, "who better than they at Google, who are immersed in this technology, to steer the Pentagon in the right direction?"); Will Knight, *Why AI Researchers Shouldn't Turn Their Backs on the Military*, MIT TECH. REV. (Aug. 14, 2018), https://www.technologyreview.com/s/611852/why-ai-researchers-shouldnt-turn-their-backs-on-the-military [https://perma.cc/XDD4-D2XK] ("AI researchers must be a part of these conversations [with governments and militaries], as their technical expertise is vital to shaping policy choices.").

not a credible model of negotiated governance between public and private representatives.

* * *

Calls for public regulatory intervention for AI are correct to recognize that the lack of a public actor is symptomatic of administrative dysfunction. But the fix is not obvious. One set of discrete administrative corrections—prescriptive regulation—is likely to work only in zones where the goal is to control a bad outcome, and where that risk is so acute that it is acceptable if sector-specific control comes at the cost of cross-sectoral principles and norm development. Those who seek a more holistic solution might understandably look to collaborative governance strategies. Yet it is premature to do so without reckoning with the current balance of public and private resources and expertise. To begin a public governance project, we need interventions that more squarely contend with both the complex combination of AI's technical attributes and the strategic context of AI development. The next Part begins this conversation.

III. IN SEARCH OF ACCOUNTABILITY

AI governance is a Gordian knot of challenges. The heart of the problem is that the United States is presently lacking a system that allocates decisionmaking authority for research and development between public and private actors in a way that is publicly accountable, from the start. This issue is compounded by the difficulty of shifting away from the status quo. As Part II addresses, AI's technical attributes[225] along with the private sector's lead in AI investment[226] will make it challenging for the public sector to suddenly spring into action as either a regulator or a collaborative partner.

As the following Sections discuss, there is no easy way to cut through this knot. Even if all practical issues were resolved, a critical theoretical issue remains: the way that AI technology is developed through code blurs the line between public and private governance choices. This theoretical challenge both compounds practical issues and will be the hardest to address. It is also where the most is at stake. In a world where private actors lead research and development, we risk slipping into a new private order of code-driven governance if we do not take steps today to preserve democratic accountability and sustain a meaningful role for public law alongside private market pressures and social norms. After further examining the nature of these challenges, this Part urges us to intervene by looking to the algorithmic fuel—data—to inject public values into the process of technical innovation

225. *See* discussion *infra* Section II.A.2.
226. *See* discussion *infra* Section II.B.2.

itself.

A. CODE AS POLICY

If we are to cultivate new governance solutions, then there must be a fundamental rethinking of what constitutes a public policy decision in the context of digital emerging technologies. When it comes to AI, we are dealing with *code as policy*: the code-based decisions that make algorithmic technologies like AI possible themselves embed values and embody normative tradeoffs. This is not to claim that code is synonymous with formal policy promulgated by the state. If we move too quickly to conclude that code is policy, then the category of "policy" risks becoming so capacious that it means nothing.[227] But formalism aside, the line between code and what we have traditionally associated with value-driven regulatory interventions is functionally blurring. Accepting Lessig's understanding of regulation as "the constraining effect of some action, or policy, whether intended by anyone or not," code for AI is its regulatory policy.[228]

Algorithmic decisions at the programming and design level carry stark regulatory implications for both human values and human safety. A version of this point holds, to be sure, in many domains. Consider a corporation's decision to move headquarters to a different location, which might have substantial economic impacts on the old and new locales as well as reshape the social norms of each neighborhood.[229] What makes AI unique is the potential for such effects through large and small business decisions and seemingly technical choices about things like statistical models and data. For algorithms, these macro- and micro-decisions are intertwined to an even greater extent with economic pressures, social norms, and physical safety.

Consider, for instance, Google's introduction of TensorFlow Federated,

227. *Cf.* Christopher T. Bavitz. *The Right to Be Forgotten and Internet Governance: Challenges and Opportunities*, 2 LATIN AMER. L. REV. 1, 8 (2019) ("If [Internet governance] is everything, [Internet governance] is nothing.") (citing Lawrence B. Solum, *Models of Internet Governance*, *in* INTERNET GOVERNANCE: INFRASTRUCTURE AND INSTITUTIONS 49 (Lee A. Bygrave & Jon Bing eds., 2009)).

228. *See* Lessig, *supra* note 44, at 662 n.1. This more capacious definition is distinct from the term "public regulation," which this Article uses to refer to regulations promulgated by a government agency. *See supra* note 44. Whether this claim extends to other emerging technology is a matter for future research. No matter where the line is ultimately drawn, AI represents a leading instance of the phenomenon discussed in this Article.

229. *See, e.g.*, Zoë Corbyn*, Is San Francisco Losing Its Soul?*, GUARDIAN (Feb. 23, 2014, 4:00 PM), https://www.theguardian.com/world/2014/feb/23/is-san-francisco-losing-its-soul [https://perma.cc/2YCD-UCG2] (assessing whether the "tech boom" in San Francisco is compromising the city's bohemian soul); Harrison Jacobs, *A Walk Through Seattle's 'Amazonia' Neighborhood Made Me Understand Why Any City Would be Wary of Hosting the Tech Giant*, BUS. INSIDER (Feb. 14, 2019, 1:41 PM), https://www.businessinsider.com/amazon-hq-photo-tour-seattle-make-any-city-wary-of-amazon-hq2-2017-12 [https://perma.cc/MA7Z-K23F] (considering "outsized impact" of Amazon's original headquarters on Seattle, Washington).

an open source AI training system that incorporates federated learning.[230] Federated learning promises to make AI development more "privacy-sensitive" because it does not expose the information in the centralized data set on which the ML algorithm is trained.[231] Instead, it maintains the information on the individual user's device, allowing the algorithm to run without exposing data and reporting just the outcome back to the centralized actor.[232] As a result, running the algorithm with such a system produces a more privacy-sensitive outcome than alternative methods.

Consider now an engineering team at a start-up that is trying to determine the most economical way to scale its AI training system. Suppose that TensorFlow is highly compatible with its existing code base and that the team is familiar with the technical syntax used in the system. Without any regulation or legal compulsion, the sensible business decision to use a federated learning template like TensorFlow will determine how privacy-protective the algorithm will be. On the other hand, countervailing economic pressures—such as, for instance, a private firm's desire to compete with Google and develop a distinct, proprietary model—might lead its developers to deploy a different training model that requires exposing all the training data to the company, making it less solicitous of individual users' informational privacy. Regardless of the choice, the result is that protecting a value (here, privacy), reflects more than formal, top-down policy interventions or administrative requirements. Choices about the design of algorithmic systems, often made for business or technical reasons, will effectively regulate the ways that this technology interacts with human values.

This same point holds when an algorithm is deployed in an application that has an even more direct impact in the physical world.[233] Consider, for instance, the use of an algorithm for medical diagnoses. To implement such an algorithm, private actors must choose what data sets and data points are sufficiently representative to serve as the training data. And if, for example,

230. *See* Laura Hautala, *Google Tool Lets Any AI App Learn Without Taking All Your Data*, CNET (Mar. 6, 2019, 8:00 AM), https://www.cnet.com/news/google-ai-tool-lets-outside-apps-get-smart-without-taking-all-your-data [https://perma.cc/EEJ9-JCCJ].

231. For an overview of federated learning, see Brendan McMahan & Daniel Ramage, *Federated Learning: Collaborative Machine Learning Without Centralized Training Data*, GOOGLE AI BLOG (Apr. 6, 2017), https://ai.googleblog.com/2017/04/federated-learning-collaborative.html [https://perma.cc/XQ89-DXBP].

232. Hautala, *supra* note 230.

233. *See, e.g.*, Matt Richtel & Conor Dougherty, *Google's Driverless Cars Run into Problem: Cars with Drivers*, N.Y. TIMES (Sept. 1, 2015), https://www.nytimes.com/2015/09/02/technology/personaltech/google-says-its-not-the-driverless-cars-fault-its-other-drivers.html [https://perma.cc/8QAK-XJNZ]; Chris Williams, *Stop Lights, Sunsets, Junctions Are Tough Work for Google's Robo-Cars*, REGISTER (Aug. 24, 2016, 8:31), https://www.theregister.co.uk/2016/08/24/google_self_driving_car_problems [https://perma.cc/P3KX-TPV7].

IBM opts to rely on Sloan-Kettering Cancer Center's data to train a ML model, then that model will, for better or worse, embed "Sloan Kettering['s] . . . particular philosophy about how to do medicine."[234] The line between initial decisions about how to construct the technology and that code's effect on human safety is thus ever-more blurred.

Such decisions, moreover, carry a hefty normative punch. Consider AVs once more. By definition, AVs require integrating algorithmically-directed vehicles into human movement patterns and traffic flows.[235] Accordingly, developers must decide how to incorporate machine processing alongside existing human driving norms.[236] Imagine, for instance, an AV getting stuck at a four-way intersection because it cannot make eye contact with a human driver to negotiate a right of way. A stalled AV may seem harmless. But it is not: the AV's attempt to follow hard-coded rules in a situation where human drivers would deviate from such rules and subscribe to dynamic rules of the road instead could easily cause collisions.

Nor does this issue end with stop signs and road directives. Deeper normative questions are at stake. In the context of AVs or other physical examples of AI technology, whose safety is to be maximized? Should the developers protect the interests of the passenger in the AV, those of other drivers, or some other individual or set of people with whom the car interacts? Similar questions apply in other domains, from algorithmic medicine to algorithmic risk assessment in criminal justice.[237] The challenge for AI is that in almost every context in which AI is applied, private entities are likely to have vital knowledge and expertise to make these normative calls at the same time that delegating authority to them creates a regime that is less democratically accountable to the citizens who must contend with any consequences.

This situation has important parallels to the "governance-by-design"

234. Frederic Lardinois, *IBM Watson CTO Rob High on Bias and Other Challenges in Machine Learning*, TECH CRUNCH (Feb. 27, 2018, 1:35 PM), https://techcrunch.com/2018/02/27/ibm-watson-cto-rob-high-on-bias-and-other-challenges-in-machine-learning [https://perma.cc/Z7NQ-MSJX].

235. This analysis assumes that, at least in the near term, there will be some amount of necessary interaction between AVs and human actors. *Cf. Automated Vehicles for Safety*, NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety [https://perma.cc/4DY4-P9Y4] (detailing five levels of automation for AVs, four of which entail less than complete automation). A complete replacement of human drivers (level five of the NHTSA guidance) would present a different picture. It also assumes that, in the near term, the architecture of roadways and pathways will not be completely reconstructed so as to avoid the need for any human/AV interactions.

236. *See, e.g.*, Richtel & Dougherty, *supra* note 233; Williams, *supra* note 233.

237. Humans may also be uncertain how to answer these questions, in which case it is possible that different companies would select different reasonable options. This resolution might be substantively acceptable. Still, without some form of public oversight or public input, the process is procedurally anti-democratic. *See* discussion *infra* Section III.B.

literature, which focuses on how technical decisions by public actors are implementing particular directives. As Mulligan & Bamberger argue, " 'governance-by-design'—the purposeful effort to use technology to embed values—is becoming a central mode of policymaking, and . . . our existing regulatory system is fundamentally ill-equipped to prevent that phenomenon from subverting public governance."[238]

This Article's distinct contribution is to recognize that a version of the same underlying problem exists when a *private* actor uses technology to embed values in algorithmic applications that implicate public wellbeing. Given private control over the design and development of AI technologies and the force of code as policy, decisions by private actors de facto regulate human behavior. And since most developers of this digital technology sit within entities that are not democratically accountable, what this Article terms *private governance* choices are, functionally, policymaking choices. These effects are cross-cutting. AI has the potential not only to alter traditional societal institutions in intangible ways (such as by affecting democratic electoral processes through virtual spaces), but also to affect physical wellbeing and safety (by interacting with humans in physical space, in applications ranging from AVs to military algorithms). And where micro-level design choices about AI regulate outcomes in these critical domains, yet remain the province of private governance, it erodes our baseline understanding of governance as democratically—not commercially—responsive.

B. THE ALGORITHMIC GOVERNANCE DILEMMA

The rise of private governance raises fundamental questions about what democratic algorithmic governance demands. The following Sections offer that the most auspicious long-term strategy is to recast the issues to focus on the bedrock public values that we want the algorithmic governance system to protect and then take affirmative steps to embed those values in the design and development of the technology.

1. The Public-Private Dilemma

The force of code as policy and the reality that commercial actors are essential to the current trajectory of AI development are leading us into an

238. Mulligan & Bamberger, *Saving Governance*, *supra* note 78, at 697; *see also* Mulligan & Bamberger, *Procurement as Policy*, *supra* note 38, at 788 (arguing, in the administrative law context, that design choices about algorithmic machine learning systems "make policy," but these "[d]esign decisions are left to private third-party developers" in a way that "abdicate[s]" governmental responsibility over that policy). As detailed below, this Article builds from these arguments to consider how code functionally acts as policy outside of the formal administrative agency context—and why the status quo augurs an era of private governance.

increasingly stark governance dilemma. Given the ways that AI applications touch life in almost every sector and given the private lead in AI development, doing nothing is likely to amplify the private sector's influence. If we wait, let a thousand self-regulatory flowers bloom, and see what flourishes in the free market, then where does public accountability come in? Relying on consumers to vote with their dollars is not an adequate check in the face of an extraordinarily complex technology like AI, particularly when the end user may not even be aware of the underlying algorithmic choices.[239]

Without public regulation or oversight, commercial actors will not necessarily be incentivized to answer more directly to individuals as democratic citizens. Imagine that you are a member of a group, for instance, that a potentially lifesaving technology systematically underserves. This was initially the case, in fact, for a University of Chicago Medicine algorithm that "would have led to the paradoxical result of the hospital providing additional case management resources to a predominantly white, more educated, more affluent population to get them out of the hospital earlier, instead of to a more socially at-risk population who really should be the ones that receive more help."[240] The private actor may sometimes catch these issues, whether out of a desire to do the right thing or to avoid public opprobrium. But without some form of more consistent, ex ante, public check on code-based policy choices, there is no guarantee that private governance will protect core civil liberties, balance risks and benefits in a way that considers normative concerns as well as economic efficiencies, or look after the interests of marginalized populations. This concern with public accountability, moreover, holds even if private products are beneficial for many members of the public.

On the other hand, the opposite approach—prescriptive regulation—is unlikely to suffice in the long term, given private expertise, resources, and

239. Corporate trade secret claims, which can further obscure what is happening under the algorithmic hood, make technological opacity even more acute. Particularly in criminal justice settings, moreover, proprietary algorithms come with additional costs. *See* Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 659 (2018) (contending "that [private law] secrecy is pervasive, problematic, and ultimately unnecessary for the production of well-designed criminal justice tools"); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1343 (2018) (arguing "that trade secrets should not be privileged in criminal proceedings"). More disclosure about algorithmic choices might help. *See, e.g.*, *Technical Solution*, DATA NUTRITION PROJECT, https://datanutrition.org [https://perma.cc/4QCC-7K3S] (proposing a "nutrition label" for datasets to "drive the creation of better, more inclusive algorithms"). But relying on transparency alone to empower consumers is not sufficient as a governance strategy.

240. Matt Wood, *How to Make Software Algorithms for Health Care Fair and Equal for Everyone*, U. CHI. MED. (Dec. 2, 2018), https://www.uchicagomedicine.org/forefront/patient-care-articles/2018/december/how-to-make-software-algorithms-for-health-care-fair-and-equal-for-everyone [https://perma.cc/YM3V-VXH6].

the classic pacing problem in regulating emerging technologies.[241] It also risks squelching promising innovations that might improve human wellbeing. For instance, if AI "can spot the warning signs of disease before we even know we are ill ourselves," and we cannot necessarily predict or discern the patterns the AI is detecting,[242] then there may be good reason to hesitate before we prescribe what the algorithm is cleared to do. In the abstract, a collaborative governance solution does seem like the best fit.[243] And yet the public presence necessary for accountable collaborative governance to succeed does not presently exist.

This conundrum is what this Article terms the *public-private dilemma*. Rather than try to escape the dilemma or control it with traditional regulatory tools, a more auspicious long-term strategy is to reframe why the dynamics of the dilemma worry us. The bedrock issues may in fact turn less on control by public as opposed to private actors, and, as the following Sections address, more on what underlying values and/or physical concerns are at stake when a public or private entity makes the call in a particular context. For instance, in some domains—like, perhaps, criminal justice—we might be especially concerned about profit motives entering the equation because of the liberty interests at stake. And in such domains, the locus of conversation might need to be whether we should implement an AI solution in a given setting at all, and if so, with what public governance guardrails for any private contributions.[244]

This sort of dialogue permits a different set of questions to emerge. In a particular setting, what defines a "good (enough)" substantive outcome, and what metrics are we applying? If a series of private firms offer a set of, say, reasonably safe AV products, is it sufficient for the public to pick and choose among them ex post? Or do we insist on more ex ante oversight of micro-level coding decisions themselves—perhaps by enumerating a core underlying value and then demanding auditing trails or independent auditors?[245] Even more fundamentally, are there some contexts in which

241. *See supra* note 106 and accompanying text.

242. Leah Kaminsky, *The Invisible Warning Signs that Predict Your Future Health*, BBC FUTURE (Jan. 16, 2019), http://www.bbc.com/future/story/20190116-the-invisible-warning-signs-that-predict-your-future-health [https://perma.cc/W4RS-MK7G].

243. *See* Kaminski, *supra* note 24, at 1557–62 (discussing benefits of collaborative governance in general and its fit for algorithmic technologies in particular).

244. *Cf.* Frank Pasquale, *The Second Wave of Algorithmic Accountability*, LAW & POL. ECON. BLOG (Nov. 25, 2019), https://lpeblog.org/2019/11/25/the-second-wave-of-algorithmic-accountability [https://perma.cc/ZK5N-JVA8] (discussing a "second wave" of algorithmic accountability that asks whether algorithmic systems "should be used at all—and, if so, who gets to govern them").

245. For an example of such a regime, albeit one initiated and run by non-state actors, consider the Global Network Initiative ("GNI"). GNI is a multi-stakeholder group of academic organizations, civil society organizations, and telephony and Internet companies, including Facebook, Google, and Microsoft.

public input would bar the development of a private technology, wholesale? And at bottom: under what conditions, if any, are we comfortable delegating choices that affect traditionally public matters to commercial firms?

These questions turn on the relationship among the public, commercial firms, and the state, and depend to a large extent on how much responsibility we expect individual citizens to bear in evaluating AI options in the marketplace. This is an important debate, and this Article insists that it should properly be a matter of public discourse. That said, these are extraordinarily complex technologies for technically adept experts, let alone for laypersons. A truly inclusive dialogue about AI is therefore even more challenging, and the United States would do well to double down on public education about AI as an initial step. Finland, for example, announced a national initiative to create "Real AI for Real People in the Real World" through online classes and programs that promote AI literacy for all.[246]

The problem, though, is that education and public discourse take time, and we need public accountability now. Nor is it realistic to assume that citizens have the time, energy, or desire to vote on every algorithmic governance question. Assuming that we are not ready to concede the end of democratic governance as we know it, we need to think creatively about structural steps to keep governance responsive to public inputs and public-minded priorities in the immediate term.

* * *

The search for accountability now requires more directly confronting the basic private governance dilemma. The contemporary barrier to a potential third way—more collaborative public-private governance—comes from a combination of practical and theoretical obstacles. Practically, there is inadequate civilian-facing AI research and development. And theoretically, code choices are policy choices, such that private choices affect public-facing outcomes to an even greater degree than in other contexts. Perhaps, then, a fundamentally different tactic is required: in lieu of formal interventions through law, we should consider other regulatory modalities to develop internal public checks, encoded in the design of the technology

See About GNI, GLOBAL NETWORK INITIATIVE, https://globalnetworkinitiative.org/about-gni [https://perma.cc/Z77Q-4HRJ]. Member companies pledge to "protect and advance freedom of expression and privacy" by following agreed-upon principles, *see The GNI Principles*, GLOBAL NETWORK INITIATIVE, https://globalnetworkinitiative.org/gni-principles [https://perma.cc/55JC-J5MD], and implementation guidelines, which include regular corporate assessments through a pre-approved auditing procedure, *see Company Assessments*, GLOBAL NETWORK INITIATIVE, https://globalnetworkinitiative.org/company-assessments [https://perma.cc/V42B-BL8M].

246. *See Our Mission*, FINNISH CTR. FOR ARTIFICIAL INTELLIGENCE, https://fcai.fi/mission [https://perma.cc/9PLQ-6RA2]; *FCAI Education*, FINNISH CTR. FOR ARTIFICIAL INTELLIGENCE, https://fcai.fi/education [https://perma.cc/M9WZ-2MFZ].

itself.[247]

2. Recasting the Terms of the Dilemma: Protecting Values by Design

Recall Lessig's four regulatory modalities: law, norms, markets, and architecture. This Article has examined both the challenges of regulating AI through law alone and the ways in which private control through technical architecture alone may be insufficient, from the perspective of someone concerned with the ways in which AI is touching physical life, affecting human safety, and inflecting public norms. But to date, less attention has been paid to regulation through the market and regulation through norms, and to what role the state might play in fostering market developments or in shaping social norms that take public as well as commercial interests into account.

Taking a step back to consider what inputs are necessary for AI research and development, there are three broad categories of resource needs: computing power, human expertise, and data.[248] Rather than attempt to govern AI as a monolithic unit, a more prudent strategy is to target each of these inputs with an eye to making sure that the development process reflects public voices and values, ex ante. To begin this strategic discourse, this Section focuses on data—the fuel that powers ML—as an example of how this approach might work and considers how we might leverage markets and norms to implement more publicly-accountable data-centered policy.[249]

a. Market-Driven Data Governance

To bolster the public presence in AI, one clear lesson is that the American public sector must invest far more in basic AI research and development, and not merely in AI applications or "technological

247. This solution is also rooted in geopolitical realities. In the international era of AI, no one nation state can hope to promulgate overarching standards. But it might be possible to introduce, for instance, market incentives or employee protections in ways that influence the culture of AI creation, and which thereby carry global effects.

248. *See* Dario Amodei & Danny Hernandez, *AI and Compute*, OPENAI: BLOG (May 16, 2018) https://openai.com/blog/ai-and-compute [https://perma.cc/45ST-3554] ("Three factors drive the advance of AI: algorithmic innovation, data (which can be either supervised data or interactive environments), and the amount of compute available for training."); Kate Crawford & Vladen Joler, *Anatomy of an AI System*, https://anatomyof.ai [https://perma.cc/47UP-GQTS] (discussing "three central, extractive processes that are required to run a large-scale artificial intelligence system: material resources, human labor, and data").

249. In keeping with the suggestions offered here, others have advocated a "public option" for AI, to include "a public data pool that would make data accessible to registered users," increased research and development spending, and the expansion of government attention to AI beyond the military and toward "health care, transportation, energy and other areas." Ben Gansky et al., *Artificial Intelligence Is Too Important to Leave to Google and Facebook Alone*, N.Y. TIMES (Nov. 10, 2019), https://www.nytimes.com/2019/11/10/opinion/artificial-intelligence-facebook-google.html [https://perma.cc/95P2-4W2G]. These and related proposals underscore the pressing need for a data-driven conversation around how to bolster public AI governance, which this Article hopes to catalyze.

revolutions that could . . . create vast new wealth . . . ."[250] It is heartening that the Trump Administration designated AI R&D as a priority in the 2019 budget.[251] But a pledge to invest alone is not sufficient. First, the dollar amount invested in R&D must increase to compete with the private sector.[252] Such basic research should extend beyond national security and intelligence settings. In addition, framing America's national AI plan as "Artificial Intelligence for American Industry" is a categorical error.[253] Research investments should not be calculated in terms of their potential impact on industry and "innovation," with the assumption that we can and should "remov[e] regulatory barriers to the deployment of AI-powered technologies."[254] Rather, the public sector must act as a body that adds a non-economically motivated research agenda to the mix.

Most ambitiously, the state itself could develop public-data backed algorithms, thereby ensuring that publicly-accountable actors control a vital input for the technology.[255] Recall that the dominant AI method at present, ML, relies on access to extremely large data sets, and that data scientists' choices about what data to use to train an algorithm can either mitigate or perpetuate underlying societal biases. By taking on the responsibility to ensure that the data used is not objectionably biased, the state could provide a bulwark against this form of harm. Consider, for example, private facial recognition algorithms sold to police forces despite the fact that they have higher error rates for female and minority faces.[256] Assuming arguendo that such an AI application continues, if the state instead provided the data for the algorithm in such a sensitive setting, then—trusting that the public sector would take care to avoid data bias—this intervention would be an institutionalized check.[257] Though underlying normative questions as well as public "governance-by-design" concerns would remain because opaque

250. WHITE HOUSE OFFICE OF SCI. & TECH. POLICY, SUMMARY OF THE 2018 WHITE HOUSE SUMMIT ON *ARTIFICIAL INTELLIGENCE FOR AMERICAN INDUSTRY* 5 (2018).

251. *See id.*

252. *See* sources cited *supra* notes 186–210 and accompanying text.

253. *See* OFFICE OF SCI. & TECH. POLICY, *supra* note 250, at 1.

254. *See id.* at 5.

255. *See* Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH L. REV. 242, 285 (2019) (discussing "public option" in criminal justice risk assessment algorithms).

256. *See* James Vincent, *AI Researchers Tell Amazon to Stop Selling 'Flawed' Facial Recognition to the Police*, VERGE (Apr. 3, 2019, 12:00 PM), https://www.theverge.com/2019/4/3/18291995/amazon-facial-recognition-technology-rekognition-police-ai-researchers-ban-flawed [https://perma.cc/JXQ6-ZLDE].

257. This point assumes, for purposes of illustration and to engage with facts on the ground, that such use of facial recognition technology is permitted, without endorsing the use of facial recognition technology as a normative matter.

technical choices could still embed values,[258] this step could nonetheless permit greater public control of the technology's development.

Alternatively, the state could filter public interests into the market by offering approved public data sets[259] for private actors to use, paired with carrots or sticks to incentivize them to do so. Public investments in approved datasets could support more publicly-minded AI products, especially if public resources were also invested in differential privacy or other technical solutions to permit access to data without disclosing personal information.[260] Such a move, furthermore, might also be paired with measures such as an ex post sanction if a firm opts to use proprietary data and there is a subsequent safety or ethical issue that use of the public data could have avoided.[261] The bottom line is how to shift the marketplace, such that more than profits drive it, thereby creating space for the public to experience AI's development and articulate norms over time.

b. Norm-Driven Data Governance

A complementary set of interventions would shift the culture of AI development by shaping the professional norms of those working within the industry to foreground data values. One area of particular promise is the cultivation of norms that heed the source of the data.[262] To appreciate why, consider once more the need for datasets—often very large datasets—to train a working ML algorithm. Where does that data come from? Researchers in search of training data will often scrape publicly available datasets, a practice

258. *See* Mulligan & Bamberger, *Saving Governance*, *supra* note 78, at 698 ("Far from being a panacea, governance-by-design has undermined important governance norms and chipped away at our voting, speech, privacy, and equality rights.").

259. *Cf.* INT'L TELECOMM. UNION, AI FOR GOOD GLOBAL SUMMIT REPORT 68 (2017), https://www.itu.int/en/itu-t/ai/documents/report/ai_for_good_global_summit_report_2017.pdf [https://perma.cc/A46V-7T6F] ("Multi-stakeholder partnerships should agree on workable incentives and governance structures for data sharing and should encourage global open standardization and availability of open-source software relevant for AI applications.").

260. For discussion of some ways in which technology might protect informational privacy, see Urs Gasser, Commentary, *Recoding Privacy Law: Reflections on the Future Relationship Among Law, Technology, and Privacy*, 130 HARV. L. REV. F. 61, 66–67 (2016).

261. Again, reserving a more complete exploration of these suggestions and how they could map to different contexts for further research, this solution is put forth as one that might be appropriate in particular contexts. It is not meant to suggest public ownership of all data or a socialized approach to data across the board. In addition, it would require further technical research to implement, given explainability challenges in AI. This Article saves for another day the question of what specific steps would be necessary, as well as questions about how to ensure the security of any such data set.

262. In the past few years, there has been growing attention to the origin of the data that powers AI. *See generally, e.g.*, CRAWFORD & JOLER, *supra* note 248; Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 NYU L. REV. ONLINE 15 (2019). This Article builds from the insights of such scholarship to suggest that the data source is a good place to intervene.

that continues a long-standing industry norm.[263]

The problem, though, is that this industry norm may no longer align with public expectations. Someone who disclosed an image for one purpose might not have agreed to its repurposing for another context.[264] As Accenture Applied Intelligence's global lead for responsible AI puts the problem, "[t]here are ways to use our data today that we were not aware of five, [ten] years ago . . . . How could we [the public] possibly have agreed to a capability that did not exist?"[265] But outright banning the use of any publicly available dataset would be a blunt rule that not only halts much research in its tracks, but also runs against the accepted principle that there is no reasonable expectation of privacy in public.[266] A norm that challenges AI creators to think about data use, in context, is a more principled way to address any gap between data practices and public expectations, while preserving room for ongoing evolution of norms.

The present point is not to articulate the content of such a norm, which should properly evolve over time with the involvement of the communities of interest, but rather to suggest that state and nonprofit actors could actively cultivate a unified professional ethos around AI. These steps might take several forms, from a public consortium modelled after the European Lab for Learning and Intelligent Systems[267] to economic or other incentives for interdisciplinary university initiatives and educational training.[268] The

263. *See* Olivia Solon, *Facial Recognition's 'Dirty Little Secret': Millions of Online Photos Scraped Without Consent*, NBC NEWS (Mar. 12, 2019, 1:32 AM) https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921 [https://perma.cc/XC4H-ZFGD] ("This is the dirty little secret of AI training sets. Researchers often just grab whatever images are available in the wild.").

264. *See* Karen Hao, *IBM's Photo-Scraping Scandal Shows What a Weird Bubble AI Researchers Live in*, MIT TECH. REV. (Mar. 15, 2019), https://www.technologyreview.com/f/613131/ibms-photo-scraping-scandal-shows-what-a-weird-bubble-ai-researchers-live-in [https://perma.cc/4U7Z-UUWJ].

265. *Id.*

266. *See* Katz v. United States, 389 U.S. 347, 351 (1967) ("What a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection."). Though beyond this Article's scope, it is likely that such a ban would also benefit large market players with access to proprietary datasets, thereby further entrenching the market position of leading firms.

267. *See* Press Release, Max Plank Inst. for Intelligent Sys., European Laboratory for Learning and Intelligent Systems (ELLIS) Launches (Dec. 6, 2018), https://www.is.mpg.de/en/news/european-laboratory-for-learning-and-intelligent-systems-ellis-launches [https://perma.cc/SS5G-B4DQ] ("The comprehensive plan for ELLIS includes the creation of a network to advance breakthroughs in AI, a pan-European PhD program to educate the next generation of AI researchers, and a focal point for industrial engagements to boost economic growth by leveraging AI technologies."); *see also Ellis Letter*, ELLIS SOC'Y, https://ellis-open-letter.eu/letter.html [https://perma.cc/3GMM-VMLC] (letter from scientists urging creation of EU-wide institute).

268. *See, e.g.*, Amy Adams, *Stanford University Launches the Institute for Human-Centered Artificial Intelligence*, STAN. NEWS. (Mar. 18, 2019), https://news.stanford.edu/2019/03/18/stanford_university_launches_human-centered_ai [https://perma.cc/N6T3-TKN2]; David Culler, *Berkeley Announces Transformative New Division*, UC BERKELEY (Nov. 1, 2018), https://data.berkeley.edu/news/berkeley-announces-transformative-new-division [https://perma.cc/6MPE-DNQ4]; *MIT*

challenge is to ensure that there are not so many splintered professionalization groups or training modules that it creates more signal than noise. This is a substantial challenge because the ML development process frequently involves data scientists, software engineers, policymakers, and executives, each of whom may be subject to distinct and diffuse professional norms.[269]

To unify these norms, the government might act as a convener for professional standard-setting efforts—even if the government does not endorse any one particular effort. For instance, public-facing, civilian forums, perhaps hosted by the NSCAI, could help to identify ML individuals as members of a unified profession who are subject to the same institutional and cultural norms.[270] This proposal is consonant with, but goes beyond, the

Reshapes Itself to Shape the Future, MIT NEWS (Oct. 15, 2018), http://news.mit.edu/2018/mit-reshapes-itself-stephen-schwarzman-college-of-computing-1015 [https://perma.cc/YG2J-AG8G]; Byron Spice, *Carnegie Mellon Launches Undergraduate Degree in Artificial Intelligence*, CARNEGIE MELLON (May 10, 2018), https://www.cmu.edu/news/stories/archives/2018/may/ai-undergraduate-degee.html [https://perma.cc/F9U6-UKKK]; *cf.* Paul Scharre & Michael C. Horowitz, *Congress Can Help the United States Lead in Artificial Intelligence*, FOREIGN POL'Y (Dec. 10, 2018, 5:46 PM), https://foreignpolicy.com/2018/12/10/congress-can-help-the-united-states-lead-in-artificial-intelligence [https://perma.cc/S4LD-V2TS] ("The federal government plan for STEM education recently released by President Donald Trump's administration takes some important steps [to increase the pipeline of AI talent], and the commission should work with university administrators to better understand what the government and industry can further do to expand the pipeline . . . ."). To succeed, these efforts must formulate a new kind of interdisciplinary model, and not merely repackage old departmental divisions with new institutional labels.

269. Though beyond the scope of this Article to canvass in full, a slew of professional and ethical protocols proliferate both in the general information technology ("IT") sector and specifically around AI. *See generally, e.g.*, Don Gotterbarn et al., *Software Engineering Code of Ethics*, 40 COMM. OF THE ACM 110 (1997) (joint project between the Association for Computing Machinery ("ACM") and Institute of Electrical and Electronics Engineers ("IEEE")); *ACM Code of Ethics and Professional Conduct*, ASS'N FOR COMPUTING MACHINERY (2018), https://www.acm.org/about-acm/code-of-ethics [https://perma.cc/3TFQ-QFYS]; *Candidate Code of Ethics*, COMPTIA, https://certification.comptia.org/testing/test-policies/continuing-education-policies/candidate-code-of-ethics [https://perma.cc/3TFQ-QFYS]; *IT Code of Ethics*, SANS (Apr. 24, 2004), https://www.sans.org/security-resources/ethics [https://perma.cc/MT2V-YM2C]. In addition, the International Federation of Information Processing ("IFIP"), another tech coalition, at one time had a Special Interest Group, SIG 9.2.2, on the Framework on Ethics of Computing. *See IFIP TC9 ICT and Society*, INT'L FED'N INFO. PROCESSING, http://ifiptc9.org [https://perma.cc/EUA8-X6D2]. This project concluded in 1996 with the publication of a handbook, *see* ETHICS OF COMPUTING: CODES, SPACES FOR DISCUSSION AND LAW (Jacques Berleur & Klaus Brunnstein eds., 1996), that "outline[d] that there are certain principles that all might want to consider and take account of in their codes." *Id.* at 257. Since 1989, IFIP has included a Technical Committee, TC12, devoted to AI, which consists of "members representing 33 national computer societies, together with representatives of the ACM and the IEEE, and has several working groups covering major topics in AI." *Welcome*, INT'L FED'N INFO. PROCESSING TC12, http://www.ifiptc12.org [https://perma.cc/WA95-93ZL]. Such a bevy of overlapping standards arguably stymies consensus around a shared professional ethic.

270. Nonprofit and civil society organizations have already launched initiatives of this sort. For instance, ABOUT ML ("Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles") "is a multi-year, multi-stakeholder initiative led by [the Partnership on AI (PAI)]" that "aims to bring together a diverse range of perspectives to develop, test, and implement machine learning system documentation practices at scale. " *About ML*, PARTNERSHIP ON AI,

National Institute of Standards and Technology's ("NIST") 2019 "Plan for Federal Engagement in Developing Technical Standards and Related Tools," which focuses on "technical," or "documentary," standards.[271] In addition to greater technical specification, public actors can and should place a greater emphasis on the "non-technical" standards—from "societal and ethical considerations" to "governance" to "privacy"—that "inform policy and human decision-making."[272] A failure to do so risks blinking the reality of code as policy and the manner in which "technical" choices amount to micro-level governance choices to an extent not present in other domains.

Again, this is not to say that the government should necessarily craft the substantive content of any such standards. But it is to urge a greater emphasis on unified cultivation of norms, as a procedural matter, and to suggest that the state could play an invaluable convening role regarding all standards. Consider how the Hippocratic oath—without formally enshrining any mandate in law—has supported a powerful set of professional norms for medicine. A similar unified process of professional development around data in AI might foster a shared identity and clarify the minimum cultural expectations for those who participate on such teams, before they begin to construct the architectural code that is affecting life in the real world today.[273] These and related cultural shifts provide safeguards for a world in which code operates as policy.

---

https://www.partnershiponai.org/about-ml [https://perma.cc/6MDZ-Y76C]. ABOUT ML includes the opportunity for public notice and comment on draft documents, providing an important channel for public accountability. *See id.* Notwithstanding the merits of such initiatives, given the contemporary imbalance of public-private resources and authority, this Article offers that there is a valuable instrumental signal to be sent by government-backed convenings.

271. NAT'L INST. OF STANDARDS & TECH., U.S. LEADERSHIP IN AI: A PLAN FOR FEDERAL ENGAGEMENT IN DEVELOPING TECHNICAL STANDARDS AND RELATED TOOLS 7–9 (2019) ("United States global leadership in AI will benefit from the Federal government playing an active and purpose-driven role in AI standards development.").

272. *Id.* at 12–13.

273. Others have called for a Hippocratic oath for data scientists. *See, e.g.*, Tom Upchurch, *To Work for Society, Data Scientists Need a Hippocratic Oath with Teeth*, WIRED (Apr. 8, 2018), https://www.wired.co.uk/article/data-ai-ethics-hippocratic-oath-cathy-o-neil-weapons-of-math-destruction [https://perma.cc/492K-NCG8] (discussing data scientist Cathy O'Neil's proposal to create an ethical code of conduct for data scientists to follow: "The idea is to imbue data scientists with a moral conscience which would guide their thinking when designing systems and force them to consider the wider societal impact of their designs"). Critically, if these starting measures prove inadequate, there is still an opportunity to identify areas of society—from AVs to lethal autonomous weapons to criminal justice sentencing algorithms—where AI may acutely threaten public safety or core democratic values, and where more stringent top-down intervention may be necessary. For instance, one tactic that awaits expansion in future work might be a federal statutory or administrative guidance to set a temporary "safety floor" in the form of non-negotiable procedural checks for AI applications, before they can be brought to market. This suggestion reflects the fact that government-dictated standards, commanded top-down, are a poor fit, as discussed *infra* Part II. It is distinct from true premarket clearance in mode of FDA because it would not assess outcomes based on empirical scientific testing. Rather, it would focus on compliance with a set of procedures, perhaps with an initial focus on safety testing and auditing trails.

CONCLUSION

This Article has argued that we need new strategies to grapple with the manner in which digital technology regulates contemporary society. In the era of AI, more than ever before, digital technology's impact is not neatly cabined to virtual spaces. Rather, from smart devices, to the Internet of Things, to machine learning, technical development and associated programming decisions are a form of technical policy that mediate our way of life in the physical world. These technologies directly shape our universe—and yet the tendency to consider them as technical and not social or political forces does not account for the myriad ways in which they affect traditionally public interests, at the potential expense of democracy, social norms, and the individual citizen's safety and security.

A comparison of AI and past administrative law challenges in technocratic domains reveals that AI's governance challenges come from the interaction between the technology's unique attributes (speed, complexity, and unpredictability) that strain traditional administrative law tactics, and the strategic context, defined as the institutional settings and market, political, and social incentives, for its development and deployment. At present, the decision points over AI rest predominantly in private hands. And yet something as seemingly mundane as a firm's choice about whether to notify a safety driver in an AV that the car's software is having trouble identifying an object in the road can result in the death of a human being. Digital code has visceral physical impacts. The current balance of authority over its development stymies public accountability. This is a public policy problem.

But, as this Article has discussed, the solution is not obvious. A domain-specific, more traditional prescriptive response is a poor fit for a general use technology like AI. Any broader procedural oversight agency would require vast public expertise and resources that are improbable given the private sector's current lead in AI research and development, and which may be impossible in light of the way that ML relies on connections in data and not cause-and-effect relationships. Nor can one possible solution—more collaborative governance response patterns—succeed without a strong, democratically accountable partner that does not presently exist.

Maximizing the potential of AI, preserving space for private innovation, and protecting public wellbeing requires rethinking the AI governance paradigm to recognize how this emerging technology is an especially poor fit for typical regulatory models. Much of this work entails theoretical reframing. Algorithmic and programming decisions structure human behavior. These choices are in fact policy decisions that function at the most essential levels of democratic governance and public interests. Put simply:

AI development is an especially stark example of how private coding choices are governance choices, embedded within products that affect both norms and physical wellbeing. Contending with this dynamic requires long-term dialogue about the contexts in which we are, or are not, willing to accept private governance by code in products offered on the open market. And in the short term, new tactical approaches are needed. To account for the technical architecture of AI (its code), this Article suggests that policymakers should focus on how to filter public input and instill public values through alternative regulatory modalities, such as markets and norms, rather than attempting direct control of algorithmic technologies like AI through the law.

The stakes for lawyers and policymakers in particular are anything but theoretical. As emerging digital technologies continue to permeate contemporary society, a failure to rethink innovation and regulation risks undermining the role of law. If we continue to support private innovation without thinking about how AI fits within our governance theories and practices, then it is not clear what role legal systems or values can play, lest they interfere with what is touted as free private ordering. At best, this path would miss an unusual opportunity for collaboration in AI, wherein many innovators are seeking policy guidance. At worst, it would allow private actors to encode particular values into AI technologies in ways that clash with the normative ideals of democratic self-governance—or even erode the vitality of democratic governance itself.