Administering Artificial Intelligence

Alicia Solow-Niederman

This draft is a work in progress. Please do not quote or cite without advance permission. Thank you!

Calls for sector-specific regulation or the creation of a federal agency or commission to guide and constrain artificial intelligence, or AI, development are increasing. This turn to administrative law is understandable because AI's regulatory challenges seem similar to those in other technocratic domains, such as the pharmaceutical industry or environmental law. But an "FDA for algorithms" or federal robotics commission is not a cross-cutting AI solution. AI is unique, even if it is not entirely different. AI's distinctiveness comes from technical attributes (speed, complexity, and unpredictability) that strain traditional administrative law tactics, in combination with institutional settings and incentives, or strategic context*, that affect its development path.*

This Article puts American AI governance in strategic context. Today, there is an imbalance of state and non-state AI authority. Commercial actors dominate research and development and private resources outstrip public investments. Even if we could redress this baseline, a fundamental, yet under-recognized problem remains. Any governance strategy must contend with the ways in which algorithmic applications permit seemingly technical decisions to de facto regulate human behavior, with a greater potential for physical and social impact than ever before. When coding choices functionally operate as policy in this manner, the current trajectory of AI development augurs an era of private governance. Without rethinking our regulatory strategies, we risk losing the democratic accountability that is at the heart of public governance.

Table of Contents

In spring 2018, a group of over thirty artificial intelligence, or AI,[1] specialists compiled a series of research results that were altogether different from what data scientists had expected.[2] Some of the outcomes were surprising in innocuous and even amusing ways. Like Amelia Bedelia, one sorting algorithm took its directive a bit too literally and deleted all the data so that it was not, technically speaking, unsorted.[3] Another simulation embraced its inner child by spontaneously learning how to do somersaults.[4] Such algorithmic creativity may be exciting and generative. But childlike evasion of a directive or playful exploration isn't always funny. For instance, researchers programmed an aircraft landing simulation to identify ways to decelerate planes. Instead, it discovered that generating extremely large force calculations at a particular point in landing would cause the force to read out as nearly zero.[5] The algorithm thus appeared to be rapidly decelerating when it was actually exploiting an unforeseen loophole in the programming[6] and leaving the plane speeding into what might be a crash landing.[7] Fortunately, the researchers caught the error, and the harm remained virtual.

---

[1] This Article defines artificial intelligence as a class of technologies that rely on some form of automated decision-making executed by a computer. AI, as used in this Article, includes both robots and AI algorithms that lack a bodied form, whether they employ machine learning or another method. *Accord* Jack Balkin, *The Path of Robotics Law*, 72 CAL. REV. CIRCUIT 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."). *Cf.* Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN L. REV. 83, 85 n.2 (2016) (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 includes autonomous and intelligent systems and associated "overlapping concerns about the design, development, deployment, decommissioning, and adoption of autonomous or intelligent software when installed into other software and/or hardware systems that are able to exercise independent reasoning, decision-making, intention forming, and motivating skills according to self-defined principles." *General Principles V2*, IEEE, https://standards.ieee.org/develop/indconn/ec/ead_general_principles_v2.pdf (last visited Jul. 7, 2018)). *See also Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2*, IEEE 20 (2017), http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf (proposing a definition of "autonomous and intelligent systems" that applies "regardless of whether they 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)").

[2] Joel Lehman et al., *The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities*, ARXIV (manuscript at 7) (Mar. 29, 2018), https://arxiv.org/pdf/1803.03453.pdf.

[3] *Id. See Meet Amelia Bedelia*, HARPERCOLLINS PUB., http://www.harpercollinschildrens.com/kids/gamesandcontests/features/amelia/meet-ab.aspx ("She does exactly what you tell her to do, so be careful what you say") (last visited Dec. 31, 2018).

[4] The simulated robot was instructed to move as quickly as possible across a finish line. Rather than discovering locomotive activities like walking or running and growing limbs to race ahead, it grew into tall, rigid towers that toppled end-over-end, which maximized the initial potential energy and propelled it forward. *See* Evolving AI Lab, *Why Walk When You Can Somersault*, YOUTUBE (Mar. 15, 2018), https://www.youtube.com/watch?v=TaXUZfwACVE&index=8&t=0s&list=PL5278ezwmoxQODgYB0hWnC0-Ob09GZGe2.

[5] Lehman et al., *supra* note 2 (manuscript at 10–11).

[6] *Id.* ("[E]volution discovered a loophole in the force calculation for when the aircraft's hook attaches to the braking cable. By overflowing the calculation, i.e. exploiting that numbers too large to store in memory 'roll-over' to zero, the resulting force was sometimes estimated to be zero. This, in turn, would lead to a perfect score.").

[7] This example is not sui generis. *See also* Devin Coldewey, *This Clever AI Hid Data from Its Creators to Cheat at its Appointed Task*, TECHCRUNCH (Dec. 31, 2018), https://techcrunch.com/2018/12/31/this-clever-ai-hid-data-from-its-creators-to-cheat-at-its-appointed-task/ ("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-

But imagine that the algorithm is not a simulation. Rather, it is part of the operating code for an autonomous vehicle, or AV, that a private company is using for grocery delivery.[8] What trust can we have 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.[9] But it is non-binding. 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 at best. The same firms that 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) are the leaders of AI development.[12] After the "Techlash"[13] against social media platforms' poor privacy and data security practices, not to

---

frequency signal.'" (quoting Casey Chu et al., *CycleGAN, a Master of Steganography*, ARXIV (Dec. 16, 2017), https://arxiv.org/pdf/1712.02950.pdf)).

[8] *See* Megan Rose Dickey, *Nuro and Kroger are Deploying Self-Driving Cars for Grocery Delivery in Arizona Today*, TECH CRUNCH (Aug. 16, 2018), https://techcrunch.com/2018/08/16/nuro-and-kroger-are-deploying-self-driving-cars-for-grocery-delivery-in-arizona-today/.

[9] *See* U.S. DEP'T OF TRANSPORTATION, PREPARING FOR THE FUTURE OF TRANSPORTATION: AUTOMATED VEHICLES 3.0 viii (2018) ("AV 3.0 [p]rovides [n]ew [m]ultimodal [s]afety [g]uidance [that] . . . affirms 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.").

[10] Facebook's internal motto in its early days was "move fast and break things." Facebook CEO Mark Zuckerberg continued with a less well-known second part of the motto: "Unless you are breaking stuff, you are not moving fast enough." *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/#BuyyGYcrYmqP.

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

[12] Google is the leading employer of AI talent, and AI intellectual property is highly concentrated among just eight global companies: Alibaba, Amazon, Apple, Baidu, Facebook, Google, Microsoft, and Tencent. *See* Nathan Benaich & Ian Hogarth, *State of AI* (June 29, 2018), https://www.stateof.ai/.

[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 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. 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. H. REP. PERMANENT SELECT COMM. ON INTELLIGENCE, https://democrats-intelligence.house.gov/social-media-content/ (last visited July 16, 2018). *See also* Editorial Board, *'Facebook Cannot Be Trusted to Regulate Itself',* N.Y. TIMES (Nov. 15, 2018), https://www.nytimes.com/2018/11/15/opinion/facebook-data-congress-russia-election.html ("Facebook has, perhaps uniquely, demonstrated a staggering lack of corporate responsibility and civic duty in the wake of this [Russian influence campaign] crisis."). 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; Timothy B. Lee, *Zuckerberg: Cambridge Analytica Leak a "Breach Of Trust" With Users*, ARS TECHNICA (Mar. 21, 2018, 2:24 PM); Douglas MacMillian, *Tech's 'Dirty Secret': The App Developers Sifting Through Your Gmail*, WSJ (July 2, 2018, 11:14 AM), https://www.wsj.com/articles/techs-dirty-secret-the-app-developers-sifting-through-your-gmail-1530544442; Janko Roettgers, *Facebook Data Backlash Reveals Tech's Biggest Challenge: Trust*, VARIETY (Mar. 27, 2018, 7:00 AM), https://variety.com/2018/digital/news/facebook-data-crisis-tech-trust-mark-zuckerberg-1202736903/. *See also* Mark Zuckerberg, FACEBOOK (Mar. 21, 2018), https://www.facebook.com/zuck/posts/10104712037900071 ("This was a breach of trust between [Cambridge

mention Russia's interference in the 2016 U.S. election via social media, it is not clear why the public should trust that technology companies can or will "do the right thing." The same issues repeat at the level of the industry. Though third-party standard-setting organizations might in theory step in, in practice, there are too many emergent, overlapping standards that operate at too high a level of abstraction to reliably guide or constrain commercial actors.[14]

Moreover, potential legal interventions are unlikely to cabin these actors in a systematic way. It is possible, to be sure, that the threat of ex post sanctions through the tort or criminal system will influence firms or entire industries, at least in egregious cases. And it is possible that the Techlash will subside and public attention will shift from a concern with commercial actions to a concern with state authority. But these outcomes do not provide consistent public accountability or more ex ante democratic control over the development of emerging technologies like AI. As a rapidly-growing body of technical and legal scholarship recognizes,[15] the impact of algorithms can be a significant problem from a fairness, accountability, and transparency perspective. Consider Google's "Smart Compose" email feature, which relies on the dominant AI method, "machine learning."[16] This "smart" product wouldn't 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.[17] This same kind of demographic bias is likely to emerge whenever an AI

University researcher Aleksandr] Kogan, Cambridge Analytica and Facebook. But it was also a breach of trust between Facebook and the people who share their data with us and expect us to protect it.").

[14] *See* discussion *infra* Section II.B.

[15] *See* Margot E. Kaminski, *Binary Governance; A Two-Part Approach to Accountable Algorithms*, S. CAL. L. REV. (forthcoming 2019) (manuscript at 2 & n.4) ("A quickly growing literature addresses how to regulate algorithmic or AI decision-making to mitigate potential harms." (citing 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 & D. Westreich, *Incomprehensible Discrimination*, 7 CALIF. L REV ONLINE 164 (2017); Michael Guihot et al., *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence*, 20 VAND. J. ENT. & TECH. L. 385 (2017); Pauline 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); Tutt, *supra* note 1; Mike Ananny & Kate Crawford, *Seeing without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability*, NEW MEDIA & SOC. 1 (2016); Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016); 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); W. Nicholson Price II, *Regulating Black Box Medicine*, 116 MICH. L. REV. 421 (2017); Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393 (2014); Tal Z. Zarsky, *Transparent Predictions*, 2013 4 U. ILL. L. REV. 1503 (2013); Mireille Hildebrandt, *The Dawn of a Critical Transparency Right for the Profiling Era*, Bus, J. (ed.), DIGITAL ENLIGHTENMENT YEARBOOK 41 (2012); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008); Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1, 23 (2005); Lee A. Bygrave, *Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, 17 COMPUTER L. & SECURITY REP. 17 (2001); Paul Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Compute*r, 43 HASTINGS L.J. 1321 (1992)).

[16] The popular AI method of machine learning, or ML, is a subset of AI. In ML, a system learns without ex ante, explicit programming. *See Some Studies in Machine Learning Using the Game of Checkers*, 3 IBM J. RES. & DEV. (1959). For more discussion of machine learning in general, see IAN GOODFELLOW ET AL., DEEP LEARNING (2016), http://www.deeplearningbook.org/. For more detail on machine learning 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). *See also* Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. (forthcoming 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=314483 (manuscript at 11–13); Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 89–101 (2014).

[17] Paresh Dave, *Fearful of Bias, Google Blocks Gender-Based Pronouns From New AI Tool*, REUTERS (Nov. 26, 2018, 10:16 PM), https://www.reuters.com/article/us-alphabet-google-ai-gender/fearful-of-bias-google-blocks-gender-based-pronouns-from-new-ai-tool-idUSKCN1NW0EF. According to publicly-available blog posts, this tool

system relies on data sets that draw connections from past human behavior,[18] 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 pre-trial detention and sentencing decisions.[19] 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.[20]

In short, technical decisions about algorithms are encoding values already, without any uniform oversight, normative requirements, or public accountability.[21] And if we are collectively trying to improve our social, economic, and political discourse with increased sensitivity to problems such as implicit bias,[22] it is a mistake to hand over more and more discretion to algorithms that are even less "woke." Furthermore, the same limitations and issues recur when an algorithm carries a more direct physical impact—from an algorithmic medical intervention to an autonomous vehicle—that acutely threatens human safety. Take, for example, the July 2018 finding that IBM's AI-powered Watson supercomputer recommended "unsafe and incorrect" cancer treatments.[23] As AI technologies are embedded in more and more applications, calls for public regulatory guidance and oversight will only increase.[24] If existing regulatory agencies are not up to the task, then perhaps it is time for a new agency or bureau dedicated to AI, robotics, and/or algorithms, or a more process-driven agency that focuses exclusively on protecting consumers from specified algorithmic harms.[25]

---

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, 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* Yonghu Wu, *Smart Compose: 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; Paul Lambert, *SUBJECT: Write emails faster with Smart Compose in Gmail*, KEYWORD BLOG, GOOGLE (May 8, 2018), https://www.blog.google/products/gmail/subject-write-emails-faster-smart-compose-gmail/.

[18] *See, e.g.*, Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. (forthcoming 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3257004 ("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 of the algorithmic tools now replacing it."); Aylin Caliskan, *Semantics Derived Automatically from Language Corpora Contain Human-Like Biases*, ARXIV (May 25, 2017), https://arxiv.org/abs/1608.07187 (describing risk of bias in natural language processing, given human biases).

[19] *See Algorithms in the Criminal Justice System*, EPIC, https://epic.org/algorithmic-transparency/crim-justice/ (last visited Dec. 4, 2018) (surveying states that require or recommend risk assessment algorithms).

[20] *See, e.g.*, SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2017)

[21] *See* discussion *infra* Part IV.

[22] *See, e.g.*, Al Baker, *Confronting Implicit Bias in the New York Police Department*, N.Y. TIMES (July15, 2018), https://www.nytimes.com/2018/07/15/nyregion/bias-training-police.html; Yuki Noguchi, *Starbucks Training Focuses on the Evolving Study of Unconscious Bias*, NPR (May 17, 2018, 3:30 PM), https://www.npr.org/2018/05/17/611909506/starbucks-training-focuses-on-the-evolving-study-of-unconscious-bias.

[23] Casey Ross & Ike Swelitz, *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.

[24] For example, the first recommendation in a recent 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*, AI NOW 1 (Dec. 2018), https://ainowinstitute.org/AI_Now_2018_Report.pdf.

[25] *See* Tutt, *supra* note 1; Ben Schneiderman, Turing Lecture: Algorithmic Accountability, ALAN TURING INST. (May 30, 2017), https://www.turing.ac.uk/events/turing-lecture-algorithmic-accountability/; Michael Seigel, *We*

But this sort of public regulatory response is no panacea for AI pathologies. First, policymakers are likely to lack the expertise or resources to make informed governance choices about highly technical digital code. Second, even assuming the requisite expertise, AI algorithms are a poor conceptual fit for top-down, prescriptive regulation.[26] 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 of "machine learning" is not fixed in the same way as, for instance, the molecules in a pharmaceutical compound. In machine learning, or 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.[27] 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.[28] This dynamism contrasts markedly with static regulatory objects that may be more amenable to prescriptive controls.

What about a solution that entails collaborative public-private efforts over time?[29] The necessary conditions for this form of governance do not presently exist.[30] A strong public force is integral for accountable governance, and this prerequisite is missing because of the extent to which contemporary AI development is centered outside of the state, particularly in America.[31] True,

---

Need an FDA For Algorithms, NAUTILUS (Nov. 1, 2018), http://nautil.us/issue/66/clockwork/we-need-an-fda-for-algorithms. *Cf.* 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/.

[26] *See* discussion *infra* Section III.A.2.

[27] For an accessible overview of ML, see Karen Hao, *What Is Machine Learning? We Drew You Another Flowchart*, MIT TECH. REV. (Nov. 17, 2018), https://www.technologyreview.com/s/612437/what-is-machine-learning-we-drew-you-another-flowchart/; *see also* VISHAL MAINI & SAMER SABRI, MACHINE LEARNING FOR HUMANS (2017), https://www.dropbox.com/s/e38nil1dnl7481q/machine_learning.pdf?dl=0.

[28] *See* Lehr & Ohm, *supra* note 16, at 701–02 ("Many machine-learning algorithms when deployed are not run merely occasionally, but continuously. . . . [which requires] continuously feed[ing] new data into the trained algorithm. . . . . Machine-learning algorithms running at scale may also be turned into online learning systems — 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 re-training of the entire model. It has already been deployed by companies in commercial applications. *See, e.g.*, *See, e.g.*, Braden Hancock et al., *Learning from Dialogue after Deployment: Feed Yourself, Chatbot!*, ARXIV, (Jan. 16, 2019), https://arxiv.org/abs/1901.05415 (discussing "a dialogue agent with the ability to extract new training examples from the conversations it participates in" that Facebook has adopted); *Karen* Hao, *Car-Hailing Firm Didi Has A New Dispatching Algorithm That Adapts To Rider Demand*, MIT TECH. REV. (Dec. 12, 2018, 4:32 AM), https://www.technologyreview.com/the-download/612568/car-hailing-firm-didi-has-a-new-dispatching-algorithm-that-adapts-to-rider/ (discussing major Chinese AV firm's adaptive dispatching algorithm).

[29] Technology law scholars have not yet robustly explored collaborative governance for AI. As Kaminski observes, the literature on collaborative governance and algorithms tends to be limited to a specific context, such as health law, *see, e.g.*, Price, *supra* note 15, or copyright law, *see, e.g.*, Perel & Elkin-Koren, *supra* note 15. *See* Kaminski, *supra* note 15, at manuscript 4 n.15. As of late 2018, there appear to be only two publications that specifically discuss collaborative governance and algorithms. *See* Guihot et al., *supra* note 15; Kaminski, *supra* note 15. 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 engage in 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.

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

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

there are recent national security investments in AI research and development.[32] And on the civilian side, the legislative and executive branches have recently signaled a renewed interest in AI policy.[33] But even assuming massive public resource investments, both the pace of development and the highly specialized nature of the technology[34] will make it challenging to shift the center of gravity away from the private sector.[35]

This public-private imbalance, moreover, suggests that a functional theory of AI governance requires us to pay more attention to the actions of private entities and individuals. In regulating AI, we should assess AI's technical attributes[36] and associated policy challenges[37] alongside the institutional settings and motivations, or "strategic context," that inform AI development.[38] Return to the IBM Watson example. The report 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.[39] And the same reporting suggests that IBM executives pushed forward to market it, despite awareness of its flaws.[40] 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

---

[32] *See* discussion *infra* Section II.B.2.

[33] Scholars and policymakers have also advocated 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 AMER. AFF. 3 (2018); K. Sameel Rahman, *The New Octopus*, LOGIC MAG. Apr. 2018, https://logicmag.io/04-the-new-octopus/; Lina Khan, Note, *Amazon's Antitrust Paradox*, 126 YALE L.J. 564 (2017). 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 (White Paper, July 30, 2018), https://graphics.axios.com/pdf/PlatformPolicyPaper.pdf.

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 decision-making 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 small number of technology companies. See discussion of code as policy, *infra* Part IV.

[34] See discussion of pace and the regulatory challenges of complexity and unpredictability *infra* Section II.A.2.

[35] *See infra* Section II.B.2.

[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.

[37] This Article uses the term "policy challenge" or "regulatory challenge" to underscore that any technology emerges in social context. *Accord* 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.").

[38] This framing is adapted from a May 2018 conference hosted by the Program for 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*, PULSE (May 7, 2018), https://law.ucla.edu/centers/interdisciplinary-studies/pulse/news/2018/05/ai-in-strategic-context-development-paths-impacts-and-governance/.

[39] The errors are reportedly due to the use of hypothetical patient data, not real data, in training the algorithm. *See* Ross & Swelitz, *supra* note 23, at 1–2.

[40] *Id.* at 6. Doctors interviewed for the report did not mince words: "[T]hey [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."

among technological attributes, regulatory constraints, particular institutional settings and incentives, and individuals' choices about how to develop and use the algorithm, in context.

Looking at the development and deployment of algorithms in strategic context reveals that 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 offers that technical decisions by private actors are also making policy when it comes to AI. As contemporary decisions about how to design and deploy an algorithm are functionally regulating[41] human behavior, it is time to recognize what this Article terms *code as policy*. And 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,[42] may evade democratic checks at the same time that it fails to provide a clear regulatory rule.

Indeed, this state of affairs is already emerging. AI development is presently subject to myriad non-binding technical and ethical standards—standards with no obvious order of authority, with considerable risk of overlap or inconsistency, and with no guarantee that commercial incentives will align with any sort of democratic consensus or be subject to any sort of democratic accountability. And we lack a public-facing, democratically accountable check on commercial development paths for AI technologies. Within the terms of traditional governance and administrative law models, we thus face what seems to be 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 in which governance is the province of commercial entities and not the state. We cannot contend with this challenge or consider the viability of alternative governance approaches unless we examine the bedrock relationship between citizens, firms, and the state in the world we wish to inhabit.

The following analysis of our governance options in the age of artificial intelligence proceeds in four parts. Part I offers a brief survey of governance theory in general and theories of regulation of digital technologies in particular. Part II then canvasses the current state of AI regulation and discusses contemporary self-regulatory efforts. Following Section II.A's description of present-day AI regulation in the United States, Section II.B. takes an analytic turn to explore how myriad AI standards that are developing without any uniformity or order of authority are unlikely to supply meaningful guidance to or constraints on commercial AI actors.

Next, Part III considers whether existing public regulatory approaches might supply necessary consistency and increase accountability. Section III.A discusses prescriptive[43] regulation such as FDA's premarket drug clearance regime, and then rebuts the case for an "FDA for algorithms." Section III.B considers governance as a public regulatory alternative, looking to environmental regulation as an example of this paradigm. It concludes that AI's highly dynamic, complex, and

[41] "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.

[42] *See infra* text accompanying note 75 and sources cited therein.

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

interdisciplinary challenges are similar to ecosystem management in particular, 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 IV 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, it calls for recognition of the power of code as policy and—assuming we are not ready to let governance get away from the public—closes with a series of suggestions to filter public input through alternative regulatory modalities, such as markets and norms, rather than relying predominantly on direct intervention via law.

I. Beyond Formal Regulation

A. From Regulation to Collaboration

Since the late 20th century, administrative law scholars have grown skeptical of traditional paradigms of regulatory administration.[44] According to this school of thought, 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.[45] These scholars contend that public regulation in the modern state requires active participation by both governmental and non-governmental actors.[46]

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

---

[44] See JON MICHAELS, CONSTITUTIONAL COUP 39-50 (describing 20th Century "*pax administrativa*" and detailing how far we have deviated from "administrative era" of the 1930s–1970s). For a classic account of the 20th century model, see JAMES M. LANDIS, THE ADMINISTRATIVE PROCESS (1938).

[45] *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 . . . There are important differences among these models, but all 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 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. 22 (1997) [hereinafter Freeman, *Collaborative Governance*]; '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)." *Id.*

[46] *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) [hereinafter Lobel, *Renew Deal*] ("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 orthogonal to the core premise: practically speaking, new scholarship around governance has emerged.

"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."[47] Jody Freeman uses the term "collaborative governance"[48] to describe contemporary administrative efforts and invokes the metaphor of contracts to situate governance as "a set of negotiated relationships."[49] 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."[50] 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 other contexts. 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—or "code"—function as a regulatory modality that both creates and controls the online world.[51] These scholars, particularly Lawrence Lessig, established a theory of regulation that flows through code, both directly and indirectly. In Lessig's words: "in real space, we recognize how laws regulate— through constitutions, statutes, and other legal codes. In cyberspace we must 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."[52] In cyberspace, "code is law."[53]

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[54] as forces that affect the

---

[47] *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) [hereinafter Lobel, *Renew Deal*] ("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 orthogonal to the core premise: practically speaking, new scholarship around governance has emerged.

[48] Freeman, *Collaborative Governance*, *supra* note 45.

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

[50] Lobel, *Renew Deal*, *supra* note 46 at 344.

[51] *See* LAWRENCE LESSIG, CODE: AND OTHER LAWS OF CYBERSPACE, VERSION 2.0 (2005) [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.

[52] Lessig, CODE V2.0, *supra* note 51, at 5 (citing WILLIAM J. MITCHELL, CITY OF BITS 111 (1999) and Joel Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553 (1998)).

[53] *Id.*

[54] 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, *The New Chicago School,* 27 J. LEGAL STUD. 661, 663 (1998) ("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.").

human beings (the "pathetic dot") at the center of the regulatory equation.[55] 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."[56] 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 the "pathetic dot."

In this model, code operates as architecture as well as law.[57] Code constructs digital realms in a literal sense: the digital bits that make up strings of code create online environments.[58] 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.[59] As Lessig explains, there are no laws of physics in cyberspace.[60] And because there are no laws of nature in virtual space, code also creates all of the fundamental parameters that apply to life within that environment.[61] Code, then, could be said both to constitute the online world and to represent the constraints and controls that govern cyberspace.[62]

* * *

Code today is not merely a "law of cyberspace" that can be confined to a "virtual" screen. More than ever before,[63] it reaches the physical, or "real," world.[64] Algorithms are pervasive: 65% of

---

[55] *Id.* at 664. For an overview of each of the modalities, see *id.* at 662–63.
[56] *Id.* at 662 n.1.
[57] *See* Lessig, CODE V2.0, *supra* note 51, at 4–6.
[58] As Lessig explains, "[c]ode is a regulator in cyberspace because it defines the terms upon which cyberspace is offered." *See* CODE V2.0, *supra* note 51, 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.
[59] *See id.* at 114 ("Code embeds values. It enables, or not, certain control. And it is also a tool of control.").
[60] *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.").
[61] *See id.*
[62] 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 136.
[63] 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.*, VIRGINIA EUBANKS, AUTOMATING INEQUALITY (2017); RACE AFTER THE INTERNET (Lisa Nakamara & Peter A. Chow-White eds., 2012); RACE IN CYBERSPACE (Lisa Nakamura et al. eds., 2000); CYBERGHETTO OR CYBERTOPIA (Bosah Ebo ed., 1998).
[64] Lessig's original account does contemplate how the outcomes of virtual interactions may affect individuals in the physical world. *See, e.g.*, Lessig CODE V2.0, *supra* note 51, 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 Characterizing Cyberspace: Past, Present, and Future*, MIT CSAIL (Mar. 12, 2010), http://docshare01.docshare.tips/files/9608/96080638.pdf ("[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.'"). Within the original cyberlaw framework, however, "code" is an online, or cyberspace, modality of regulation, with incidental impacts in offline, or physical space. Architecture is "real-space code," Lessig, CODE V2.0, *supra* note 51, at 342, and code is digital-space architecture, with the two realms remaining descriptively distinct.

American adults walk around with a "smart" cellular device[65] that contains more computing power than NASA had on hand during the Apollo missions.[66] This constant connectivity may make the world feel smaller, yet the global impact is not always for the good. For a tragic recent example of the "real" world impact of code, consider how a social media platform's choices about how to construct its website and what content to display online can contribute to genocidal violence.[67] Yet cyberlaw theory and technology law have not amply accounted for the ways that digital technologies 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[68] because it combines "the promiscuity of information" with a "physical" impact. 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.[69] This is a lived reality today because contemporary life is not virtual or physical. It is simultaneously both, and neither. And algorithmic technology's potency is particularly stark when it comes to artificial intelligence, or AI, which is defined broadly in this Article to refer to a class of technologies that rely on some form of automated decision-making executed by a computer.[70] 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.

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, to be sure, 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,[71] and prior legal scholarship assessing how existing law applies to the internet[72] as well as how cyberspace might affect traditional notions of state sovereignty and jurisdiction.[73] Moreover, scholars such as Mireille Hildebrandt have sparked important conversations about the

---

[65] *See* Monica Anderson, *Technology Device Ownership: 2015*, PEW RESEARCH CTR. (Oct. 29, 2015), http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/.
[66] *See* David Grossman, *How Do NASA's Apollo Computers Stack Up to an iPhone?*, POPULAR MECHANIC (Mar. 13, 2017), https://www.popularmechanics.com/space/moon-mars/a25655/nasa-computer-iphone-comparison/.
[67] *See* Matthew Ingram, *Facebook Slammed by UN for its Role in Myanmar Genocide*, COLUM. JOURNALISM REV. (Nov. 8, 2018), https://www.cjr.org/the_media_today/facebook-un-myanmar-genocide.php.
[68] *See, e.g.*, Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CAL. 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.").
[69] *See* discussion *supra* note 1 and sources cited therein. *See also* KAI-FU LEE, AI SUPERPOWERS (2018) (discussing what he terms "online-merge-of-offline" as the digital and physical worlds combine).
[70] For more detail, see discussion supra note 1 and sources cited therein.
[71] *See, e.g.*, *Governance of Emerging Technologies & Science (GETS) Conference*, Arizona State University, http://events.asucollegeoflaw.com/gets/ (last visited Feb. 1, 2019).
[72] For the classic debate, compare Lawrence Lessig, Commentary, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501 (1999) with Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207.
[73] *See, e.g.*, David G. Post, *Against "Against Cyberanarchy*," 17 BERKLEY TECH. L.J. 1365 (2002); Jack L. Goldsmith, *Against Cyberanarchy*, 65 U. CHI. L. REV. 1199 (1998); Timothy S. Wu, Notes, *Cyberspace Sovereignty—The Internet and the International System*, 10 HARV. J. L. & TECH. 648 (1997).

ways in which technology may demand fundamental changes in our conception of the law.[74] And others such as 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.[75] However, governance and cyberlaw analyses tend to occur in parallel streams of legal scholarship.[76] In particular, the broader collaborative governance discussion has not amply engaged with cyberlaw theory or its implications for emerging digital technologies, especially when it comes to algorithmic interventions. These dialogues can and should intersect, and this Article's analysis of AI governance options begins this conversation.

II. AI Today: A Self-Regulation Story

A. Standards Development for AI

Especially in the United States, self-regulatory approaches presently dominate AI governance. As of early 2019, the federal government has not promulgated a structured set of initiatives in support of AI development.[77] Though the Obama Administration began a number of federal

---

[74] *See generally* MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW (2015). Hildebrandt uses the term "onlife" 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 "networked self" functions in contemporary society); Julie E. Cohen, *Cyberspace and/as Space*, 107 COLUM. L. REV. 210 (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").

[75] *See* Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-by-Design*, 106 CALIF. L. REV. 697, 705–14 (2018) (offering detailed literature review of science, technology, and society (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, 771 & n. 231 (2009) at 711 & n. 231 (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" (citing Martin Heidegger, *The Question Concerning Technology*, *in* THE QUESTION CONCERNING TECHNOLOGY AND OTHER ESSAYS 3, 19–21 (William Lovitt trans, 1977))). *See also* THE OXFORD HANDBOOK OF LAW, REGULATION, AND TECHNOLOGY (Roger Brownsword et al. eds., 2017) (compiling a cross-disciplinary set of essays centered on "the 'disruptive' potential of technological innovation," *id.* at 7–14); 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 (Modern Technology: Problem or Opportunity?) 121, 122–23 (1980) (situating technology as socially mediated and discussing the ways in which "artifacts can contain political properties").

[76] There are conversations about governance *by* technology, *see* sources cited *supra* note 15 (assessing need for accountability in algorithmic decisions) and *supra* note 75 (analyzing use of technology to regulate), and the internet itself represents a decentralized form of governance, *cf.* Jonathan Masters, *What Is Internet Governance?*, COUNCIL FOR. REL. (Apr. 23, 2014), https://www.cfr.org/backgrounder/what-internet-governance. And there are some moves to assess challenges of emerging technologies. *See, e.g.*, *Governance of Emerging Technologies & Science (GETS) Conference*, *supra* note 71. 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.

[77] *See* Tim Dutton et al., *Building an AI World: Report on National and Regional AI Strategies*, CIFAR (2019), https://www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf (manuscript at 5) (describing American AI strategy as "uncoordinated" and lacking an "overarching strategy to guide policymakers"); Michael

government efforts to research and regulate AI development in 2016, the Trump Administration invested in other priorities from 2016–2018, and nascent AI initiatives in the Executive Branch were not continued.[78] This dynamic may have begun to change in 2018 with increasing attention to AI in both the executive[79] and legislative[80] branches. For instance, a July 2018 policy memorandum from the Executive Office of the President highlighted "American Leadership in Artificial Intelligence, Quantum Information Sciences, and Strategic Computing" as the second-most important area for R&D investment.[81] This publication followed a May 2018 "White House AI for American Industry Summit" at which the President chartered a National Science and Technology Council Select Committee on Artificial Intelligence.[82] This Select Committee held its first meeting in July 2018[83] and issued a fall 2018 "Request For Information on Update to the 2016 National Artificial Intelligence Research and Development Strategic Plan," which had been developed by the prior administration.[84] And most recently, the President issued an executive order outlining an "American AI Initiative."[85] For the time being, however, there is no

Horowitz et al., *Strategic Competition in an Era of Artificial Intelligence*, CTR. FOR A NEW AMER. SECURITY (July 25, 2018), https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf (manuscript at 9) ("[T]he United States does not currently have a structured national strategy for how to approach artificial intelligence.").

[78] For critical analysis of the Trump administration's approach to AI as of early 2018, see John R. Allen, *Trump's 1st State of the Union: Artificial Intelligence and the Future of America*, BROOKINGS (Jan. 30, 2018), https://www.brookings.edu/blog/fixgov/2018/01/30/trumps-1st-sotu-artificial-intelligence-and-the-future-of-america/ ("[E]vidently missing from [Trump's first State of the Union] speech is the clarion call to develop the full potential of artificial intelligence. . . .S]uch epic undertakings as the Manhattan Project, the Marshall Plan, and the space program will not have as great an individual or collective influence on America and the world as AI.").

[79] For an optimistic summary of these developments, see *AI Policy – United States*, FUTURE OF LIFE INSTIT., https://futureoflife.org/ai-policy-united-states (last visited Nov. 20. 2018).

[80] *See* Press Release, Senator Brian Schatz, *Schatz, Gardner Introduce Legislation To Improve Federal Government's Use Of Artificial Intelligence* (Sept. 26, 2018), https://www.schatz.senate.gov/press-releases/schatz-gardner-introduce-legislation-to-improve-federal-governments-use-of-artificial-intelligence. Given the current reality of partisan gridlock, this Article is skeptical about Congress's ability to pass meaningful legislation in a timely fashion, particularly in ways that deviate from more traditional administrative or statutory prescriptions—as this Article contends will be necessary.

[81] Memorandum on FY 2020 Administration Research and Development Budget Priorities, EXEC. OFFICE OF THE PRESIDENT, (July 31, 2018), https://www.whitehouse.gov/wp-content/uploads/2018/07/M-18-22.pdf.

[82] *See* SUMMARY OF THE 2018 WHITE HOUSE SUMMIT ON ARTIFICIAL INTELLIGENCE FOR AMERICAN INDUSTRY, OFFICE OF SCI. & TECH. POL'Y (2018), https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf [hereinafter SUMMARY OF THE 2018 WHITE HOUSE SUMMIT ON AI]. *See also* Fact Sheet, *Artificial Intelligence for the American People,* WHITEHOUSE.GOV (May 10, 2018),https://www.whitehouse.gov/briefings-statements/artificial-intelligence-american-people/ (White House briefing announcing "funding for fundamental AI research and computing infrastructure, machine learning, and autonomous systems" as an executive priority).

[83] *Readout from the Inaugural Meeting of the Select Committee on Artificial Intelligence*, OFFICE OF SCI. & TECH. POL'Y (June 27, 2018), *available at* https://epic.org/privacy/ai/WH-AI-Select-Committee-First-Meeting.pdf.

[84] Request for Information on Update to the 2016 National Artificial Intelligence Research And Development Strategic Plan, NETWORKING & INFO. TECH. RES. & DEV. PROGRAM (Sept. 26, 2018) https://www.nitrd.gov/news/RFI-National-AI-Strategic-Plan.aspx.

[85] Executive Order on Maintaining American Leadership in Artificial Intelligence, WHITEHOUSE.GOV (Feb. 11, 2019), https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/. *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/.

comprehensive national strategy for AI development.[86] Nor is there a democratically-accountable set of development standards.

In the meantime, the commercial sector has led the most recent surge of AI development and deployment in the United States,[87] and federal agencies are relying on standards-setting organizations and private firms. Consider AVs. In the United States, at least at the federal level, regulatory agencies and lawmakers have taken a backseat approach in order to avoid prematurely chilling innovation.[88] DOT has declined to promulgate binding standards while simultaneously exhorting states to be modest with the reach of their legislation and espousing a voluntary self-certification regime for private manufacturers.[89] This Article does not assess this strategy on the merits, instead focusing on an antecedent procedural question that has received too little attention: what standards, normative and technical, are to be used in this self-certification regime?

Self-regulatory efforts abound as a number of AI researchers, non-governmental organizations, and industry members pursue ethical and/or technical standards to guide the technology's development.[90] The following description summarizes two leading private AI standard-setting efforts by, respectively, the IEEE[91] Global Initiative on the Ethics of Autonomous and Intelligent Systems and the joint committee of the International Organization for Standardization and the International Electrotechnical Commission, known as ISO/IEC/JTC 1.[92]

---

[86] *See* Knight, *supra* note 85 (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*, SCIENCE (Feb. 11, 2019, 12:01 AM), https://www.sciencemag.org/news/2019/02/trump-launch-artificial-intelligence-initiative-many-details-lacking.

[87] Though a full history of AI is beyond this Article's scope, non-commercial actors have played notable roles in previous cycles of AI research and development. For instance, the modern move to explore AI is often attributed to a 1956 convening at Dartmouth College, at which researchers coined the term "artificial intelligence." *See* J. McCarthy et al., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (Aug. 31, 1955) (proposing a 2-month, 10-person study of artificial intelligence), *available at* http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html. AI progress since that period tended to occur in boom-and-bust cycles, with periods of major excitement, research, and government investment followed by disillusionment with a particular technological approach, defunding, and so-called "AI Winters." For a detailed account, see STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 16–28 (3d. ed. 2014). This Article begins its narrative in the current upswing that started with increasing access to the very large datasets required for machine learning, see *id.* at 27–28, with an emphasis on twenty-first century developments.

[88] *See* U.S. DOT, *supra* note 9, at viii ("The right approach to achieving safety improvements begins with a focus on removing unnecessary barriers and issuing voluntary guidance, rather than regulations that could stifle innovation.").

[89] *See id.* at ix.

[90] Though this Article focuses on industry-wide efforts, some individual companies have also developed their own ethical standards for AI. *See, e.g.*, *Artificial Intelligence at Google: Our Principles*, GOOGLE AI, https://ai.google/principles/ (last visited Dec. 27, 2018).

[91] The IEEE, or Institute of Electronic and Electrical Engineers, is a technical professional organization that was created in 1884 and now includes engineers, computer scientists, doctors, physicists, and IT professionals. *See About IEEE*, IEEE, https://www.ieee.org/about/index.html (last visited Nov. 21, 2018).

[92] *See ISO/IEC JTC 1 — Information Technology*, INT'L ORG. FOR STANDARDIZATION, https://www.iso.org/isoiec-jtc-1.html (last visited July 31, 2018). This joint committee has operated since 1987 as a "consensus-based, globally relevant, voluntary international standards group," with members from 163 countries. *See Welcome*, ISO/IEC JTC 1, https://jtc1info.org/ (last visited Nov. 21, 2018). It builds from ISO's earlier work on ICT standardization. *Id.* ISO is an "independent, non-governmental organization made up of members from the national standards bodies of 162 countries" that was founded in 1946 to facilitate the international coordination and unification of industrial standards. *Structure and Governance*, ISO, https://www.iso.org/structure.html (last visited Nov. 21, 2018).

The IEEE's initiative identifies general principles such as human rights, prioritizing wellbeing, accountability, and transparency "that apply to all types of autonomous and intelligent systems (A/IS*), regardless of whether they 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)."[93] The IEEE has divided this complex field into 14 standards development efforts.[94] Each standards development effort is associated with a working group that is sponsored by an IEEE committee. IEEE is also concurrently developing an Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) intended to help implement these standards by "creat[ing] specifications for certification and marking processes that advance transparency, accountability and reduction in algorithmic bias."[95]

As one example of what such a process entails, consider the "IEEE P7002™, Data Privacy Process" working group, which aims to "have one overall methodological approach that specifies practices to manage privacy issues within the systems/software engineering life cycle processes." According to the proposal for the working group, it addresses data privacy matters by "defin[ing] requirements for a systems/software engineering process for privacy oriented considerations regarding products, services, and systems utilizing employee, customer or other external user's personal data."[96] It operates by providing specific examples and "includes a use case and data model (including metadata)," aiming to provide "specific procedures, diagrams, and checklists [so that] users of this standard will be able to perform a conformity assessment on their specific privacy practices." It also contemplates the use of privacy impact assessments. In addition, to the extent that these issues raise questions of algorithmic bias, the newly launched ECPAIS program will ostensibly provide particular certification requirements that vendors must satisfy.

Compare the IEEE's program with the work of another enterprise: the joint committee of the International Organization for Standardization and the International Electrotechnical Commission known as ISO/IEC/JTC 1.[97] This joint committee is presently developing ISO/IEC JTC 1/SC 42, Artificial Intelligence, which aims to "[s]erve as the focus and proponent for JTC 1's standardization program on Artificial Intelligence" and guide "JTC 1, IEC, and ISO committees developing Artificial Intelligence applications."[98] ISO/IEC/JTC 1 has published materials that include a "Big Data Reference Architecture" and an accompanying set of use cases

---

[93] *See General Principles V2*, IEEE, https://standards.ieee.org/develop/indconn/ec/ead_general_principles_v2.pdf (last visited Jul. 7, 2018).

[94] *See The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, IEEE STANDARDS ASS'N, https://standards.ieee.org/industry-connections/ec/autonomous-systems.html (last visited Jan. 13, 2019) (listing and linking to working groups).

[95] *The Ethics Certification Program for Autonomous and Intelligent Systems* (ECPAIS), IEEE, https://standards.ieee.org/industry-connections/ecpais.html (last visited Dec. 27, 2018).

[96] *P7002 Project Allocation Request*, IEEE STANDARDS ASS'N, https://development.standards.ieee.org/get-file/P7002.pdf?t=91648400003 (last visited July 16, 2018).

[97] *See ISO/IEC JTC 1 — Information Technology*, INT'L ORG. FOR STANDARDIZATION, https://www.iso.org/isoiec-jtc-1.html (last visited July 31, 2018). This joint committee has operated since 1987 as a "consensus-based, globally relevant, voluntary international standards group," with members from 163 countries. *See Welcome*, ISO/IEC JTC 1, https://jtc1info.org/ (last visited Nov. 21, 2018). It builds from ISO's earlier work on ICT standardization. *Id.* ISO is an "independent, non-governmental organization made up of members from the national standards bodies of 162 countries" that was founded in 1946 to facilitate the international coordination and unification of industrial standards." *Structure and Governance*, ISO, https://www.iso.org/structure.html (last visited Nov. 21, 2018).

[98] *ISO/IEC JTC 1/SC 42*, INT'L ORG. FOR STANDARDIZATION, https://www.iso.org/committee/6794475.html (last visited Jan. 30, 2019). This work is proceeding within seven working groups: Dissemination and Outreach; a joint working group on Governance implications of AI; Computational Approaches and Characteristics of Artificial Intelligence Systems; Foundational Standards; Big Data; Trustworthiness; and Use cases and Applications. *Id.*

and technical standards. According to the ISO/IEC publication, this document aims to summarize the standards currently in place to support big data, how big data affects existing standards, and the "central scientific, technological, and standardization challenges that need to be addressed to accelerate the deployment of robust big data solutions."[99]

How different or similar are these two standards, in practice? Until the IEEE text is finalized, it is not possible to say with certainty. And regardless of the substantive merits of any individual standard, there remains a procedural problem: the current panoply of initiatives and associated guidance fragments the conversation and further confounds already-difficult questions about which normative and technical standards should guide AI development.[100]

B. More Standards, More Problems

In theory, self-regulatory standards might serve as a norm-driven institutional check, operating in lieu of a formal check-and-balance by the state.[101] Putting to the side concerns about the democratic legitimacy of this outcome, the present problem for AI is that a proliferation of standards and lack of uniformity or ordering of authority among or even within them is likely to confuse rather than elucidate any such norm-driven governance requirements.

Organizationally, there are so many stakeholders that it is difficult to unravel who has input, at which points, in ways that facilitate meaningful oversight or public input. The IEEE's 14 working groups are currently developing AI standards on topics as complex and diverse as, for example, a "model process for addressing ethical concerns during system design" and an "ontological standard for ethically driven robotics and automation systems."[102] The sheer range of knowledge and expertise required for even a single one of these initiatives makes it difficult to trace the development of these standards. Or consider again the ISO/IEC JTC 1/SC 42 project, which consists of 21 participating countries and 10 observing members.[103] One of the participating members is the American National Standards Institute (ANSI), which has operated since 1918 and which "accredits standards developers that will establish consensus among qualified groups," with "220 distinct entities currently accredited to develop and maintain nearly 10,000 American National Standards (ANS)."[104] These projects are vast in both their ambitions and their membership.

[99] *Id.*
[100] *Cf.* Gary E Marchant & Wendell Wallach, *Coordinating Technology Governance*, 31 ISSUES SCI. & TECH. (Summer 2015), https://issues.org/coordinating-technology-governance/ (discussing "big coordination problems" that are "all too common" when there are multiple soft law initiatives for the same emerging technology and proposing Governance Coordinating Committees as a remedy).
[101] *Cf.* Lessig, *The New Chicago School*, *supra* note 54, at 662–63 (detailing role of norms as a regulatory modality); AYRES & BRAITHEWAITE, *supra* note 45, at 13–14 ("Where one institutional order is weak in a society, massive and often unsuccessful efforts are made to compensate through other institutions. . . . [T]he institutions of community, market, state, and associational order are to some extent both mutually constituting and mutually constraining.").
[102] *See Ethics in Action*, IEEE, https://ethicsinaction.ieee.org/ (last visited Nov. 21, 2018).
[103] *See Participation, ISO/IEC JTC 1/SC 42*, ISO, https://www.iso.org/committee/6794475.html?view=participation (last visited Nov. 21, 2018).
[104] *ANSI*, ISO, https://www.iso.org/member/2188.html (last visited Nov. 21, 2018).

If the end goal is for a state actor to fold the standard into a formal regulation,[105] then voluntary consensus and guidance might be enough to facilitate interoperable technical standards that permit free trade, in the manner that the ISO/IEC JTC 1 project seems to envision. And substantive specialization in narrower domain areas, as the IEEE working groups appear to contemplate, might make good functional sense if the public can have faith that the output has been subjected to democratic checks within the political process.

But neither of these conditions holds for AI. Consider AVs once more and imagine a private firm seeking a standard for the sort of voluntary self-certification that the DOT recommends. Recall that the DOT has declined to provide a more formal regulatory prescription, such that the standards are the sole guidance. But how is a firm to select among the many standards, and what guarantee does the public have that its choice reflects anything other than self-interest? The traditional standard-setting model may, at best, create more noise than signal. At worst, it permits private companies to claim they are adopting industry best practices, when in fact there is no public check on the substance of these practices—particularly because both the IEEE and ISO standards require users to pay non-trivial fees to access the standards[106] and there are no penalties other than the speculative risk of a less commercially desirable product if a firm chooses not to adopt a measure that becomes the industry standard.

Moreover, even assuming the best intentions among private actors, the risk of substantive confusion remains to the extent that different initiatives frame similar issues in distinct terms. For instance, the IEEE text appears to address the issue in terms of a sociotechnical approach to data privacy "considerations," whereas the ISO/IEC effort appears to address the issue by foregrounding the "scientific, technological, and standardization challenges" associated with the data itself.[107] The emphasis is placed on different syllables in the data privacy conversation, with no clear governance structure within which to address any problematic areas of distinction. Perhaps such heterogeneity permits only the best guidance to stick. But with so many multifaceted, interdisciplinary considerations, confusion seems a more likely outcome.

If consistency and clarity are the missing pieces of standardization efforts, then perhaps it is time for more formal government intervention. A turn to public law might seem natural. In theory, at

[105] *See id.* ("Though all ANS are developed as voluntary documents, U.S. federal, state, or local bodies are increasingly referring to ANS for regulatory or procurement purposes. Many ANS are also national adoptions of globally relevant international standards.").

[106] IEEE pricing is not yet available because the standards are in development, but past compilations of technical standards have been priced at $2000-3000. *See, e.g.*, *IEEE Smart Grid Research: Computing*, IEEE (Apr. 30, 2013), https://www.techstreet.com/ieee/standards/ieee-smart-grid-research-computing?gateway_code=ieee&vendor_id=5644&product_id=1857774. The ISO/IEC standards available as of late 2018 are priced at $88-$198, respectively. *See ISO/IEC TR 20547-5:2018, Information Technology – Big Data Reference Architecture – Part 5: Standards Roadmap*, INT'L ORG. FOR STANDARDIZATION (2018), https://www.iso.org/standard/72826.html?browse=tc; *ISO/IEC TR 20547-2:2018, Information Technology – Big Data Reference Architecture – Part 2: Use Cases and Derived Requirements*, INT'L ORG. FOR STANDARDIZATION (2018), https://www.iso.org/standard/71276.html?browse=tc.

[107] The two already-published standards suggest a focus on technical architectures and application-specific considerations, as opposed to professional or ethical issues. *See ISO/IEC TR 20547-5:2018*, *supra* note 106 ("ISO/IEC TR 20547-5:2018 describes big data relevant standards, both in existence and under development, along with priorities for future big data standards development based on gap analysis."); *ISO/IEC TR 20547-2:2018*, *supra* note 106 ("ISO/IEC TR 20547-2:2018 provides examples of big data use cases with application domains and technical considerations derived from the contributed use cases.").

least, what is more uniform and binding across the board than legislation or public administrative regulation?

III. Alternative Administrative Paradigms

This Part considers two alternative administrative approaches that public actors might apply to an emerging technology like AI: prescriptive regulation and collaborative governance. It provides a stylized summary of each model, then explores their limits for AI, underscoring how the strategic context for AI development combines with the same technical attributes that complicate prescriptive interventions to create a different kind of AI governance challenge.

A. Prescriptive Regulation

1. Pharmaceutical Clearance by FDA

FDA, as the name suggests, regulates food and drugs.[108] Federal regulation of the American food and drug industry in fact predates FDA's creation and is traceable to the Pure Food and Drug Act of 1906,[109] inspired in part by "muckraking" journalism that exposed unsavory factory conditions and indicted "snake oil" medical hucksters. By the 1930s, a new wave of concerns and controversies[110] led Congress to pass the Food, Drug, and Cosmetics Act of 1938, which established FDA as a federal "citizen-protection agency."[111]

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.[112] FDA's drug regulation involves a command-and-control tactic that requires industry players to meet certain safety and efficacy thresholds before

[108] 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 (2004). The stylized summary presented in Section III.A focuses primarily on policy challenges and patterns of administrative law responses.

[109] *See Part I: The 1906 Food and Drugs Act and Its Enforcement*, FDA's Evolving Regulatory Powers, FDA (Feb. 2, 2018), https://www.fda.gov/AboutFDA/History/FOrgsHistory/EvolvingPowers/ucm054819.htm (("This act, which the Bureau of Chemistry was charged to administer, 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.")

[110] 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 100 victims, including many children. *See* Rebecca S. Eisenberg, *The Role of the FDA in Innovation Policy*, 13 MICH. TELECOMM. TECH. L. REV. 345, 345 & nn.1–2 (2007) (citing PHILIP J. HILTS, PROTECTING AMERICA'S HEALTH (2003) and sources cited therein); *Part II: 1938, Food, Drug, Cosmetic Act*, FDA's Evolving Regulatory Powers, FDA (Feb. 2, 2018), https://www.fda.gov/AboutFDA/History/FOrgsHistory/EvolvingPowers/ucm054826.htm.

[111] HILTS, *supra* note 110, at xi.

[112] *See id.* at xii ("The new agency was the people's investigator, with the specific mission of intervening on behalf of citizens and against businesses when necessary. . . . Roosevelt and Congress were establishing 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?*, FDA, https://www.fda.gov/AboutFDA/Transparency/Basics/ucm214416.htm (last updated Aug. 22, 2018).

permitting them to market drugs to the general public.[113] 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 their physical well-being.[114] 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,[115] with statutory expansions of FDA's authority over time. Notably, responding in part to a spate of deaths and birth defects linked to thalidomide,[116] the Kefauver-Harris Amendments of 1962 increased FDA's pre-market authority. 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 pre-market notification system to a pre-market approval system.[117] Approval, moreover, required "adequate and well-controlled" scientific experiments carried out by "experts qualified by scientific training."[118] The responsibility of establishing a drug's safety was thus shifted to the private manufacturer, with the company required to demonstrate affirmatively a product's safety and efficacy for its stated purpose in order to obtain market clearance.[119] 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.

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

The FDA example illustrates 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. But before concluding that at least some categories of AI technology[120] are good candidates for a similar preclearance regime, three further technical attributes and associated policy challenges[121] merit special consideration. First, the potential speed of algorithmic development alters the resource demands to create the product and, in turn, to change the underlying algorithm. 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

---

[113] *See* W. Nicholson Price II, *Regulating Black-Box Medicine,* 116 MICH. L. REV. 421, 424 (2017); Eric R. Claeys, *The Food and Drug Administration and the Command-and-Control Model of Regulation*, 49 ST. LOUIS U. L.J. 105 (2004); Gilhooley, *supra* note 108.

[114] *See* HILTS, *supra* note 110, 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. 2013).

[115] *See* Claeys, *The Food and Drug Administration and the Command-and-Control Model of Regulation*, *supra* note 113, at 106.

[116] *See* HILTS, *supra* note 110, at 154–58 (describing thalidomide controversy and reporting evidence that drug caused birth defects in up to 8000 babies worldwide, with an estimated 5000 to 7000 additional pre-birth deaths).

[117] *See* Richard A. Merrill, *The Architecture of Government Regulation of Medical Products*, 82 VA. L. REV. 1753, 1764–65 (1996).

[118] HILTS, *supra* note 110, at 160, 164–65.

[119] *Id.* at 164.

[120] To permit more precise analysis, this discussion focuses on machine learning, the dominant AI method at present. *See supra* note 16; *see also* text accompanying note 27 and sources cited therein.

[121] *See* discussion *supra* note 36.

uncertainty and emergence.[122] While recognizing how 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 for AI-based products.

a. Speed

The speed of AI development, adjustment, and deployment makes AI applications especially ill-suited to prescriptive methods. Speed matters specifically for AI in at least three ways: (i) the resources to create the product or service; (ii) the resources to affect the overall product or service by adjusting the algorithm; and (iii) algorithmic computation, or "compute." The relationship between AI and "compute" power warrants separate analysis,[123] and the focus of this Article is on the first two dimensions, with an emphasis on settings in which AI is applied within a tangible product.[124]

Algorithmic development can be more rapid than past innovation cycles in part because the strings of programming text, or software, on which AI runs can be developed, erased, and re-created with relatively fewer investments in physical infrastructure and resources.[125] True, considerable capital investment and research is required to build an AI algorithm, and the system requires a computer on which to run. Yet the core resources required to implement it are digital, not physical.

Consider the high costs of, first, pharmaceutical research and development, and, second, drug manufacturing. FDA's pre-market clearance regime means that any would-be manufacturer must make sizeable resource investments to create and test the drug,[126] including paying for laboratories and scientists' time, in order to generate adequate empirical evidence of that drug's safety and efficacy for a particular use. 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.[127]

---

[122] *Cf.* Calo, *Robotics and the Lessons of Cyberlaw*, *supra* note 68, at 538–45.

[123] 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*, SSRN (Mar. 24, 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=314797. *Cf. Speed Conference*, CORNELL TECH, https://www.dli.tech.cornell.edu/speed (last visited July 9, 2018) (focusing on computing speed in particular applications and domains). Recent research underscores the link between hardware configuration and AI advances. *See* ImportAI: 122, *Google Obtains New Imagenet State-Of-The-Art Accuracy with Mammoth Networks Trained Via 'Gpipe' Infrastructure*, (Nov. 26, 2018), *available at* https://us13.campaign-archive.com/?u=67bd06787e84d73db24fb0aa5&id=6d61d65ae0&e=[UNIQID] (quoting Yanping Huang et al., *GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism*, ARXIV (Nov. 20. 2018), https://arxiv.org/abs/1811.06965 ("Our work validates the hypothesis that bigger models and more computation would lead to higher model quality.")).

[124] 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 dialog between this Article and recent legal scholarship that has focused on embodied algorithms, or robots. *See supra* note 1.

[125] 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 54 and surrounding text. The category "code" thus includes the algorithms discussed here.

[126] *See* Joseph A. DiMasi et al., *The Price of Innovation: New Estimates of Drug Development Costs*, 22 J. HEALTH ECON. 151 (2003). *But see* MERRILL GOOZNER, THE $800 MILLION PILL 238-39 (2004) (contesting this figure).

[127] *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.").

In contrast, the inputs needed to produce software are far less resource-intensive. 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. Even where physical hardware exists, moreover, software updates can often change the manner in which they function without needing to start from scratch.[128] Assuming access to adequate data on which to run an algorithm, it is possible to create and change the algorithm far more quickly than in other sectors, which require large investments of capital and sustained periods of development to implement a similar scale of change.

Such algorithmic development also affects the nature of the consumer-facing product. As Paul Ohm explains, the "change-per-effort ratio" is far lower for software, or algorithmic, construction as compared to past industrial processes.[129] Since AI algorithms typically consist of building blocks for other goods or services, as opposed to products on their own terms, the comparatively lower resource investment required for software adjustment also permits relatively faster and easier changes to the final product or service. Consider, for instance, an algorithm that gathers data from on-board 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. In practice, the utility-like nature of AI means a manufacturer can fundamentally 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.[130] 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. This adjustment is far faster and easier than, say, attempting to modify the active molecules in a drug.

Given the potential for rapid change at relatively low cost, it might seem even more reasonable to impose prescriptive regulatory requirements for algorithms. 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 exactly right? The general problem with this argument for ML technologies[131] is that it fails to account for the aggregate cost of what is likely to be a very large number of small interventions required to reach the regulatory standard. Adjusting the algorithm is relatively cheaper and easier than adjusting a physical object. But that does not mean that a single adjustment will deliver a product that meets the regulatory requirement. Rather, particularly if a system is "offline" rather than dynamic in its learning, a data scientist might need to acquire an entirely new data set, clean it, retrain the model, and develop an entirely new working algorithm.

---

[128] *See* Paul Ohm, Commentary, *We Couldn't Kill the Internet If We Tried*, 130 HARV. L. REV. F. 79, 81 (2017).
[129] *Id.* ("Software is nothing like the industrial processes it has begun to replace. 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. The 'change-per-effort ratio' is thus much lower in software construction.").
[130] *Cf.* Colin Dwyer, *NTSB: Uber Self-Driving Car Had Disabled Emergency Brake System Before Fatal Crash*, NPR (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 ("As to why the software did not engage the brakes on its own, NTSB noted that this passive approach is actually an intentional part of the design. The agency explained that the vehicle, a modified 2017 Volvo XC90, comes 'factory equipped' with automatic emergency braking — but that Uber's system disables this function and others when it's in use.").
[131] ML is a subset of AI in which a system learns without ex ante, explicit programming. *See supra* note 16; text accompanying note 27 and sources cited therein.

The "cost-per-effort ratio" might be lower, yet the net cost could still be quite high. A version of the classic "pacing problem" thus emerges.[132] There is an inevitable tradeoff between slowing down technical development to meet a regulatory prescription and allowing innovation to occur without potentially burdensome regulatory safeguards. Moreover, particularly because other algorithmic interventions are likely to speed ahead, the more rapid pace of technical change risks compounding any gap between law on the books and the state of technology in the world.[133]

b. Complexity

The technical complexity of AI introduces a number of additional potential regulatory challenges. Two dimensions are especially salient: (i) domain expertise, given the specialized nature of the technology itself and the knowledge required to conduct AI research or create applications and (ii) interpretability, including both inexplicability and incommensurability with traditional ways of understanding the world.

(i) Domain Expertise. 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.[134] And even before seeking this more specialized expertise, individuals trained to tackle AI in general or ML in particular are in short supply. Again, according to the most recent available figures, there are just "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."[135] Moreover, formal training is inadequate because "developing successful machine learning applications requires a substantial amount of 'black art' that is hard to find in textbooks."[136]

True, the "black art" of functional domain expertise is a prerequisite in many technocratic disciplines. As Arthur C. Clarke's third law holds, "any sufficiently advanced technology is indistinguishable from magic."[137] 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.

---

[132] *See* Gary E. Marchant, *The Growing Gap Between Emerging Technologies and the Law*, *in* THE GROWING GAP BETWEEN EMERGING TECHNOLOGIES AND LEGAL-ETHICAL OVERSIGHT 19 (Gary E. Marchant et al. eds., 2011) ("In contrast to th[e] accelerating pace of technology, the 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 processes to regulate." (citations omitted)).

[133] 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.

[134] For a description of ML stages targeted at legal scholars, see Lehr & Ohm, *supra* note 16.

[135] *Global AI Talent Report 2018*, JFG, http://www.jfgagne.ai/talent (last visited Dec. 4, 2018).

[136] Pedro Domingos, *A Few Useful Things to Know about Machine Learning*, 55 COMMS. ACM 78, 78 (2012).

[137] *Clarke's Three Laws*, WIKIPEDIA, https://en.wikipedia.org/wiki/Clarke%27s_three_laws (last visited Dec. 27, 2018).

But FDA can point to a specific regulatory object, the drug, and craft a regime around that object. Though these delineations are not perfect,[138] FDA can nonetheless bound its authority in ways that do not naturally map onto the general-purpose, dynamic nature of AI technology.

Trying similarly to bound AI within particular regulatory sectors presents both conceptual and normative wrinkles. First, imagine that the government attempts to narrow the field by adopting a sectoral approach in areas where development is deemed especially risky. This could be for any number of reasons, from an application that takes an embodied form and is deemed more likely to result in a physical harm to a criminal justice intervention that implicates core constitutional rights and is deemed more likely to result in a harm to life and liberty. Such a tack in fact bears a family resemblance to American privacy law's statutory framework to protect especially sensitive kinds of information.[139] It is also close to the AI status quo, in which no single U.S. federal agency has clear jurisdiction over all elements of contemporary AI.[140] Having identified the right sectors, imagine, further, an agency with pre-market clearance authority for algorithms, akin to FDA.[141] Opting for domain expertise within a single administrative agency, such as DOT, might initially seem like a good strategy. It might not only leverage sector-specific domain expertise an agency has cultivated, but also permit focused treatment of one category of issues, such as physical safety risks posed by AVs.[142]

But there is a conceptual problem with a sector-specific approach. A critical issue that is not explored elsewhere is that AI is not like a drug, for which the molecules are ostensibly stable once taken out of the lab. Rather, especially for ML applications, data scientists must constantly make choices about how to "play with the data" in order to run a particular model.[143]

Accordingly, even recognizing that drugs also involve emergent properties, running 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 pre-clearance regime.

---

[138] 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* W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419 (2015) [hereinafter Price, *Black-Box Medicine*] This Article reserves further analysis of what factors may make an emerging technology especially ill-suited for traditional administrative law paradigms for future work.

[139] *See* Alicia Solow-Niederman, *Beyond the Privacy Torts: Reinvigorating a Common Law Approach for Data Breaches*, 127 YALE L.J. FORUM 617–18 n.13 (2018) (summarizing American sector-by-sector approach to privacy regulation).

[140] *See* discussion *supra* Part II.A. Presently, for example, the Security and Exchange Commission (SEC) regulates automation in the financial arena; the Department of Transportation coordinates autonomous vehicle regulation across the Federal Highway Administration, the Federal Railroad Administration, the National Highway Traffic Safety Administration, the Federal Motor Carrier Safety Administration, the Federal Transit Administration, the Pipeline and Hazardous Materials Safety Administration, the Federal Aviation Administration, the Maritime Administration, and relevant state and local bodies; and so forth. *See* U.S. DOT, *supra* note 9.

[141] Andrew Tutt has in fact proposed just such an approach, though he appears to endorse an overarching process agency as opposed to one targeted at a particular substantive issue area. *See* Tutt, *supra* note 2.

[142] *See* Meredith Whittaker et al., *supra* note 20, at 4 ("Governments need to regulate AI by expanding the powers of sector-specific agencies to oversee, audit, and monitor these technologies by domain. . . . We need a sector-specific approach that does not prioritize the technology, but focuses on its application within a given domain.").

[143] *See generally* Lehr & Ohm, *supra* note 16. *See also* sources cited *supra* note 16.

Moreover, the scale and complexity of the problem compounds if the institutional intervention is an overarching agency to oversee algorithms across different sectors.[144] Even if the goal is simply to provide advice and oversight, such as, for instance, a Federal Robotics Commission[145] or a National Algorithm Safety Board,[146] the sheer number of domains affected is daunting. For instance, the Obama Administration's 2016 Machine Learning and Artificial Intelligence Subcommittee, led by White House Office of Science and Technology Policy and National Institute of Standards and Technology, included members from 20 federal departments and agencies, from Department of Commerce to Department of Veterans Affairs to Central Intelligence Agency, as well as 7 offices of the Executive Office of the President.[147]

Furthermore, the broad sweep of potential applications for AI becomes even more complex because the technology often implicates a range of public interests in the "real world," 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[148] and emotional state,[149] compiles 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[150] (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 well-being of the driver and that of other individuals might be in conflict).[151] Other cross-cutting questions, such as cybersecurity, also arise in multiple sectors.[152] And since ML

---

[144] *Accord* Meredith Whittaker et al., *supra* note 20, 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.").
[145] *See, e.g.*, Ryan Calo, *The Case for a Federal Robotics Commission*, *supra* note 25.
[146] *See* Schneiderman, *supra* note 25.
[147] *See* NAT'L SCI. & TECH. COUNCIL COMM. ON TECH., EXEC. OFFICE OF THE PRESIDENT, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016).
[148] Vehicles are already being equipped with such systems. *See* Joann Muller, *Driver Monitoring Systems Are Here — And So Are Privacy Concerns*, AXIOS (Oct. 26, 2018), https://www.axios.com/driver-cameras-bring-privacy-concerns-873804d2-8897-468b-82f4-b3586bdfea31.html (discussing "Super Cruise" driver monitoring system now available in GM Cadillac CT6).
[149] Again, manufacturers are already developing this capability. *See* Viknesh Vijayenthiran, *Kia concept for 2019 CES Previews a World with Self-Driving Cars* (Jan. 3, 2019), https://www.motorauthority.com/news/1120570_kia-concept-for-2019-ces-previews-a-world-with-self-driving-cars (discussing Kia's READ (Real-time Emotion Adaptive Driving) system, a not-yet-deployed feature that monitors and responds to the driver's emotional state).
[150] *See* Keith Naughton, *Ford Breaks With GM, Toyota on Future of Talking-Car Technology*, BLOOMBERG (Jan. 7, 2019, 5:00 PM), https://www.bloomberg.com/news/articles/2019-01-07/ford-breaks-with-gm-toyota-on-future-of-talking-car-technology (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"); 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 (discussing need to embed sensors city-wide for AVs to function).
[151] *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 2018 AV START bill on grounds it did not amply address cybersecurity and privacy issues).
[152] 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-

requires massive data sets to train them to the requisite level of performance, overlapping ground-level discussions about fairness in data sets undergird any application of the technology.[153]

Cabining these and similar issues in the context of questions about a single domain risks obscuring or wholly eliding these cross-cutting considerations. The non-trivial cost of an isolated sectoral approach, moreover, compounds to the extent that it prevents the development of cross-sectoral principles or precludes the emergence of publicly-shared consensus on the values that are not to be compromised in any sector.[154] If we are contending with a brave new algorithmic world, then we should have first principles that reflect the actual ground conditions.

(ii) Interpretability. Turning from specialized knowledge that ML requires to the technology itself, ML's interpretability challenges further compound its complexity. 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.[155] 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.[156] Second, the manner in which the model is connecting the data to identify a pattern might be non-intuitive, particularly when compared to the cause-and-effect reasoning that drives the scientific method.[157] It is possible, to be sure, that advances in interpretable machine learning might decrease inscrutability,[158] and a researcher may be able to increase

---

over-the-air-software-update-vehicles-117d0b7d-cb13-4b63-aa0f-0dae365f97dc.html (discussing cybersecurity risks when "over-the-air" software updates are sent to vehicles)

[153] 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, referred to as FAT or FATE, addresses these questions. *See* Conference on Fairness, Accountability, and Transparency (FAT*), https://fatconference.org/index.html (last visited Dec. 4, 2018). *Cf.* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016).

[154] Privacy law scholars may be reminded of the distinction between the U.S. regime of protecting informational privacy rights sector-by-sector, *see* Solow-Niederman, *supra* note 139, at 617–18 n.13, as compared to the European Union's protection of privacy as a fundamental human right across all sectors.

[155] 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.

[156] *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3126971.

[157] *Id.*

[158] There has been a flurry of recent work to develop "interpretable" and "explainable" AI models. *See, e.g.*, Chris Olah et al., *The Building Blocks of Interpretability*, DISTILL (2018), https://distill.pub/2018/building-blocks/; Finale Doshi-Velez & Been Kim, *Towards A Rigorous Science of Interpretable Machine Learning*, ARXIV (Mar. 2, 2017), https://arxiv.org/abs/1702.08608; David Gunning, *Explainable Artificial Intelligence*, DARPA https://www.darpa.mil/program/explainable-artificial-intelligence (last visited Nov. 11, 2018).

interpretability or implement auditing procedures that make a particular outcome less opaque.[159] Nonetheless, these are coping strategies, not cures.

The non-intuitiveness problem, moreover, may be harder to resolve. For instance, in one well-noted example, a neural network built to 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.[160] Along similar lines, in a possibly apocryphal story, the 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.[161] This sort of non-intuitive, yet rationalizable correlation abounds in ML; indeed, the technique's appeal comes from its capacity to identify patterns that humans would not necessarily discern. 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, or 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,[162] the specific "mechanism of action" through which they operate remains unknown.[163] In other words, the precise cause-and-effect process that makes these drugs efficacious remains inscrutable, even if it is effective for many patients. And FDA has seemingly contended with these challenges while maintaining its command-and-control framework for premarket clearance.

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."[164] For 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.[165] 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

---

[159] *See* Lehr & Ohm, *supra* note 16, at 656–58.
[160] Marco Tulio Ribeiro et al., *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, ARXIV (Feb. 16, 2016), https://arxiv.org/abs/1602.04938. *See also* Selbst & Barocas, *supra* note 156, at 47–48 (discussing example).
[161] Eliezer Yudkowsky, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, *in* GLOBAL CATASTROPHIC RISKS 308, 323–24 (Nick Bostrom & Milan M. Ćirković eds., 2008).
[162] *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 (last updated Feb. 20, 2018).
[163] *See* Full Prescribing Information 20, http://pi.lilly.com/us/prozac.pdf (last visited Nov. 16, 2018) ("Although the exact mechanism of PROZAC is unknown, it is presumed to be linked to its inhibition of CNS neuronal uptake of serotonin.").
[164] W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421 (2017) (quoting Price, *Black-Box Medicine*, *supra* note 138, at 421, 429–34).
[165] *See* Price, *Black-Box Medicine*, *supra* note 164, at 436–37.

authority. To date, the attitude has been that FDA can respond by adjusting its regulatory practices.[166] 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. It is far from a first-best solution, particularly in an emerging field that is as cross-cutting as AI algorithms.

c. Unpredictability

In addition to questions about the predictability of a model's inner workings, an additional and seemingly novel twist emerges when a model interacts with the real world in unpredictable ways. Unpredictability encompasses two broad categories of problems: (i) uncertainty and (ii) emergence, each of which applies for both algorithms and embodied AI, or robots.

(i) Uncertainty. The inability to predict AI outcomes with complete certainty may produce accidents, particularly when a real-world AI system is poorly designed in ways that lead to unintended and harmful behavior.[167] 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."[168] 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, like the flight simulator that generated such a gigantic force that it overflowed the counter and made it seem like the force was zero.[169]

AI researchers have begun to explore how careful study of categories of common issues such as reward hacking might minimize such accidents. If categories of accidents are par for the AI course, then it may be possible to implement mechanisms or strategies to decrease their likelihood or mitigate their impact. Proposed technical solutions to reward 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 "intentionally 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.[170]

---

[166] *See* Price, *Regulating Black-Box Medicine*, *supra* note 164, at 452 (expressing concern that FDA lacks expertise to contend with complex black-box algorithms, yet doubting that another government agency is better positioned).
[167] *See* Dario Amodei et al., *Concrete Problems in AI Safety*, ARXIV (July 25, 2016), https://arxiv.org/abs/1606.06565 (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").
[168] *Id.* (manuscript 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-ofai-systems/.
[169] *See supra* text accompanying notes 2–7; *see also* Amodei et al., *supra* note 167 ((manuscript 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")).
[170] *See id.* (manuscript at 7–11).

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,[171] but also an important substrate of risk management of emerging technologies in particular,[172] including legal scholarship on point.[173] In fact, the principles of this growing field undergird pharmaceutical regulation.

Consider FDA's premarket clearance requirements for drugs.[174] 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.[175] 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.[176] 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 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. Without much finer-grained specification of a particular regulatory object, AI's technical attributes do not fit neatly within risk management paradigms. At the same time, however, narrowing the regulatory scope to a single sector again compromises an important opportunity to develop grounding principles that would apply across sectoral applications.[177]

(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.[178] Emergence is, on one hand, a desirable property insofar as

[171] *See, e.g.*, DANIEL M. BYRD III & C. RICHARD COTHERN, INTRODUCTION TO RISK ANALYSIS (2005).
[172] *See, e.g.*, *Governance of Emerging Technologies & Science (GETS) Conference*, *supra* note 76.
[173] *See, e.g.*, Gary E. Marchant & Yvonne A. Stevens, *Resilience: A New Tool in the Risk Governance Toolbox for Emerging Technologies*, 51 U.C. DAVIS 233, 241–47 (2017); Gary E. Marchant, *Advancing Resilience through Law* (citing S.A. Shapiro, & R.L. Glicksman, *Improving Regulation Through Incremental Adjustment*, 52 U. KANSAS L. REV. 1179 (2004)), *in* INTERNATIONAL RISK GOVERNANCE CENTER, RESOURCE GUIDE ON RESILIENCE 158, 159 (vol. 1 2016)).
[174] *See supra* text accompanying notes 116–119 and sources cited therein.
[175] *See* Bara Fintel et al., *The Thalidomide Tragedy: Lessons For Drug Safety and Regulation*, HELIX MAG. (Jul. 28, 2009), https://helix.northwestern.edu/article/thalidomide-tragedy-lessons-drug-safety-and-regulation.
[176] *See supra* Section III.A.1.
[177] *See supra* text accompanying notes 150–154.
[178] *See* STEVEN JOHNSON, EMERGENCE 18 (2001); *see also* Calo, *Robotics and the Lessons of Cyberlaw*, *supra* note 68, at 538–45.

it catalyzes creative outcomes that human programmers would not necessarily have considered.[179] Indeed, in some contexts, it may be a new form of intelligence.[180]

But this intelligence has two faces. In an algorithm, each line of programming code operates as a low-level element of an emergent system.[181] 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 instance, in an application such as an AV, each line of code will come together to form a working algorithm that interacts with real-world 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 a preliminary investigation by the National Transportation Safety Board (NTSB), the software was confused by the pedestrian and bicycle walking in tandem and "classified the pedestrian as an unknown object, [then] as a vehicle, and then as a bicycle with varying expectations of future travel path."[182] Because the vehicle's software could not determine what it was driving toward, its waffling ate up precious response time. And once it made a final determination, it had inadequate time to avert the collision. In theory, there might still have been other ways to stop the vehicle in time. In practice, the company had turned off the emergency braking function "to reduce the potential for erratic behavior"[183] and make the ride less turbulent.[184] The software system also did not provide a warning to the human safety driver in time to intervene. 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 which the human safety driver responded to a late warning. AV-human interactions are complex systems whose interactions cannot necessarily be predicted by focusing on any single subpart.

Even so, 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.[185] Or a system of stricter ex post sanctions in tort and/or criminal law could incentivize manufacturers to proceed more cautiously.

---

[179] *See* Lehman et al., *The Surprising Creativity of Digital Evolution*, *supra* note 2; Calo, *Robotics and the Lessons of Cyberlaw*, *supra* note 68at 539–40 (noting potential for "useful but unexpected problem solving by machines").
[180] *See* JOHNSON, *supra* note 178; Steven Johnson, *Only Connect*, GUARDIAN (Oct. 15, 2001), https://www.theguardian.com/books/2001/oct/15/society (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's] much more fluid and nuanced than the logic we traditionally expect from our computers").
[181] 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.
[182] NAT'L TRANSP. SAFETY BOARD, PRELIMINARY REPORT, HIGHWAY HWY18MH010 (MAY 25, 2018), https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf.
[183] *Id.* The NTSB investigation is ongoing. *See Car With Automated Vehicle Controls Crashes Into Pedestrian*, NTSB, https://www.ntsb.gov/investigations/Pages/HWY18FH010.aspx (last visited July 30, 2018).
[184] *See* 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/.
[185] *See* David Weinberger, *Optimization over Explanation*, Medium (Jan. 28, 2018), https://medium.com/berkman-klein-center/optimization-over-explanation-41ecb135763d (advocating a turn to existing policy processes to "to decide what we want [AI] systems optimized for").

But such uses of law to target organizational protocols cannot resolve 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. Consider the fact that it took software experts almost two years to disentangle the tangled mess of coding errors that led Toyota Camry automobiles to mistakenly accelerate and kill a number of drivers. And this finding was after NASA engineers had found no error in six months of research.[186] The Camry incident, moreover, did not even involve the complexity of AV software. As computer scientist Ellen Ullman explains, "[i]n some ways we've lost agency. When programs pass into code and code passes into algorithms and then algorithms start to create new algorithms, it gets farther and farther from human agency. Software is released into a code universe which no one can fully understand."[187]

Again, it might be tempting to develop a public regulatory response to try to mitigate the most critical policy concerns. After all, FDA's pharmaceutical regulatory framework has also needed to contend with emergent properties of drugs to ensure their safe usage. 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]."[188] 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 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.[189] This interaction is now fairly well-publicized by prominent medical clinics,[190] yet FDA does not regulate the safety or efficacy of herbal or botanical remedies that are used as dietary supplements.[191] Accordingly, the question of whether FDA should clear the SSRI as "safe" and "effective" based on clinical data can address 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.

---

[186] *See U.S. Department of Transportation Releases Results from NHTSA-NASA Study of Unintended Acceleration in Toyota Vehicles*, NHTSA (June 2011), https://one.nhtsa.gov/About-NHTSA/Press-Releases/2011/U.S.-Department-of-Transportation-Releases-Results-from-NHTSA%E2%80%93NASA-Study-of-Unintended-Acceleration-in-Toyota-Vehicles ("We enlisted the best and brightest engineers to study Toyota's electronics systems, and the verdict is in. There is no electronic-based cause for unintended high-speed acceleration in Toyotas." (quoting U.S. Transportation Secretary)). *But see* Andrew Smith, *Franken-Algorithms: The Deadly Consequences of Unpredictable Code*, GUARDIAN (Aug. 30, 2018), https://www.theguardian.com/technology/2018/aug/29/coding-algorithms-frankenalgos-program-danger (discussing coding errors ultimately detected in Toyota vehicles). The prospect of such undesirable outcomes is only multiplied by the risk that the millions of lines of code in an AV algorithm will contain at least some "spaghetti code," a derogatory computer science term that refers to tangled masses of programming inputs.

[187] Smith, *supra* note 186.

[188] *Selective Serotonin Reuptake Inhibitors (SSRIs)*, NHS INFORM, *supra* note 162.

[189] *Selective Serotonin Reuptake Inhibitors (SSRIs)*, MAYO CLINIC, https://www.mayoclinic.org/diseases-conditions/depression/in-depth/ssris/art-20044825 (last updated May 17. 2018); *see also St. John's Wort*, NAT'L CTR. COMPLEMENTARY & INTEGRATIVE HEALTH, NIH, https://nccih.nih.gov/health/stjohnswort/ataglance.htm (last updated Dec. 1, 2016) (describing uses of St. John's Wort and potentially dangerous interaction between herbal remedy and SSRIs).

[190] *See Selective Serotonin Reuptake Inhibitors (SSRIs)*, MAYO CLINIC, *supra* note 189.

[191] *See* Dietary Supplement Health and Education Act of 1994, Pub. L. No. 103-417 (codified as amended at scattered provisions of 21 U.S.C.).s

It is a mistake, however, to extrapolate from the FDA example to algorithmic contexts because FDA's regulatory attitude towards emergence is meaningfully distinct from AI. In the pharmaceutical context, the idea is to control the drug, such that emergent properties are failures of control. For instance, many acid reflux drugs contain a chemical compound, omeprazole, that has been shown to increase an individual's serum exposure to the SSRI.[192] 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 has historically contended with this class of drug interaction challenges through alternative methods of control, in the form of a combination of labeling requirements that mandate disclosure of known interactions[193] and reliance on doctors' counsel in prescribing medical interventions in a way that accounts for the patient's full medical history.[194]

For AI, however, there is a crucial difference: 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.[195] At times, these solutions can be *better* than the human would have predicted.[196] 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

---

[192] *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 (2015).

[193] 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 110, 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 Labeling*, Guidances (Drugs), FDA (Sept. 19, 2018), https://www.fda.gov/drugs/guidancecomplianceregulatoryinformation/guidances/ucm065010.htm. *See also* Eisenberg, *supra* note 110 (describing differences in disclosure requirements for prescription versus over-the-counter drugs).

[194] 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." *See* Full Prescribing Information, *supra* note 163, at 26. Moreover, recognizing that these methods may not be enough, FDA is presently developing additional guidance on clinical drug interactions. *See Guidance Agenda New & Revised Draft Guidances CDER Plans to Publish During Calendar Year 2018,* FDA (Jan. 19, 2018) https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM417290.pdf (reporting forthcoming "Clinical Drug Interactions Studies: Study, Design, Data Analysis, Implications for Dosing and Labeling Recommendations, Revised Draft").

[195] Lehman et al., *supra* note 2, at 5–24 (discussing "Surprise from Algorithms and Simulations," 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/ ("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.").

[196] *See* Lehman et al., *supra* note 2, at 13–17.

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 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. 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. However, coupled with the complexity and unpredictability of the technology, AI's combined policy challenges are likely to strain the capacity of such an administrative framework.

B. Collaboration and Negotiation

1. From Regulation to Governance in Environmental Law

Governance might seem to offer a contrasting public regulatory strategy. Indeed, "collaborative governance" models have emerged as an alternative to either state-driven prescriptive frameworks or wholly market-driven, deregulatory paradigms[197] in environmental law,[198] another complex and dynamic domain.

Environmental law emerged as a distinct field in the 1970s as the state expanded the set of legally protected rights.[199] 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, and a new agency, the Environmental Protection Agency (EPA).[200] 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

---

[197] *See, e.g.*, Orly Lobel, *New Governance as Regulatory Governance*, *in* THE OXFORD HANDBOOK OF GOVERNANCE 65 (David Levi-Four ed., 2012) [hereinafter Lobel, *Regulatory Governance*] (positioning new governance as a "third way vision between unregulated markets and top-down government controls"); Jody Freeman & Daniel A. Farber, *Modular Environmental Regulation*, 54 DUKE L.J. 795, 800 n.4 (2005) (Thirty-Fourth Annual Administrative Law Issue) (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); 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).

[198] 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 197, at 800 n.4.

[199] *See* CASS SUNSTEIN, AFTER THE RIGHTS REVOLUTION 1 (1990).

[200] *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).

pollution."[201] The "unprecedented" substantive reach of these statutes reflected the vast number of implicated domains[202] 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 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 hydro-carbons and carbon monoxide by 1975 and nitrogen oxides by 1976."[203] And the Clean Air Act is just one of eighteen major federal environmental protection statutes enacted in the 1970s.[204]

Yet since the early 1980s,[205] there has been a shift away from such stringent, top-down statutory command and toward "reform,"[206] "rethinking,"[207] or "reinvention"[208] of traditional regulatory approaches for environmental law.[209] 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]."[210]

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.[211] 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

[201] Richard J. Lazarus, *supra* note 200, at 78.
[202] *See* ROGER W. FINDLEY ET AL., CASES AND MATERIALS ON ENVIRONMENTAL LAW (6th ed. 2003) 1–2 (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).
[203] RICHARD J. LAZARUS, THE MAKING OF ENVIRONMENTAL LAW 70 (2008).
[204] *Id.*
[205] Detailed documentation of the phases of this revolution is outside the scope of this Article. For exposition, see *id.* at 85–99.
[206] *See, e.g.*, Bruce A. Ackerman & Richard B. Stewart, Comment, *Reforming Environmental Law*, 37 STAN. L. REV. 1333 (1985).
[207] *See, e.g.*, Fiorino, *supra* note 197.
[208] *See, e.g.*, *id.*; Farber, *supra* note 197.
[209] See sources cited *supra* note 197.
[210] Freeman & Farber, *supra* note 197, at 795–96.
[211] *See id.* at 837–76.

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."[212] 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 criteria."[213] 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. 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.[214] 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."[215] 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."[216] Faced with such dynamic and uncertain conditions, it can be incredibly challenging to predict how a particular intervention will unfold—such that in at least some instances, the result may well be surprising or downright non-intuitive. 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.

---

[212] *Id.* at 842.
[213] *Id.*
[214] Karkkainen, *supra* note 197, at 207–08 and sources cited therein (describing Chesapeake Bay Program).
[215] *Id.* at 195 (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).
[216] *Id.*

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 state partner, then collaboration is an oxymoron.[217] Any negotiation will occur in an unregulated market, without democratically-accountable coordination or enforceable checks on commercial profit motives.[218]

And for AI, there is no democratically accountable state body with which to create public-private partnerships. As discussed previously, there is presently no comprehensive national strategy for AI, nor any overarching civilian policy for AI standards development or regulatory intervention.[219] In addition, following the money underscores the extent to which the public sector has lagged the private sector in AI investment.[220] The February 2019 "American AI Initiative" does not include any lump sum funding for AI, instead directing federal funding agencies to prioritize AI investments at their own discretion.[221] Yet public expenditures thus far do not come close to the scale and scope of private-side research and development, or R&D, expenditures. According to a recent report by Stanford University's AI Index, just two firms (Amazon and Alphabet) invested a combined $30 billion in R&D in 2017. This sum is more than five times the *combined* 2019 budgets for the National Science Foundation, DARPA,[222] and DOT investments in autonomous and unmanned systems.[223] And in addition to corporate spending, consider, for example, that a single individual pledged $125 million over three years

[217] 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 Barry M. Leiner et al., *Brief History of the Internet*, INTERNET SOC. (1997), https://www.internetsociety.org/wp-content/uploads/2017/09/ISOC-History-of-the-Internet_1997.pdf. *See also 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).
[218] For a discussion of how self-regulation in AI creates coordination and oversight problems, see supra Section II.B.
[219] *See supra* text accompanying notes 77–84.
[220] 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.
[221] *See* sources cited *supra* note 86.
[222] The Defense Advanced Research Projects Agency, or DARPA, sits within the Department of Defense and also supports and funds AI research. *See Our Research*, DARPA, https://www.darpa.mil/our-research?tFilter=73&oFilter=&sort=undefined (last visited June 18, 2018).
[223] YOAV SHOHAM ET AL., THE AI INDEX 2018 ANNUAL REPORT 58 (2018) ("Private companies play the central role in AI development / investment in the U.S. In 2017, private technology companies like Amazon and Alphabet invested $16.1B and $13.9B, 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.").

for a "common sense AI" initiative, [224] and that there are scores of such initiatives[225] shifting the epicenter of AI R&D outside of the government sector.

Federal AI investments have not supported sustained basic research, instead prioritizing the military and intelligence sectors. For instance, the Department of Defense, or DOD, spent an unspecified $7.4 billion on AI in 2017[226] and has presumably made additional classified expenditures. DOD made two further notable investments in 2018: First, in April 2018, it announced a "Joint Artificial Intelligence Center" (JAIC) dedicated to AI production and prototyping.[227] Second, in September 2018, it announced a two-billion-dollar campaign to develop the "next wave" of AI technologies. Furthermore, in addition to AI research by DARPA in the Department of Defense, the Office of the Director of National Intelligence's IARPA (Intelligence Advanced Research Projects Activity) has several AI research projects.[228] Any military research might be dual-use in the sense that it could crossover to civilian applications.[229] Yet a much longer timeline is forecast for such R&D efforts,[230] 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 programmatic support that comes from an institution that is not motivated by military or intelligence concerns.

Nor do these trends seem likely to change anytime soon. Consider a new government investment in AI: the 2019 National Defense Authorization Act's allocation of up to $10 million in support of an independent executive body, the Security Commission on Artificial Intelligence. This Commission is expected to issue recommendations on "action by the executive branch and Congress related to artificial intelligence, machine learning, and associated technologies,

---

[224] The individual is Paul Allen, who launched the Allen Institute for AI, or AI2, in 2014. *See About*, AI2, https://allenai.org/about.html (last visited Dec. 4, 2018). In early 2018, he pledged another $125 million over three years for a new "common sense AI" initiative, Project Alexandria. Press Release, 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.

[225] *See* Tate Williams, *As Concern Grows, Another Philanthropy-Backed AI Watchdog Launches*, INSIDE PHILANTHROPY, https://www.insidephilanthropy.com/home/2017/1/13/as-concern-grows-another-philanthropy-backed-ai-watchdog-launches (Jan. 13, 2017) (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).

[226] GOVINI, DEPARTMENT OF DEFENSE ARTIFICIAL INTELLIGENCE, BIG DATA AND CLOUD TAXONOMY 7 (2018), *available at* https://www.govini.com/home/insights/.

[227] *Joint Office on Artificial Intelligence Announced By DOD*, FEDMANAGER (Apr. 17, 2018), https://www.fedmanager.com/featured/3015-joint-office-on-artificial-intelligence-announced-by-dod.

[228] *See, e.g.*, *Deep Intermodal Video Analytics (DIVA)*, IARPA, https://www.iarpa.gov/index.php/research-programs/diva (last visited June 18, 2018); *Integrated Cognitive-Neuroscience Architectures for Understanding Sensemaking (ICArUS)*, IARPA, https://www.iarpa.gov/index.php/research-programs/icarus (last visited June 18, 2018); *Knowledge Representation in Neural Systems* (KRNS), https://www.iarpa.gov/index.php/research-programs/krns (last visited June 18. 2018).

[229] DARPA's explainable AI work in particular might have significant civilian applications. *See Explainable Artificial Intelligence* (XAI), *supra* note 158.

[230] *See The Public Policy Challenges of Artificial Intelligence*, HARVARD IOP, (Feb. 15, 2018), http://iop.harvard.edu/forum/public-policy-challenges-artificial-intelligence (interview with head of IARPA Jason Matheny, who predicts a 10-year timeline for IARPA's AI R&D). 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. 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 that would be a direct counterpoint to actions by private firms.

including recommendations to more effectively organize the Federal Government."[231] Given the inclusion of this funding in a defense authorization bill, it is likely that the focus will on national security implications of AI—particularly because two-thirds of the Commission were chosen by congressmembers who sit on armed services and intelligence committees,[232] and who ostensibly selected individuals whom they believe will advance their institutional objectives as representatives of those governing bodies.

If money talks, then the federal government in general and its non-military officials and entities in particular have not been sufficiently forceful in their AI speech. The strategic context for AI development 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.[233] Furthermore, even where ostensibly public AI entities do exist, they continue to rely on private support both for technical assistance[234] and for policy counsel.[235] There is nothing inherently wrong with such private-side involvement. To the contrary, private expertise may be needed,[236] 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

---

[231] Tajha Chappellet-Lanier, *Alphabet, Microsoft Leaders named to National Security Commission on Artificial Intelligence*, FEDSCOOP (Nov. 15, 2018), https://www.fedscoop.com/alphabet-microsoft-leaders-named-national-security-commission-artificial-intelligence/.

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

[233] 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 . *See generally* Kaveh Waddell, *1 Big Thing: A Feud Atop AI's Commanding Heights*, AXIOS (Sept. 6, 2018), https://www.axios.com/newsletters/axios-future-f9f455cd-d89c-4406-a4b4-acca49281d12.html?chunk=0#story0.

[234] *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/ ("Following the [Project] Maven example, the military will rely mostly on contractors and third parties for its AI, and the center [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,* FOR. POL'Y (June 29, 2018, 4:11 PM), https://foreignpolicy.com/2018/06/29/google-protest-wont-stop-pentagons-a-i-revolution/ ("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.").

[235] For instance, corporate representatives form the largest bloc of the fifteen seats on the newly created Security Commission. 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.

[236] *See* Ash Carter, *Shaping Disruptive Technological Change for Public Good*, BELFER CTR. (Aug. 2018), https://www.belfercenter.org/publication/shaping-disruptive-technological-change-public-good (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/ ("AI researchers must be a part of these conversations [with governments and militaries], as their technical expertise is vital to shaping policy choices.").

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 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. The problem is that focusing on the regulatory challenges of AI in isolation leads to the wrong kinds of interventions. True, the technical attributes of AI are salient insofar as they may make a model like command and-control a poor choice. But they are not the only symptoms—and targeting them is likely to produce an incomplete cure. 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. We need interventions that more squarely contend with both the complex combination of AI's technical attributes and the strategic context of AI development.

IV. In Search of Accountability[237]

What we are presently lacking is a system that allocates decision-making authority between public and private actors in a way that remains publicly accountable. Furthermore, AI's technical attributes[238] along with the private sector's lead in AI investment[239] will make it challenging for the public sector to suddenly engage as either a regulator or a collaborative partner. And even if these and related practical issues are resolved, a critical theoretical issue remains: the way that AI technology is developed through code blurs the line between public and private governance choices.

Put simply, the role of what this Article terms *code as policy* complicates traditional public regulatory models in which the state controls or negotiates with private actors. The following two Sections offer that the code-based decisions that make algorithmic technologies like AI possible 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.[240] But it is to suggest that 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.[241] 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. And when

[237] Special thanks to Richard Re, whose comments were especially helpful in developing this Part.
[238] *See* discussion *supra* Section III.A.2.
[239] *See supra* Section III.B.2.
[240] *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))).
[241] 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 41.

it comes to AI in particular, without some form of public control over that code's development, myriad micro-level decision points that affect human safety and implicate democratic norms are decided without democratic accountability. Before we slip unwittingly into a new order of governance by private actors, we must either consciously decide that the potential substantive gains are worth it—or else affirmatively inject public input into the process of technical innovation itself.

A. Code as Policy

There must be a fundamental rethinking of what constitutes a public policy decision in the context of digital emerging technologies. AI policy turns on more than formal decisions by public officials. A version of this point holds, to be sure, in many domains. Consider a business executive's choice to move manufacturing to a different location, which might have substantial economic impacts and policy interactions. But AI's policy impact is harder to trace back to discrete, top-down business or policy choices.

For AI, technical decisions at the programming and design level carry starker regulatory implications for both human values and human safety. These choices are already regulating how AI algorithms operate in real-world applications. Consider, for instance, Google's March 2019 introduction of TensorFlow Federated, an open source AI training system that incorporates federated learning.[242] 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.[243] Instead, it maintains the information on the individual device on which the algorithm is run while reporting just the outcome back to the centralized actor.[244] Accordingly, a seemingly technical design choice such as whether to use the TensorFlow Federated template will determine how privacy-protective the algorithm will be. Alternatively, a private firm's desire to compete with Google and develop a distinct, proprietary model might lead their 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. Protecting a value (here, privacy), then, reflects more than formal, top-down policy interventions or administrative requirements. Choices about the design of algorithmic systems will effectively regulate the ways that this technology interacts with human values.

This same point applies when an algorithm is applied in an application that has an even more direct impact in the physical world.[245] Consider, for instance, the use of an algorithm for medical diagnoses. Private actors must make key choices such as which data sets are sufficiently representative to serve as the training data. And if, for instance, IBM opts to rely on Sloan-Kettering Cancer Center's data to train a ML model, then that model will embed "Sloan

---

[242] *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/.
[243] 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.
[244] Hautala, *supra* note 242.
[245] *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; 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/.

Kettering's particular philosophy about how to do medicine."[246] The line between initial decisions about how to construct the technology and that code's effect on human safety is thus ever-more blurred.[247]

Such technical 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.[248] Accordingly, developers must decide how to incorporate machine processing alongside existing human driving norms.[249] 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 the situation is less innocuous if a collision results from 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. The problem, then, is not just the technical translation of formal safety rules into code. It is also how to code a machine to make adaptations that allow the technology to interact with social norms—lest the disjuncture between the two compromise humans' physical wellbeing.

It is not possible to resolve these sorts of normative questions without invoking public values. In the context of AVs or other physical examples of AI technology, whose safety is to be maximized?[250] Should the developers protect the interests of the passenger in the autonomous vehicle, those of other drivers, or some other individual or set of people with whom the car interacts? Whether this potential for widespread physical and social effects is a feature or a bug depends on how much faith we place in private choices. The challenge for AI is that private entities may have vital knowledge and expertise to make these 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" 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

---

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

[247] Again, even if these dynamics are not novel, they are nonetheless unique. *See supra* Section III.A.2.

[248] 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*, NHTSA, https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety (last visited June 18, 2018) (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.

[249] *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; 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/.

[250] Humans may also be uncertain how to answer this question, 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. Whether this is an acceptable outcome turns on bedrock questions of the proper relationship between the state, commercial actors, and citizens. *See* discussion *infra* Section IV.B.

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."[251]

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, 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. This is true at the highest level: decisions made by AI developers today are creating constitutional principles for the technology. And it is also true in specific ways that are more akin to the impact of statutes or administrative regulations. 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 Private Governance Dilemma

The rise of private governance thus raises fundamental questions about what democratic governance demands. Given the ways that AI applications touch life in almost every sector and given the private lead in AI development, it would be a mistake to let a thousand self-regulatory flowers bloom and see what flourishes in the free market. But it would also be a mistake to assume that prescriptive public regulation alone will suffice in the long term, given private expertise, resources, and the classic "pacing problem" in regulating emerging technologies.[252] In the abstract, a collaborative governance solution does seem like the best fit.[253] And yet the public presence necessary for *accountable* collaborative governance to succeed does not presently exist. This Section explores the apparent public-private governance dilemma, offers a different way to think about this Gordian knot, and closes with some preliminary suggestions to slice the issues differently.

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 increasingly stark private governance dilemma: Despite any conceptual mismatch between prescriptive regulation and AI, is the "least worst" option for the state to clamp down and control more, lest governance get away from the public? Or must it open up and cede control to non-state actors, accepting that technology has gotten away from government? In other words, must we do more to regulate and constrain firms ex ante, despite the technical attributes and theoretical problems with this tack? Or should we step away and accept

[251] Mulligan & Bamberger, *supra* note 75, at 697.
[252] *See supra* note 132 and surrounding text.
[253] *Accord* Kaminski, *supra* note 15, at manuscript 5, 22 (discussing benefits of collaborative governance in general and its fit for algorithmic technologies in particular).

AI as a harbinger of a new order in which governance is the work of commercial entities and not the state?

Each of these options, moreover, carries a substantial cost. On the one hand, if we pursue traditional "command-and-control" solutions, then we cannot escape the pacing problem. We also risk squelching promising innovations that might improve human well-being. 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,[254] then do we necessarily want to prescribe its permissible uses?

On the other hand, a lack of public regulation or oversight means that 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 life-saving 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."[255] 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 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.

2. Recasting the Dilemma

The way out of the dilemma is to reframe it. The most auspicious reframing requires recognizing that the bedrock issues are less about control by public as opposed to commercial actors, and more about identifying the public and private contexts in which we care as much about procedural checks as we do about substantive outcomes. This reframing creates space for more precise discussions about whether we should implement an AI solution in a given setting. Framed this way, a different set of questions emerge. 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?[256] Even more

[254] Leah Kaminsky, *The Invisible Warning Signs that Predict Your Future Health*, BBC FUTURE (Jan. 17, 2019), http://www.bbc.com/future/story/20190116-the-invisible-warning-signs-that-predict-your-future-health.
[255] Matt Wood, *How To Make Software Algorithms for Health Care Fair and Equal for Everyone*, U. CHI. MED. (Dec. 3, 2018), https://www.uchicagomedicine.org/forefront/patient-care-articles/2018/december/how-to-make-software-algorithms-for-health-care-fair-and-equal-for-everyone.
[256] 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. *See About GNI*, GNI, https://globalnetworkinitiative.org/about-gni/ (last visited Jan. 23, 2019). Member companies pledge to "promote and protect freedom of expression and privacy" by following agreed-upon principles, *see The GNI Principles*, GNI, https://globalnetworkinitiative.org/gni-principles/ (last visited Jan. 23, 2019), and implementation guidelines, which

fundamentally, are there some contexts in which 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. But 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 U.S. would do well to double down on public education about AI as an initial step. Finland, for example, recently 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."[257]

But education and public discourse take time. Assuming that we are not ready to concede the end of democratic governance as we know it, it is well worth thinking creatively about structural steps to keep governance responsive to public inputs and public-minded priorities in the immediate term. Returning to 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 turn to other regulatory modalities to develop internal public checks, encoded in the very design of the technology itself.[258]

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 unsatisfying, 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 norms and regulation through the market, and to what role the state might play in shaping social norms or in fostering market developments that take public as well as commercial interests into account.

Though a full articulation of these tactics awaits future research, an initial public policy agenda might consist of two prongs targeted at, respectively, markets and norms. To bolster the public presence in AI, one clear lesson is that the American public sector must invest far more in basic AI R&D, and not merely in AI applications or "technical revolutions that could . . . create vast new wealth."[259] It is heartening that the Trump Administration designated AI R&D as a priority in the 2019 budget.[260] But a pledge to invest alone is not sufficient. First, the dollar amount

include regular corporate assessments through a pre-approved auditing procedure, *see Company Assessments*, GNI, https://globalnetworkinitiative.org/company-assessments/ (last visited Jan. 23, 2019).

[257] *See* Finnish Center for Artificial Intelligence, https://fcai.fi/ (last visited Jan. 23, 2018).

[258] 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.

[259] SUMMARY OF THE 2018 WHITE HOUSE SUMMIT ON AI, *supra* note 82, at 5.

[260] *Id.*

invested in R&D must increase to compete with the private sector.[261] 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.[262] Research investments should not be calculated in terms of their potential impact on industry and "innovation," with the assumption that we can and should "remove regulatory barriers to the deployment of AI-powered technologies."[263] 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 offer a "public option" in at least some settings, thereby ensuring that publicly-accountably actors create the technology in the first instance.[264] Though public "governance-by-design" concerns would remain because opaque technical choices could still embed values,[265] this step could nonetheless permit greater public control of the technology's development. Such a solution might be most promising in sensitive settings, such as criminal justice, where concerns with procedural justice and fundamental rights are paramount, and there are clearer, constitutionally defined minimal protections.

Short of a full public option, the state could also filter public interests into the market by offering, for example, approved public data sets.[266] Recall that the dominant AI method at present, ML, relies on access to extremely large data sets, and that data scientists' choices about what training data to use are instrumental in either mitigating or perpetuating any underlying structural biases. 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.[267] Such a move, furthermore might also be paired with an ex ante requirement that commercial actors rely on and contribute to the common public resources and/or an ex post sanction if a firm uses proprietary data and there is a subsequent safety or ethical issue that the public data could have avoided.[268] The bottom line is how to shift the marketplace, such that more than profits drive it.

To this end, a second set of "softer" interventions would work to change the culture of AI development by shaping the professional norms of those working within the industry. These steps might take several forms, from a public consortium modelled after the recently-created

---

[261] *See* discussion *supra* text accompanying notes 220–236 and sources cited therein.
[262] SUMMARY OF THE 2018 WHITE HOUSE SUMMIT ON AI, *supra* note 82, at 1.
[263] *Id.*
[264] *See* Richard R. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice* (manuscript on file with author) (discussing "public option" in criminal justice risk assessment algorithms).
[265] *See* Mulligan & Bamberger, *supra* note 75, 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.").
[266] *Cf.* ITU, *AI For Good Global Summit Report* (2017) (manuscript at 68), https://www.itu.int/en/itu-t/ai/documents/report/ai_for_good_global_summit_report_2017.pdf ("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.").
[267] For discussion of ways that 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 (2017).
[268] This solution would require further technical research to implement, given "explainability" challenges in AI. This Article reserves the question of what specific steps would be necessary, as well as questions about how to ensure the security of any such data set.

European Lab for Learning and Intelligent Systems[269] to economic or other incentives for interdisciplinary university initiatives and educational training that cultivate a unified professional ethos around AI.[270] In tandem with these efforts, the government might consider how to act as a convener for professional standard-setting efforts. At present, 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.[271] Accordingly, public-facing fora, perhaps hosted by the newly-created Select Committee on AI, could identify ML individuals as members of a unified profession who are subject to the same institutional norms. 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 for 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.[272] These and related cultural shifts might provide safeguards for a world in which code operates as policy.

---

[269] *See* Press Release, *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 ("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 Initiative to Establish a European Lab for Learning & Intelligent Systems*, ELLIS SOC., https://ellis-open-letter.eu/letter.html (last visited Dec. 14, 2018) (letter from scientists urging creation of EU-wide institute).

[270] *See, e.g.*, David Culler, Berkeley Announces Transformative New Division, UC Berkeley (Nov. 1, 2018), https://data.berkeley.edu/news/berkeley-announces-transformative-new-division; Fei-Fei Li & John Etchemendy, *Introducing Stanford's Human-Centered AI Initiative*, STAN. UNIV. (Oct. 19, 2018), https://hai.stanford.edu/news/introducing_stanfords_human_centered_ai_initiative/; Press Release, MIT Reshapes Itself to Shape the Future, MIT (Oct. 15, 2018), http://news.mit.edu/2018/mit-reshapes-itself-stephen-schwarzman-college-of-computing-1015; Byron Splice, *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. *Cf.* Paul Scharre & Michael C. Horowitz, *Congress Can Help the United States Lead in Artificial Intelligence*, FOR. POL. (Dec. 10, 2018, 5:46 PM), ("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 of students acquiring advanced STEM degrees."). To succeed, these efforts must formulate a new kind of interdisciplinary model, and not merely repackage old departmental divisions in new institutional labels.

[271] 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, e.g.*, John Gotterbarn et al., *Software Engineering Code of Ethics*, 40 COMMUN. ACM 11, 110 (1997) (joint project between ACM and IEEE); *ACM Code of Ethics*, ACM (Oct. 16, 1992), https://www.acm.org/about-acm/code-of-ethics; *Candidate Code of Ethics*, COMPTIA, https://certification.comptia.org/testing/test-policies/continuing-education-policies/candidate-code-of-ethics (last visited June 18, 2018); *IT Code of Ethic*s, SANS (Apr. 24, 2004), https://www.sans.org/security-resources/ethics. 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, IFIP, http://ifiptc9.org/ (last visited June 18, 2018). This project concluded in 1996 with the publication of a handbook, 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." 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 six working groups covering major topics in AI." IFIP TC12, IFIP, http://www.ifiptc12.org/ (last visited June 18. 2018). Such a bevy of overlapping standards arguably stymies consensus around a shared professional ethic.

[272] *Cf.* 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 (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

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 AI, 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 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 and its 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 the solution is not obvious. A domain-specific 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 R&D. Nor can 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 may not fit within traditional regulatory operating 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. Seen this way, new tactical approaches become available. Accounting for the technical architecture of AI—its code— may require filtering public input and instilling public values through alternative regulatory modalities, such as markets and norms, rather than direct mediation 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

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 *supra* 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.

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. 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.