University of Chicago Law School
Chicago Unbound

Public Law and Legal Theory Working Papers | Working Papers

2020

Toward the Democratic Regulation of AI Systems: A Prolegomenon

Mariano-Florentino Cuéllar

Aziz Z. Huq

Follow this and additional works at: https://chicagounbound.uchicago.edu/public_law_and_legal_theory

 Part of the Law Commons

Chicago Unbound includes both works in progress and final versions of articles. Please be aware that a more recent version of this article may be available on Chicago Unbound, SSRN or elsewhere.

Recommended Citation

Cuéllar, Mariano-Florentino and Huq, Aziz Z., "Toward the Democratic Regulation of AI Systems: A Prolegomenon" (2020). *Public Law and Legal Theory Working Papers*. .
https://chicagounbound.uchicago.edu/public_law_and_legal_theory/754

This Working Paper is brought to you for free and open access by the Working Papers at Chicago Unbound. It has been accepted for inclusion in Public Law and Legal Theory Working Papers by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

[Draft July 14, 2020]

Toward the Democratic Regulation of AI Systems: A Prolegomenon

Mariano-Florentino Cuéllar and Aziz Z. Huq *

I.

Where, then, is this "artificial intelligence," or "AI" of which so many speak? At a first blush, some version of AI technology — by which contemporary observers often seem to mean, at a minimum, technology relying on computing algorithms to discern patterns in data, and then trigger actions or recommendations in response[1] — seems to be everywhere. Roughly four out of every ten American adults get their news through Facebook's news-feed algorithm.[2] The algorithm often directs even users with mainstream political views to QAnon and other conspiracy theorists bent on undermining civic trust.[3] At the same time, AI is also being increasingly being used to monitor and remove material from social media platforms.[4] By one recent measure, automated "bots" generate just over half of status updates on Twitter while comprising 43 percent of all accounts.[5] This year, the Covid-19 crisis led to the postponement of the International Baccalaureate exam for high school students. Students instead received an algorithmically-predicted exam score generated from their pre-exam academic performance.[6] In medical settings, the Food and Drug Administration has approved more than 30 "AI algorithms" for clinical use on the ground that they can provide "equivalent levels of diagnostic accuracy compared with health care professionals."[7] A deep-learning tool introduced in the United Kingdom for routine mammogram screenings

* Justice, California Supreme Court, Herman Phleger Visiting Professor of Law, Stanford Law School, and affiliated scholar, Freeman Spogli Institute for International Studies at Stanford University. Frank and Bernice J. Greenberg Professor of Law and Mark C. Mamolen Teaching Scholar, University of Chicago Law School.

[1] See, e.g., DAVID FREEMAN ENGSTROM ET AL., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 16 (Feb. 2020).

[2] John Gramlich, *10 facts about Americans and Facebook*, PEW RESEARCH CENTER (May 16, 2019), https://tinyurl.com/y4apu58j.

[3] Julia Carrie Wong, *Down the Rabbit Hole: How QAnon Conspiracies Thrive on Facebook*, GUARDIAN (June 25, 2020), https://www.theguardian.com/technology/2020/jun/25/qanon-facebook-conspiracy-theories-algorithm/

[4] TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 97-110 (2018)

[5] Zafar Gilani et al., *A large-scale behavioural analysis of bots and humans on twitter*, 13 ACM TRANSACTIONS ON THE WEB 1, 10 (2019). For a more skeptical view of the prevalence of bots, see Siobhan Roberts, *Who's a Bot? Who's Not?*, N.Y. TIMES (June 16, 2020), https://tinyurl.com/ybwvbyhp.

[6] Andrew Jack, *Students and Teachers Hit at International Baccalaureate Grading*, FIN. TIMES (Jul. 9, 2020)(critics of the algorithm insist it's already produced "really appalling injustices").

[7] Xiaoxuan Liu et al., *A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis*, 1 LANCET DIGITAL HEALTH e271, e271 (2019).

Electronic copy available at: https://ssrn.com/abstract=3671011

improves on the accuracy of human screening by roughly 10 percent by one measure.[8] In government, some 64 different bodies within the civilian wings of the national government employing 157 "AI/ML" tools.[9] Nor does the federal government have a monopoly on these tools. Starting with the Los Angeles Police Department, some 60 police forces around the country have adopted 'Predpol,' a controversial predictive tool that is designed to mine historical crime data to identify where crimes will occur in the subsequent 12 hours.[10] In the financial markets, some large trading platforms have long been dominated by AI. A 2017 study found that some 80 percent of Forex trades are executed by algorithm.[11] The proportion is likely higher now. Finally, there is the home: The AI "Siri" is in active use on more than a half billion devices globally.[12] In the United States, as of 2019 some 69 percent of U.S. homes used "smart" devices, such as home networking, home security, smart thermostats, smart lighting, or video doorbells.[13]

Like dew forming after cool, clear evenings, AI can seem both ubiquitous and elusive. Its apparent pervasiveness in both private and public hands is apt to breed confusion and concern. For one thing, there is a fair amount of uncertainty among the public, policymakers, and even scholars about what basic terms such as "bot" and "AI" even mean. Our reading suggests that the term "bot" is used promiscuously to criticize not only entirely automated producers of social media posts, but even humans using technologies to amplify the size of their audience or its level of interest, without clear distinctions about the underlying technologies used.[14] Our experience reading about and writing on "artificial intelligence" also leaves us with the impression of parallel confusion about that phrase. For decades people in the private sector, government, and academia have worked on the project of understanding and harnessing intelligence by creating techniques and applications with at least some of the properties we associate with human intelligence. As this project has matured into a range of different technologies, some better understood than others, the term "artificial intelligence" has become increasingly difficult to pin down. It could, for example, refer to the continuing project of simulating intelligence, to analytical techniques that constitute the building blocks for specific applications of AI in applied settings, or to computing systems (whether instantiated in software or hardware, or in networked or stand-alone systems) that deploy AI analytical techniques to achieve particular functions.

Alternately a science fiction storyline, a tech industry buzzword, and a catchall referent for any and all technologies that appear to mimic some element of human reasoning (whether they do

---

[8] Scott Mayer McKinney, et al., *International evaluation of an AI system for breast cancer screening*, 577 NATURE 89, 90 (2020).

[9] Engstrom et al., *supra* note 1.

[10] Mark Puente, *LAPD pioneered predicting crime with data. Many police don't think it works*, L.A. TIMES (July 3, 2019), https://tinyurl.com/y8rb2mtt.

[11] Alessandro Bigiotti and Alfredo Navarra, *Optimizing Automated Trading Systems*, in THE 2018 INTERNATIONAL CONFERENCE ON DIGITAL SCIENCE 254, 254 (2018).

[12] *HomePod arrives February 9, Available to Order this Friday*, APPLE (Jan. 23, 2018), https://www.apple.com/newsroom/2018/01/homepod-arrives-february-9-available-to-order-this-friday/

[13] Chuck Martin, *Smart Home Technology Hits 69% Penetration in U.S.,* MEDIAPOST (Sept. 30, 2019), https://www.mediapost.com/publications/article/341320/smart-home-technology-hits-69-penetration-in-us.html/

[14] Darius Kazemi, *The Bot Scare* (Dec. 31, 2019), https://tinysubversions.com/notes/the-bot-scare/.

Electronic copy available at: https://ssrn.com/abstract=3671011

so or not), the term AI hence yields in common parlance a fractious and motley bunch of applications. As a result, it can hide as much as it clarifies, and so it can easily disserve careful democratic debate. Uncertain if we're imagining a superintelligence beyond human reckoning, a software object that "learns" a personality, or a spruced-up OLS regression, we are at peril of losing sight of what (if anything) is distinctively at stake in recent technology change if we aren't even talking about the same thing. And there is another risk: Faced with a rapid pace of change in both technology and its use by society, a public vocabulary for technology that is too imprecise or woolly at the edges creates a serious risk that we will fail to perceive qualitative changes that do raise serious concerns, or that we will misperceive and thus miss their importance, or that we will just overreact to changes that are in fact minor and uninteresting.[15]

Yet even recognizing this ardent confusion, it is hard to deny that a gaggle of troubling concerns with AI now exist that are plainly are worth thinking through—and perhaps meeting with responsive democratic action. Scholars in law, information technology, and sociology have been active on this point. During the past five years, a lively cottage industry has emerged to condemn the effect of new technologies (including, but not limited to AI) in terms of democracy, power (both economic and political), race, and liberal notions of individual autonomy. Just as the definition of AI is occasionally sketched with a cloudy and imprecise line, so the normative case against its adoption is often painted with broad, evocative brushstrokes instead of pointillist precision. But even the impressionistic can land with force. Hence, in 2015, Frank Pasquale warned that "authority is increasingly expressed algorithmically,"[16] while Barnard Harcourt cautioned against becoming "dulled to the perils of digital transparency," a risk that remained "largely invisible to democratic theory and practice."[17] More recently, Shoshona Zuboff condemned tech and social media companies having "scraped, torn, and taken" the very stuff of "human nature" itself.[18] Focusing on a different cluster of equality-related concerns, Ruha Benjamin has sounded an alarm about "biased bots, altruistic algorithms, and their many coded cousins" that produce what she calls "coded inequity" or the "new Jim Code."[19] And Carl Benedikt Frey has explored the possibility that up to 47 percent of American jobs could be "susceptible to automation" thanks to advances in AI.[20] Each of these critiques picks up on an important normative concern. But not all of them are defined with the clarity needful to pursue an effective treatment.[21]

---

[15] For a bracing fictional account of technological change that dramatizes this problem, see Ted Chiang, *The Lifecycle of Software Objects*, in EXHALATION 62 (2019).

[16] FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 8 (2015).

[17] BERNARD E. HARCOURT, EXPOSED: DESIRE AND DISOBEDIENCE IN THE DIGITAL AGE 19, 261 (2015).

[18] SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 94 (2018)

[19] RUHA BENJAMIN, RACE AFTER TECHNOLOGY 7 (2019).

[20] CARL BENEDIKT FREY, THE TECHNOLOGY TRAP: CAPITAL, LABOR, AND POWER IN THE AGE OF AUTOMATION 320-21 (2018).

[21] For criticisms along these lines, see Mariano-Florentino Cuéllar and Aziz Z. Huq, *Economies of Surveillance*, 133 HARV. L. REV. 1280 (2019), and Aziz Z. Huq, *Apps in Black and White*, 61 EURO. J. SOCIOLOGY – (forthcoming 2020).

Electronic copy available at: https://ssrn.com/abstract=3671011

Surprisingly absent from these treatments, moreover, is a serious engagement with the question of how at an effective public response would proceed. Harcourt, at one extreme, seems to shrug off the very possibility of democratic intervention entirely in favor of individual efforts to "diminish our own visibility" and just "encrypt better."[22] Zuboff, meanwhile, looks to "law" as a remedy but doesn't fill in any details of what effective regulation might look like.[23] To immerse oneself in this literature, therefore, is to be overwhelmed by a sense of moral and political dissolution without a commensurate remedy in view.

We are not ready to conclude that that the project of democratic regulation of AI should be abandoned in anticipation of its failure. We acknowledge difficult threshold questions of what counts as a democratic arrangement in the first instance.[24] But let us put those to one side, and assume we're talking about "democracy" as that word is used in the demotic to capture the form of governance practiced at least since the Second Reconstruction in the United States. Democracies identified as such in common parlance often are deeply flawed by economic inequality, party polarization, and polluted public sphere, and the like. And they all face seemingly overwhelming challenges—from geostrategic competitors in autocracies, economic catastrophes, pathogens, or internal atrophy. All too often, even a legitimately democratic response will be confused, incomplete, regressive, tardy, or even evasive. Democracies, if they prevail, do so by muddling through, rather than rushing to clasp quick triumphs.[25]

So Harcourt may be right to associate at least the initial the democratic reaction to crisis with "apathy," "complacency," and even "despotism."[26] But this is no cause to abandon the project of collective, democratic regulation altogether, and too optimistic about the prospects for some vague sort of cyber-libertarian utopia. At the same time, we think that it is not enough to call for new "law" as Zuboff does without thinking carefully about the frictions and constraints on the actual implementing institutions that we have to hand. We fear that in the absence of a careful analysis of how democratic regulation of AI might proceed the most likely outcome will be governance through private, corporate instruments, such as Facebook's "Supreme Court." However great our respect for certain members of that body, its substitution of democratic regulation by corporate simulacra operating beyond the shadow of law raises for us some tricky normative questions about the legitimacy and political economy of regulatory power in the digital technologies space.[27]

---

[22] HARCOURT, *supra* note 17, at 270.

[23] ZUBOFF, *supra* note 18, at 486.

[24] For a parsimonious definition focused on the institutional preconditions of democracy, see TOM GINSBURG AND AZIZ Z. HUQ, HOW TO SAVE A CONSTITUTIONAL DEMOCRACY 9-10 (2018).

[25] *See* DAVID RUNCIMAN, THE DEMOCRACY TRAP: A HISTORY OF DEMOCRACY IN CRISIS FROM 1914 TO THE PRESENT (2013).

[26] HARCOURT, *supra* note 17, at 258.

[27] It is, to be sure, also nothing entirely new. There is a long history of "government and firms competed over marginal spaces-- namely, those domains that could conceivably be subject to greater or lesser state coercive regulation (or private market ordering)." Jon D. Michaels, *We the Shareholders: Government Market Participation in the Postliberal U.S. Political Economy*, 120 Colum. L. Rev. 465, 473 (2020).

Electronic copy available at: https://ssrn.com/abstract=3671011

The space left unfilled in the literature in our view calls for clear thinking about what it means for a democracy to regulate AI in the first instance. We think that an initial purchase on the challenge of regulating AI by taking two steps. First, we should begin with a careful definition of what is being regulated. In contrast to the relatively narrow attention to the technical details of computational tools, we suggest that it is more useful to identify "AI systems" as the appropriate object of regulation. Further, a democratic regulation of these systems should examine primarily how they embed a forward-looking 'policy,' rather than the deontologically flavored question whether they violate 'rights,' say to privacy or non-discrimination. Policy rather than rights provides a more tractable and useful object of inquiry—one that draws attention to the important questions of how AI systems alter the distribution of resources and respect.

Second, to get a sense of the challenges that the project of democratic regulation of AI faces, we must canvas its enemies. We distinguish here between two 'flavors' of obstacle, which demand subtly different evaluations. On the one hand, there are *institutional* impediments to an effectual democratic response. These sound in the register of political economy, and the path-dependent states of regulatory potential embodied at the national and substate levels. On the other hand, there are *ontological* impediment. By this, we mean to capture the sense—stressed by critics such as Zuboff—that AI systems can be constitutive of human subjectivity in ways that make the very project of identifying democratic preferences incoherent, or at least subject to subversion. Here, we draw attention to the endogeneity of democratic preferences and institutions to the design and operation of AI in public life.

In sum, our contribution here is best thought of as ground-clearing. We don't set here a program for the democratic regulation of AI. More modestly, our ambition is to light a pathway toward that laudable, and even essential, goal.

II.

What does it mean to regulate "*AI*"? It is useful to begin with a clear sense of the appropriate object of democratic concern. How a problem is conceptualized matters greatly to how it is addressed. For example, there's a big difference between talking about a "war on poverty" and a "war on crime," and the jump from one to another has been very consequential.[28] Here, we propose the term "AI system" as the appropriate unit of analysis and regulation. At least when used in any civic context, we define an AI system as *a sociotechnical embodiment of public policy codified in an appropriate computational learning tool and embedded in a specific institutional context*. The key terms here, which we will carefully define and then gloss, are "system" and "policy." Both draw attention to opportunities for democratic regulation that are to date underappreciated. In contrast, a careful reader will notice that our definition embeds a measure of ambiguity in respect to the precise range of computational tools at stake. This isn't an oversight on our part. Let's first defend (or at least explain!) our ambiguity, before spinning out why we think terms "system" and "policy" are handy analytic tools.

1. "AI" as Moving Target

[28] As shown in ELIZABETH HINTON, FROM THE WAR ON POVERTY TO THE WAR ON CRIME (2018).

Electronic copy available at: https://ssrn.com/abstract=3671011

To begin with, we are concerned with a range of computational tools that generally share a 'family resemblance,' rather than being strictly defined in terms of a set of functions or outcomes.[29] As many others have pointed out, the term AI does not map onto a strictly defined set of characteristics; even more precise terms, such as "machine learning" allow for some ambiguity. Very colloquially, Hannah Fry has usefully suggested that the term AI be used when "[y]ou give the machine data, a goal and feedback when it's on the right track – and leave it to work out the best way of achieving the end."[30] In effect, these instruments operate as "incredibly skilled mimics, finding correlations and responding to novel inputs as if to say, 'This reminds me of . . . ,' and in doing so imitate successful strategies gleaned from a large collection of examples."[31] For our purpose, Fry's nicely stated account of a "computational learning tool" is sufficiently precise to provide traction without inducing needless confusion over technical details.

One reason for being cautious about being too precise about a term such as "AI," or even the more technical sounding idea of 'machine learning,' is the breakneck pace of technological innovation and social change in the use of technology. On the one hand, a simple machine learning tool is akin to the sort of ordinary least squares regression that many encountered in college. But at the other end of the technological spectrum are tools such as reinforcement learning and the use of synthesized rather than historical data.[32] Reinforcement learning, for example, entails learning to solve a task by trial and error by interacting with the environment and receiving rewards for successful interactions.[33] At the same time, rapid change continues in the public uptake of software with applications ranging from mapping to ride-sharing to dating to real-time feedback based on physical reactions. Not all such applications are generally assumed by the public to incorporate AI and some line-drawing might be necessary even under our rubric — yet another example of how technologies become so commonplace that it becomes difficult to remember people once would have considered to embody a capacity to perform functions that, for humans, require intelligence. With this in mind, we think it does not make sense to define "AI systems" strictly in terms of a specific technical form, rather than the more informal definition of a computational learning tool as described by Fry.

Still, even this colloquial approach allows us to flag a number of common features shared by the relevant set of computational tools deployed in the early twenty-first century. Not every tool has every one of the following qualities; but all have a few. First, these instruments often rely on a set of 'training data' that can be analyzed to gauge how different variables relate to one another. In some cases, this is historical data, such as past medical records, the past crime data for a municipality, or the list of Internet searches typed in by a given population. Alternatively, an AI

---

[29] Hence, a leading textbook offering a series of alternative definitions of AI that encompass simply thinking and acting humanly as well as rationally. STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 2-14 (3d ed. 2013).

[30] HANNAH FRY, HELLO WORLD: BEING HUMAN IN THE AGE OF ALGORITHMS 11 (2019); *see also* JERRY KAPLAN, ARTIFICIAL INTELLIGENCE: WHAT EVERYONE NEEDS TO KNOW 32 (2016) (providing a similar colloquial description).

[31] KAPLAN, *supra* note 30, at 32,

[32] On the latter, see Lei Xu et al., *Information security in big data: privacy and data mining*, 2 IEEE ACCESS 1149, 1155 (2014).

[33] Robert Moni, *Reinforcement Learning algorithms — an intuitive overview*, MEDIUM (Feb. 18, 2019), https://medium.com/@SmartLabAI/reinforcement-learning-algorithms-an-intuitive-overview-904e2dff5bbc.

Electronic copy available at: https://ssrn.com/abstract=3671011

instrument can also generate its own training data through repeatedly attempting a task, such as playing a game like Chess or Go.[34] Second, the instrument is tasked with developing a model that can be used to estimate an outcome variable based on a set of inputs. In constructing this model, the instrument will be asked to follow a cost function or a reward function, which defines the sort of inference that the machine should make. For example, an instrument might be asked to construct a model of the relationship between past employment history, demographic details, and the likelihood of success as a teacher, minimizing the rate of false positives but also tolerating a higher rate of false negatives. The resulting model can be predictive—in the sense of offering inferences for events that have not happened—or descriptive—in the sense of drawing to human attention correlations or relationships that would have otherwise gone unnoticed. Third, the model is applied to new, 'out-of-sample' data that is not part of the training data-set.[35] Here is the essence of 'learning' being applied. Today, such systems appear recondite; soon, they will not just be common-place, but, more importantly, will cease to be perceived as 'technologies' fraught with ethical and social implications at all.

With these common features in hand, we think it is possible to presume– at least for present purposes, in most currently-relevant applied settings — such systems are sufficiently similar in terms of the legal and policy problems that they pose.

2. "AI Systems"

More important than the technical details of an AI tool narrowly-defined, in our view, is the institutional setting of its adoption. In key respects, AI is a "general purpose technology" much like electricity or the transistor.[36] Like other general purpose technologies, AI is necessarily adopted and integrated into the design and operation of freestanding contexts. An AI instrument of the sort we have just described never stands in isolation. Rather, it is almost always embedded in a specific institutional context that is, to a greater or lesser degree, the object of conscious design or perhaps Burkean evolution. It contains "affordances": this refers to a set of "fundamental properties that determine just how the thing could possibly be used."[37] But it can also contain pernicious "disaffordances"—for instance when a person with dark skin cannot trigger an automatic soap dispenser because of the calibration of the light sensor.[38] Design in general embeds judgments, conscious or not. about how users interact with the tool, how decision-makers with relevant legal authority and expertise receive an instrument's outputs, and whether opportunities exist for revising or second-guessing that output. In contrast, an approach that takes an AI tool in isolation as "a technical and self-contained object that exists as a distinct presence is likely to be a

[34] Tom Simonite, *This More Powerful Version of AlphaGo Learns on Its Own*, WIRED (Oct. 18, 2017), https://www.wired.com/story/this-more-powerful-version-of-alphago-learns-on-its-own [https://perma.cc/L38N-D94H].

[35] Sendhil Mullainathan & Jann Spiess, *Machine learning: an applied econometric approach*, 31 J. ECON. PERSP. 87, 88 (2017) (defining machine learning in terms of its capacity for "out of sample" prediction); *see generally* ETHEM ALPAYDIN, MACHINE LEARNING: THE NEW AI (2016).

[36] FREY, *supra* note 20, at 305.

[37] DONALD A. NORMAN, THE PSYCHOLOGY OF EVERYDAY THINGS 8 (1988).

[38] SASHA COSTANZA-CHOCK, DESIGN JUSTICE: COMMUNITY-LED PRACTICES TO BUILD THE WORLDS WE NEED 45-45 (2020).

Electronic copy available at: https://ssrn.com/abstract=3671011

mistake."[39] It is far better to recognized their embedded quality—in part to appreciate the complex normative choices that go into that embedding, and in part to perceive opportunities for regulatory intervention that otherwise might go unnoticed.

Let's make this more concrete with an example. In 2013, the governor of Michigan Rick Snyder introduced an algorithmic tool called Midas to detect fraudulent applications for unemployment benefits as part of a larger overhaul of the information technology by the state.[40] This AI tool was introduced as part of a conscious policy—remember that word!—of austerity on Governor Snyder's, a sort of junior-varsity 'starve the beast.'[41] Within the first years of adoption, the system had racked up a denial rate of 93 percent, all the while falsely accusing 40,000 Michigan residents of fraudulently claiming benefits. Until the spike of unemployment claims associated with the Covid-19 pandemic, the state benefits agency also employed only 12 people to resolve and correct fraud allegations. Even as the pandemic accelerated, calls to the agency would result in applicants being connected not with a state employee, but with another benefit claimant who had been denied. Claimants who allegedly have been wrongly denied a benefit report calling state office more than a thousand times a day—and still not being able to get through.[42] The state of Michigan thus chose to provide a user interface with relatively limited opportunities for submitting information, and relatively few opportunities for revision or correction after the fact. Whatever the formal status of an instrument's predictions as advisory or only presumptively valid, the institutional context of the Michigan algorithm made its predictions de facto binding for tens of thousands of people and all but guaranteed an exorbitant false positive rate when it came to fraud detection.

We think it is sensible not to look solely at the technical specifications of the Midas algorithm to evaluate its consequences. Instead, we think it is necessary to view the instrument as entangled in a specific institutional context to understand and evaluate its consequences. That is, we must look at *AI systems* and not just instruments in splendid isolation. This inquiry is necessarily *sociotechnical* in character insofar as it demands attention not just at choices embedded in code, but also to the range and nature of affordances and interactions between an instrument and human actors at both the front end and the back end.

AI systems widely vary in their design. But the systems of interest here tend to have certain distinctive characteristics. We can pick out five that strike us as particularly important for the project of democratic regulation. *First*, AI these systems are at least ostensibly designed to add either private or social value by facilitating decisions or operations in particular settings. Facebook's feed algorithm, for example, advances the private value of increasing people's engagement with the social network. The Midas algorithm is intended to advance the social value

---

[39] David Beer, *The Social Power of Algorithms*, 20 INFO. COMM. & SOC. 1. 4 (2017).

[40] Robert N. Charette, *Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold*, IEEE SPECTRUM, Jan. 24, 2018, https://spectrum.ieee.org/riskfactor/computing/software/michigans-midas-unemployment-system-algorithm-alchemy-that-created-lead-not-gold [https://perma.cc/ZLZ9-T29S].

[41] Lee Saunders, *Government Didn't Fail Flint, Austerity Did*, GOVERNING, Feb. 16, 2016, https://www.governing.com/gov-institute/voices/col-flint-water-austerity-public-services.html

[42] Taylor Desormeau, *Michigan Ineffective Unemployment System is Nothing New*, GOVERNING, June 17, 2020, https://www.governing.com/work/Michigans-Ineffective-Unemployment-System-Is-Nothing-New.html.

Electronic copy available at: https://ssrn.com/abstract=3671011

that the Snyder Administration associated with a winnowing of the social state. Clearly, in both cases it is possible to contest whether the instrument is 'really' advancing a social value. Nevertheless, public authorities and corporate actors who create AI systems commonly appeal to these gains when justifying the elimination of human discretion.

Second, many applied settings where AI systems are embedded involve collective decision-making. Social networks such as Facebook require decisions about what will be jointly discussed and debated by people on the site. The focus of shared discussion are made with algorithmic assistance. (Interestingly, and relevant to our discussion below in Part IV, Facebook users seem to underestimate the exist, and even the existence, of such AI nudging[43]). The risk prediction instruments used in pretrial bail and sentencing context can be thought of as devices for pooling the information and coordinating the collective inputs of probation officers, prosecutors, and judges. When AI systems will operate in such settings, they will change the nature—and likely the outcomes—of collective processes. Many of those processes are supposed to produce democratic outputs. Facebook and like social-media platforms, for example, can be conceptualized as part of the public sphere in which "society engage[s] in critical public debate."[44] Yet it seems the outcomes of such debate will be necessarily endogenous to the operation of their algorithmic arrangements.

Third, an AI tool will generally include a user interface designed to abstract analytical conclusions and facilitate interaction. Just as Facebook has a particular site architecture, so the Midas tool used in Michigan deployed a particular visual matrix to gather information from applicants and a different one to display its outputs to state officials. The design of such interfaces will commonly entail some normatively freighted choices. Among other things, for instance, the interface's designer can try to leverage her knowledge of behavioral psychology to 'nudge' a user in ways that either facilitate interaction to even push toward a particular outcome. It is hence not just the *content* of the instrument but also its *context* that will "shape organisation, institutional, commercial and governmental decision-making."[45] A risk assessment tool used in a criminal justice setting, for example, might supply judges with a simple numerical score. That score might distil information about a plurality of risks, including the possibility of violent crime, non-violent crime, and flight from the jurisdiction. Some of these risks might be more amenable to prediction than others.[46] The instrument then must array that information in terms of a scale—say, from 1 to 10, or 1 to 100, or the letters "A" through "D." The choices between the different ways in which risk can be represented—will it be cardinal or ordinal? Will it foreground one sort of risk, violent or nonviolent, or try to aggregate together different risks—are all consequential. These choices, it should be emphasized, are between various permutations of an interface—and not simply

[43] Blake Hallinan, Jed R. Brubaker, and Casey Fiesler, *Unexpected expectations: Public reaction to the Facebook emotional contagion study*, 22 NEW MEDIA & SOC. 1076 (2019).

[44] JÜRGEN HABERMAS. THE STRUCTURAL TRANSFORMATION OF THE PUBLIC SPHERE: AN INQUIRY INTO A CATEGORY OF BOURGEOIS SOCIETY 52 (Thomas Burger trans., 1991).

[45] Beer, *supra* note 39, at 5.

[46] For example, there is an argument that the prediction is violence is infeasible. *Technical Flaws of Pretrial Risk Assessment Tools Raise Grave Concerns* 2 (2019), https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf.

Electronic copy available at: https://ssrn.com/abstract=3671011

questions about technical elements of an algorithmic tool, such as the choice of outcome variable, but also decisions about the manner in which predictive outputs are presented.

Fourth, it will often be the case that neither the interface nor output is likely to supply the information necessary for someone with technical expertise to evaluate performance or to facilitate comparisons to some sort of 'ground truth.' Rather, choices must be made about how 'transparent' the operation of an instrument is to those who rely upon its output, and indeed what kind of 'transparency' is desirable.[47] This nuanced choice about transparency—or, perhaps better, between different forms of transparency—involves a rather subtle layered principal-agent problem that can emerges here in the public context: Officials in the judicial system, as in the Michigan welfare bureaucracy, are delegated authority by the people to execute the law, but then have delegated out from under them a set of policy choices embodied in code and interface design. The management of these agency-cost problems must be addressed through the design of an institutional context.

Finally, it is increasingly the case that AI systems have, or are presented as having, some capacity to adapt and improve performance over time. In that regard, they differ from the application of a static and predefined categorization rubric or unchanging statistical function to a set of data. The possibility of perpetual recalibration is presently illustrated vividly by commercial applications such as Google's PageRank algorithm. This search tool is subject to "an iterative process of feedback and change to accommodate the shifting environments" and users' changing needs.[48] As reinforcement learning tools are increasingly adopted, such continual refinement will itself become automated, and likely more common. The possibility of dynamic adaption, or algorithmic updating, in turn opens the horizon of questions as to what precisely such a process of adaptation maximizes, and whether such adaptation should extend not only to the pursuit of particular goals but to the definition of those goals. Search engines such as Google, for example, have been criticized because search engines have at times generated results that reflect racist associations.[49] When the PageRank algorithm is adjusted to change these outcomes, Google is introducing a (laudable) antiracist normative consideration into the algorithm's dynamic design.

But what does this actually mean in practice? There is not obviously one and one way of mitigating racial bias in an algorithm's operation.[50] What if the social forms of such bias change over time? Rectifying for racial bias in search results means having some conceptualization of what counts as 'bias' at a given moment in time, and also some account of what a 'neutral' search result looks like. Where a search engines is operating on a textual corpus that likely embodies and reflects the biases embedded in common human behavior and speech, this may be no easy task. Indeed, it is even possible for an anti-racism modification to operate in normatively troubling ways. Consider the (related) example of Twitter's use of a machine learning tool to identify and block hate speech and abusive speech. These instruments in operation show "substantial racial

[47] For debates about the meaning of the term, see Tim Miller, *Explanation in artificial intelligence: Insights from the social sciences*. 267 ARTIFICIAL INTEL. 1, 1-2 (2019), and Michael Gleicher, *A Framework for Considering Comprehensibility Modeling*, 4 BIG DATA 75, 77–84 (2016).

[48] Michele Willson, *Algorithms (and the) Everyday*, 20 INFO. COMM. & SOC. 137, 142 (2017).

[49] SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM 66-80 (2018).

[50] Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L. J. 1043 (2019) [hereinafter "Huq, *Racial Equity*"].

Electronic copy available at: https://ssrn.com/abstract=3671011

bias" in that they are far more likely to flag the speech of African-American Twitter users than white users for blocking or other penalties.[51]

To be clear, these five common features—the promotion of social or private value; the integration of AI tools into ongoing processes of collective decision-making; the ubiquity of value-laden user interfaces; the resolution of layered principle-agent problems via choices between different kinds of transparency; and the calibration of dynamic updating—are not the only design margins imaginable. But all involve choices about institutional context that bear heavily on the impact that an instrument will have on human outcomes. We emphasize them here, in addition, because they are also points of leverage for democratic regulators—points of leverage that would be elided or lost if one were to focus more narrowly on an AI instrument standing in isolation.

3. AI Systems as Embodied Policy

Having identified the appropriate object of regulation, we need to decide what it is that a democratic system should focus upon when intervening in AI systems. A tempting answer is the pronation of individual 'rights,' say to privacy, nondiscrimination, and due process. We don't want to deny that thinking about the interaction between AI systems and rights is a useful goal.[52] But we think that there is another useful lens.

The distinction between "rights" and "policy" is set forth in a crisp form by Karen Orren and Stephen Skowronek. They define a policy as a "commitment to a designated goal or course of action, made authoritatively on behalf of a given entity or collectivity, and accompanied by guidelines for its accomplishment."[53] In contrast they define rights as "claims that the person, inside or outside of government, may make on the action or person or another, enforceable in a court of law."[54] A focus on rights draws attention to discrete, perhaps even binary relationships between distinct and identifiable people. It is discrete, interpersonal, and largely blinded to larger context. In contrast, a focus on policy draws attention to the manner in which an assemblage of actors, nested within an institution, produce either a specified goal or course of action, or possibly an unintended set of consequences.

A focus on policy rather than rights as those terms are defined by Orren and Skowronek has a number of benefits. Not least, the governance of AI systems is not well pursued through the management of binary interpersonal relations. Changes to a reward function or an interface, for example, are almost certain to have complex and plural effects. Efforts to reduce rates of false negatives, for example, are mathematically certain to change the rate (and the distribution) of false positives.[55] As has long been apparent, rights—especially when enforced by courts—are not an

[51] Thomas Davidson, Debasmita Bhattacharya, and Ingmar Weber, *Racial bias in hate speech and abusive language detection datasets*, ARXIV PREPRINT ARXIV:1905.12516 (2019).

[52] And it would hypocritical to do so, at least for one of us. Aziz Z. Huq, *Litigating Constitutional Rights in the Machine Learning State*, 106 CORNELL L. REV. – (forthcoming 2020).

[53] KAREN ORREN AND STEPHEN SKOWRONEK, THE POLICY STATE 27 (2017).

[54] *Id.* at 29.

[55] Huq, *Racial Equity*, *supra* note 50 (discussing these complexity).

Electronic copy available at: https://ssrn.com/abstract=3671011

ideal vehicle for managing what Lon Fuller called polycentric disputes.[56] There's every reason to think Fuller's worries have equal force in this novel technological context.

Moreover, the manner in which normative concerns about equality, privacy, and due process arise out of AI systems is not well captured by rights on its own. As we have suggested, the technical choices of algorithmic design and also their embedding in institutional consequences can entail contestable normative judgments. The manner in which predictions are reported, the feasibility of verifying the basis for predictions, and the nature of any dynamic updating all depend on normative judgments as much as the choice of training data and reward function. Worse, technical judgments (say, about what reward function is used) can be entangled in complex ways with system design choices (say, the manner in which predictions are expressed in a user interface). Picking out a single thread of interaction between the state and an individual as a 'right' may not even be sensible—let alone practically effective. Rather, the effects of an AI system are often spread out across aggregations who experience a classification, rather than concentrated on individuals. At the margin, the size of those effects will also depend on the prior institutional and policy landscape in place when an AI system is adopted.

Let's unpack that a bit. It is by now familiar fare that historical training data might reflect implicit or explicit biases. Related but subtly distinct questions can also arise about whether prediction is appropriate and whether some form of correction should be made. Where a reward function optimizes for a certain goal, challenges can be lodged as to how to appropriately capture and operationalize social or private value while the accounting of externalities. Think here about the manner in which Facebook's ambition to maximize the time that users spend on the site while managing the spread of 'fake news' via QAnon and like groups. Yet it is less commonly noted that normative judgments are also embedded in the material, institutional context in which an AI tool is implemented. The Midas system, for example, embodied a normative view of the social state in terms of when and how individual could submit information and then seek reconsideration.[57] Governor Snyder expressed his normative view about the public good not just in the calibration of the threshold for benefits denials, but also through his decisions about the staffing of the agency, and the manner in which such reconsiderations would proceed. Trying to cleanly separate technical choices from institutional context—or the marginal effect of adopting Midas—doesn't seem a smart move. And while the Midas system was successfully challenged on due process grounds,[58] that challenge does not really convey the profound worry many had: That the Snyder administration was pursuing an unpopular and perhaps morally indefensible policy of 'starving the beat,' a policy that became only less attractive in an age of pandemic.

As a consequence, to capture these normative judgments, we think it is appropriate to ask what *policy* an AI *system* embodies. What can we glean from the architecture and features of the system (or even the incentives affecting its designers, controllers, and users) about the suite of effects the system appears intended or likely to sow in the world? And what class of effects is it likely to achieve, independent of what its (perhaps dimly-seeing) designers might have intended?

---

[56] Lon L. Fuller, *The forms and limits of adjudication*, 92 HARV. L. REV. 353 (1978).

[57] For a critical look at the right to seek a do-over as one that is far more complicated than commonly supposed, See Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611 (2020).

[58] Zynda v. Arwood, 175 F. Supp. 3d 791, 799 (E.D. Mich. 2016).

Electronic copy available at: https://ssrn.com/abstract=3671011

III.

What then does it mean to *regulate* AI systems? We get a start on that question by canvassing the considerable hurdles a project of democratic regulation confronts. These come in two different flavors: institutional and ontological. The existing literature pays a good deal of attention to the latter, but we think it is insufficiently attentive to the former. Perhaps this is because there's a certain drudgery in slogging through the mundane and technocratic details of designing regulatory environments—but we think the exercise is worthwhile, and even essential. Our focus here on the barriers to effective regulation should not be taken as skepticism about the project of bringing AI systems to heel on democratic terms. We instead aim to clarify the stakes of embarking on this task.

1. Institutional Barriers

The institutional barriers to effective regulation arise from an interaction between the common qualities of AI systems on the one hand, and the institutional capacity of the American federal and state governments on the other. We think they are significant and generally underappreciated.

Law has frequently confronted situations where technological change occurs rapidly. Think of the first years of Moore's Law, the early history of aircraft, or even electrification. But changes in the enabling technologies for AI systems and the underlying analytical techniques are occurring with exceptional speed — particularly the growth in available computing power at a manageable cost is particularly rapid.[59] Increasing the epistemic burden of regulation, the general utility of AI as a technology means it is likely to be adopted widely by a varied range of both private and public actors, and then applied to rather different ends. Much of the relevant technological innovation, in addition, happens in the private sector because of "declining investment in basic and foundational research, combined with lack of access to computational resources and large datasets."[60] This is in stark contrast to the development of Cold War technologies such as nuclear power and remote sensing. AI instruments are in consequence likely to be nested within larger corporate operations—whether it be a social network or a home security system—in ways that make it difficult to know whether or how an instrument is changing outcomes at the margin. Intellectual property rights often shield the results from public scrutiny. And even if more readily examined by the public or elected representatives, AI instruments will often remain opaque in operation. The harm from an AI system is quite unlike in salience the harm from, say, unlawful police brutality—although notice that the first might lead to the second by framing the underlying policy task in a certain way. An AI system such as the Facebook feed algorithm or the Midas tool distributes results across large populations. It will often be difficult to infer from any one case whether there are *systemic* problems with the tool. Patterns of false positives and false negatives, for instance, can only be discerned by looking across aggregates, in

[59] *See* THE 2019 AI INDEX REPORT, https://hai.stanford.edu/research/ai-index-2019.

[60] John Etchemendy and Fei-Fei Li, *National Research Cloud: Ensuring the Continuation of American Innovation*, HAI BLOG, March 20, 2020, https://live-stanford-hai.pantheonsite.io/blog/national-research-cloud-ensuring-continuation-american-innovation.

Electronic copy available at: https://ssrn.com/abstract=3671011

lieu of individual cases. Even then, we lack a common metric for judging how much inaccuracy, or what kinds of racial and gender imbalances, are unacceptable.[61]

Precisely because AI is a general-purpose technology likely to be widely adopted to discrete ends across the economic and social landscape, it is simply infeasible to cleave off its regulation into a special-purpose regulatory vehicle.[62] To the contrary, grappling with AI systems, whether as a matter of internal organization or as a matter of externally oriented regulation, will be an inescapable obligation at both the state and federal level for courts of general jurisdiction, administrative agencies tasked with the provision of social services, regulatory bodies and attorney general offices, and chief executives. In any case, the idea of a single, centralized regulator with wide-ranging power over a new, general-purpose technology doesn't seem a good one either from a political-economy, a historical, or even a constitutional one. As the pandemic has so forcefully reminded us, the United States remains a defiantly decentralized federal state, one beset by profound and at times disabling coordination problems and ensnared in paralyzing partisan polarization. Despite the revival in historical scholarship of the idea of a "strong" American state,[63] we think it would be a mistake to assume strength in this particular regulatory domain.

Once we recognize that the regulation of AI systems will be inevitably dispersed, we face yet other difficulty. Regulation requires internal expertise both on the technical and sociotechnical elements of AI systems. At a time of immense fiscal strain—again, a consequence exacerbated by the pandemic—this may be hard for state and local bodies in particular to acquire. (That said, a recent study found that a majority (53 percent) of the AI applications in nonmilitary use at the federal level were "the product of in-house efforts."[64]) Government bodies already use AI systems in quite varied ways to "prioritize enforcement … engage with the public [and] conduct regulatory research, analysis, and monitoring."[65] At the same time, government decisions about how to employ and how to regulate AI systems occur in a political pressure-cooker. Corporate vendors and lobbies applying great pressuring toward adoption of certain technologies, while pressing hard against tighter regulation from the Beltway. Elizabeth Joh has thus mapped what she plausibly terms the "undue influence" surveillance and analytic technology manufacturers have on police departments.[66] In Congress, Pawel Popiel has found, tech companies such as Google and Twitter were "among the top 20 biggest lobbying industries, spending nearly as much as the defense and telecommunications in 2017, and outspending the commercial banking sector."[67] Captured agencies, revolving doors (in both stripes of administration), and iron triangles—the full

---

[61] Huq, *Racial Equity*, *supra* note 50.

[62] The best version of this argument is to be found in Andrew Tutt, *An FDA for algorithms*, 69 ADMIN. L. REV. 83 (2017).

[63] For a brilliant treatment, see William J. Novak, *The Myth of the "Weak" American State*, 113 AM. HIST. REV. 752, 762 (2008).

[64] ENGSTROM ET AL., *supra* note 1, at 7.

[65] *Id.* at 17.

[66] Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies in Policing*, 92 N.Y.U. L. REV. ONLINE 19, 20 (2017).

[67] Pawel Popiel, *The Tech Lobby: Tracing the Contours of New Media Elite Lobbying Power*, 11 COMM. CULTURE & CRITIQUE 566, 572 (2018),

Electronic copy available at: https://ssrn.com/abstract=3671011

complement of cynical metaphors—are all appropriate. Finally, at the federal level, regulatory options are constrained by the perception that China is obtaining a geopolitical edge through research in AI and quantum computing.[68] In an influential book, Kai-Fu Lee argues that China's comparative advantage over the United States in access to "abundant data, hungry entrepreneurs, AI scientists, and an AI-friendly policy environment" will provide it with an edge in ginning up and then harnessing AI systems to domestic and foreign policy ends.[69] Particularly when it comes to military and public-security adoptions, the perception of geostrategic competition can lead to "calls for restraint, reflection, and regulation as a strategic disadvantage to U.S. national interests."[70]

Finally, there is a troublesome paradox embedded in the project of regulating AI systems. An implication of our analysis here is that a more robust set of state institutions is a necessary step toward the regulation of AI systems. Even if one doesn't credit the direst accounts of AI's impact on society and the individual, even if one thinks that AI presents merely an 'ordinary' problem of regulation, still a *stronger* state is required. Yet it is equally clear that the stronger the state's capacity to deploy and control AI systems, the greater the threat of state overreaching that pinches important human interests. Authoritarian states, such as China, for example, have deployed a range of AI tools from facial recognition to content moderation of social media postings to suppress political dissent and maintain ideological conformity.[71] That is, there is an extent to which the threats from private control of technology *trade off* against the threats from an AI-empowered state.

Of course, the paradox dissolves if the state is well-regulated and tightly leashed by the rule of law. But states do not choose when to confront AI systems. The timing of their confrontation is rather determined by the pace of technological change and diffusion. For nations such as China, AI has by tragic fortuity arisen at an opportune time for those wishing to consolidate one-party rule. It thus entrenches authoritarian elements of the Chinese regime. On the other hand, where a regime is generally well-ordered and leashed by clear, public laws, AI system regulation is likely to be feasible without any subtle threat to democratic values. The force of the paradox thus turns on the ex ante quality of democratic control of the state through rule-of-law mechanisms.

And the U.S.? We leave the reader to decide where the United States falls in the ensuing spectrum of possibilities as a useful exercise in self-criticism.

2. Ontological Barriers

Profound as these challenges might seem, they might seem at first blush to pale in comparison to a more foundational critique leveled by social critics of AI system. This critique, rather than the institutional argument, has formed the crux of the case against new digital

---

[68] Mariano-Florentino Cuéllar & Aziz Z. Huq, *Privacy's Political Economy and the State of Machine Learning: An Essay in Honor of Stephen J. Schulhofer*, __ N.Y.U. ANN. SURV. AM. L. (forthcoming 2020).

[69] KAI-FU LEE, AI SUPERPOWERS: CHINA, SILICON VALLEY, AND THE NEW WORLD ORDER 14-15 (2018).

[70] AI NOW 2019 REPORT, Dec. 2019, at 43, https://ainowinstitute.org/AI_Now_2019_Report.pdf.

[71] MARGARET E. ROBERTS, CENSORED: DISTRACTION AND DIVERSION INSIDE CHINA'S GREAT FIREWALL (2018); Paul Mozur, *Inside China's Dystopian Dreams: A.I., Shame, and Lots of Cameras*, N.Y. TIMES, July 8, 2018.

Electronic copy available at: https://ssrn.com/abstract=3671011

technologies such as AI in the existing literature. It has been given eloquent voice, for example, by Shoshana Zuboff. She argues that digital technologies enable the acquisition of intimate, private information about people, and then the exploitation of that information to manipulate their preferences and behavior. She hence warns of the "assertion of decision rights over the expropriation of human experience," and prophesizes "the dispossession of human experience" through "datafication."[72] She describes in particular three broad strategies of "tuning," "herding," and "conditioning" through which human agency is undermined.[73] A variant on this critiques focuses specifically on the shaping of political preferences through behavioral advertising.[74] Philip Howard, for example, examines spending by the "Vote Leave" campaign prior to the Brexit referendum in the United Kingdom, and estimates that such advertising may have changed the votes of eight million people and elicited time or money contributions from 800,000 people.[75] The clear implication of Howard's work is a shadow cast on a putatively democratic referendum.

We call these 'ontological' challenges because they go to very basic assumptions about the reality of human agency and action—assumptions that seem necessary for a democracy to be a meaningful goal. Yet we think these are not as severe as critics have made them out to be. Even if AI systems can shape preferences, the ensuing effects do not undermine the possibility of democratic control. The ontological challenge, therefore, may be more manageable than the institutional one.

AI systems, on this account, present a challenge because they assail the building blocks of individual agency that lie at the base of a democratic and inclusive political order. This concern might be theorized in two different ways. First, the worry might concern the fabrication of inauthentic preferences. Howard, for example, flags Vote Leave's "consistent and simple political lies."[76] If AI systems are especially good as eliciting false beliefs, perhaps their availability undermines the very possibility of individual agency as a positive good: If people are easily duped, then democracy in particular doesn't seem a good idea[77] (and notice that the same might be said of consumer choice in free markets). The second version of the argument might would focus on power rather than autonomy. It diagnoses the problem in terms of the ability of those few who possess the means of technological production to shape the preferences of the many wanting access to digital products and services. This is a problem of power, not liberal autonomy. It is a matter of who calibrates the Overton window and sets the menu of public-policy options. It is a matter of

[72] ZUBOFF, *supra* note 18, at 128, 233-34.

[73] *Id*. at 294-96.

[74] Online behavioral advertising involves monitoring people's actions online and then showing them individually targeted ads tailored on the basis of those actions. Sophie C., Boerman, Sanne Kruikemeier, and Frederik J. Zuiderveen Borgesius, *Online behavioral advertising: A literature review and research agenda*, 46 J. ADVERTISING 363, 363 (2017).

[75] PHILIP N. HOWARD, LIE MACHINES: HOW TO SAVE DEMOCRACY FROM TROLL ARMIES, DECEITFUL ROBOTS, JINK NEWS OPERATIONS, AND POLITICAL OPERATIVES 128-29 (2020).

[76] *Id*. at 123.

[77] Indeed, this is Jason Brennan's position. JASON BRENNAN, AGAINST DEMOCRACY 170 (2017) (arguing that "democracy systematically violates the competency principle").

Electronic copy available at: https://ssrn.com/abstract=3671011

how AI enables a "producing and refining informational persons who are subject to the operations of fastening."[78]

Neither of these lines of criticism are groundless, but at the same time we should be careful before lumping influence together with control, or assuming that the powers exercised by AI systems wholly usurp the possibility of independent judgment. To be sure, many private applications of AI operate on a terrain of preconscious and half-glimpsed emotional states. Famously, Facebook has run A/B experiments on its users to show the effects of changes in the balance of positive and negative news in its newsfeed.[79] But, as we have argued elsewhere, the more stark arguments about "dispossession" or "expropriation" miss the mark.[80] Rather, we think that Marion Fourcade and Daniel Kluttz capture the problematic better when they argue that the acquisition of personal data through web-based interactions but on an exploitation of social structures of trust, consent, and gift giving.[81] Social networks, and other applications that generate large volumes of consumer data, exploit a "natural compulsion to reciprocate" and "existing solidaristic bond[s]" to generate a circulation-system of interaction ripe with personal data to be harvest. Fourcade and Kluttz persuasively argue that the economic logic underneath the 'big data' economy is one that relies on a subtle form of emotional manipulation. In this regard, it is not qualitatively distinct from earlier forms of private and public governance that "constructs individuals who are capable of choice and action, shapes them as active subject, and seeks to align their choices with the objectives of governing authorities."[82]

More generally, we should distinguish between the classificatory pressure imposed by state AI systems and the "apparatuses of security" that actively and physically coerce.[83] As scholars such as James Scott and Colin Koopman have amply demonstrated, states and private actors have used schemes of classification and knowledge-managements as instruments of control long before the transistor was a thing. Koopman, for example, has recently developed a compelling account of the federal Old Age Insurance (OIA) system—today, social security—as an early "big data" problem that entailed "a massive information harvesting that depended on physically going door to door," and the technological breakthrough entailed in the use of "automated record keeping for millions of workers in the form of punching holes into cards."[84] Systems of "datafication" (to use Koopman's neologism) can certainly embed and obscure normative projects in troubling ways. Michigan's Midas system is an example of a seemingly technocratic intervention that hid a deregulatory, neoliberal ambition. Machine learning in the criminal justice context that are touted as devices of 'reform' may on the ground end up working as means for recapitulating old

---

[78] COLIN KOOPMAN, HOW WE BECAME OUR DATA: A GENEALOGY OF THE INFORMATIONAL PERSON 12 (2019).

[79] Adam Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, *Experimental evidence of massive-scale emotional contagion through social networks*, 111 PROC. NAT'L ACA. SCI. 8788 (2014).

[80] ZUBOFF, *supra* note 18, at 126, 233; *see* Huq & Cuéllar, *supra* note 21.

[81] Marion Fourcade and Daniel N. Kluttz, *A Maussian bargain: Accumulation by gift in the digital economy*, 7 BIG DATA & SOC. 1, 10 (2020).

[82] David Garland, *'Governmentality' and the problem of crime: Foucault, criminology, sociology*, 1 THEO. CRIM. 173, 175 (1997).

[83] Michael Foucault, *Governmentality*, *in* POWER 201, 220 (James D. Faubion, ed., 1997).

[84] KOOPMAN, *supra* note 78, at 59-61.

Electronic copy available at: https://ssrn.com/abstract=3671011

prejudices about communities and individuals in new, less facially offensive forms. Democracy is no doubt degraded by some AI systems. Hence, the Facebook feed algorithm's tendency to promote QAnon conspiracies is a troubling example of technological skewing of the public sphere.[85] But it does not undermine a user's ability to seek out a range of other news sources to evaluate the veracity of what they have found. AI systems' more important effect may be the mystification of policy choices than the vitiation of human agency.

So yes: AI systems shape preferences and skew political choices, just as they become objects of democratic regulation. But this circularity does not distinguish them from a long line of public and private procedures of governance. There is, in any case, no immaculate 'ground truth' of individual preference that exists in isolation, wholly immured from social, state, or market pressure. Even if Google and Facebook want your attention, Amazon wants you to buy more, and politicians want you to believe their lies—and, yes, they all do—we still should not simply assume that their projects will overwhelm a fragile human subjectivity even as we maintain a critical scrutiny of the obscurantist and manipulative elements of their projects.[86] Rather, we need to think critically about the mutuality, and entanglement, of democratic subjectivity and the democratic project of taming AI systems.

IV.

No silver bullet will meet the institutional and ontological challenges to regulating AI systems. Most solutions instead will necessarily be local or at least contextual, tailored to the specifics of different institutions and environments. A state judicial system is not going to approach these challenges in the same way as the Federal Trade Commission. Precisely how that is one will vary per local context. Nevertheless, we think that there are certain common ambitions that might usefully link the different versions of the project of democratic regulation of AI systems regardless of their institutional and historical context. Across the board, advancing that project requires some dissolving of institutional and ontological barriers to democratic regulation.

As a foundational matter, we agree with Harcourt that there is a need to empower individuals, although we think that term is more nuanced and tricky than he seems to believe.[87] Where he sees empowerment in large part as a cultivation by the few of the "art of not being governed,"[88] we would urge a search for ways in which the wider rank and file of citizens can better understand the moral and political choices embedded not just in code but in the design choices of AI systems. Rather than searching for exit routes for some, we would ask how to educate and empower individuals and thus invite mobilizations within and around AI systems. Rather than facilitating opt outs, we would search for ways to deepen public understanding of AI systems as

[85] *See* Wong, *supra* note 3.

[86] Indeed, we think that many of the criticisms that are made of libertarian paternalism can be usefully transferred to the context of AI systems. *See, e.g.,* Christopher McCrudden and Jeff King, *The dark side of nudging: the ethics, political economy, and law of libertarian paternalism*, in CHOICE ARCHITECTURE IN DEMOCRACIES, EXPLORING THE LEGITIMACY OF NUDGING 75 (Alexandra Kemmerer et al, eds. 2015).

[87] HARCOURT, *supra* note 17, at 270.

[88] *Cf.* JAMES C. SCOTT, THE ART OF NOT BEING GOVERNED: AN ANARCHIST HISTORY OF UPLAND SOUTHEAST ASIA (2010).

Electronic copy available at: https://ssrn.com/abstract=3671011

embedded policy, and ways to facilitate public mobilizations to challenge and reexamine them. This means empowering as many users as possible, and then giving them platforms to coordinate responses, so as to influence and even change the policies and values embedded in those systems, whether adopted in the public and the private sphere.

Not only does such education and mobilization address the ontological objections raised to AI systems, they also help mitigate the institutional ones. Where local institutions are under pressure from engaged parts of the public, they are more likely to take inclusive and ethnically defensible choices about the scope and operation of AI systems. The need for education and empowerment, moreover, is likely to grow over time. AI systems are likely to be integrated with increasing frequency and seamlessness into core institutional elements of our democracy such as education and the electoral process. This means we must anticipate—and start to theorize now—an infrastructure of democracy in which individuals continue to have opportunities for individual and collective action against the policies embedded in AI systems.

Central to the public empowerment necessary to counteracts both kinds of barriers is the social movement capable of frontally challenging policies surreptitiously advanced by AI systems. Some of these have bubbled up from inside the tech sector. Some have emerged in resistance to it. In the past two years, employees at firms such as Google and Facebook have used the tools of collective action, such as walk-outs, secondary boycotts, and go-slows, to challenge the way in which new technologies were being used. Movements like #TechWontBuildIt and #KeepFamiliesTogether have used the opportunities supplied by social media platforms as a way of raising the political morality of certain uses of new digital technologies.[89] These labor-aligned mobilizations have had an impact disproportionate to their numbers because of the "privileged position" that "[d]esigners, developers, and technologists occupy."[90] They are not confined to the industry itself. In the wake of the George Floyd killing, for example, a group of some 1,400 mathematicians issued a letter urging a boycott of work for PredPol and other vendors of predictive policing. The letter argued that "given the structural racism and brutality in US policing," that profession involvement by mathematicians made it "simply too easy to create a 'scientific' veneer for racism."[91]

Further, there are examples that do not rely on the leveraging of occupational or epistemic advantage. Organizations such as the Movement for Socially Useful Production, the Appropriate Technology Movement, and the People's Science Movement, are all examples of initiatives aligned with social movements that have aimed to repurpose technology to generate a "fairer and more sustainable" world.[92] In the context of the Covid-19 pandemic, a number of housing justice

[89] For a useful survey, see Haydn Belfield, *Activism by the AI community: Analysing recent achievements and future prospects*, PROC. AAAI/ACM CONF. ON AI, ETHICS, & SOC. (2020).

[90] COSTANZA-CHOCK, *supra* note 38, at 216,

[91] Davide Castelvecchi, *Mathematicians urge colleagues to boycott police work in wake of killings*, NATURE, June 19, 2020.

[92] ADRIAN SMITH ET AL. GRASSROOTS INNOVATION MOVEMENTS 3-5 (2016).

Electronic copy available at: https://ssrn.com/abstract=3671011

collectives have also used new computational tools to "compile data on landlords and speculators to embolden the work of housing justice organizing."[93]

Law, to be clear, will rarely be the first mover in these efforts. But still it may well have a role since these movement often entail collective action that may or may not be vulnerable to legal sanction as a matter of contract law or criminal law. We think it is useful to ask how law can create opportunities or barriers to useful collective action, especially in light of the successes and failures of the movements we have flagged. Noncompete agreements and trade secrets can impede legitimate public debate, for example, and law appropriately accounts for these costs. It is also important to ask which platforms provide the best affordances for education and mobilization. Certification systems and government procurement decisions can elicit better rather than worse choices in this respect.

Consider just one example of a legal intervention that might usefully facilitate education and empowerment: transparency and benchmarking mandates for AI systems. These should focus not on 'how they work,' but on 'what they do.' The law should treat AI systems as complex embodiments of policy, that is, and think carefully about how to air the ways in which they distribute entitlements (as in the case of the Michigan Midas system), coercion (as in the case of bail and sentencing instruments), or human attention and understanding (as in the Facebook news feed). Rather than asking directly or only about privacy, due process, or equality—legal concepts that may need some work before they can fit well into the new digital landscape[94]—the law should instead attend to how adding an AI system to an existing institution will either entrench or unravel hierarchical relationships between persons and groups. Law, for instance, might bring to the surface the many important intertemporal trade-offs implied in many AI systems—but particularly in social media context. Just as education in a democracy implicates an interplay of choice and constraint, perhaps certain core features of AI—including how it is used in social media to shape access to information, and over time, tastes and identity—necessarily implicate an interplay of choice and constraint. The law should look for ways to push users to consider actively how they want to change (or don't want to change) over time as they engage with a technology. In particular, the law can draw attention to how a digital technology will change preferences and behaviors over time.

Finally, a focus on the policies embedded in AI systems invites a sustained inquiry into the distributive and dynamic effects of those systems. Just as state and federal administrative law demands attention to whether regulations are justified in cost-benefit terms, so the regulation of AI should account for their effects on power and stratification. Sasha Costanza-Chock has, for example, proposed that designers "systematically interrogate how the unequal distribution of user experiences might be structured by the user's position within the intersecting fields of race, class gender, and disability."[95] Their claim can be generalized as a call to account carefully for the manner in which an AI system changes both the relative and the absolute position of groups and

[93] Erin McElroy, Meredith Whittaker, and Genevieve Fried, *COVID-19 Crisis Capitalism Comes to Real Estate*, BOSTON REV., May 7, 2020.

[94] For a discussion, see Huq, *Litigating, supra* note 52.

[95] COSTANZA-CHOCK, *supra* note 38, at 59.

Electronic copy available at: https://ssrn.com/abstract=3671011

individuals who are already subject to some form of structural disadvantage.[96] In the governmental context, such consideration might be built into the procurement or certification process for new state algorithmic interventions. This is a step toward, although not a substitute for, a public ethics of the AI system attuned to its potential enervating effects on democracy, and its aggravating effects on pernicious hierarchies such as those of race, gender, and class.

Finally, a sustained focus on those dynamic effects can best serve the larger projects of democracy when it also considers how AI systems can enliven civic possibilities. AI systems can lower barriers to expert, technical knowledge that can help the public navigate intricate bureaucratic procedures, or make the most of access to courts without a lawyer, or obtain access to educational opportunities that might otherwise be limited to more affluent families. At the core of most regulatory problems for intellectually honest citizens and democratic policymakers is the reality that life itself is at risk from the very world that sustains it.[97] Whether they're sculpting flows of information, deepening habits of thought, or disrupting the privileged role of technical experts even as they enrich vast companies, AI systems raise compelling regulatory questions in a democracy for an analogous reason: because in principle they can enrich civic life as easily as they can erode it. For democratic societies, the resulting regulatory challenge encompasses not only familiar technical questions about matters such as the sensitivity of certain private actors to legal sanctions, but also the broad themes we've sketched out that call for a candid awareness of how AI systems can continually shape the very process through which democratic societies decide choose their priorities in a contentious world.

Conclusion

None of this will be easy. But our goal here has neither been to sugarcoat a bitter pill or to discourage the project of democratic regulation of AI. Democracy is a product not only of societal norms and culture, but of law and institutions—including AI systems. At the same time as they become a new object of regulation, AI systems shape democratic preferences and processes. Basic cultural and social assumptions shift. In democracies, conflicts about AI regulation will transcend technical disagreements; they will also provoke value-laden disagreements about how society regulates values and practices that may initially seem to implicate "private" decisions about speech or association that obviously shape what we value.[98]

There will be no 'Sputnik moment' in this dialectical process of understanding and shaping how AI systems rearticulate what we value, and how we shape our institutions and systems (including AI ones) in response. The fight to find a satisfying and defensible equilibrium with AI will be long and difficult, with no clear end. Law will, however, play a pivotal, nuanced role. In that enterprise, nuanced, pragmatic judgments of institutional capacity and effectual and intelligent public mobilization are necessary. Else, the worst fears of technological skeptics are apt to be realized.

---

[96] *Id.* at 193-94.

[97] *See* Mariano-Florentino Cuéllar and Jerry Mashaw, *Regulatory Decision-Making and Economic Analysis,* Stanford Law and Economics Olin Working Paper No. 525 (2018).

[98] *See*, e.g., Lawrence Lessig, *The Regulation of Social Meaning*, U. CHI. L. REV. 943 (1995).

Electronic copy available at: https://ssrn.com/abstract=3671011