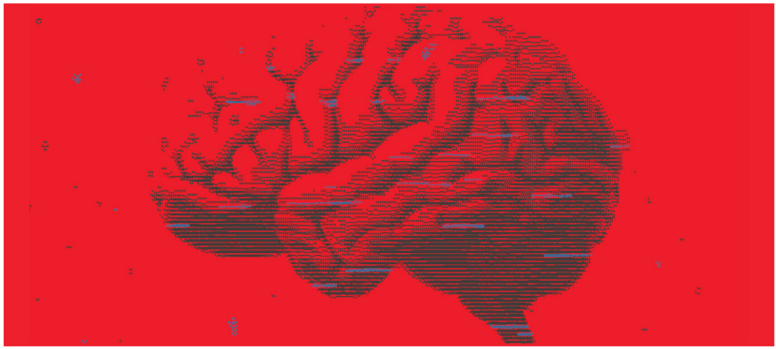
DATA AND DEMOCRACY



The Democratic Regulation of Artificial Intelligence

By Mariano-Florentino Cuéllar and Aziz Z. Huq

KNIGHT
FIRST AMENDMENT
INSTITUTE at
COLUMBIA UNIVERSITY

In October 2020, the Knight First Amendment Institute at Columbia University convened a virtual symposium, titled "Data and Democracy," to investigate how technological advances relating to the collection, analysis, and manipulation of data are affecting democratic processes, and how the law must adapt to ensure the conditions for self-government. This symposium was organized by the Institute's 2019-2020 Senior Visiting Research Scholar, Yale Law Professor Amy Kapczynski, and co-sponsored by the Law and Political Economy Project at Yale Law School.

The essays in this series were originally presented and discussed at this two-day event. Written by scholars and experts in law, computer science, information studies, political science, and other disciplines, the essays focus on three areas that are both central to democratic governance and directly affected by advancing technologies and ever-increasing data collection: 1) public opinion formation and access to information; 2) the formation and exercise of public power; and 3) the political economy of data.

The symposium was conceptualized by Knight Institute staff, including Jameel Jaffer, Executive Director; Katy Glenn Bass, Research Director; Amy Kapczynski, Senior Visiting Research Scholar; Alex Abdo, Litigation Director; and Larry Siems, Chief of Staff. The essay series was edited by Glenn Bass with additional support from Lorraine Kenny, Communications Director; A. Adam Glenn, Writer/Editor; Madeline Wood, Research Coordinator; Chand Rajendra-Nicolucci, Research Fellow; Kushal Dev, Research Fellow; and Ashwin Pillai, Intern.

The full series is available at knightcolumbia.org/research/

ABSTRACT

This essay explores the challenge of regulating AI—one of the most rapidly developing technologies, and one that simultaneously impacts both the forms and the subjects of legal regulation. Our aim is to offer a flexible and general framework for the doctrinal and normative analysis of AI. In contrast to the relatively narrow attention to the technical details of computational tools, we suggest that it is most useful to start with the concept of an "AI system" as the appropriate object of regulatory attention. We then argue that sensible regulation of such systems in democracies should examine primarily how they embed a forward-looking "policy" rather than the deontologically flavored question of whether they violate "rights," say to privacy or nondiscrimination. Next, to get a sense of the challenges that the project of democratic regulation of AI faces, we canvas the practical and theoretical obstacles that regulators confront. We distinguish here between two kinds of hurdles, 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. On the other, there are ontological impediments. By this, we mean to capture the sense 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. In closing, we draw attention to the endogeneity of democratic preferences and institutions to the design and operation of AI in public life.

INTRODUCTION

WHAT, THEN, IS THE phenomenon of "artificial intelligence" or "AI," of which so many speak? Contemporary observers often seem to mean, at a minimum, a form of technology relying on computing algorithms to discern patterns in data, and then trigger actions or recommendations in response.[1] So defined, some version of AI technology seems to be everywhere. Roughly 4 out of every 10 American adults get their news through Facebook's newsfeed algorithm.[2] This algorithm often directs even users with mainstream political views to groups managed by QAnon and other conspiracy theorists bent on undermining civic trust and sowing violent strife.[3] At the same time, AI is also increasingly being used for content moderation, which involves monitoring and removing material from social media platforms.[4] Content on those platforms is also increasingly a product of machines. By one recent measure, automated "bots" generate just over half of status updates on Twitter while comprising 43 percent of all accounts.[5] AI is also proliferating beyond the web. Last 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 claims to improve 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 employ 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, AIs play an increasingly dominant role.[11] Finally, there is the home: The AI "Siri" is in active use on more than a half billion devices globally.[12] And 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]

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" mean. Our reading, for instance, 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, all without clear distinctions about the underlying technologies used.[14]

Our experience reading about and writing on "artificial intelligence" reinforces the impression that scholars, policymakers, and members of the lay public rarely converge on a straightforward definition of the 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 so or not as a technical matter), the term AI in common parlance yields up a fractious and motley bunch of applications. In consequence, its usage can hide as much as it clarifies. As a result, use of the term can easily disserve careful democratic debate. 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] The "hype and promise" with which AI technologies are rolled out may in some cases imply capacities that extend "far past the current methodological capabilities."[16]

Such hype and uncertainty imbue discussions of AI-related legal and policy questions with a nebulous quality. Still, thoughtful observers would be hard-pressed to deny the troubling concerns raised by the reliance on AI systems in public and private bodies. These give rise to questions not just about specific regulatory responses but also about the larger structure and theory of democratic regulation of AI: How, for example, should democracies respond to the diffusion of AI in the private sector? Should they adopt AI themselves in response to private uses of AI? Should they aim instead to bar AI in both public and private domains? And are there some public functions that cannot be assigned to a machine as opposed to a human being?

Scholars in law, information technology, and sociology have been active on these questions. 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 an impressionistic argument can land with a punch. So in 2015, Frank Pasquale powerfully warned that "authority is increasingly expressed algorithmically,"[17] while Bernard Harcourt cautioned against becoming "dulled to the perils of digital transparency," a risk that remained "largely invisible to democratic theory and practice."[18] More recently, Shoshana Zuboff condemned tech and social media companies having "scraped, torn, and taken" the very stuff of "human nature" itself.[19] 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."[20] 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.[21] Each of these critiques picks up on an important normative concern. Yet not all of them are defined with the precision or clarity needful to pursue an effective treatment.[22]

Surprisingly absent from these critiques, moreover, is any serious engagement with the question of how an effective public response would be formalized or implemented. Harcourt, at one extreme, seems to shrug off casually the very possibility of democratic intervention entirely in favor of individual efforts to "diminish our own visibility" and just "encrypt better."[23] Zuboff, meanwhile, looks to an underspecified "law" as a remedy but doesn't fill in any details of what effective regulation might look like, or whence it might come.[24] Other commentators provide little further clarity. To immerse oneself in this literature is to be overwhelmed by a sense of moral and political dissolution without a commensurate remedy in view. It is to be brushed with a vague sense of democracy's inadequacy without any analysis of how democracy functions or founders.

We are far from persuaded 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.[25] 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 in the United States, at least since the civil rights movement opened the franchise across the color line.

Democracies today, as identified in common parlance, are deeply marred by economic inequality, party polarization, polluted public sphere, and the like. They all face seemingly overwhelming challenges—from geostrategic competitors in autocracies, from economic catastrophes, from viral pathogens, or from internal atrophy of political cultures. 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.[26]

In consequence, Harcourt may well be right to associate at least the initial democratic reaction to crisis with "apathy," "complacency," and even "despotism."[27] But this is no cause to abandon the project of collective, democratic regulation altogether; it is also no reason for caution-to-the-winds optimism about the prospects for some vague sort of anarchic cyberlibertarian 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 on hand. We fear that in the absence of a careful analysis of how democratic regulation of AI might proceed on the ground, 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 democratically created law raises for us difficult normative questions about the legitimacy and political economy of private regulatory power in the digital technologies space.[28]

The space left unfilled in the literature in our view thus calls for clear thinking about what it means for a democracy to regulate AI in the first instance. Because policymaking in this domain can affect so many interests and can benefit from such a broad range of analytical tools, we think an initial purchase on the challenge of regulating AI can be best achieved by isolating two distinct facets of the problem. First, we need to 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 inflected question of whether they violate "rights," say to privacy or nondiscrimination. 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 kinds of obstacles, 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 subnational levels. On the other hand, policymakers and the public also encounter *ontological* impediments. 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.

Finally, we ground our analysis by discussing two case studies concerning AI systems with very different relationships to democratic self-government. We consider first the algorithmic regulation of content on social media platforms, and then look at the use of wearable medical technology that collects and analyzes physiological data. Both these examples yield some insight into how democratic regulation might proceed, what resources exist in current law, what trade-offs exist, and what challenges are rooted in political economy. Both are illustrative, and not meant as exhaustive accounts of democratic regulation.

Along with our broader discussion of the prospects for democratic regulation of AI systems, these two case studies illustrate certain themes—such as the importance of institutions like federalism, and the relevance of existing bodies of public and private law—likely to be especially relevant in fashioning a sensible response to AI in democracies. That said, our primary goal is to clear the ground for further analysis of how and why democracies can go about regulating AI systems by defining key concepts and understanding some recurring tensions likely to affect the enterprise. By focusing more on offering tentative sketches, rather than firm predictions or conclusions,

we acknowledge both the fluid nature and complexity of the interactions between democracy, artificial intelligence, and regulation. But even these preliminary sketches make clear that prudent, carefully calibrated regulation of AI systems is possible and often desirable. We hope that our survey of the terrain helps chart a path toward that laudable, and even essential, goal.

I. THE OBJECT OF AI REGULATION

WHAT DOES IT MEAN to regulate AI, whether it be engaged in content moderation or health risk assessment? 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 the solution to urban malaise as a "war on poverty" and as a "war on crime." The jump from one to another during the 1970s and 1980s proved highly consequential.[29] 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. Both attend closely to the role that law plays (both positively and negatively) in the construction of economic and technical systems.[30] 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 try to explain) our ambiguity, before spinning out why we think terms "system" and "policy" are handy analytical tools.

A. AI as moving target

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.[31] 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 proposed 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."[32] In effect, such instruments work 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."[33] For our purpose, Fry's nicely stated account of a "computational learning tool" is sufficiently precise and open-ended to provide traction without inducing needless confusion over technical details.

One reason for caution about being too precise with 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 readers will have encountered in college. At the other end of the technological spectrum are tools such as reinforcement learning and the use of synthesized rather than historical data.[34] Reinforcement learning entails learning to solve a task by trial and error, interacting with the environment, and receiving rewards for successful interactions.[35] At the same time, rapid change continues in the public uptake of software, with applications including mapping, ride-sharing, dating, and real-time feedback based on physical reactions. Not all such applications are generally assumed by the public to incorporate AI. Some line-drawing might be necessary even under our rubric. It is also worth remembering that technologies sometimes become so commonplace that it can be difficult to remember that people performed the same tasks and thought that it was impossible for a machine to act as a substitute (think of Amazon's recommender algorithm or dating apps). 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.[36]

Still, even this colloquial definition allows us to flag a number of common features shared by the relevant set of computational tools deployed in the early 21st 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, past crime data for a municipality, or the list of internet searches typed in by a given population. Alternatively, an AI instrument can also generate its own training data through repeatedly attempting a task, such as playing a game like chess or go.[37]

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 (sometimes known as 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 certain, 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.[38] Here is the essence of "learning" being applied. Today, such systems appear recondite; soon, they will not just be commonplace but, more importantly, will cease to be perceived as "technologies" (as opposed to mere conveniences) fraught with ethical and social implications at all.

With these common features in hand, we think it is possible to presume that—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.

B. "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.[39] Like other general-purpose technologies, AI is necessarily adopted and integrated into the design and operation of free-standing 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, to a greater or lesser degree, is the object of conscious design—or perhaps a dynamic of Burkean evolution—and the object of legal regulation. It contains "affordances": This refers to a set of "fundamental properties that determine just how the thing could possibly be used."[40] 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.[41] 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 mistake."[42] It is far better to recognize its 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 then-governor of Michigan, Rick Snyder, introduced an algorithmic tool called MiDAS (for Michigan Integrated Data Automated System) to detect fraudulent applications for unemployment benefits as part of a larger overhaul of information technology by the state.[43] This AI tool was introduced as part of a conscious strategy of austerity on Gov. Snyder's part, a sort of junior-varsity "starve the beast."[44] 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 reported calling the state office more than a thousand times a day, and still not being able to get through.[45] The state of Michigan thus chose to provide a user interface with relatively limited opportunities for submitting information, and relatively few external opportunities for revision or correction after the fact. Whatever the formal status of an instrument's predictions

as advisory or only presumptively valid, it was the wider institutional context of the Michigan algorithm that made its predictions de facto binding for tens of thousands of people. It all but guaranteed an exorbitant false positive rate when it came to fraud detection.

The technical specifications of the MiDAS algorithm give us only the most tentative and incomplete picture of the consequences arising from its use. 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 to the 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. Code may be law, but law is inert without actual bodies to implement it.

AI systems widely vary in their design. The systems of interest here nevertheless tend to have certain important characteristics. We can pick out five that strike us as particularly important for the project of democratic regulation. Not all will matter in every instance of democratic regulation, but we think it would be unwise for a democratic regulator to always ignore them.

First, these systems are at least ostensibly designed to add either private or social value by facilitating decisions or operations in particular settings through the distinctive affordances of digital prediction and analysis. 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 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 and its substitution with machine tools. The normative work of "system design" occurs importantly when policymakers resolve "recurring choices of scope" and definition.[46] To design an AI system means defining a particular task (e.g., unemployment insurance as an element of fiscal policy rather than a countercyclical stabilizer). It means excluding other policy ends, and delimiting a range of policy instruments. Attention to the

instrument rather than the system risks missing these normative choices.

Second, many applied settings in which AI systems are embedded involve a sort of 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 discussions are made with algorithmic assistance. (Interestingly, and relevant to our discussion below in Part III, Facebook users seem to underestimate the effect, and even the existence, of such AI nudging.[47]) The risk-prediction instruments used in pretrial bail and sentencing contexts can also 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."[48] Hence, the outcomes of such debate will be necessarily endogenous to the operation of their algorithmic arrangements—even if users believe themselves to be acting autonomously.

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 one 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 or 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."[49]

A risk assessment tool used in a criminal justice setting, for example, might supply judges with a simple numerical score. That score might distill information about a plurality of risks, including the possibility of violent crime, nonviolent crime, and flight from the jurisdiction. Some of these risks might be more amenable to prediction than others.[50] 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 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 the 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.[51] This nuanced choice about transparency—or, perhaps better stated, between different forms of transparency—involves a rather subtle layered principal-agent problem 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, which are embodied in code and interface design. What can arise is a gap between the perception of the officials actively involved in the system's day-to-day operation and the cost function encoded in the algorithm itself. Even officials implementing a system may not, as a consequence, understand precisely what goals the algorithm is designed to pursue. Nor will they be always able to correct deviations from the system's formally stated goal. The management of this complex network of 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 dynamic recalibration is 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.[52] Updating also prevents the gaming of the search function by manipulating the context of searchable pages.

As reinforcement learning tools are increasingly adopted, such continual refinement will itself become automated, and likely more common. The possibility of dynamic adaptation, 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 also to the definition of those goals. Search engines such as Google, for example, have been criticized because they have at times generated results that reflect racist associations.[53] When the PageRank algorithm is adjusted to change these outcomes, Google is introducing a (laudable) anti-racist consideration into the algorithm's dynamic design. This consideration is normative: It may well generate search results that fail to reflect the actual universe of search behavior as a way to prevent ambient patterns of discriminatory sentiment from being baked newly into search outputs. Adaptation here is not just about making technical changes to optimize an algorithm's performance; it is also about changing what "effective performance" means in the first place.

Thoughtful decisions about how to understand, define, measure, and constrain what's defined as "effective performance" depend on recognizing the normative character of such choices. Rarely if ever is there only one viable way to mitigate racial bias in an algorithm's operation.[54] To what extent should background asymmetries in attributes (say, criminal records) be allowed to factor into search results? 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 engine 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 bias" in that they are far more likely to flag the speech of Black Twitter users than white users for blocking or other penalties.[55] Just as the operator of an

AI system does not always know what's being optimized, so its designer might not be able to predict the social consequences of its normative choices in the wild.

To be clear, none of 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 principal-agent problems via choices between different kinds of transparency; or the calibration of dynamic updating—are unique to AI. Nor are they the only possible margins that one might pick out. But all involve choices about institutional context that bear heavily on the impact that an instrument will have on human outcomes, and hence are of particular importance to a democratic regulator. We emphasize them here, in addition, because they are also points of leverage for those regulators—points of leverage that would be elided or lost if one were to focus more narrowly on an AI instrument standing in isolation.

C. 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 promotion of individual "rights"—to privacy, nondiscrimination, and due process, for example. We don't want to deny that that way of thinking about the interaction between AI systems and rights is a useful goal entirely.[56] But we think another lens is potentially more useful.

The distinction between "rights" and "policy" is set forth in a crisp form in recent work 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."[57] 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."[58] 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 rights draws attention to binary relationships between distinct and identifiable people,

or between specific people and institutions. The correlate of a right, in legal theorist Wesley Hohfeld's famous system of correlates, is a specific individual's duty. A right is thus discrete, interpersonal, and need not account for any larger social or institutional context.

Foregrounding 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 propagate out complex and plural effects on the whole population subject to regulation. Efforts to reduce rates of false negatives, for example, are mathematically certain to change the rate (and the distribution) of false positives.[59] Rights are an inapt lens for thinking about AI systems because of their entangled quality, which makes it implausible to isolate just one pair of actors for distinctive treatment. As has long been apparent, rights—especially when enforced by courts—are not an ideal vehicle for managing what Lon Fuller called polycentric disputes.[60] There's every reason to think Fuller's worries have equal weight in this novel technological context, where an intervention to improve the lot of one subject can have complex ramifications for many others.

Moreover, the manner in which normative concerns about equality, privacy, and due process arise out of AI systems is not well captured by the idea of a right standing on its own. As we have suggested, the technical choices of algorithmic design and also their embedding in institutional consequences can entail a range of 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 accounting for externalities. Think here about 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. (We will have more to say about content moderation as a general problem below in Part III). 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 an individual could submit information and then seek reconsideration.[61] Gov. Snyder expressed his normative judgment 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—is not analytically feasible. And while the MiDAS system was successfully challenged on due process grounds,[62] 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 beast," a policy that became only less attractive in an age of pandemic.

That policy, moreover, may have been easier to implement, and more difficult to challenge, because it is buried in the cost function of the algorithm—a mathematical formulation that not even frontline officers will encounter, let alone members of the public. More cynically, the MiDAS system may spark a worry that AI systems will be adopted precisely because they are opaque, resistant to public understanding and critique, and hence a means to achieving otherwise unpopular (or even immoral) policy ends.

As a consequence, to capture these normative judgments, we think it is appropriate to ask what *policy* an AI *system* embodies. That question is almost always more useful, and more tractable, than asking whether the system

violates certain *rights*. 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?

II. TO *REGULATE* AI?

WHAT THEN DOES IT mean to *regulate* AI systems, particularly given the vast range of possible types and levels of intensity of interventions in related domains, ranging from public health to environmental protection? 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.

A. 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 U.S. federal and state governments on the other. We think those barriers 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—the growth in available computing power at a manageable cost is particularly rapid.[63] Increasing the epistemic burden of regulation, the general utility of AI as a technology means it is likely to be adopted

widely by a range of both private and public actors, and then applied to widely different ends. Much of the relevant technological innovation, in addition, happens in the private sector because of "declining government investment in basic and foundational research, combined with lack of access to computational resources and large datasets."[64] This is in stark contrast to the development of Cold War technologies such as nuclear power and remote sensing. AI instruments are hence 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.

Further, the harm from an AI system is quite unlike, at least in salience, the harm from, say, unlawful police brutality—although notice that an AI system might lead to an increase in police brutality by framing the underlying policy task of police 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 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.[65]

Precisely because AI is a general-purpose technology likely to be widely adopted to quite different ends in various institutions across the economic and social landscape, it is simply infeasible to cleave off its regulation into a special-purpose regulatory vehicle.[66] 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 effective either from a political-economy, a historical, or even

a constitutional perspective. As the COVID-19 pandemic has so forcefully reminded us, the United States remains a defiantly decentralized federal state. It is one beset by profound and at times disabling coordination problems, ensnared in paralyzing partisan polarization. Despite the revival in historical scholarship of the idea of a "strong" American state,[67] 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 another difficulty. Regulation requires internal expertise, both on the technical and the 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.")[68] Government bodies already use AI systems in quite varied ways to "prioritize enforcement, … engage with the public [and] conduct regulatory research, analysis, and monitoring."[69]

At the same time, government decisions about how to employ and how to regulate AI systems occur in a political pressure cooker. Elizabeth Joh has also mapped what she plausibly terms the "undue influence" surveillance and analytic technology manufacturers have on police departments.[70] Corporate vendors and lobbies applying great pressure push state actors toward certain technologies, while bucking hard against tighter regulation. Pawel Popiel has found that tech companies such as Google and Twitter were "among the top 20 biggest lobbying industries, spending nearly as much as the defense and telecommunications sectors in 2017, and outspending the commercial banking sector."[71] Captured agencies, revolving doors (in both Republican and Democratic administrations), and iron triangles—the full complement of cynical metaphors—are all appropriate terms here.

At the federal level, moreover, regulatory options are constrained by the perception that China is obtaining a geopolitical edge through research in AI and quantum computing.[72] 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.[73] Particularly when it comes to

military and public-security adoptions, geostrategic competition can lead to the perception that "calls for restraint, reflection, and regulation [are] a strategic disadvantage to U.S. national interests."[74]

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 overreach pinching on 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.[75] 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 has sufficient technical capacity, is well-regulated, and remains tightly leashed by the rule of law when it first confronts the challenge of regulating, and regulating through, AI systems. 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 United States? We leave the reader to decide where it falls in the ensuing spectrum of possibilities. It is, as the Maoists like to say, a useful exercise in self-criticism.

B. Ontological barriers

Profound as these challenges might seem, at first blush they appear to pale in comparison to a more foundational critique leveled by social critics

of AI systems. This critique, rather than the institutional argument, has formed the crux of the case against new digital 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 also warns of the "assertion of decision rights over the expropriation of human experience," and prophesizes "the dispossession of human experience" through "datafication."[76] In particular, she describes three broad strategies of "tuning," "herding," and "conditioning" through which human agency is undermined.[77] A variant on this critique focuses specifically on the shaping of political preferences through behavioral advertising.[78] Philip Howard, for example, has examined spending by the "Vote Leave" campaign prior to the Brexit referendum in the United Kingdom, and estimated that such advertising may have changed the votes of 8 million people and elicited time or money contributions from 800,000 people.[79] 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.[80] The ontological challenge, therefore, may be more manageable than the institutional one.

AI systems, on Zuboff's account, present a challenge because as presently organized and constituted 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."[81] If AI systems are especially good at eliciting false beliefs, perhaps their availability undermines the very possibility of individual agency as a positive good: If people are easily duped, though, then democracy doesn't seem such a good idea[82] (although notice that the same might be said of consumer choice in free markets).

The second version of the argument 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 matter of power, not liberal autonomy, because it turns on who calibrates the Overton window and thus sets the menu of public policy options. It is a matter of how AI enables "producing and refining informational persons who are subject to the operations of fastening."[83]

Neither of these lines of criticism are groundless, but at the same time we should be careful before conflating influence 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.[84]

But, as we have argued elsewhere, the more stark arguments about "dispossession" or "expropriation" miss the mark.[85] 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 is built on an exploitation of social structures of trust, consent, and gift giving.[86] 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 harvested. 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 subjects, and seeks to align their choices with the objectives of governing authorities."[87]

More generally, we should distinguish between the classificatory pressure imposed by state AI systems and the "apparatuses of security" that actively and physically coerce.[88] As scholars such as James Scott and Colin Koopman have amply demonstrated, states and private actors have used schemes of classification and knowledge-management 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 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."[89] 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 tools in the criminal justice context that are touted as devices of "reform" may end up working on the ground as means for recapitulating old 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 fantastical conspiracies is a troubling example of technological skewing of the public sphere.[90] 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.[91] Rather, we need to think critically about the mutuality, and entanglement, of democratic subjectivity and the democratic project of taming AI systems.

III. THE *DEMOCRATIC* REGULATION OF AI

NO SINGLE SILVER BULLET can 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. 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. We first develop a very general observation. Then, we offer a pair of case studies to clarify our perspective.

It is a task of central importance, as Harcourt suggests, to empower individuals. But we think the term "empowerment" is rather more nuanced and tricky than he seems to believe.[92] Where he sees empowerment in large part as a cultivation by the few of the "art of not being governed,"[93] 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.

Hence, 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. Instead of facilitating opt-outs, we would search for ways to deepen public understanding of AI systems as embedded policy, and would seek 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 or 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 ethically 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 counteract both kinds of barriers is the possibility of a 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 few years, employees at firms such as Google and Facebook have used the tools of collective action, such as walkouts, secondary boycotts, and go-slows, to challenge the ways 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.[94] These labor-aligned mobilizations have had an impact disproportionate to their numbers because of the "privileged position" that "[d]esigners, developers, and technologists occupy."[95]

These movements 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 U.S. policing," professional involvement by mathematicians made it "simply too easy to create a 'scientific' veneer for racism."[96] There are also examples that do not rely on the leveraging of occupational or epistemic advantage. Movements for socially useful production, appropriate technology, and people's science are all examples of initiatives, aligned with social movements, that have aimed to repurpose technology to generate a "fairer and more sustainable" world.[97] In the context of the COVID-19 pandemic, a number of housing justice collectives have also used new computational tools to "compile data on landlords and speculators to embolden the work of housing justice organizing."[98]

Legal institutions, to be clear, will rarely be the first movers in these efforts. But still they may well have a role. These movements often entail

collective action that may or may not be vulnerable to legal sanction as a matter of contract law or criminal law. Hence, 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. Non-compete agreements and trade secrets can impede legitimate public debate, for example, and law appropriately accounts for these spillover social 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 newsfeed). 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[99]—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 the 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. It can thereby prod people to clarify for themselves their own second-order preferences, and then adjust their approach to technologies accordingly.

We think that legal interventions might focus on revealing the distributive and dynamic effects of AI systems to the public as a way of concentrating

debate on the central question of policy effects. 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 its 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."[100] 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 individuals who are already subject to some form of structural disadvantage.[101] That is, we should attend not only to how inequality is replicated; we should also attend to how AI systems such as the Michigan MiDAS tool can create it from the ground up.

In the governmental context, such considerations might be built into the procurement or certification process for new state algorithmic interventions. This is a step toward, although not a complete 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.

A sustained focus on those dynamic effects will 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 members of 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 our lives themselves are at risk from the very world that sustains them—whether this is a matter of the way our use of cars creates the risk of accidents or how our industrial infrastructure imperils the climate as a whole.[102] 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 profound regulatory questions in a democracy for an analogous reason: 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 choose their priorities in a contentious world.

To gain a sense of what this general project of democratic empowerment might look like, we must think about the configuration of both public law and also private law. A first, and very obvious, domain in which democratic regulation of AI systems may be called for is social media. At issue here is the regulation of AI moderation and AI-produced content for the promotion of a working democratic public sphere characterized by (1) norms of truthfulness in the production and dissemination of information, and (2) the absence of drivers of sharp affective polarization. As we have noted, there is ample evidence that moderation algorithms can generate perverse and undesirable channeling effects, pulling individuals away from the political center and increasing epistemic (and perhaps affective) polarization.[103] At the same time, both domestic and foreign actors are increasingly using the same algorithmic moderators to disseminate false political news. Although there is reasonable disagreement on how significant a problem this is in practice for the sound operation of a nation's democracy, few doubt that it presents a risk of serious (even if not fatal) concern for democracies' health.

A threshold challenge entails the reconceptualization of the policy crystalized in a content-moderation AI. Such systems are usually narrowly conceived by their creators as AI systems for maximizing platform traffic.[104] But as an alternative, they can be construed from the vantage point of public law, as AI systems for channeling and fashioning understandings and judgments necessary to the exercise of democratic judgment. That is, they are AI systems to promote the democratic public sphere through their ability to lower the cost of organizing and acquiring truthful information. Manipulation of the sort Howard documents targeting voting decisions, however, plainly challenges the proper functioning of the content-moderation system understood in this fashion. More subtly, so does the tendency of content-moderation tools to generate echo chambers and to increase affective polarization. Notice that a rights lens does not illuminate the concerns raised by such AI systems from a democratic perspective. Users of social media platforms, after all, can

simply switch to alternate sources of news if they wish to. To the extent that there are rights concerns in play, they concern whatever First Amendment interests the platforms have in exercising a measure of editorial control when curating their content.[105] We think these interests will often be weak.

Instead, it is more useful to ask how the law can intervene, first, to encourage users to become sophisticated in their use of the services that social media platforms provide, and second, to nudge those platforms toward a constructive role in fashioning the democratic public sphere.[106] In particular, an effective regulator would harness the broadly shared second-order preference held (we believe) by many to be well-informed and civil-minded participants in democracy. It would enable citizens to achieve this goal despite the short-term tugs of platform dysfunctionality. To recall our earlier taxonomy, this means attending to the values threaded through user interfaces, and also the manner in which collective processes of debate are organized and winnowed.

There is no necessary reason that this needs to be done all at the level of the national government. To the contrary, state governments have control over many of the levers of effective intervention, and might be more trusted by citizens because they are less geographically remote, less ideologically distant, and perhaps more attuned to local needs.[107] States are also responsible for secondary education, which can be a site for training in the responsible and skeptical use of social media. Training citizens to be careful users of AI-moderated social media content may not have immediate payoffs, but may well be the most effective long-term approach. States are further likely to be more capable, because of their smaller size, to identify troubling trends in beliefs and behavior before the federal government. (Think, for example, of the many false rumors about COVID-19 that spread in the Black community.)[108] They may have more ready access to connections within affected communities that can effectively respond to the online spread of misinformation and incitement to polarization, for example by alerting influential community actors of fake news and polarizing efforts while they are in their early stages.[109]

Yet states need not act alone. Either states or the national government might also consider affirmative obligations on sites that disseminate news to maintain mechanisms for preventing polarizing cascades, or the

dissemination of knowingly false information. Such an obligation would not require platforms to remove or limit specific voices. Instead, the legal regime would focus on the aggregate health of the platform's informational and affective ecosystem. It would impose platform-level penalties if the level of misinformation or hate speech reached specific levels. In this fashion, the law would elicit a specific policy end, without directly targeting any user's rights.[110]

The curation of social media by AI systems is a domain bearing quite a direct relationship to democracy and civic life. In contrast, medical devices functioning as AI systems are an example of how democracies face broader choices about ever-more ubiquitous technologies with potentially enormous benefits and risks. These technologies involve not only systems operated in clinics or hospitals, but mobile and wearable technology capable of inferring medical risks and even emotional conditions from users' daily behavior. Smart watches functioning as AI systems, for instance, can be used as medical monitors that advise users about their stress levels; risks of complications from physical conditions such as coronary heart failure, asthma, and chronic obstructive pulmonary disease; mental health; and the relationship between their behavior and overall well-being.[111] At the same time, networks of wearable sensors and remote servers can also facilitate the centralization of enormous amounts of data. And such information reveals far more than just users' particular requests for information or location. Scrutinized by AI systems, the data instead illuminate what most authoritarian regimes would consider a treasure trove: how users' physiological reactions–the best window we have into their feelings and subconscious dynamics–are mediated by particular settings, events such as exercising or attending a performance, or commercial transactions.[112] To the extent the psychology of persuasion and attitude change has a physiological signature, wearable sensors latched onto AI systems will render it legible, and hence perhaps make people more pliable. The question of what private or social values such systems optimize is therefore acute.

Yet the potential benefits are quite palpable: AI systems fed data from such sensors could offer users increasingly precise warnings about the need for earlier medical intervention, encouragement, and incentives to engage in health-enhancing behaviors, and insights about how they might better

align their daily routines with their desires and goals.[113] That said, many consumers, observers, and policymakers readily appreciate that the privacy concerns are quite substantial, arguably on a whole different scale relative to web searches and social media.[114] It's one thing to know what news or names a person is looking up on the internet; it's quite another to know how a particular news story or name alters a person's heartbeat. Some of these systems already have the capacity to share this data with health care providers synchronously, and certain employers offer incentives to share activity data. What's more, even in the absence of sabotage, AI systems deployed in this fashion risk creating subtly misleading or even dangerously false impressions among users about the risks or safety they face at any given moment, or the actions they should take to mitigate health risks or accept risk. Again, the nature and value choices embedded in a user interface will be highly consequential.

How should such devices be regulated? Under the federal Food, Drug, and Cosmetic Act, medical devices are subject to regulation by the Food and Drug Administration (FDA) for some of the reasons listed above.[115] Although the FDA has a considerable degree of flexibility in how it uses its regulatory authority and what guidance it provides to private entities engaged in innovative activity,[116] it would be difficult to argue that the agency could simply categorically and explicitly exempt from regulation a vast segment of technologies quite likely to fit the broad category of "medical device" simply because it seeks to foster innovation.[117] More likely, principled agency decision-makers would use—as they have in recent years—some mix of guidance and ex post risk of regulatory sanctions to calibrate the value of administrative regulation and the importance of leaving some room for innovative developments in the field.[118] In doing so, they also leave such technologies subject to the demands of state-level tort law,[119] under which even technologies nominally complying with FDA requirements could conceivably expose companies to the risk of liability in the event they breach, for example, a duty of care to avoid chronically misaligning an AI system's medical recommendations with the user's interests.

The resulting doctrinal and prudential questions about how democracies should regulate AI systems making use of wearable technology are intricate, and bear further scrutiny. But they also gesture toward three broader

themes, which each add further nuance to our argument about the democratic regulation of AI systems.

First, AI systems incorporating wearable devices to monitor medical data in real time offer a compelling example of how existing law (in this case, both quite specific substantive law governing federal regulation of medical devices, as well as tort liability doctrine cutting across domains and rooted in the common law) already provides a regulatory framework for the use of AI systems in American democracy. To frame the discussion instead as being about whether to establish a regulatory framework for AI from the ground up misses the extent to which questions about the regulation of AI systems in some settings are questions about generally applicable regulation. This observation doesn't deny that some regulatory questions about AI are distinctive even in the medical device context—instead, we emphasize that the AI-related questions don't arise entirely separately from other regulatory policy problems as a factual matter, nor do they implicate only a specific regulatory regime involving AI.

Second, while there is some interplay between public regulatory law (such as the Food, Drug, and Cosmetic Act) and private law (tort liability), they also constitute somewhat separate legal arrangements, both relevant to the regulation of AI. Given that common law tort liability (with all its costs and benefits) operates as its own form of regulation by allocating risk and shaping incentives, efforts to preempt state tort liability to advance regulatory consistency can substantially change the incentives of the private sector and redistribute risks associated with innovation and health. Administrative regulation rooted in public law, and liability regimes rooted in private law implicate different decision-makers and call for different analyses.[120] But both systems allow policymakers, lawyers, judges, and other key actors some flexibility to consider the value of leaving some room for innovation, to take account of the many benefits that this technology can yield, but also a range of risks from obvious to subtle.[121]

Finally, choices about how democracies retain or reform existing laws to regulate AI systems inevitably affect the power and competence of key institutions within society at a crucial, early moment when AI systems are becoming more ubiquitous. Choices about AI regulation can affect the role of states, for example, or the independence of the FDA from political appointees

at the Department of Health and Human Services. Given the potentially path-dependent consequences of decisions about AI regulation when the influence of such systems is growing, the conversation about AI regulation in a democracy should retain space for discussion of how particular regulatory choices either further or cut against broader commitments, such as by leaving states room for democratic experimentation and by striking a carefully calibrated balance between the role of bureau-level technical experts, political appointees, and judges safeguarding procedural integrity.

CONCLUSION

None of this will be easy. But our goal here has neither been to sugarcoat a bitter pill nor to discourage the project of democratic regulation of AI for practical or prudential reasons. Democracy is a product not only of societal norms and culture, but also 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.[122]

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 play out as reality rather than risk.

NOTES

1 *See, e.g.*, David Freeman Engstrom et al., *ACUS Study: Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies* 12 (Admin. Conf. U.S. 2020), https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf [https://perma.cc/5HE2-TPDB].

2 John Gramlich, *10 facts about Americans and Facebook*, Pew Rsch. Ctr. (June 1, 2021), https://www.pewresearch.org/fact-tank/2021/06/01/facts-about-americans-and-facebook/ [https://perma.cc/R9SZ-5YEG].

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/ [https://perma.cc/V37E-KQNX].

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), https://doi.org/10.1145/3298789 [https://perma.cc/W9AJ-HHV9]. 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://www.nytimes.com/2020/06/16/science/social-media-bots-kazemi.html [https://perma.cc/PET5-UM27].

6 Andrew Jack, *Students and teachers hit at International Baccalaureate grading*, Fin. Times (July 9, 2020), https://www.ft.com/content/eeof4d97-4eoc-4bc3-8350-19855e70focf [https://perma.cc/G3ZW-XZ7L] (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 Digit. Health e271, e272 (2019), https://doi.org/10.1016/S2589-7500(19)30123-2 [https://perma.cc/J846-KYXU].

8 Scott Mayer McKinney et al., *International evaluation of an AI system for breast cancer screening*, 577 Nature 89, 91 (2020), https://doi.org/10.1038/s41586-019-1799-6 [https://perma.cc/T9V9-8FWX].

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, 9:20 AM), https://www.latimes.com/local/lanow/la-me-lapd-precision-policing-data-20190703-story.html [https://perma.cc/H2J3-7ZG8].

11 Larry D. Wall, *Some financial regulatory implications of artificial intelligence*, 100 J. Econ. & Bus. 55 (2018), https://doi.org/10.1016/j.jeconbus.2018.05.003 [https://perma.cc/Q68Q-XA7N].

12 Press Release, Apple, HomePod arrives February 9, available to order this Friday (Jan. 23, 2018), https://www.apple.com/newsroom/2018/01/homepod-arrives-february-9-available-to-order-this-friday/ [https://perma.cc/HB82-V4TD].

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/ [https://perma.cc/XP3Z-SSQQ].

14 *See* Darius Kazemi, *The bot scare*, Tiny Subversions (Dec. 31, 2019), https://tinysubversions.com/notes/the-bot-scare/ [https://perma.cc/RH7J-UEC7].

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 M.C. Elish & danah boyd, *Situating methods in the magic of Big Data and AI*, 85 Comm. Monographs 57, 60 (2018), https://doi.org/10.1080/03637751.2017.1375130.

17 Frank Pasquale, The Black Box Society 8 (2015).

18 Bernard E. Harcourt, Exposed 19, 261 (2015).

19 Shoshana Zuboff, The Age of Surveillance Capitalism 94 (2018).

20 Ruha Benjamin, Race after Technology 7 (2019).

21 Carl Benedikt Frey, The Technology Trap 320-21 (2018). Displacement is not the only way in which AI might change the labor market; automated

scheduling, task redefinition, loss and risk prediction, and the incentivization of productivity may also be important effects. *See also* Pegah Moradi & Karen Levy, *The Future of Work in the Age of AI, in* The Oxford Handbook of Ethics of AI (Markus D. Dubber et al. eds., 2020).

22 For criticisms along these lines, see Mariano-Florentino Cuéllar & Aziz Z. Huq, *Economies of Surveillance*, 133 Harv. L. Rev. 1280 (2019), https://harvardlawreview.org/2020/02/economies-of-surveillance/ [https://perma.cc/B5HW-FLAJ]; and Aziz Z. Huq, *Apps in Black and White*, 61 Euro. J. Socio. 423 (2021), https://doi.org/10.1017/S0003975620000223 [https://perma.cc/2WN6-N5FK].

23 Harcourt, *supra* note 18, at 270.

24 Zuboff, *supra* note 19, at 486.

25 For a parsimonious definition focused on the institutional preconditions of democracy, see Tom Ginsburg & Aziz Z. Huq, How to Save a Constitutional Democracy 9-10 (2018).

26 *See* David Runciman, The Democracy Trap (2013).

27 Harcourt, *supra* note 18, at 258.

28 It is, to be sure, also nothing entirely new. There is a long history of "government and firms compet[ing] 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), https://columbialawreview.org/content/we-the-shareholders-government-market-participation-in-the-postliberal-u-s-political-economy/ [https://perma.cc/2WKD-B2FT].

29 *See* Elizabeth Hinton, From the War on Poverty to the War on Crime (2017).

30 For a programmatic statement of the value of such efforts, see Jedediah S. Britton-Purdy et al., *Building a Law-and-Political-Economy Framework: Beyond the Twentieth-Century Synthesis*, 129 Yale L. J. 1784 (2020), https://www.yalelawjournal.org/feature/building-a-law-and-political-economy-framework [https://perma.cc/DD69-PZL5].

31 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 (3rd ed. 2013).

32 Hannah Fry, Hello World: Being Human in the Age of Algorithms 11 (2019); *accord.* Jerry Kaplan, Artificial Intelligence: What Everyone Needs to Know 32 (2016) (providing a similar colloquial description).

33 Kaplan, *supra* note 32, at 32.

34 On the latter, see Lei Xu et al., *Information security in big data: privacy and data mining*, 2 IEEE Access 1149, 1155 (2014), https://doi.org/10.1109/ACCESS.2014.2362522 [https://perma.cc/5PH2-KU2E].

35 Robert Moni, *Reinforcement Learning algorithm—an intuitive overview*, Medium (Feb. 18, 2019), https://medium.com/@SmartLabAI/reinforcement-learning-algorithms-an-intuitive-overview-904e2dff5bbc [https://perma.cc/C8H5-T5RN].

36 We don't use the term "algorithmic systems" because this might induce a reader to think of non-digital systems, such as the clinical checklists or algorithms used by psychologists for much of the 20th century. "AI" might be ambiguous, but at least it excludes such wholly human systems.

37 Tom Simonite, *This More Powerful Version of AlphaGo Learns on Its Own*, Wired (Oct. 18, 2017, 1:00 PM), https://www.wired.com/story/this-more-powerful-version-of-alphago-learns-on-its-own [https://perma.cc/L38N-D94H].

38 Sendhil Mullainathan & Jann Spiess, *Machine learning: an applied econometric approach*, 31 J. Econ. Persp. 87, 89 (2017), https://doi.org/10.1257/jep.31.2.87 [https://perma.cc/5JVL-5XHT] (defining machine learning in terms of its capacity for "out-of-sample" prediction); *see generally* Ethem Alpaydin, Machine Learning: The New AI (2016).

39 Frey, *supra* note 21, at 305.

40 Donald A. Norman, The Psychology of Everyday Things 8 (1988).

41 Sasha Costanza-Chock, Design Justice: Community-Led Practices to Build the Worlds We Need 45 (2020).

42 David Beer, *The social power of algorithms*, 20 Info. Comm. & Soc. 1, 4 (2017), https://doi.org

/10.1080/1369118X.2016.1216147 [https://perma.cc/3XM3-PEVH].

43 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].

44 Lee Saunders, *Government Didn't Fail Flint, Austerity Did*, Governing (Feb. 12, 2016), https://www.governing.com/gov-institute/voices/col-flint-water-austerity-public-services.html [https://perma.cc/BND7-PL4B].

45 Taylor DesOrmeau, *Michigan's Ineffective Unemployment System Is Nothing New*, Governing (June 17, 2020), https://www.governing.com/work/Michigans-Ineffective-Unemployment-System-Is-Nothing-New.html [https://perma.cc/MY34-KNCU].

46 Bernard E. Harcourt, *The Systems Fallacy: A Genealogy and Critique of Public Policy and Cost-Benefit Analysis*, 47 J. Legal Stud. 419, 431 (2018), https://doi.org/10.1086/698135 [https://perma.cc/BZ3C-VCZK].

47 *See* Blake Hallinan et al., *Unexpected expectations: Public reaction to the Facebook emotional contagion study*, 22 New Media & Soc. 1076 (2019), https://doi.org/10.1177%2F1461444819876944 [https://perma.cc/923B-QJ5N].

48 Jürgen Habermas, The Structural Transformation of the Public Sphere 52 (Thomas Burger trans., 1991).

49 Beer, *supra* note 42, at 5.

50 For example, there is an argument that the prediction of violence is infeasible. *See* Martha Minow et al., *Technical Flaws of Pretrial Risk Assessment Tools Raise Grave Concerns* 2 (2019), Berkman Klein Ctr. for Internet & Soc'y (July 17, 2019), https://dam-prod.media.mit.edu/x/2019/07/16/TechnicalFlawsOfPretrial_ML%20site.pdf [https://perma.cc/B2TZ-GGJP].

51 For debates about the meaning of the term, see Tim Miller, *Explanation in artificial intelligence: Insights from the social sciences*, 267 A.I. 1, 1-2 (2019) https://doi.org/10.1016/j.artint.2018.07.007 [https://perma.cc/L4UB-62X2], and *see also* Michael Gleicher, *A Framework for Considering Comprehensibility Modeling*, 4 Big Data 75, 77–84 (2016), https://dx.doi.org/10.1089%2Fbig.2016.0007 [https://perma.cc/CZQ2-7DVA].

52 Michele Willson, *Algorithms (and the) everyday*, 20 Info., Commc'n. & Soc.'y 137, 142 (2017), https://doi.org/10.1080/1369118X.2016.1200645 [https://perma.cc/4Q7R-LKMB].

53 Safiya Umoja Noble, Algorithms of Oppression: How Search Engines Reinforce Racism 66-80 (2018).

54 Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 Duke L. J. 1043 (2019), https://scholarship.law.duke.edu/dlj/vol68/iss6/1/ [https://perma.cc/Q9DJ-H2LL] [hereinafter "Huq, Racial Equity"].

55 Thomas Davidson et al., *Racial Bias in Hate Speech and Abusive Language Detection Datasets*, arXiv preprint arXiv:1905.12516 (2019), https://arxiv.org/abs/1905.12516 [https://perma.cc/NU2C-6MJ8].

56 *Cf.* Aziz Z. Huq, *Constitutional Rights in the Machine Learning State*, 105 Cornell L. Rev.1875 (2020), https://dx.doi.org/10.2139/ssrn.3613282 [https://perma.cc/E5R6-6VXS].

57 Karen Orren & Stephen Skowronek, The Policy State 27 (2017).

58 *Id.* at 29.

59 Huq, *Racial Equity*, *supra* note 54 (discussing these complexities).

60 Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 Harv. L. Rev. 353 (1978), https://doi.org/10.2307/1340368 [https://perma.cc/HWY9-V3B8].

61 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), https://www.virginialawreview.org/articles/right-human-decision/ [https://perma.cc/3Y9S-4V98].

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

63 *See* AI Index 2019 Report (Human-Centered A.I. Inst. 2019), https://hai.stanford.edu/ai-index-2019 [https://perma.cc/PAL9-TB8C].

64 John Etchemendy & Fei-Fei Li, *National Research Cloud: Ensuring the Continuation of American Innovation*, HAI Blog (Mar. 28, 2020), https://live-stanford-hai.pantheonsite.io/blog/national-research-cloud-ensuring-continuation-american-innovation [https://perma.cc/CR9Q-WXHX].

65 Huq, *Racial Equity*, *supra* note 54.

66 For the best version of this argument, see Andrew Tutt, *An FDA for Algorithms*, 69 Admin. L. Rev. 83 (2017), https://dx.doi.org/10.2139/ssrn.2747994 [https://perma.cc/3QFU-46VJ].

67 For a brilliant treatment, see William J. Novak, *The Myth of the "Weak" American State*, 113 Am. Hist. Rev. 752, 762 (2008), https://doi.org/10.1086/ahr.113.3.752 [https://perma.cc/C82Q-LNT3].

68 Engstrom et al., *supra* note 1, at 7.

69 *Id*. at 17.

70 Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. Rev. Online 19, 20 (2017), https://dx.doi.org/10.2139/ssrn.2924620 [https://perma.cc/5XZJ-AZ5N].

71 Pawel Popiel, *The Tech Lobby: Tracing the Contours of New Media Elite Lobbying Power*, 11 Commc'n, Culture & Critique 566, 572 (2018), https://doi.org/10.1093/ccc/tcy027 [https://perma.cc/KUN8-5FPK].

72 *See* Mariano-Florentino Cuéllar & Aziz Z. Huq, *Privacy's Political Economy and the State of Machine Learning* (U of Chi., Pub. L. Working Paper No. 714, 2019), https://ssrn.com/abstract=3385594 [https://perma.cc/Z6EZ-X5ML].

73 Kai-Fu Lee, AI Superpowers: China, Silicon Valley, and the New World Order 14-15 (2018).

74 Kate Crawford et al., *AI Now 2019 Report* 43 (A.I. Now Inst. 2019), https://ainowinstitute.org/AI_Now_2019_Report.pdf [https://perma.cc/JT2U-DSFM].

75 *E.g.*, 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), https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html [https://perma.cc/QD4H-R3MD].

76 Zuboff, *supra* note 19, at 128, 233-34.

77 *Id*. at 294-96.

78 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 et al., *Online Behavioral Advertising: A Literature Review and Research Agenda*, 46 J. Advert. 363, 363 (2017), https://doi.org/10.1080/00913367.2017.1339368 [https://perma.cc/2ENL-Q28A].

79 Philip N. Howard, Lie Machines: How to Save Democracy From Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives 128-29 (2020).

80 We make these points at greater length in Cuéllar & Huq, *supra* note 22.

81 Howard, *supra* note 79, at 123.

82 Indeed, this is Jason Brennan's position. Jason Brennan, Against Democracy 170 (2017) (arguing that "democracy systematically violates the competency principle").

83 Colin Koopman, How We Became Our Data: A Genealogy of the Informational Person 12 (2019).

84 Adam Kramer et al., *Experimental evidence of massive-scale emotional contagion through social networks*, 111 Proc. Nat'l Aca. Sci. 8788 (2014), http://dx.doi.org/10.1073/pnas.1320040111 [https://perma.cc/LY2K-QS26].

85 Zuboff, *supra* note 19, at 126, 233; *see* Cuéllar & Huq, *supra* note 22.

86 Marion Fourcade & Daniel N. Kluttz, *A Maussian bargain: Accumulation by gift in the digital economy*, 7 Big Data & Soc.'y 1, 10 (2020), https://doi.org/10.1177%2F2053951719897092 [https://perma.cc/B768-BBEG].

87 David Garland, *'Governmentality' and the Problem of Crime: Foucault, Criminology, Sociology*, 1 Theoretical Criminology 173, 175 (1997), https://doi.org/10.1177%2F1362480697001002002 [https://perma.cc/P664-XZBP].

88 Michel Foucault, *Governmentality, in* Power 201, 220 (James D. Faubion ed., 1997).

89 Koopman, *supra* note 83, at 59-61.

90 *See* Wong, *supra* note 3.

91 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 & 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).

92 Harcourt, *supra* note 18, at 270.

93 *Cf.* James C. Scott, The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia (2010).

94 For a useful survey, see Haydn Belfield, *Activism by the AI Community: Analysing Recent Achievements and Future Prospects*, Proc. AAAI/ACM Conf. on A.I., Ethics, & Soc.'y (2020), https://doi.org/10.1145/3375627.3375814 [https://perma.cc/NB4C-MJ9R].

95 Costanza-Chock, *supra* note 41, at 216.

96 Davide Castelvecchi, *Mathematicians urge colleagues to boycott police work in wake of killings*, Nature (June 19, 2020), https://doi.org/10.1038/d41586-020-01874-9 [https://perma.cc/N5BE-J8LT].

97 Adrian Smith et al., Grassroots Innovation Movements 3-5 (2016).

98 Erin McElroy et al., *COVID-19 Crisis Capitalism Comes to Real Estate*, Bos. Rev. (May 7, 2020), http://bostonreview.net/class-inequality-science-nature/erin-mcelroy-meredith-whittaker-genevieve-fried-covid-19-crisis [https://perma.cc/C4EF-9FF3].

99 For a discussion, see Huq, *supra* note 56.

100 Costanza-Chock, *supra* note 41, at 59.

101 *Id.* at 193-94.

102 *See* Mariano-Florentino Cuéllar & Jerry Mashaw, *Regulatory Decision-Making and Economic Analysis* (Stanford L. and Econ. Olin Working Paper, Paper No. 525, 2018), https://dx.doi.org/10.2139/ssrn.3238749 [https://perma.cc/4YXS-PJKR].

103 Howard, *supra* note 79, at 128-29.

104 Gillespie, *supra* note 4, at 64.

105 For an example of an analysis framed in terms of the First Amendment interests of users and moderators, see Kyle Langvardt, *Regulating Online Content Moderation*, 106 Geo. L. J. 1353 (2018), https://dx.doi.org/10.2139/ssrn.3024739 [https://perma.cc/23SE-T9CE].

106 Giovanni De Gregorio, *Democratising online content moderation: A constitutional framework*, 36 Comput. L. & Sec. Rev. 105374 (2020), https://doi.org/10.1016/j.clsr.2019.105374 [https://perma.cc/Y99S-XRKN].

107 Certainly, the greater responsiveness and efficacy of state governments during the COVID-19 pandemic is likely to increase the perceived gap between the national and state governments.

108 Erica Weintraub Austin et al., *COVID-19 disinformation and political engagement among communities of color: The role of media literacy*, Harv. Kennedy Sch. Misinformation Rev. (2021), https://misinforeview.hks.harvard.edu/article/covid-19-disinformation-and-political-engagement-among-communities-of-color-the-role-of-media-literacy/ [https://perma.cc/GF5W-4A72].

109 For a suggestion along these lines in the counterterrorism space, see Aziz Z. Huq, *The Social Production of National Security*, 98 Cornell L. Rev. 637 (2013), https://scholarship.law.cornell.edu/clr/vol98/iss3/3 [https://perma.cc/XN3N-TDSU].

110 The proposal here, though, is in tension with a strong view of platforms' First Amendment rights.

111 Chun Yu, *A Review of AI Technologies for Wearable Devices*, 688 IOP Conf. Series: Materials Science & Eng. (2019), http://dx.doi.org/10.1088/1757-899X/688/4/044072 [https://perma.cc/2S62-M2SJ].

112 Ctr. for Long-Term Cybersecurity, *Cybersecurity Futures* 2020 92-113 (Ctr. for Long-Term Cybersecurity, Univ. of Calif., Berkeley 2016), https://cltc.berkeley.edu/wp-content/uploads/2016/04/cltcReport_04-27-04a_pages.pdf [https://perma.cc/45WD-TLZP].

113 Blaine Reeder & Alexandria David, *Health at Hand: A systemic review of smart watch uses for health and wellness*, 63 J. Biomedical Informatics 269 (2016), https://doi.org/10.1016/j.jbi.2016.09.001 [https://perma.cc/E9M9-GHCR].

114 *See* Nathan Cortez, *The Mobile Health Revolution?* 47 U.C. Davis L. Rev. 1173 (2014), https://dx.doi.org/10.2139/ssrn.2284448 [https://perma.cc/F4ME-J7SL]. These concerns are not entirely new. *See, e.g.*, Stephen S. Intille & Amy M. Intille, *New Challenges for Privacy Law: Wearable Computers that Create Electronic Digital Diaries*, MIT House_n Tech. Rep. (Sept. 15, 2003), http://web.media.mit.edu/~intille/papers-files/IntilleIntille03.pdf [https://perma.cc/YL9X-GM5J]. What's changed is the scale of use of such technology, and the myriad ways in which the data can be aggregated, analyzed, and redeployed using machine-learning techniques.

115 S. 3744, 115th Cong. (2018).

116 *See* Heckler v. Chaney, 470 U.S. 821 (1985).

117 Artificial Intelligence and Machine Learning in Software as a Medical Device, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#regulation [https://perma.cc/UWD2-HDCQ] (last visited July 20, 2021). The dynamic of flexibility and constraint obviously also implicates the companies' behavior–and creates regulatory enforcement challenges. Many devices choose to not define themselves as medical devices and instead call themselves patient education tools to exempt themselves from the FDA when in fact they are likely used for more explicitly medical purposes.

118 Food & Drug Admin., *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback* (Apr. 2, 2020), https://www.fda.gov/media/122535/download [https://perma.cc/2RBS-R2QU].

119 *See, e.g.*, Scott v. C.R. Bard, Inc., 231 Cal. App. 4th 763 (2014).

120 *See* Mariano-Florentino Cuéllar, *A Common Law for the Age of Artificial Intelligence: Incremental Adjudication, Institutions, and Relational Non-Arbitrariness*, 119 Colum. L. Rev. 1773 (2019), https://ssrn.com/abstract=3522733 [https://perma.cc/5S6Z-ERD3].

121 *See, e.g.*, Expert Group on Liability and New Technologies – New Technologies Formation, *Liability for Artificial Intelligence and Other Emerging Digital Technologies* (Nov. 27, 2019), https://op.europa.eu/en/publication-detail/-/publication/1c5e-30be-1197-11ea-8c1f-01aa75ed71a1 [https://perma.cc/C8DQ-P8AP].

122 *See, e.g.*, Lawrence Lessig, *The Regulation of Social Meaning*, 62 U. Chi. L. Rev. 943 (1995), https://chicagounbound.uchicago.edu/uclrev/vol62/iss3/1/ [https://perma.cc/GYV6-JAKJ].

About the Authors

Mariano-Florentino (Tino) Cuéllar—a law professor and public servant with broad experience in international and domestic policy, the justice system, education, and philanthropy—is the president of the Carnegie Endowment for International Peace. A scholar of transnational regulatory and security problems, American institutions, and technology's impact on law and government, he previously served as a justice on the Supreme Court of California, the highest court of America's largest judiciary. He is the first Mexican immigrant ever to serve in this capacity. Previously, he was the Stanley Morrison Professor at Stanford Law School and director of Stanford University's Freeman Spogli Institute for International Studies. He served two U.S. presidents in a variety of roles in the federal government, including as special assistant to the president for justice and regulatory policy in the Obama administration.

Aziz Z. Huq is the Frank and Bernice J. Greenberg Professor of Law and the Mark Claster Mamolan Teaching Scholar at the University of Chicago. His scholarship focuses broadly on how institutional design influences individual rights and liberties. His book *How to Save a Constitutional Democracy* (with Tom Ginsburg) was published by the University of Chicago Press in October 2018, and won the International Society of Public Law I-CON Book Prize (2019). Before teaching, Huq worked as associate counsel and then director of the Liberty and National Security Project of the Brennan Center for Justice at NYU School of Law. He was also a senior consultant analyst for the International Crisis Group. He clerked for Judge Robert D. Sack of the U.S. Court of Appeals for the Second Circuit and for Justice Ruth Bader Ginsburg of the U.S. Supreme Court.

© 2022, Mariano-Florentino Cuéllar and Aziz Z. Huq.

About the Knight First Amendment Institute

The Knight First Amendment Institute at Columbia University defends the freedoms of speech and the press in the digital age through strategic litigation, research, and public education. It promotes a system of free expression that is open and inclusive, that broadens and elevates public discourse, and that fosters creativity, accountability, and effective self-government.

knightcolumbia.org

Design: Point Five
Illustration: ©Erik Carter



KNIGHT
FIRST AMENDMENT
INSTITUTE at
COLUMBIA UNIVERSITY