University of Pennsylvania Law School

ILE

INSTITUTE FOR LAW AND ECONOMICS

A Joint Research Center of the Law School, the Wharton School,
and the Department of Economics in the School of Arts and Sciences
at the University of Pennsylvania

RESEARCH PAPER NO. 17-8

Regulating by Robot: Administrative Decision Making in the Machine-Learning Era

Cary Coglianese
UNIVERSITY OF PENNSYLVANIA

David Lehr
RESEARCH AFFILIATE, PENN PROGRAM ON REGULATION,
UNIVERSITY OF PENNSYLVANIA LAW SCHOOL

ARTICLES

Regulating by Robot: Administrative Decision Making in the Machine-Learning Era

CARY COGLIANESE* AND DAVID LEHR**

Machine-learning algorithms are transforming large segments of the economy as they fuel innovation in search engines, self-driving cars, product marketing, and medical imaging, among many other technologies. As machine learning's use expands across all facets of society, anxiety has emerged about the intrusion of algorithmic machines into facets of life previously dependent on human judgment. Alarm bells sounding over the diffusion of artificial intelligence throughout the private sector only portend greater anxiety about digital robots replacing humans in the governmental sphere. A few administrative agencies have already begun to adopt this technology, while others have clear potential in the near term to use algorithms to shape official decisions over both rulemaking and adjudication. It is no longer fanciful to envision a future in which government agencies could effectively make law by robot, a prospect that understandably conjures up dystopian images of individuals surrendering their liberty to the control of computerized overlords. Should society be alarmed by governmental use of machine-learning applications? We examine this question by considering whether the use of robotic decision tools by government agencies can pass muster under core, time-honored doctrines of administrative and constitutional law. At first glance, the idea of algorithmic regulation might appear to offend one or more traditional doctrines, such as the nondelegation doctrine, procedural due process, equal protection, or principles of reason-giving and transparency. We conclude, however, that when machine-learning

* Edward B. Shils Professor of Law and Professor of Political Science, University of Pennsylvania; Director, Penn Program on Regulation.

** Research Affiliate, Penn Program on Regulation; Yale Law School, J.D. expected 2020. Research underlying this Article was generously supported by a grant from the Robert Wood Johnson Foundation. For helpful comments and discussions on an earlier version of this manuscript, we are grateful to Kaiya Arroyo, Tom Baker, Richard Berk, Tanner Bowen, Ronald Cass, Bryan Choi, Peter Conti-Brown, Adam Finkel, Claire Finkelstein, Jill Fisch, Josh Gold, Chelsey Hanson, Kelly Joyce, Joseph Kable, Seth Kreimer, Sandra Mayson, Paul Ohm, Aaron Roth, Gabriel Scheffler, Shana Starobin, Kenneth Steif, Kristene Unsworth, Jeffrey Vagle, R. Polk Wagner, Michael Weisberg, Christopher Yoo, and David Zaring, along with participants in the Regulatory Law and Policy Seminar at the University of Pennsylvania Law School and the editors of this journal. The views expressed in this Article are not necessarily those of the individuals or institutions we gratefully acknowledge here; we alone bear responsibility for our conclusions.

technology is properly understood, its use by government agencies can comfortably fit within these conventional legal parameters. We recognize, of course, that the legality of regulation by robot is only one criterion by which its use should be assessed. Agencies should not apply algorithms cavalierly, even if doing so might not run afoul of the law; in some cases, safeguards may be needed for machine learning to satisfy broader, good-governance aspirations. Yet, in contrast with the emerging alarmism, we resist any categorical dismissal of a future administrative state in which algorithmic automation guides, and even at times makes, key decisions. Instead, we urge that governmental reliance on machine learning should be approached with measured optimism about the potential benefits such technology can offer society by making government smarter and its decisions more efficient and just.

TABLE OF CONTENTS

INTRODUCTION

Algorithms are not new. For decades, they have served as integral components of every computer program.[1] But today, advanced machine-learning algorithms are creating a vastly automated society, transforming many facets of life. Many products and services, including email spam filters,[2] medical diagnoses,[3] product marketing,[4] and self-driving cars,[5] now depend on machine-learning algorithms and their ability to deliver astonishing forecasting power and speed. Today's algorithms are digital "robots"[6] that possess effectively autonomous abilities to adapt and learn.[7] As this type of artificial intelligence becomes more intricately woven into the economy, a new kind of information

1. In the most general of terms, algorithms can be defined as "well-defined set[s] of steps for accomplishing a certain goal." Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 640 n.14 (2017).

2. *See* Thiago S. Guzella & Walmir M. Caminhas, *A Review of Machine Learning Approaches to Spam Filtering*, 36 EXPERT SYSTEMS WITH APPLICATIONS 10206 (2009).

3. *See* W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 432–34 (2015).

4. *See* Cade Metz, *Now Anyone Can Tap the AI Behind Amazon's Recommendations*, WIRED (Apr. 9, 2015, 8:09 PM), http://www.wired.com/2015/04/now-anyone-can-tap-ai-behind-amazons-recommendations/ [https://perma.cc/PQ33-JBG2].

5. *See* Alexis C. Madrigal, *The Trick That Makes Google's Self-Driving Cars Work*, ATLANTIC (May 15, 2014), http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-that-makes-googles-self-driving-cars-work/370871/ [https://perma.cc/XBM4-V7FK].

6. The word "robot" has historically evoked an image of a physical entity directly interacting with the world, but recent advances in artificial intelligence have expanded that narrow conception of robotics. Here we primarily discuss virtual robots contained within computer programs, but the machine-learning algorithms underpinning them also increasingly support the functioning of physical robots; self-driving cars developed by Google, for instance, can predict the behavior of other cars and pedestrians on the road using machine learning. *Id.*

7. The CEO of Nvidia recently stated that "[i]t wasn't until the last few years that AI could do things that people can't do." Andrew Nusca, *The Current State of Artificial Intelligence, According to Nvidia's CEO*, FORTUNE (Mar. 22, 2016, 9:00 AM), http://fortune.com/2016/03/22/artificial-intelligence-nvidia/ [https://perma.cc/Q6YJ-LHG6]. He also noted how, over the past two years, the number of companies using deep learning, a prominent kind of machine learning, has grown thirty-five times larger. *Id.*

revolution is beginning that may lead to fundamental changes in society.[8] The ability of increasingly intelligent systems to replace human workers, for instance, has prompted widespread concern about the impact of machine learning on employment opportunities across a variety of occupational and professional fields.[9] According to some projections, artificial intelligence threatens to displace workers in "all routinized jobs and skill-based jobs that require the ability to perform diverse kinds of 'cognitive' labor, from physicians to reporters to stockbrokers."[10]

Even more ominously, some commentators worry about potentially new forms of human oppression that could stem from greater reliance on artificial intelligence. High-tech entrepreneur Elon Musk, for example, has warned that artificial intelligence presents the "biggest existential threat" to society, likening it to "summoning the demon."[11] Renowned physicist Stephen Hawking has

---

8. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 7 (2013).

9. *See generally* ERIK BRYNJOLFSSON & ANDREW MCAFEE, THE SECOND MACHINE AGE: WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES (2014) (describing how digital technologies have prompted shifts in employment and life and how individuals can respond to these shifts); TYLER COWEN, AVERAGE IS OVER: POWERING AMERICA BEYOND THE AGE OF THE GREAT STAGNATION (2013) (describing recent developments in the labor force and how technology has played into those changes); MARTIN FORD, RISE OF THE ROBOTS: TECHNOLOGY AND THE THREAT OF A JOBLESS FUTURE (2015); Claire Cain Miller, *As Robots Grow Smarter, American Workers Struggle to Keep Up*, N.Y. TIMES (Dec. 15, 2014), http://www.nytimes.com/2014/12/16/upshot/as-robots-grow-smarter-american-workers-struggle-to-keep-up.html [https://perma.cc/X2X3-3T8V]; Eduardo Porter, *Jobs Threatened by Machines: A Once 'Stupid' Concern Gains Respect*, N.Y. TIMES (June 7, 2016), http://www.nytimes.com/2016/06/08/business/economy/threatened-by-machines-a-once-stupid-concern-gains-respect.html [https://perma.cc/APL4-6JJH]; Jason Furman, Chairman, Council of Econ. Advisers, Is This Time Different? The Opportunities and Challenges of Artificial Intelligence, Remarks at AI Now: The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term (July 7, 2016), https://www.whitehouse.gov/sites/default/files/page/files/20160707_cea_ai_furman.pdf [https://perma.cc/D3M8-77QW].

10. JOHN MARKOFF, MACHINES OF LOVING GRACE: THE QUEST FOR COMMON GROUND BETWEEN HUMANS AND ROBOTS 327 (2015). Debates have begun to emerge over other scenarios made possible by machine learning, such as a future transportation network that may be based on self-driving cars. *See* Robert Hutchinson, *Driverless Cars: What Could Possibly Go Wrong?*, HARV. BUS. REV. (Jan. 15, 2016), https://hbr.org/2016/01/driverless-cars-what-could-possibly-go-wrong [https://perma.cc/XJ7U-DQVL]; Claire Cain Miller, *If Robots Drove, How Much Safer Would Roads Be?*, N.Y. TIMES (June 10, 2014), http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html [https://perma.cc/8NRU-LPJY]; Tom Vanderbilt, *Let the Robot Drive: The Autonomous Car of the Future Is Here*, WIRED (Jan. 20, 2012, 3:24 PM), http://www.wired.com/2012/01/ff_autonomouscars/all [https://perma.cc/YGN9-CSAK]. In addition, there is worry that algorithms are serving an inappropriate, editorial role in selecting which news items to display on social media feeds. *See* Jeffrey Herbst, *The Algorithm Is an Editor*, WALL ST. J. (Apr. 13, 2016, 6:09 PM), http://www.wsj.com/articles/the-algorithm-is-an-editor-1460585346 [https://perma.cc/FGW6-3BNF]. Finally, deep concerns about personal privacy have been implicated by increasing reliance on big data analytics; the amount of potentially personally identifiable information being collected by commercial firms keeps growing exponentially, and even if this information is maintained by separate entities, new methods of analysis make it increasingly easy to piece together disparate information to form highly personal portraits. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 8, at 152–57.

11. Samuel Gibbs, *Elon Musk: Artificial Intelligence Is Our Biggest Existential Threat*, GUARDIAN (Oct. 27, 2014, 6:26 AM), http://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat [https://perma.cc/SZD4-7WCL].

eerily forecasted that "[t]he development of full artificial intelligence could spell the end of the human race."[12] *New York Times* reporter John Markoff summarizes the views of a growing number of observers who worry that emerging "smart machines" risk constituting a new set of "masters" controlling humanity.[13]

These worries about artificial intelligence's impact on human liberty have surfaced over the use of machine learning by private-sector institutions such as banks and media companies, but presumably the prospect of governmental reliance on autonomous, self-learning robots only magnifies any perceived risks of digital oppression. Machine learning uses by defense, homeland security, and criminal law enforcement authorities have understandably begun to trigger alarm and prompt serious ethical consideration. Waging war through robots, for example, presents obviously novel concerns.[14] The use of algorithms to analyze telephone records and other personal information has also sparked extensive public controversy over privacy issues.[15] Similar controversies surround law enforcement agencies' use of machine learning to detect, respond to, and perhaps even predict crime.[16]

---

12. Rory Cellan-Jones, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC NEWS (Dec. 2, 2014), http://www.bbc.com/news/technology-30290540 [https://perma.cc/Q4YD-ZURC].

13. MARKOFF, *supra* note 10, at iv. One recent book characterizes modern algorithms as "weapons." *See generally* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016).

14. *See* Bill Keller, Opinion, *Smart Drones*, N.Y. TIMES (Mar. 16, 2013), http://www.nytimes.com/2013/03/17/opinion/sunday/keller-smart-drones.html [https://perma.cc/R3N9-ZTUH]; John Markoff, *Fearing Bombs That Can Pick Whom to Kill*, N.Y. TIMES (Nov. 11, 2014), http://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html [https://perma.cc/735E-LNZR].

15. *See* ACLU v. Clapper, 785 F.3d 787, 826 (2d Cir. 2015) (holding that the National Security Agency's bulk telephone metadata collection exceeded the scope of what Congress had authorized in the USA PATRIOT Act). On the political debate over national security surveillance, see, for example, Glenn Greenwald & Ewen MacAskill, *NSA Prism Program Taps in to User Data of Apple, Google and Others*, GUARDIAN (June 7, 2013, 3:23 PM), http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data [https://perma.cc/XM4Q-QE32]; Andrea Peterson, *NSA Reform Bill Passes House, Despite Loss of Support from Privacy Advocates*, WASH. POST (May 22, 2014), https://www.washingtonpost.com/news/the-switch/wp/2014/05/22/nsa-reform-bill-passes-house-despite-loss-of-support-from-privacy-advocates/ [https://perma.cc/Z52M-MTBP]. On legal issues raised by post-9/11 governmental surveillance, see, for example, Bruce Ackerman, *The Emergency Constitution*, 113 YALE L.J. 1029 (2004); Kim Lane Scheppele, *Law in a Time of Emergency: States of Exception and the Temptations of 9/11*, 6 U. PA. J. CONST. L. 1001 (2004); Daniel J. Solove, *Data Mining and the Security-Liberty Debate*, 75 U. CHI. L. REV. 343 (2008).

16. *See* Steven M. Bellovin et al., *When Enough Is Enough: Location Tracking, Mosaic Theory, and Machine Learning*, 8 N.Y.U. J.L. & LIBERTY 556, 575–76 (2014) (advancing the proposition that the use of machine learning to infer information about individuals is a new form of privacy invasion that needs to be squared with the Fourth Amendment). In addition, law enforcement authorities' increasing use of machine learning to predict crime—so-called predictive policing—has similarly raised concerns about racially-biased targeting as well as questions about the very possibility that individuals with a propensity toward criminality could be identified and punished for crimes that they have not yet committed. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 8, at 157–63; Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. (forthcoming 2017); Maurice Chammah with Mark

Notwithstanding the extensive attention given to the use of machine-learning algorithms by national security and criminal law enforcement agencies, the use of such artificial intelligence by other governmental institutions in the day-to-day operation of government has escaped sustained analysis. Granted, commentators have occasionally speculated about a fanciful future in which institutions like the Supreme Court[17] or the presidency[18] might be replaced by artificial intelligence. But such speculations are usually intended as absurdities. Most people believe that core governmental institutions and their decisions must be grounded in judgments made by real human beings: a "government of the people, by the people."[19] Even if machine-learning algorithms come to be widely accepted as substitutes for human control over automobiles or other functions in private life, a deeper, more fundamental suspicion about artificial intelligence will presumably remain about the use of machine learning in the governmental sphere. When it comes to making laws and other governmental decisions, the notion of using algorithms as substitutes for human decisions would appear to create a serious threat to democratic governance, conjuring images of unaccountable, computerized overlords.[20]

And yet, despite dire warnings about the dangers of runaway algorithmic governance, many aspects of public administration could undoubtedly benefit from the application of machine-learning algorithms, both today and in years to come. The vast work of administrative agencies, with their many routine regulatory responsibilities and adjudicatory processes, would seem ripe to

---

Hansen, *Policing the Future*, VERGE (Feb. 3, 2016, 8:01 AM), http://www.theverge.com/2016/2/3/10895804/st-louis-police-hunchlab-predictive-policing-marshall-project [https://perma.cc/2T87-VJCQ]; Matt Stroud, *The Minority Report: Chicago's New Police Computer Predicts Crimes, But Is It Racist?*, VERGE (Feb. 19, 2014, 9:31 AM), http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist [https://perma.cc/NX4G-5A99]; *see also* Maria Konnikova, *The Future of Fraud-Busting*, ATLANTIC (Mar. 2016), http://www.theatlantic.com/magazine/archive/2016/03/the-future-of-fraud-busting/426867/ [https://perma.cc/6CQM-X8QH] (discussing use of predictive analytics to identify future "con artists, and then interven[ing] before they cause trouble"). There have also been worries that current Fourth Amendment doctrine may be ill-equipped to handle law enforcement's prediction of criminal activity using machine learning. *See, e.g.*, Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871 (2016).

17. Sean Braswell, *All Rise for Chief Justice Robot!*, OZY (June 7, 2015), http://www.ozy.com/immodest-proposal/all-rise-for-chief-justice-robot/41131 [https://perma.cc/G86F-9WYP].

18. Maureen Dowd, *Beware Our Mind Children*, N.Y. TIMES (Apr. 25, 2015), http://www.nytimes.com/2015/04/26/opinion/sunday/maureen-dowd-beware-our-mind-children.html [https://perma.cc/NK4G-4896] ("Can [Alex Garland] envision an A.I. president, even more sleek and less emotive than the one we have now? 'There could be an A.I. president; there could,' he replies."); *Studio 360: The Computer as Artist*, WNYC RADIO & PRI (Nov. 12, 2015), http://www.wnyc.org/story/the-computer-as-artist/ [https://perma.cc/LEK8-GTWE] (stating that his so-called "creativity machine," an artificial intelligence system, should be used "in the White House").

19. Abraham Lincoln, Gettysburg Address (Nov. 19, 1863).

20. *Cf.* Ed Felten, *Accountable Algorithms*, FREEDOM TO TINKER (Sept. 12, 2012), https://freedom-to-tinker.com/2012/09/12/accountable-algorithms/ [https://perma.cc/TV9N-CGYA] (describing how the use of algorithmic decision making in lieu of human decision making may necessitate making these algorithms more technically accountable).

benefit from such automation. These agencies fulfill important oversight responsibilities over vital aspects of everyday life, from regulating the safety of the food we eat to setting the rules by which financial markets operate. Officials in these agencies must make an array of crucial judgments on a daily basis that are not unlike the kinds of judgments that machine learning has so clearly helped improve in the private sector.[21] Moreover, with the private sector increasingly relying on algorithms to make faster, more precise decisions, the increased speed and complexity of economic activity in the machine-learning era surely demands that government agencies keep pace and make use of the same analytic tools in order to regulate the private sector more effectively.[22] If machine learning can help regulatory agencies make smarter, more accurate decisions, the benefits to society could be considerable.

But can the prospect of the government regulating by robot, or adjudicating by algorithm, be accommodated within prevailing legal norms? Fitting machine learning into the regulatory state may turn out to be one of the most fundamental challenges facing the U.S. governmental system in the decades to come. Regulating by robot would hardly seem, at first glance, to fit naturally within prevailing principles of administrative law. That law, after all, is built on the assumption that governmental decisions will be made by humans. Such an assumption has led to a variety of legal constraints aimed at ensuring administrative decisions are democratically accountable and fair. Indeed, even with humans at the helm, administrative agencies are already thought to constitute a vast bureaucratic "machinery" of government with an ever-present potential to grow detached from and unaccountable to the rest of society.[23] If many bureaucrats' jobs come to be replaced by automated systems, the likelihood of an even more detached, possibly despotic, administrative government would only seem to increase.[24]

---

21. *See supra* notes 2–7.

22. Cary Coglianese, *Optimizing Government for an Optimizing Economy*, *in* SECTION 8: POLITICS, NEW ENTREPRENEURIAL GROWTH AGENDA (2016), http://www.kauffman.org/neg/section-8#optimizinggovernmentforanoptimizingeconomy [https://perma.cc/EUK2-EYSR]. Governments themselves are increasingly recognizing the need to oversee artificial intelligence itself. For example, the UK's House of Commons Science and Technology Committee recently launched an inquiry into "the social, legal, and ethical issues raised by developments of . . . artificial intelligence." Science and Technology Committee (Commons), *Robotics and Artificial Intelligence Inquiry Launched*, U.K. PARLIAMENT (Mar. 24, 2016), http://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2015/robotics-and-artificial-intelligence-inquiry-launch-15-16/ [https://perma.cc/VQ49-9VQM]. Also, the White House Office of Science and Technology Policy hosted a series of public workshops in 2016 about the risks of machine learning, including its uses in government. Ed Felten, *Preparing for the Future of Artificial Intelligence*, WHITE HOUSE BLOG (May 3, 2016, 3:01 PM), https://www.whitehouse.gov/blog/2016/05/03/preparing-future-artificial-intelligence [https://perma.cc/3G6G-KRZF].

23. *See generally* F.A. HAYEK, THE ROAD TO SERFDOM (1944); THEODORE J. LOWI, THE END OF LIBERALISM: THE SECOND REPUBLIC OF THE UNITED STATES (2d ed. 1979); DAVID SCHOENBROD, POWER WITHOUT RESPONSIBILITY: HOW CONGRESS ABUSES THE PEOPLE THROUGH DELEGATION (1993).

24. *Cf.* Free Enter. Fund v. Pub. Co. Accounting Oversight Bd., 561 U.S. 477, 497 (2010) ("The diffusion of power carries with it a diffusion of accountability.").

When contemplating the use of robotic algorithms, public officials, lawyers, and judges should ask how well the use of machine learning will conform to well-established legal principles of constitutional and administrative law. In this Article, we address significant but previously unanalyzed legal questions raised by machine learning. In particular, we consider how nonhuman decision tools would have to be used to comport with the nondelegation doctrine and with rules about due process, antidiscrimination, and governmental transparency. Although administrative agencies are still only beginning to use machine-learning algorithms, today's widespread concern about the robotic control of other facets of life makes it opportune to examine carefully the potential for machine learning's use by government agencies.[25]

Given the relative complexity of modern machine-learning algorithms, we begin in Part I by providing a brief, non-technical summary of how these algorithms operate. Drawing on this background, we also identify existing and likely future applications of machine learning within the administrative state. In Part II, we provide a legal analysis of such algorithmic applications by federal administrative agencies, considering whether various methods of implementing machine learning will likely offend principles of nondelegation, due process, antidiscrimination, and transparency. We conclude that government agencies should be able to rely on autonomous, learning decision tools without running afoul of legal standards. Contrary to popular perceptions, machine learning will not lead to a runaway government, as a series of technical limitations preclude a future in which complete regulatory or adjudicatory power can be ceded to autonomous algorithms over which humans exert little control. When used thoughtfully, these machine-learning applications should not offend the core legal foundations of the regulatory state.

---

25. We know of no other work of scholarship that comprehensively analyzes the constitutional and administrative law issues implicated by machine learning in the regulatory state. Admittedly, for more than a decade it has been clear that "many possibilities exist for applying information technologies in new ways to government rulemaking." Cary Coglianese, *E-Rulemaking: Information Technology and the Regulatory Process*, 56 ADMIN. L. REV. 353, 369 (2004). A few scholars have recently identified ways that machine learning could be used by regulatory agencies to improve regulatory outcomes. *See* Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards* (Univ. of Chicago Public Law Working Paper No. 550, 2015), http://ssrn.com/abstract=2693826 [https://perma.cc/9YDG-RBRC] (noting how agencies may implement machine learning to let entities know whether their discrete actions would be aligned with broad legislatively-established standards); Joshua Mitts, Predictive Regulation (June 27, 2014) (unpublished manuscript), http://ssrn.com/abstract=2411816 [https://perma.cc/GVU5-VU5D] (discussing the use of machine learning to identify market trends indicating future need for regulation). Moreover, some recent or forthcoming papers provide some initial consideration of accountability or fairness issues related to the use of algorithms, broadly defined, across a range of decision-making settings. *See, e.g.*, Kroll et al., *supra* note 1 (briefly taking up machine-learning algorithms in the context of discrimination, but mainly focusing on algorithms more generally); Mariano-Florentino Cuéllar, *Cyberdelegation and the Administrative State* 1 (Stanford Public Law Working Paper No. 2754385, 2016), http://ssrn.com/abstract=2754385 [https://perma.cc/3EYG-62WJ]) (exploring "some of the questions and trade-offs associated with delegating administrative agency decisions to computer algorithms").

After analyzing the administrative law implications of machine learning, we step back in Part III to reflect on the policy merits of applying machine learning to administrative tasks. We do recognize that machine learning could be implemented irresponsibly in ways that, even though legal, might still offend more conventional notions of good government.[26] As with any tool, artificial intelligence could be misused. And by no means should anyone think that machine learning constitutes a panacea for government. The use of machine learning in specific settings and for particular purposes may well prove inadvisable once all things are considered. The way algorithms are used will matter too, and at times some safeguards may need to be implemented to ensure consonance with the broader purposes standing behind the administrative law doctrines we consider here. In general, though, governmental decision making in the machine-learning era can and, in appropriate circumstances, should take advantage of the enhanced public value that can be achieved from regulating by robot and adjudicating by algorithm. Our detailed examination of the legal issues and policy tradeoffs leads us to be reasonably optimistic about machine learning and the instrumental role it can play in making a more promising future for administrative government.

I. MACHINE LEARNING AND ITS ADMINISTRATIVE APPLICATIONS

Some of the most prominent examples of private-sector growth today—from Amazon[27] to Zillow[28]—depend on the use of machine learning to optimize production processes, supply chains, marketing, and the pricing of goods and services. Machine learning undergirds future growth across a wide range of sectors, from the introduction of "fintech" firms in the financial industry[29] to advances in healthcare delivery via precision medicine.[30] Machine learning's main attraction stems from how well it "outperforms human intelligence."[31] As

26. *Cf.* Adrian Vermeule, *Conventions of Agency Independence*, 113 COLUM. L. REV. 1163 (2013) (describing conventions and norms surrounding the administrative state, particularly agency independence).

27. *See* Metz, *supra* note 4 (describing Amazon's use of machine learning to recommend products to consumers).

28. *See* Jessica Davis, *Zillow Uses Analytics, Machine Learning to Disrupt with Data*, INFORMATIONWEEK (Oct. 14, 2016, 11:06 AM), http://www.informationweek.com/big-data/zillow-uses-analytics-machine-learning-to-disrupt-with-data/d/d-id/1327175 [https://perma.cc/8UTD-QX6S] (describing Zillow's use of machine learning to provide forecasts of housing prices).

29. *See* Falguni Desai, *The Age of Artificial Intelligence in Fintech*, FORBES (June 30, 2016, 10:42 PM), http://www.forbes.com/sites/falgunidesai/2016/06/30/the-age-of-artificial-intelligence-in-fintech [https://perma.cc/EK89-DD3Y] (describing how fintech firms use artificial intelligence to improve investment strategies and analyze consumer financial activity).

30. For a discussion of these and other examples of an increasingly optimizing economy, see Coglianese, *supra* note 22.

31. NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 11 (2014). Of note, machine learning recently bested human intelligence in the incredibly complex game of Go. *See* David Silver et al., *Mastering the Game of Go with Deep Neural Networks and Tree Search*, 529 NATURE 484, 484 (2016).

private firms pursue significant efficiency gains through the kind of smarter and more contextualized decisions made possible by algorithmic analysis of big data, the government will undoubtedly need to follow suit, not merely to keep up with new risks these private-sector uses of machine learning might bring, but also to improve government's ability to address a host of existing risks and regulatory problems.[32] Machine learning promises to make the government, like the private sector, smarter and more efficient. In this Part, we introduce machine learning and discuss how government agencies are already beginning to explore its use to optimize administrative tasks, an endeavor that is likely to grow both in size and scope in the years ahead. We first explain what machine learning is and describe its distinguishing features. We then discuss how agencies are already using machine learning. Finally, we show how this technology could, in the future, potentially transform the administrative state through what we call "rulemaking by robot" and "adjudicating by algorithm."

A. WHAT IS MACHINE LEARNING?

Fundamentally, machine-learning algorithms are used to make predictions. This emphasis on *prediction* contrasts markedly with traditional statistical techniques which seek to *model* underlying data-generating processes in the real world. Although traditional statistical techniques can also generate predictions, they do so only when the model created by the analyst fits well with the underlying processes being modeled. These traditional techniques require the analyst first to specify a mathematical equation expressing an outcome variable as a function of selected explanatory variables put together in a particular way, and then to see how well the data fit with the analyst's choices. For example, when analysts employ the traditional techniques of ordinary least squares regression or logistic regression, they specify equations that represent their a priori beliefs about the functional relationships that exist between *independent* (or explanatory) and *dependent* (or outcome) variables. What regression does, in essence, is estimate the magnitude and direction of these relationships between the two types of variables that are selected and specified by the analyst. The relationships in the statistical model ostensibly represent the relationships in the real world, which is why regression results are often used to support causal inferences.

By contrast, machine learning is nonparametric in that it does not require the researcher to specify any particular functional form of a mathematical model in advance. Instead, these algorithms allow the data themselves to dictate how information contained in *input* variables is put together to forecast the value of

32. *See* Coglianese, *supra* note 22. A senior strategist at the National Security Agency has noted how advanced analytic techniques are particularly essential for agencies that will soon be forced "to operate at cyberspeed and at scale." GOV'T BUS. COUNCIL, GOV'T EXEC. MEDIA GRP., DATA ANALYTICS: A STRATEGIC ASSET TO GOVERNMENT 2 (2015), http://cdn.govexec.com/media/gbc/docs/2015-09-03_qlik_issue_brief_designed_4.pdf [https://perma.cc/LY76-FY5V].

an *output* variable.[33] Machine-learning algorithms do not generate quite the same kind of information on the magnitude or direction of the effects that might be associated with any single input variable on the output variable, controlling for the other variables. The functional relationships in machine learning are not necessarily the complete set of those in nature's true data-generating process. As a result, no claim can be made that the machine-learning process represents any set of true relationships in the world, and thus none of the causal inferences that typically characterize statistical modeling can be applied to results of machine learning. In short, with machine-learning results, causal relationships between inputs and outputs may simply not exist, no matter how intuitive such relationships might look on the surface. If a machine-learning algorithm tends to forecast that older individuals commit fewer crimes than younger individuals, for example, it cannot be claimed on the basis of the machine-learning process that older age causes any reduction in the propensity to commit crimes.[34]

Nevertheless, from a technical standpoint, machine learning's distinctive predictive and nonparametric focus turns out to be paramount to its impressive usefulness in generating reliable forecasts. Also of central importance, and what gives machine learning its name, is how such algorithms mathematically "learn" to generate their predictions. There are many machine-learning algorithms that do so in different mathematical ways, but they all attempt, as one textbook explains, to "optimize a performance criterion using example data or past experience."[35] In other words, these algorithms make repeated passes through data sets, progressively modifying or averaging their predictions to optimize specified criteria.

To illustrate this functioning, consider a common application of machine learning that has proven critical to improvements in the government's handling of postal mail and other paperwork-processing tasks—the recognition and classification of handwritten digits.[36] In this simple application, an algorithm's performance criterion, or objective function, is classification accuracy—that is, how often it correctly recognizes, say, a handwritten number two as a two. To

33. *See* RICHARD A. BERK, STATISTICAL LEARNING FROM A REGRESSION PERSPECTIVE 13 (2008).

34. For a discussion of the inferential value of outputs from machine-learning algorithms, see *id.* at 9–17.

35. ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING 3 (2d ed. 2010). In the broader field, the varied types of machine learning are referred to by a dizzying array of different terms, some technical, some colloquial, for example: smart machines, expert systems, neural networks, deep learning, hierarchical learning, reinforcement learning, structured learning, and more. Although we explain some of these different terms in this Part, for the most part throughout this Article we use the terms "machine learning," "algorithms," and "artificial intelligence" for convenience to capture all possible variations in terms, as we are concerned primarily with the legal issues surrounding the general use of this family of techniques.

36. *See, e.g.*, Cheng-Lin Liu et al., *Handwritten Digit Recognition: Benchmarking of State-of-the-Art Techniques*, 36 PATTERN RECOGNITION 2271 (2003); Y. LeCun et al., Comparison of Learning Algorithms for Handwritten Digit Recognition, Presented at the International Conference on Artificial Neural Networks (1995), http://yann.lecun.com/exdb/publis/pdf/lecun-95b.pdf [https://perma.cc/2NGQ-85FS].

perform this classification, an algorithm must "learn" what aspects of a handwritten digit make it likely to be a two. Over the course of iterative passes through the data, such an algorithm tries to use many different mathematical descriptions of shapes, as well as relationships of shapes, in the pictures of handwritten digits to make its classifications. If a particular descriptive method is optimal, the algorithm will be "rewarded" with a low error rate; if the descriptions are not optimal, the algorithm will be "punished" with a high error rate. It can learn, for example, that a handwritten digit is likely to be a two if the topmost section of the digit depicted is semicircular and facing downward. Ultimately, the algorithm will seek to make classifications based on mathematical descriptions of shapes that yield the lowest error rates.[37]

This handwriting recognition example provides an illustration of machine-learning algorithms applied to classification problems, where the goal is to sort objects into classes. But classification problems represent only some of the diverse applications of machine-learning techniques. Machine-learning algorithms can also be used to predict numerical values, such as house prices or stock market index values—endeavors that are often termed regression problems.[38] They also can be applied to scenarios, such as playing chess, where an algorithm can be used to determine the optimal sequence of actions.[39] Variety in the types of machine-learning algorithms means that they can be used in a wide variety of predictive endeavors.

Admittedly, many "non-learning" techniques have long been used to pursue these same endeavors. For example, ordinary least squares regression can estimate numerical outcomes, and logistic regression is commonly used as a binary classifier. Given the existence of these alternative statistical techniques, what advantages do machine-learning algorithms offer? Put simply, they outperform standard procedures in terms of predictive accuracy and statistical efficiency (that is, the increased ability to obtain predictions with both low bias and low variance).[40] Furthermore, many phenomena that analysts want to forecast

37. This is an example of supervised machine learning, where each handwritten digit is labeled with its correct digit so that the algorithm knows when it is making errors. Unsupervised learning uses unlabeled data, so that performance criteria being optimized are not measures of error rates, because the truth is not known, but measures of similarity between digits determined by the algorithm to be the same. *See* ALPAYDIN, *supra* note 35, at 11–13.

38. *See id.* at 9–11.

39. *See id.* at 13–14; *see also* Cade Metz, *In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go*, WIRED (Jan. 27, 2016, 1:00 PM), http://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/ [https://perma.cc/9YQ9-LHM4].

40. These benefits result from the mathematical techniques of boosting or bagging (or both). For demonstrations of these benefits resulting from boosting, see Robert E. Schapire, The Boosting Approach to Machine Learning: An Overview, Presented at the MSRI Workshop on Nonlinear Estimation and Classification (2002), https://www.cs.princeton.edu/picasso/mats/schapire02boosting_schapire.pdf [https://perma.cc/64PZ-PY57]. For an illustration of similar benefits resulting from bagging, see Leo Breiman, *Some Infinity Theory for Predictor Ensembles* (U.C. Berkeley Technical Report 577, 2000), https://www.stat.berkeley.edu/breiman/some_theory2000.pdf [https://perma.cc/KZ3D-RWMY]; Peter Bühlmann & Bin Yu, Explaining Bagging, Presented at the Seminar für Statistik (2000),

are extraordinarily complex, and analysts often lack the a priori knowledge necessary to specify an accurately forecasting conventional model. By eschewing this dependency on existing knowledge and the need to identify the functional form of any relationships, machine learning can apply to a wider range of problems and yield vastly enhanced accuracy over its alternatives, whether human intuition, expert judgment, or traditional statistical techniques.[41] Learning algorithms can also adapt more dynamically; as new data become available, they can search for new patterns and thereby improve forecasting accuracy.

Although machine-learning algorithms are known and prized for their accuracy, this benefit does come at an interpretive cost. This cost is frequently invoked by references to machine-learning algorithms as "black-box" procedures.[42] The black-box nature of machine learning holds important implications for administrative law, so to understand this feature of machine learning consider again the classification of handwritten digits. We said that an algorithm might learn that certain geometric characteristics of the shapes of handwritten digits are useful for determining which digits they represent—yet we cannot really know what precise characteristics any machine-learning algorithm is keying in on. Machine-learning algorithms transform a series of inputs to a series of outputs by optimizing a performance criterion, but that is where the analyst's easy ability to interpret the algorithms' workings comes to an end. The user of an algorithm cannot really discern which particular relationships between variables factor into the algorithm's classification, or at which point in the algorithm they do, nor can the user determine how exactly the algorithm puts together various relationships to yield its classifications.[43] For this reason, machine-learning algorithms are often described as transforming inputs to outputs through a black box. An analyst cannot look inside the black box to understand how that transformation occurs or describe the relationships with the same intuitive and causal language often applied to traditional statistical

---

http://e-collection.library.ethz.ch/eserv/eth:23905/eth-23905-01.pdf [https://perma.cc/CPE2-QEAF]. For a demonstration of the benefits resulting from combined classifiers achieved through both bagging and boosting, see V. Koltchinskii & D. Panchenko, *Empirical Margin Distributions and Bounding the Generalization Error of Combined Classifiers*, 30 ANNALS STAT. 1 (2002).

41. *See, e.g.*, Volodymyr Mnih et al., *Human-Level Control Through Deep Reinforcement Learning*, 518 NATURE 529, 529 (2015); Kaiming He et al., Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, Presented at the IEEE International Conference on Computer Vision (2015), http://arxiv.org/abs/1502.01852 [https://perma.cc/5ERE-L2HK]; Bo Pang et al., Thumbs Up? Sentiment Classification Using Machine Learning Techniques, Presented at the Conference on Empirical Methods in Natural Language Processing (2002), http://www.cs.cornell.edu/home/llee/papers/sentiment.pdf [https://perma.cc/G4VS-XYM2].

42. *See, e.g.*, Leo Breiman, *Statistical Modeling: The Two Cultures*, 16 STAT. SCI. 199, 199 (2001).

43. For example, a common machine-learning algorithm known as random forests generates its predictions by, roughly speaking, producing thousands of decision trees (called classification or regression trees) and then averaging predictions across all trees. *See* Leo Breiman, *Random Forests*, 45 MACHINE LEARNING 5, 5 (2001). The analyst can examine the structure of a particular tree and determine to some extent how variables and interactions between variables functionally affect the predictions, but this will tell the analyst nothing about all such processes in the forest as a whole.

modeling.[44]

Despite this interpretive limitation, machine-learning algorithms have been implemented widely in private-sector settings. Companies desire the savings in costs and efficiency gleaned from these techniques, and the lack of intuitive interpretability is of little concern in endeavors where accuracy, not causality, is the valued metric. Netflix, for instance, employs a form of machine learning called "artificial neural networks" to suggest entertainment options to its customers based on their prior viewing habits.[45] Google uses machine learning to identify house numbers in its Street View imagery,[46] to save energy in its data centers,[47] and to keep its self-driving cars from crashing.[48] Machine learning has also shown great utility in the financial sector, where it is employed to predict the value of investments and financial instruments.[49] The benefits of learning algorithms have also promoted their adoption in academic research in disciplines closely connected to policymaking, where predictive accuracy is critical. For example, researchers have shown that machine-learning algorithms can help predict the propensity of probationers and parolees to commit violent crimes,[50] estimate population densities of homeless persons in cities,[51] and forecast student retention at universities.[52] In these ways, both private businesses and academic researchers have embraced machine learning, and machine-learning applications in a wide variety of settings are already actively shaping society.

B. EXISTING ADMINISTRATIVE APPLICATIONS

For much the same reason that machine learning has been exploited in the private sector, its use holds potentially great value to government agencies. We

---

44. *See* Breiman, *supra* note 42, at 199–200.

45. Alex Chen et al., *Distributed Neural Networks with GPUs in the AWS Cloud*, NETFLIX TECH BLOG (Feb. 10, 2014), http://techblog.netflix.com/2014/02/distributed-neural-networks-with-gpus.html [https://perma.cc/T7CR-7QGM].

46. Ian J. Goodfellow et al., *Multi-Digit Number Recognition from Street View Imagery Using Deep Convolutional Neural Networks*, CORNELL UNIV. LIB. COMP. VISION & PATTERN RECOGNITION (2013), http://arxiv.org/abs/1312.6082 [https://perma.cc/H7KK-DZJT].

47. Joe Kava, *Better Data Centers Through Machine Learning*, GOOGLE OFFICIAL BLOG (May 28, 2014), http://googleblog.blogspot.com/2014/05/better-data-centers-through-machine.html [https://perma.cc/ZEP4-2Z8M].

48. *See* Madrigal, *supra* note 5.

49. *See* Quentin Hardy, *Wealth Managers Enlist Spy Tools to Map Portfolios*, N.Y. TIMES, (Aug. 3, 2014), http://www.nytimes.com/2014/08/04/technology/wealth-managers-enlist-spy-tools-to-map-portfolios.html [https://perma.cc/K4F4-VZEE]; *see also* MICHAEL LEWIS, FLASH BOYS: A WALL STREET REVOLT 36 (2014).

50. *See, e.g.*, Richard Berk et al., *Forecasting Murder Within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning*, 172 J. ROYAL STAT. SOC'Y SERIES A (STAT. IN SOC'Y) 191 (2009).

51. *See, e.g.*, Brian Kriegler & Richard Berk, *Small Area Estimation of the Homeless in Los Angeles: An Application of Cost-Sensitive Stochastic Gradient Boosting*, 4 ANNALS APPLIED STAT. 1234 (2010).

52. *See, e.g.*, Dursun Delen, *A Comparative Analysis of Machine Learning Techniques for Student Retention Management*, 49 DECISION SUPPORT SYS. 498 (2010).

have already noted that national security and law enforcement agencies are starting to rely on machine learning to support functions as varied as assessing risks of street crime and automating weapons delivery systems. Outside the security and law enforcement context, other government agencies have also begun to explore uses of machine learning, revealing growing recognition of its promise across a variety of policy settings and at all levels of government.[53]

Although we mainly focus in this Article on the use of machine learning by the federal government, the nation's largest cities have received much attention so far for their embrace of machine learning and its potential to improve governmental efficiency and effectiveness.[54] The City of Chicago, for example, has established an award-winning SmartData Platform initiative through which city officials are using machine learning to support a range of city services, from identifying restaurants that should be inspected[55] to predicting where and when rodent control bait should be placed throughout the city.[56] New York City has established a Mayor's Office of Data Analytics,[57] which, among other things, is working with the city's fire department to use machine learning to decide where to send building inspectors.[58] Flint, Michigan has partnered with Google and the University of Michigan to address its recent water crisis by targeting pipe replacements based on machine-learning predictions of lead contamination.[59] The City of Los Angeles has installed sensors in all of its streets that continuously feed data into a machine-learning system that automatically determines when traffic signals should turn red or green to optimize traffic

---

53. Although outside our scope, we note that courts are also increasingly looking to machine learning as a tool for discovery in the litigation process. *See, e.g.*, Wallis M. Hampton, *Predictive Coding: It's Here to Stay*, SKADDEN (May 5, 2014), https://www.skadden.com/sites/default/files/publications/LIT_JuneJuly14_EDiscoveryBulletin.pdf [https://perma.cc/4WHG-ZLKF] (noting courts' increasing interest in and favorable inclination toward the use of machine learning to assist in e-discovery).

54. *See, e.g.*, STEPHEN GOLDSMITH & SUSAN CRAWFORD, THE RESPONSIVE CITY: ENGAGING COMMUNITIES THROUGH DATA-SMART GOVERNANCE (2014); Bechara Choucair, Jay Bhatt & Raed Mansour, *How Cities Are Using Analytics to Improve Public Health*, HARV. BUS. REV. (Sept. 15, 2014), https://hbr.org/2014/09/how-cities-are-using-analytics-to-improve-public-health/ [https://perma.cc/R26N-7RU2].

55. Nick Rojas, *Chicago and Big Data*, TECHCRUNCH (Oct. 22, 2014), http://techcrunch.com/2014/10/22/chicago-and-big-data/ [https://perma.cc/P47F-9JKV]; *see also* Edward L. Glaeser et al., *Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy*, 106 AM. ECON. REV. 114, 114 (2016). Interestingly, similar restaurant hygiene algorithms deployed in Boston were developed via crowdsourcing, a potentially cost-effective alternative to private contracting in certain situations. *Id.*

56. Ash Center Mayors Challenge Research Team, *Chicago's SmartData Platform: Pioneering Open Source Municipal Analytics*, DATA-SMART CITY SOLUTIONS (Jan. 8, 2014), http://datasmart.ash.harvard.edu/news/article/chicago-mayors-challenge-367 [https://perma.cc/MY8X-PDD6].

57. THE MAYOR'S OFFICE OF DATA ANALYTICS, CITY OF NEW YORK, http://www1.nyc.gov/site/analytics/index.page [https://perma.cc/4QTF-M3UL].

58. Brian Heaton, *New York City Fights Fire with Data*, GOV'T TECH. (May 15, 2015), http://www.govtech.com/public-safety/New-York-City-Fights-Fire-with-Data.html [https://perma.cc/XG7W-5UYH].

59. Gabe Cherry, *Google, U-M to Build Digital Tools for Flint Water Crisis*, U. MICH. NEWS (May 3, 2016), http://ns.umich.edu/new/multimedia/videos/23780-google-u-m-to-build-digital-tools-for-flint-water-crisis [https://perma.cc/GV4C-LR6N].

flow.[60]

At the federal level, one of the earliest domestic applications of machine learning came from, as we already noted, the U.S. Postal Service's need for a method to sort mail automatically by predicting the zip codes written on envelopes.[61] Meteorologists within the National Oceanic and Atmospheric Administration have explored the use of machine learning to improve forecasts of severe weather events.[62] Other federal agencies have also started to rely on machine learning to support various regulatory and administrative activities.[63]

Analysts at the U.S. Environmental Protection Agency (EPA), for example, have developed a program called ToxCast to help the agency predict toxicities of chemical compounds.[64] Chemical toxicity has traditionally been established using animal testing, but these laboratory techniques are costly and time consuming, not to mention often harmful to animals. Faced with tens of thousands of chemicals that could be potentially subject to EPA regulation, the agency developed ToxCast to prioritize which of the multitude of chemicals in production should undergo more in-depth testing. ToxCast applies machine-learning algorithms—specifically, linear discriminant analysis—to data on chemicals' interactions obtained from *in vitro* testing to predict their toxicities.[65] In one

60. Ian Lovett, *To Fight Gridlock, Los Angeles Synchronizes Every Red Light*, N.Y. TIMES (Apr. 1, 2013), http://www.nytimes.com/2013/04/02/us/to-fight-gridlock-los-angeles-synchronizes-every-red-light.html [https://perma.cc/25T2-DCFG]; David Z. Morris, *How Swarming Traffic Lights Could Save Drivers Billions of Dollars*, FORTUNE (July 13, 2015, 4:47 PM), http://fortune.com/2015/07/13/swarming-traffic-lights [https://perma.cc/KP8C-N88S].

61. By 1988, USPS contractors had developed one of the first methods for extracting visual images from envelopes and compiling them into an analyzable data set. *See* Ching-Huei Wang & Sargur N. Srihari, *A Framework for Object Recognition in a Visually Complex Environment and Its Application to Locating Address Blocks on Mail Pieces*, 2 INT'L J. COMP. VISION 125, 125 (1988). This data set then enabled the development of early algorithms analyzing handwritten zip codes. *See* O. Matan et al., Handwritten Character Recognition Using Neural Network Architectures, Presented at the 4th USPS Advanced Technology Conference (1990), http://yann.lecun.com/exdb/publis/pdf/matan-90.pdf [https://perma.cc/P4LS-5HZH].

62. David John Gagne II et al., Day-Ahead Hail Prediction Integrating Machine Learning with Storm-Scale Numerical Weather Models, Presented at the Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence (2015), http://www.aaai.org/ocs/index.php/IAAI/IAAI15/paper/view/9724/9898 [https://perma.cc/ZZ97-3UCJ].

63. More than a decade ago, the General Accountability Office surveyed 128 federal agencies and found that fifty-two of them were engaged in "data mining" activities, defined broadly "as the application of database technology and techniques—such as statistical analysis and modeling—to uncover hidden patterns and subtle relationships in data and to infer rules that allow for the prediction of future results." U.S. GEN. ACCOUNTING OFFICE, GAO-04-548, DATA MINING: FEDERAL EFFORTS COVER A WIDE RANGE OF USES 4 (2004). It is not clear from the report how many of these efforts involved machine-learning techniques as opposed to more traditional statistical methods. We know of no comparable effort to survey agencies across government about their use of machine learning.

64. U.S. ENVTL. PROT. AGENCY, TOXCAST FACT SHEET (2013), http://www.epa.gov/sites/production/files/2013-12/documents/toxcast-fact-sheet.pdf [https://perma.cc/P3UU-YL4X].

65. Robert Kavlock et al., *Update on EPA's ToxCast Program: Providing High Throughput Decision Support Tools for Chemical Risk Management*, 25 CHEMISTRY RES. TOXICOLOGY 1287, 1295 (2012). On the use of machine learning in toxicology, see Huanxiang Liu, Xiaojun Yao & Paola Gramatica, *The Applications of Machine Learning Algorithms in the Modeling of Estrogen-Like Chemicals*, 12 COMBINATORIAL CHEM. & HIGH THROUGHPUT SCREENING 490 (2009).

application during ToxCast's first phase, analysts estimated that using machine learning could save the government $980,000 per toxic chemical positively identified.[66] Although the EPA presently uses ToxCast to identify chemicals for additional testing through more traditional means, its underlying predictive approach could eventually form an independent basis for justifying the imposition of regulatory controls.[67]

The U.S. Internal Revenue Service (IRS) has also used machine-learning algorithms to aid its auditing and enforcement functions. In 2001, it began developing a "risk-based collection model" that prioritized the IRS's collection cases for small businesses and self-employed taxpayers by using machine-learning algorithms, including neural networks, to predict risk of nonpayment.[68] In that same year, the agency began to use support vector machines, another type of machine-learning algorithm, to predict abuse and fraud in tax returns and to allocate cases for human review based on the probability of abuse and the magnitude of the dollar amount of the abuse.[69] More recently, in 2009, the IRS launched an Information Reporting and Document Matching program, which applies algorithms to credit card and other third-party data to predict tax underreporting and non-filing by businesses.[70] The IRS increased its requested funding for enforcement targeting from $1.4 million in 2012[71] to over $39 million in 2016,[72] specifically to develop better ways to use machine-learning algorithms, including neural networks, to "identify emerging areas of non-compliance."[73]

---

66. Matthew T. Martin et al., *Economic Benefits of Using Adaptive Predictive Models of Reproductive Toxicity in the Context of a Tiered Testing Program*, 58 SYS. BIOLOGY REPROD. MED. 3, 4–6 (2012).

67. Richard S. Judson et al., *Estimating Toxicity-Related Biological Pathway Altering Doses for High-Throughput Chemical Risk Assessment*, 24 CHEM. RES. TOXICOLOGY 451, 457–60 (2011).

68. Jane Martin & Rick Stephenson, Risk-Based Collection Model Development and Testing, Presented at the Internal Revenue Service Research Conference (2005), http://www.irs.gov/pub/irs-soi/05stephenson.pdf [https://perma.cc/M65K-D9D7].

69. David DeBarr & Maury Harwood, Relational Mining for Compliance Risk, Presented at the Internal Revenue Service Research Conference (2004), http://www.irs.gov/pub/irs-soi/04debarr.pdf [https://perma.cc/Y9F8-RWNK].

70. *See* CHRIS WAGNER ET AL., TAXPAYER ADVOCATE SERV., *IRS Policy Implementation Through Systems Programming Lacks Transparency and Precludes Adequate Review*, *in* 2010 ANNUAL REPORT TO CONGRESS 71, 76, http://www.irs.gov/pub/irs-utl/2010arcmsp5_policythruprogramming.pdf [https://perma.cc/3DHR-ZXAT]. Interestingly, this program was temporarily halted in 2014 not because of issues with its predictive algorithms, but due to continued reliance on human tax examiners; the Automated Underreporter System that predicts underreporting was successfully deployed, but the Case Management System that then handed flagged cases to humans for further examination did not meet technical requirements. *See* TREASURY INSPECTOR GEN. FOR TAX ADMIN., 2014-20-088, THE INFORMATION REPORTING AND DOCUMENT MATCHING CASE MANAGEMENT SYSTEM COULD NOT BE DEPLOYED (2014), https://www.treasury.gov/tigta/auditreports/2014reports/201420088fr.pdf [https://perma.cc/E3CS-VLEL].

71. U.S. INTERNAL REVENUE SERV., FY 2012 BUDGET REQUEST 63 (2011), https://www.treasury.gov/about/budget-performance/Documents/CJ_FY2012_IRS_508.pdf [https://perma.cc/GB72-GXW5].

72. U.S. INTERNAL REVENUE SERV., FY 2016 PRESIDENT'S BUDGET 88 (2015), https://www.treasury.gov/about/budget-performance/CJ16/02-06.%20IRS%20FY%202016%20CJ.pdf [https://perma.cc/2BXV-DPFQ].

73. U.S. INTERNAL REVENUE SERV., *supra* note 71, at 63.

In addition to the EPA's and the IRS's use of machine learning, the U.S. Food and Drug Administration (FDA) has conducted research on the use of machine-learning techniques to extract information about known equipment failures, errors, or other adverse events from medical device reports.[74] This safety agency is also currently engaged in a five-year collaborative research agreement with the Massachusetts Institute of Technology (MIT) focusing on "artificial intelligence, advanced statistical machine learning and data mining methods."[75] MIT researchers have also recently collaborated with researchers at the U.S. Department of the Treasury's Office of Financial Research (OFR) to survey methods of evaluating systemic risk in consumer credit markets, including the use of classification and regression trees.[76] Separately, academic researchers have demonstrated how machine-learning algorithms can be used to predict cases of financial statement fraud,[77] electoral fraud,[78] and even illegal fishing practices.[79] Agencies like the Commodity Futures Trading Commission and the Securities and Exchange Commission (SEC) have also taken note of these new approaches to fraud detection.[80]

For machine-learning algorithms to work, they depend on accessible and analyzable data. Toward that end, many agencies are beginning to recognize the importance of so-called big data—or large volumes of information—in ways

74. *Commissioner's Fellowship Program: Final Report Abstracts*, U.S. FOOD & DRUG ADMIN., https://www.fda.gov/AboutFDA/WorkingatFDA/FellowshipInternshipGraduateFacultyPrograms/CommissionersFellowshipProgram/ucm413253.htm [https://perma.cc/67MG-UXJL]. Work by the same researcher has similarly used machine-learning algorithms to extract laboratory test information from FDA decision summaries of device premarket notifications. Yanna Shen Kang & Mehmet Kayaalp, *Extracting Laboratory Test Information from Biomedical Text*, 4 J. PATHOL. INFORM. 23 (2013).

75. U.S. FOOD & DRUG ADMIN., MOU 225-12-0010, MEMORANDUM OF UNDERSTANDING BETWEEN THE UNITED STATES FOOD AND DRUG ADMINISTRATION AND MASSACHUSETTS INSTITUTE OF TECHNOLOGY (2012), http://www.fda.gov/AboutFDA/PartnershipsCollaborations/MemorandaofUnderstandingMOUs/AcademiaMOUs/ucm318476.htm [https://perma.cc/DUA8-9KM4].

76. For the research surveyed, see Amir E. Khandani et al., *Consumer Credit Risk Models via Machine-Learning Algorithms*, 34 J. BANKING & FIN. 2767 (2010).

77. *See, e.g.*, Johan Perols, *Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms*, 30 AUDITING 19 (2011).

78. Francisco Cantu & Sebastian M. Saiegh, *A Supervised Machine Learning Procedure to Detect Electoral Fraud Using Digital Analysis* (Caltech/MIT Voting Technology Project, Working Paper No. 11, 2010), http://ssrn.com/abstract=1594406 [https://perma.cc/NSJ2-9FD9].

79. Cleridy E. Lennert-Cody & Richard A. Berk, *Statistical Learning Procedures for Monitoring Regulatory Compliance: An Application to Fisheries Data*, 170 J. ROYAL STAT. SOC'Y SERIES A (STAT. IN SOC'Y) 671 (2007); *see also* Richard Berk, *Forecasting Consumer Safety Violations and Violators*, *in* IMPORT SAFETY: REGULATORY GOVERNANCE IN THE GLOBAL ECONOMY 131, 135–36 (Cary Coglianese, Adam M. Finkel & David Zaring eds., 2009).

80. *See, e.g.*, Scott W. Bauguess, Deputy Chief Economist, U.S. Sec. & Exch. Comm'n, The Hope and Limitations of Machine Learning in Market Risk Assessment (Mar. 6, 2015), http://cfe.columbia.edu/files/seasieor/center-financial-engineering/presentations/MachineLearningSECRiskAssessment030615public.pdf [https://perma.cc/HY5T-GKXH]; Scott D. O'Malia, Commissioner, U.S. Commodity Futures Trading Comm'n, Opening Statement at the 12th Meeting of the Technology Advisory Committee (June 3, 2014), http://www.cftc.gov/PressRoom/SpeechesTestimony/omaliastatement060314 [https://perma.cc/RK4T-UK84].

that suggest that the analytical infrastructure needed to use machine learning more extensively may soon be realized. Officials at the U.S. Federal Aviation Administration (FAA), for example, have recognized that in the service of aviation safety "there is significantly more potential" for the use of big data.[81] The U.S. Federal Deposit Insurance Corporation (FDIC) has included as a component of a recent Business Technology Strategic Plan the maturation of "the back-end disciplines of in-memory analytics, big data, and data quality."[82] Similarly, the U.S. Federal Communications Commission (FCC) has developed a Data Innovation Initiative to support the goal of improving its data analytic capacity.[83] Throughout the Obama Administration, the White House prioritized big data use across the executive branch through a Big Data Research and Development Initiative,[84] with President Obama's 2016 budget calling for a $1 billion increase in funding for statistical programs.[85]

Efforts remain underway not only to create large data sets to support agency functions but also to make big data more readily analyzable. One example can be found in the creation of the global Legal Entity Identifier (LEI), a universal reference code for each entity active in financial markets.[86] Treasury's OFR launched an effort to establish LEI in 2010,[87] and by 2014 the LEI Regulatory Oversight Committee had assumed operational responsibility for its development.[88] Having such a unique identifier will enhance regulators' ability "to identify parties to financial transactions instantly and precisely," allowing the

81. Letter from R. John Hansman, Chairman, U.S. Fed. Aviation Admin. Res., Eng'g & Dev. Advisory Comm., to Michael P. Huerta, Administrator, U.S. Fed. Aviation Admin. (Oct. 2, 2013), http://www.faa.gov/about/office_org/headquarters_offices/ang/offices/tc/about/campus/faa_host/rdm/media/pdf/Guidance-FY2016.pdf [https://perma.cc/2VDQ-SVZ3]; *see also* U.S. FED. AVIATION ADMIN., MEETING MINUTES OF THE RESEARCH, ENGINEERING, AND DEVELOPMENT ADVISORY COMMITTEE 7 (2012), http://www.faa.gov/about/office_org/headquarters_offices/ang/offices/tc/about/campus/faa_host/rdm/media/pdf/minutes-FullComm_09262012.pdf [https://perma.cc/DNG9-GET4] (noting that the agency is "taking a look at overarching data management").

82. U.S. FED. DEPOSIT INS. CORP., BUSINESS TECHNOLOGY STRATEGIC PLAN 2013–2017, at 8 (2013), https://www.fdic.gov/about/strategic/it_plan/BusinessTechnologyStrategicPlan2013-2017.pdf [https://perma.cc/L7KE-KPT4].

83. Michael Byrne, *Big Data*, FCC BLOG (Oct. 28, 2010, 1:06 PM), https://www.fcc.gov/news-events/blog/2010/10/28/big-data [https://perma.cc/AU7Q-J6XA].

84. *See* Press Release, Office of Sci. and Tech. Policy, Obama Administration Unveils "Big Data" Initiative: Announces $200 Million in New R&D Investments (Mar. 29, 2012), http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf [https://perma.cc/GX57-AB83].

85. Aaron Boyd, *Obama Budget Pushes Better Decisions Using Open Data*, FED. TIMES (Feb. 3, 2015), http://www.federaltimes.com/story/government/management/budget/2015/02/03/open-data-evidence-based-decisions-funded-2016-budget/22802323 [https://perma.cc/63U6-CQUG].

86. *See generally Legal Entity Identifier–Frequently Asked Questions*, U.S. OFF. OF FIN. RES., http://financialresearch.gov/data/legal-entity-identifier-faqs [https://perma.cc/LH9B-YUG9] (answering frequently asked questions about the legal entity identifier).

87. *See* Statement on Legal Entity Identification for Financial Contracts, 75 Fed. Reg. 74,146 (Nov. 30, 2010).

88. Matthew Reed, *Legal Entity Identifier System Turns a Corner*, U.S. OFF. OF FIN. RES. (July 3, 2014), https://financialresearch.gov/from-the-management-team/2014/07/03/legal-entity-identifier-system-turns-a-corner/ [https://perma.cc/6H64-VFFL].

authorities to apply machine learning to larger data sets.[89]

Agencies are also actively working toward development of the cloud storage systems necessary to exploit the power of machine learning.[90] Such storage that takes place via distributed networks of computers proves to be better suited to running computationally intensive algorithms, and its availability better facilitates interagency sharing of big data. The FDA, for example, has leveraged cloud computing to store information on foodborne pathogens, giving the agency "the ongoing, simultaneous capacity to collect, control and analyze enormous data sets."[91] Similarly, the EPA created a Cross-Agency Data Analytics and Visualization Program intended to foster the creation of databases that will permit the analysis of data from many different agencies and organizations.[92] The SEC is implementing cloud computing to store and process its one billion daily records of financial market activities, often time-stamped to the microsecond, allowing the SEC to "perform analyses of thousands of stocks . . . involving 100 billion records at a time."[93] The proliferation of such efforts to capture, share, and analyze vast quantities of data makes it easy to envision, for example, an extension of the SEC's cloud computing program that would eventually allow agency computers to monitor trading activities in real time, predicting in milliseconds whether a financial transaction is the result of insider trading and then automatically stopping or reversing trades based on those predictions.[94]

---

89. U.S. OFF. OF FIN. RES., *supra* note 86.

90. *See Response to–Request for Information: Preparing for the Future of Artificial Intelligence*, IBM http://research.ibm.com/cognitive-computing/ostp/rfi-response.shtml [https://perma.cc/5PGS-J24T] (describing how AI systems deployed at scale will require "high-performance distributed cloud systems, new computing architectures such as neuromorphic and approximate computing, and new devices such as quantum and new types of memory devices").

91. Taha A. Kass-Hout, *FDA Leverages Big Data Via Cloud Computing*, FDA VOICE (June 19, 2014), http://blogs.fda.gov/fdavoice/index.php/2014/06/fda-leverages-big-data-via-cloud-computing [https://perma.cc/DT6V-N5JK].

92. *See EPA's Cross-Agency Data Analytics and Visualization Program*, U.S. ENVTL. PROT. AGENCY, https://web.archive.org/web/20160414154548/ https://www.epa.gov/toxics-release-inventory-tri-program/epas-cross-agency-data-analytics-and-visualization-program [https://perma.cc/P769-LGB6].

93. *Market Information Data Analytics System*, U.S. SEC. & EXCH. COMM'N, http://www.sec.gov/marketstructure/midas.html [https://perma.cc/2YYE-3LKN].

94. Some academic research has already been conducted both to use machine-learning algorithms to predict such trading violations and to call for their use to make such predictions. *See, e.g.*, Steve Donoho, Early Detection of Insider Trading in Option Markets, Presented at the Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2004); Andrei A. Kirilenko & Andrew W. Lo, *Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents*, 27 J. ECON. PERSP. 51 (2013); Shawn Mankad et al., *Discovering the Ecosystem of an Electronic Financial Market with a Dynamic Machine-Learning Method*, 2 ALGORITHMIC FIN. 151 (2013); Gregory Scopino, *Preparing Financial Regulation for the Second Machine Age: The Need for Oversight of Digital Intermediaries in the Futures Markets*, 2015 COLUM. BUS. L. REV. 439, 443–44 (2015); *see also Foresight: The Future of Computer Trading in Financial Markets*, GOV'T OFF. FOR SCI. 42 (2012), http://www.cftc.gov/idc/groups/public/@aboutcftc/documents/file/tacfuturecomputertrading1012.pdf [https://perma.cc/NW9Z-EXGG].

With these various efforts underway, the government is well on its way into the era of machine learning. Before turning to the legal implications of this new era, we next develop more precisely what machine learning portends for government agencies and why its use might raise questions under prevailing administrative law doctrines.

C. ADJUDICATING BY ALGORITHM, RULEMAKING BY ROBOT

What exactly might be problematic about an era in which government embraces machine learning? Up to this point in our discussion, perhaps the answer will not be obvious. Were it not for dire warnings in the popular press of impending artificially intelligent oppression, it might seem that machine learning simply represents a more sophisticated, data-rich, and predictively useful version of the kind of analytic methods that government agencies have long used. If that is what machine learning is, and if government can use new statistical techniques to improve its performance of various functions from weather forecasting to identifying potentially hazardous chemicals, then presumably a machine-learning era in government should not only be completely unproblematic but also positively encouraged.

Three principal properties of machine learning combine, however, to distinguish it from other analytical techniques and give rise to potential concerns about the greater reliance on machine learning by governmental authorities. The first is machine learning's self-learning property. The results of algorithms do not depend on humans specifying in advance how each variable is to be factored into the predictions; indeed, as long as learning algorithms are running, humans are not really controlling how they are combining and comparing data. These algorithms effectively look for patterns on their own. The second key property is machine learning's "black box" nature. The results of machine learning analysis are not intuitively explainable and cannot support causal explanations of the kind that underlie the reasons traditionally offered to justify governmental action. Finally, machine learning, as with other computational strategies in today's digital era, can be fast and automatic, supporting uses in which the algorithm produces results that can shorten or potentially bypass human deliberation and decision making. All three of these factors combine to make machine-learning techniques appear qualitatively more independent from humans when compared to other statistical techniques.

To illustrate these features of machine learning and what they portend for government, consider the challenges the U.S. Pipeline and Hazardous Materials Safety Administration (PHMSA) faces in deciding how to allocate limited inspection resources to oversee the many thousands of miles of gas, oil, and chemical pipelines throughout the United States. Major leaks as well as explosions from leaky pipelines in recent years have made palpable the significance of effective governmental oversight of pipeline

safety.[95] In recent years, PHMSA has explored using a traditional regression approach to predict risks of pipeline accidents and decide how to target the agency's inspections.[96] Although such an attempt to engage in quantitative decision making is certainly laudable, much more efficient inspection targeting could result if PHMSA generated its risk predictions using machine learning. This could be possible in the near future once big data sets are shared in real time between different agencies and information streams could be provided by remote-sensing technologies. Instead of being limited to analyzing a dozen or so variables that PHMSA's analysts have predetermined should be included in their regression analysis, machine-learning algorithms could work their way through massive amounts of data containing hundreds of potentially predictively useful variables, ranging from pipeline operators' tax returns to their firms' workforce diversity. The resulting predictions of pipeline accident risk could be used not only to target inspections but also potentially, if such an algorithm could be supplied with real-time data from remote sensors, to order preemptive shutdowns of pipeline segments that the algorithm predicts are at risk of imminent failure. To the extent that modern pipeline systems are equipped with computerized, remote shut-off capabilities, a machine-learning algorithm could even be programmed to send an automatic order to pipeline operator's system calling for an immediate, automatic shutdown of a section of pipeline based on real-time forecasts produced by machine learning, all potentially without any human intervention.

Machine learning is well suited for automating these kinds of decisions, given its emphasis on accuracy and the government's overwhelming need to use its limited resources to prevent dangers from arising. But notice that a shift to a machine-learning approach in this context could come along with some qualitative loss of human involvement. Under machine learning, PHMSA analysts would no longer predetermine which variables should be included in the agency's risk models; indeed, they would not even create any risk *models* at all, in the sense of building equations specifying exactly how various variables might impact pipeline risk. Machine learning also does not afford a ready means of explaining why any section of pipeline should be inspected or shut down. The computerized nature of machine learning also means that it can automate decisions currently made by humans, such as the dispatching of inspectors or even the inspections themselves. It is not difficult to imagine a future in which a

95. *See, e.g.*, David R. Baker, *L.A. Gas Leak Plugged, but California Pipelines Regularly Leak*, S.F. CHRON. (Feb. 14, 2016), http://www.sfchronicle.com/business/article/L-A-gas-leak-plugged-but-California-pipelines-6830717.php [https://perma.cc/6HQA-X6FU]; Deirdre Fulton, *More Than 300 a Year: New Analysis Shows Devastating Impact of Pipeline Spills*, COMMON DREAMS (Nov. 17, 2014), http://www.commondreams.org/news/2014/11/17/more-300-year-new-analysis-shows-devastating-impact-pipeline-spills [https://perma.cc/L7MN-S7UH].

96. RICK KOWALEWSKI & PEG YOUNG, BUREAU OF TRANSP. STATISTICS, SR-010, DATA-DRIVEN RISK MODELS COULD HELP TARGET PIPELINE SAFETY INSPECTIONS 3–4 (2008), http://www.rita.dot.gov/ bts/sites/rita.dot.gov.bts/files/publications/special_reports_and_issue_briefs/special_report/2008_010/pdf/entire.pdf [https://perma.cc/4825-D39L].

machine-learning system based in PHMSA's headquarters in Washington, D.C. could be used to automatically dispatch agency drones distributed throughout the country, having them fly over sections of pipeline to take video images or collect air quality samples, thereby removing altogether the need to send human inspectors to the scene.

Admittedly, even with this potential future scenario in mind, it still might not be self-evident why machine learning may be problematic. After all, machines have long supported governmental functions in the administrative state. Although not flown by drones, machines currently collect air quality samples in a network of fixed sites around the country, informing state and federal environmental regulatory decision making.[97] Moreover, when it comes to inspecting potentially hazardous sites, eliminating the need for humans to enter high-risk areas should surely be a positive advance, not a reason for alarm.[98] Furthermore, decisions about the allocation of inspection resources have long been treated—as a matter of well-accepted law—as falling entirely within an agency's discretion.[99] If agencies can legally allocate inspection resources by the flip of a coin—that is, sending inspectors to sites at random, as some agencies do—then they should be able legally to rely on more sophisticated algorithms that deploy scarce inspection resources automatically but more efficiently.[100] For this reason, we foresee comparatively little resistance, as a matter of law, to applications of machine learning that aim to make more efficient use of scarce inspection resources.

Many uses of machine learning by administrative agencies will be like the use of machine learning to decide where to send inspectors in that they will inform actions committed to agency discretion. Most of these uses will be unproblematic from the standpoint of administrative law.[101] Surely the U.S. Postal Service's reliance on machine-learning algorithms to sort mail hardly constitutes any grave threat to society, either existential or constitutional. In

97. *See, e.g.*, *Ambient Air Monitoring*, U.S. ENVTL. PROT. AGENCY, https://www.epa.gov/air-quality-management-process/ambient-air-monitoring [https://perma.cc/GQ2N-SYJE].

98. *Cf.* Suzanne Goldenberg, *Deepwater Horizon Oil Spill: Underwater Robots Trying to Seal Well*, GUARDIAN (Apr. 26, 2010, 2:31 PM), http://www.theguardian.com/environment/2010/apr/26/deepwater-horizon-spill-underwater-robots [https://perma.cc/NAC5-RGKX].

99. Heckler v. Chaney, 470 U.S. 821, 831 (1985) (noting that the Supreme Court has recognized that "an agency's decision not to prosecute or enforce, whether through civil or criminal process, is a decision generally committed to an agency's absolute discretion").

100. In addition to New York City's use of machine learning to determine where to send its building inspectors, the City of Chicago is using algorithms to allocate food safety inspectors, Mohana Ravindranath, *In Chicago, Food Inspectors Are Guided by Big Data*, WASH. POST (Sept. 28, 2014), https://www.washingtonpost.com/business/on-it/in-chicago-food-inspectors-are-guided-by-big-data/2014/09/27/96be8c68-44e0-11e4-b47c-f5889e061e5f_story.html [https://perma.cc/V25Y-53CG].

101. We recognize, of course, that it might be possible for agencies to abuse their discretion, which is why we characterize uses of machine learning for discretionary purposes to be *virtually* unproblematic. *See, e.g.*, 5 U.S.C. § 706(2)(A) (2012). One way that such discretion could possibly be abused would be if it were to be deployed in an unlawfully discriminatory fashion. In Section II.C, we consider whether the use of machine learning even in discretionary enforcement allocation decisions might offend equal protection.

addition, even when agency officials use learning algorithms to support actions that are not committed to agency discretion, if they use them simply to inform their own independent judgments, this too should be unremarkable. Such use would be indistinguishable from any other research support or informational input into agency decision making. The non-shaded parts of Table 1 highlight several general types of agency uses of machine learning that should easily be viewed as beyond reproach, at least from the standpoint of existing general principles of structural law governing the administrative state.

The domains in which machine learning might be of concern, at least as a prima facie matter, will be those in which artificial intelligence is used more for determining, rather than just supporting, decisions that are not otherwise committed to agency discretion by law. As shown in the shaded portions of Table 1, that leaves two important realms in which machine learning could be incorporated into the administrative state: *adjudicating by algorithm* and *rulemaking by robot*.[102]

One example of *adjudicating by algorithm* would be our posited PHMSA pipeline safety machine-learning system that automatically issues shut-off orders when the system forecasts a heightened risk. It is not difficult to imagine

Table 1. Applications of Machine Learning in the Administrative State

Role of Machine Learning in Agency Decision-Making	Type of Administrative Action		
	"Discretionary"	"Non-Discretionary"	
		Adjudication	Rulemaking
Supportive			
↕			
Determinative		*Adjudicating by Algorithm*	*Rulemaking by Robot*

102. Adjudication and rulemaking, of course, are the two canonical types of actions that agencies may take under the Administrative Procedure Act. 5 U.S.C. § 551(5), (7) (2012). We have labeled these two types of administrative action as "non-discretionary" not because agencies are mandated to take these actions (although sometimes they can be). Rather, we have labeled them this way because these actions will be surrounded by "law to apply" that will subject these actions to judicial review under the Administrative Procedure Act. Citizens to Preserve Overton Park v. Volpe, 401 U.S. 402 (1971). By "discretionary" in Table 1, we mean simply that an action is "committed to agency discretion" and thus not subject to judicial review. 5 U.S.C. § 701(a)(2). We do recognize, of course, that on occasion there may be law to apply even to supportive uses of analytic techniques or to the use of other factors that support decisions, such as would be the case if a statute were to prohibit an agency from using machine learning in even a non-determinative role. *Cf.* Whitman v. American Trucking Assocs., Inc., 531 U.S. 457 (2001) (holding that the Clean Air Act prohibits the EPA Administrator from considering costs when setting air quality standards). For these reasons, Table 1 should be viewed simply as a heuristic intended to illustrate some generalizations about the administrative state.

other examples of adjudicatory decisions that could be automated by algorithms,[103] especially when the relevant criteria for an adjudicatory action are forward-looking and thus dependent on accurate predictions. At some point, for example, the FAA might be able to license pilots through an entirely automated process relying on risk-based machine learning forecasts of individual applicants' overall level of safety.[104] The Federal Trade Commission or the Department of Justice's Antitrust Division might conceivably come to rely on machine learning to predict what effects a proposed merger would have on future competition and market pricing, perhaps entirely automating the antitrust review process.

When it comes to *rulemaking by robot*, we need not rely entirely on the imagination. The City of Los Angeles' current traffic signaling system illustrates a very simple but still real-world application of rulemaking by robot. Although deciding the color of traffic lights may seem like a trivial example, a traffic signal does determine what rule applies to anyone who wants to drive along a city street at a given period of time. Yet with the system in place in Los Angeles, just as no human determines when a traffic light should be red or green, no government official can really explain why the city's machine-learning system sets any given traffic light (that is, rule) when it does. We can expect it will not be long before more government authorities, at the local and federal levels, will be able to develop similar systems in their own domains that are conceptually equivalent to Los Angeles' traffic control system.

It is not difficult to imagine more complex and consequential examples of regulating by robot. Consider the possibility that the SEC might find it beneficial, even necessary, to govern the world of high-speed electronic trading by making nimble and equally high-speed adjustments to the rules of market transactions, perhaps modifying stock exchanges' current, rigid trading circuit breakers with ones that adjust in real time.[105] The U.S. Department of the Treasury, for similar reasons, might plausibly seek to establish a dynamic, automated process according to which certain macro-prudential rules governing financial institutions respond to real-time market changes indicative of systemic

---

103. Imagination may not be required for much longer. IBM is currently developing machine-learning algorithms to predict smog levels in China, predictions that may soon be used to determine governmental shutdowns of factories or limits on traffic volumes. Will Knight, *Can Machine Learning Help Lift China's Smog?*, MIT TECH. REV. (Mar. 28, 2016), https://www.technologyreview.com/s/600993/can-machine-learning-help-lift-chinas-smog/ [https://perma.cc/SL77-VJVK].

104. Currently, such forward-looking adjudicatory decisions like licensing are based on rules, making the issue one of whether applicants comply with the applicable rule, or the criteria contained in a rule, and thus qualify to receive a license. Machine learning makes an alternative adjudicatory framework possible, one that considers forecasted risk based on an algorithmic analysis of potentially hundreds of variables. Machine learning has been shown to be an effective tool in making certain forward-looking adjudicatory decisions in the criminal law system. Richard A. Berk et al., *Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions*, 13 J. EMPIRICAL L. STUD. 94, 110 (2016) [hereinafter Berk, *Forecasting Domestic Violence*]; Berk et al., *supra* note 50, at 208.

105. *See supra* note 94.

risk.[106] Even when time is not so critical and the "good cause" exemption to the standard rulemaking process might not apply, it is hardly unimaginable today that agencies could automate entirely the notice-and-comment rulemaking process, especially for the kinds of routine rules that make up the bulk of government rules.[107] Natural language processing programs could even conceivably read and summarize any public comments submitted on proposed rules and potentially even craft some of the regulatory language.[108]

For anything but perhaps the simplest rules, like traffic signals, rulemaking by robot will require that machine learning be combined with other analytic techniques. Rules are forward-looking, but they also involve complex normative judgments, not merely predictive ones. Determining the content of rules often

106. For a discussion of how predictions of systemic risk can affect rulemaking by regulatory agencies, see Dimitrios Bisias et al., *A Survey of Systemic Risk Analytics* 2, 10–11 (U.S. Dep't of the Treasury Office of Fin. Research, Working Paper No. 0001, 2012), https://www.treasury.gov/initiatives/wsr/ofr/Documents/OFRwp0001_BisiasFloodLoValavanis_ASurveyOfSystemicRiskAnalytics.pdf [https://perma.cc/D8LJ-8EYJ].

107. The notice-and-comment process, and the good cause exception to it, is provided at 5 U.S.C. § 553 (2012). It is important to recognize that we have adopted a formulation of algorithm-created rules resembling that of rules as they exist today; an algorithm would promulgate rules specifying a particular course of action, a particular safety standard, a particular acceptable emissions level, or so forth. Under this formulation, any changes to an algorithm that result in a different prescription, including merely re-running the algorithm as specified on new data, would necessitate a new rulemaking process, absent a good cause exemption. But, because machine learning is likely to be of most utility when engaged to regulate dynamic, time-sensitive environments, it is probable that such an exemption could often or even categorically be claimed. An alternative formulation might be one in which rules state merely that an algorithm will be used to promulgate prescriptions continuously; instead of a rule reciting a particular course of action or safety standard on the basis of algorithmic output, the rule would say that a future algorithm will run continuously and be updated dynamically to decide the appropriate course of action or safety standard in the future. *Cf.* Coglianese, *supra* note 25, at 370–71 (noting the possibility of "a reconceptualization of the form in which rules are promulgated"). Such a formulation probably would not necessitate that a new rulemaking be commenced whenever the algorithm is updated, but it might be legally problematic given the need for reason-giving and transparency. *See infra* Section II.D.

108. Already agencies use digital tools to sort and identify duplicates in comments submitted in rulemaking proceedings that have generated large volumes of public submissions. *See, e.g.*, Jane E. Fountain, *Prospects for Improving the Regulatory Process Using E-Rulemaking*, 46 COMMC'NS ACM 63, 63–64 (2003) (discussing federal agency use of automated tools to sort comments beginning as early as 1997). Similar programs are now capable of processing data and automatically writing prose, at least for now in the context of sports reports and fiction. *See* Ian Crouch, *The Sportswriting Machine*, NEW YORKER (Mar. 26, 2015), http://www.newyorker.com/news/sporting-scene/the-sportswriting-machine [https://perma.cc/25FB-UBFS]; Matt McFarland, *A Computer Program Is Writing New 'Friends' Episodes. Are They Any Good?*, WASH. POST (Jan. 21, 2016), https://www.washingtonpost.com/news/innovations/wp/2016/01/21/a-computer-program-is-writing-new-friends-episodes-are-they-any-good/ [https://perma.cc/57FC-UY86]. By itself, an algorithm might not be capable of writing the entire content of a final rule document, as sections such as the summary would presumably require a nuanced explanation of the background of a rule and its purposes, and natural language processing cannot generate such complete thoughts de novo. Natural language processing could, however, probably write sections of a rule document that require statements of facts. Such algorithms can rely on previous examples of how factual statements are worded to create sentences that describe new facts in a fill-in-the-blank manner. Other sections of rule documents require human input related to the goal of a potential rule, but such sections could plausibly be written in advance by humans and then the rest could be filled in with algorithm-written content.

requires making difficult choices about the entities to be regulated, the conduct or outcome that the rule tells these entities to achieve or avoid, and the nature and degree of the consequences that follow from adhering or not adhering to the rule's commands.[109] Machine-learning algorithms cannot directly make the choices about these different aspects of a rule's content not only because some of these choices are normative ones, but also because learning algorithms are merely predictive and thus unable to overlay causal interpretations on the relationship between possible regulations and estimated effects.[110] The justification for new rules depends, after all, on the effects that their adoption and implementation are likely to cause.[111]

Nevertheless, it may be possible for machine learning to make rules in this fashion when used in conjunction with procedures known as agent-based models (ABM) or multi-agent systems (MAS).[112] Agent-based modeling refers to the use of an algorithm consisting of a mathematically-defined environment that includes agents that observe the overall environment and take actions designed to reach a specified goal.[113] Multi-agent systems are similar to agent-based models but with multiple autonomous agents interacting with each other.[114] With either of these agent-based techniques, the agents—which, in the rulemaking context, would include the regulator and the regulated entities—must have some defined decision-making processes that allow them to translate observations of the environment into actions. These decision-making processes can be specified a priori by the researcher or regulatory official, but such a priori knowledge often does not exist or is not sophisticated enough to mimic how real-world agents make their decisions. Therefore, machine learning—often called reinforcement learning in these applications (or what we will, for ease of reference, call "embedded machine learning")—is incorporated into agent-based models' decision-making processes of individual agents. The mathematical agents within these systems, in other words, learn how to make decisions.

---

109. *See* Cary Coglianese, *Engaging Business in the Regulation of Nanotechnology*, *in* GOVERNING UNCERTAINTY: ENVIRONMENTAL REGULATION IN THE AGE OF NANOTECHNOLOGY 46, 50–51 (Christopher J. Bosso ed., 2010).

110. *See* BERK, *supra* note 33, at 9–17. This difficulty also often faces even conventional techniques in attempting to claim causal inference. *See* Richard A. Berk et al., *What You Can Learn from Wrong Causal Models*, *in* HANDBOOK OF CAUSAL ANALYSIS FOR SOCIAL RESEARCH 403, 422–23 (Stephen L. Morgan ed., 2013).

111. CARY COGLIANESE, MEASURING REGULATORY PERFORMANCE: EVALUATING THE IMPACT OF REGULATION AND REGULATORY POLICY, OECD Expert Paper No. 1 (Aug. 2012), http://www.oecd.org/gov/regulatory-policy/1_coglianese%20web.pdf [https://perma.cc/6ZZZ-NERP].

112. For incorporation of machine learning into ABM, see, for example, W. Rand, *Machine Learning Meets Agent-Based Modeling: When Not to Go to a Bar* 2 (Northwestern Univ. Working Paper, 2006), https://ccl.northwestern.edu/papers/agent2006rand.pdf [https://perma.cc/SU6T-AB9V]. For use of machine learning in MAS, see Lucian Buşoniu et al., *A Comprehensive Survey of Multiagent Reinforcement Learning*, 38 IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, PART C: APPLICATIONS AND REVIEWS 156, 156 (2008).

113. For a more formal definition of ABM, see Nigel Gilbert, *Agent-Based Models*, *in* 153 QUANTITATIVE APPLICATIONS IN THE SOCIAL SCIENCES (John Fox ed., 2008).

114. For a more formal definition of MAS, see Buşoniu et al., *supra* note 112, at 156.

To translate these embedded machine-learning techniques to possible rulemaking applications, consider how the Occupational Safety and Health Administration (OSHA) might proceed if it were to create an automated process for determining whether to implement a new workplace safety regulation. OSHA could implement an algorithm in which the modeled agents are the employers being regulated. The environment in which these agents operate would include mathematically-specified factors capable of influencing agent behavior, including a possible regulation. The employer-agents in the model would "observe" the environment, which would include different regulatory alternatives (including an environment with no regulation), and then "take" actions, such as complying with the regulation, to reach their own goals, perhaps defined as profit maximization. Now, although OSHA would like to use this agent-based model to see how employers respond to the potential new regulation and, consequently, what effects the regulation may have, OSHA does not know a priori how employers will decide how to respond to any regulation. The agent-based model would therefore use a machine-learning technique to select employers' optimal responses to the regulation given their profit maximization goal.

This example suggests how OSHA might use machine learning embedded within an agent-based model of the effects of a proposed regulation. But the techniques' real potential to inform the content of regulations comes from the ability of OSHA to include an agent representing itself in the ABM. This mathematically-represented agent would "issue" multiple different possible regulations—formulated in advance by human programmers—and then "select" the regulatory alternative that yields those effects, as defined in relation to observable components of the environment, that maximize an objective function (or goal) established by the real-world OSHA. The possible regulations analyzed in this fashion could assume any number of different combinations of regulatory targets, commands, and consequences, with the forecasted effect of these regulations on the actions of regulated entities being observed through the modeling exercise. Unlike in the adjudicatory context, where machine learning directly makes individualized forecasts and where an adjudication can be "determined" simply by whether an algorithm forecasts risks or other outcomes above a threshold level, in the rulemaking context machine learning would need to be nested within a larger decision-making model to support automated regulatory decisions. Machine learning predictions would, within an agent-based simulation, inform agents' actions, which in turn would generate predicted outcomes from different regulatory permutations.

This fusion of agent-based or multi-agent models with machine learning may hold great potential for assisting in certain kinds of rulemaking, but, even with this fusion, governmental reliance on algorithms would still not cede entirely the involvement of humans. As already indicated, at a foundational level, humans will still need to choose and then input into embedded machine-learning systems the data, as well as overarching goals to be maximized and

constraints to be minimized. Moreover, due to data limitations as well as core uncertainties, many rulemaking decisions will still by necessity call for human judgment and thus be incapable of automation.

As with any statistical technique, the algorithms that could be embedded in automated rulemaking models will require data. Because all historical data arise within a world with a different rule than the one proposed (even if that is no rule at all), regulators will seldom (if ever) find enough data to correspond to all possible forms a future regulation and resulting environmental state might take. This is often a challenge in applications of agent-based models in other contexts, such as healthcare provision. In those other contexts, the lack of data is often addressed through the creation of simulated environmental data.[115] Generation of simulated data, however, requires that the architecture of the environment being modeled, and the relationships between components of that environment, be sufficiently well known a priori as to be specifiable. Embedded machine-learning techniques have been successfully developed for applications like modeling how infectious patients should be moved around a hospital.[116] In that context, the environment of interest can be reasonably well specified. The actors and parameters are limited—for example, healthcare professionals, infected patients, uninfected patients, and rooms—and the analyst knows a priori enough about how diseases are transmitted to generate simulated data using probabilities of infection based on proximity and time spent near infected patients. This kind of a priori knowledge would seem to be less likely to exist in the more complex or uncertain situations that many regulators address, where the relevant causal relationships do not stem from processes as law-like as biological disease transmission. If the system being modeled is extremely complex—as with many forms of regulation, whether of complex financial instruments or advanced industrial operations[117]—the regulator may not know enough about the underlying causal architecture to generate simulated environmental data bearing any resemblance to real-world data.

Of course, despite these difficulties, the conditions for using embedded machine learning for rulemaking may still sometimes exist. In a comment letter to the SEC, for example, academic and business experts in agent-based modeling and financial markets have advocated the use of such models in regulating equity markets, arguing that algorithms in this context would be sufficiently specifiable.[118] Although we take no position on these specific claims, we raise

115. *See, e.g.*, Marek Laskowski, *A Prototype Agent Based Model and Machine Learning Hybrid System for Healthcare Decision Support*, *in* DIGITAL ADVANCES IN MEDICINE, E-HEALTH, AND COMMUNICATION TECHNOLOGIES 230, 231 (2013).

116. *Id.* at 235–36.

117. *See generally* CHARLES PERROW, NORMAL ACCIDENTS: LIVING WITH HIGH-RISK TECHNOLOGIES (2d ed. 1999).

118. W. Brian Arthur et al., Comment Letter to Elizabeth M. Murphy on File Number S7-02010 "Concept Release on Equity Market Structure" (Apr. 16, 2010), https://www.sec.gov/comments/s7-02-10/s70210-109.pdf [https://perma.cc/GTE3-W2M8].

them to suggest the plausibility of using embedded machine learning to automate the process of selecting and designing regulations in some settings. Agency officials will need to determine the applicability of any embedded machine-learning rulemaking tool on a case-by-case basis.

Our point is to show that, even if many applications of machine learning will be completely benign as a matter of administrative law, agencies may soon be able, for the first time, to set the content of certain types of rules by automated artificial intelligence techniques. Whether in making individualized forecasts or in feeding into more generalized modeling results, machine-learning algorithms have the potential to transform key governmental functions in ways that not only augment human judgment but replace it with automated, algorithmic analysis. For some observers, this prospect will trigger loud alarm bells of the kind set off by the use of artificial intelligence more generally. At a minimum, the prospect of either robotic rulemaking or algorithmic adjudication raises important questions about whether such automated techniques can be squared with core principles of constitutional and administrative law.

II. THE LEGALITY OF MACHINE LEARNING IN THE ADMINISTRATIVE STATE

As government agencies continue on the path toward increased reliance on machine learning in administrative decision making, public officials, lawyers, and scholars will confront choices about whether to encourage or constrain this technology. Making these choices will depend, at least in the first instance, on assessing how agency use of machine learning would conform to the cornerstones of constitutional and administrative law: principles of nondelegation, due process, antidiscrimination, and transparency.[119] These core legal principles, against which we assess machine learning in this Part, present issues that arise from the mathematical distinctiveness of machine learning, in particular its self-learning, black-box, and automated properties. Especially when machine learning is used to determine outcomes that are judicially reviewable, its

119. A further concern widely aired in discussions of big data and machine learning centers on privacy. Privacy concerns are not trivial, but we do not take them up in detail here in part because they have been widely considered elsewhere. *See, e.g.*, Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 96 (2014); Roger Allan Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 MICH. TELECOMM. & TECH. L. REV. 1, 3–4 (2016); Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125, 1170–71 (2015); Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 395–97 (2014); Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 42–43 (2013); Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63, 65 (2012). We also do not focus on privacy concerns because they are not unique to administrative agencies' use of machine learning. Our principal concern here is with issues that are distinctively applicable to the use of machine learning in the administrative state. Furthermore, many of machine learning's privacy implications could be more accurately characterized as implications of the use of big data; although algorithms themselves may make possible new kinds of inferences, this ability is manifested only in large data sets, so the underlying privacy problems that others have flagged appear to have been raised more by the collection of big data than by the use of machine-learning algorithms per se.

properties combine to surface some first-order questions stemming from a legal system that has been historically premised on the existence of governmental decision makers who are human beings.

Our answers to the legal questions presented in this Part must, by necessity, assume a degree of generality. Machine learning is not a singular entity with one prescribed method of implementation, so we cannot pretend to offer a definitive legal analysis of all possible applications of artificial intelligence in the administrative process. How machine learning will come to be used in particular contexts, by particular agencies, will no doubt prove pivotal to determining its legality under certain doctrines. Neither the technique of machine learning nor its uses are completely uniform—even while sharing general properties—and even some of those general properties are far from absolute. Still, learning algorithms do exhibit to a sufficient degree some core features that make it possible to offer some overarching conclusions. For example, even though algorithms can learn on their own and support automated decisions, humans still must decide how algorithms are specified, deployed, and integrated into broader administrative processes. Machine-learning algorithms are thus "autonomous" only in the sense that they can run continuously and have the potential to translate their outputs automatically into regulatory actions.[120] They do not set their own objective functions nor are they completely outside human control. An algorithm, by its very definition, must have its parameters and uses specified by humans, and this property will likely prove pivotal in the legal assessment of specific applications of artificial intelligence by federal administrative agencies.

A. NONDELEGATION

Military weapons systems driven by algorithms raise serious concerns because, with these systems, decisions possessing life-and-death consequences may no longer be made directly by humans.[121] For similar reasons, concerns could be raised about administrative agencies' use of artificial intelligence. Although officials at administrative agencies do not fire missiles, they are responsible for choices that can carry equally important consequences, such as determining the safety of food, water, and drugs. If concerns exist over delegating too much control to non-humans in the military context, then perhaps at least a prima facie case exists against the use of algorithms to make administrative and regulatory decisions.

120. Note that this limited "autonomy" also implies, given our description of how machine learning operates, the potential for algorithms to yield unexpected, seemingly original outcomes. Also, although we have chosen to use and qualify the term "autonomy," other legal literature on robotics refers, with good reason, to similarly limited independence as "emergence." *See, e.g.*, Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CAL. L. REV. 513, 539–40 (2015).

121. *See* Markoff, *supra* note 14.

The U.S. Constitution provides that "[a]ll legislative powers" of the federal government "shall be vested" in Congress.[122] Courts have long accepted that Congress can, within limits, delegate authority to administrative agencies headed by appointed officers who, although unelected, are overseen in various ways by members of Congress and the President.[123] Yet the nondelegation doctrine, still a fixture in American constitutional and administrative law, places some theoretical limits on those delegations, which must, for example, be accompanied by an intelligible principle. Although this doctrine has long accepted even broad delegations of authority to administrative agencies, the law has always assumed that the recipient of that authority would be a human being, such as an officer of the United States or, on occasion, a private individual or group of individuals. As machine learning becomes more advanced and government agencies use it more extensively, decision-making authority could effectively become delegated still further—to computerized algorithms. Yet if government actions should be undertaken by humans, then delegation to autonomously learning machines could potentially transfer governmental power outside the bounds that the Constitution permits. Such an objection under the nondelegation doctrine has never been squarely contemplated under previous judicial rulings, but it bears clear conceptual affinity with the spirit and tradition of the nondelegation doctrine. The underlying concern is the same: the improper transfer of legislative authority.

Given that the nondelegation doctrine has traditionally governed the granting of authority by *Congress* to *agencies*, we confront two possible scenarios. The first assumes that Congress has granted authority to an agency to deploy machine-learning algorithms to make administrative decisions. The second is a scenario in which Congress has delegated some kind of administrative authority to the head of an agency in a conventional manner, but the relevant agency official then seeks out and relies upon a machine-learning system to exercise that authority. We will address both of these possible situations in turn, even though the second scenario might be the one that is more likely to arise in practice, given that statutes have not been drafted with algorithmic administration in mind. However, the second scenario raises different, even if related, questions of statutory interpretation—namely, whether a statute granting an agency certain governmental authority precludes the agency from subdelegating its decision making to machine-learning algorithms.[124] We will put such a statutory question to the side initially, focusing on the first scenario, in which Congress delegates authority to agencies to regulate by robot. This scenario allows us to focus directly on the constitutional question raised by what Justice

122. U.S. CONST. art. 1, § 1.

123. A.L.A. Schechter Poultry Corp. v. United States, 295 U.S. 495, 505 (1935).

124. Executive exercise of legislative authority by administrative agencies must conform to statutory provisions. *See* Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 585 (1952) ("The President's power, if any, to issue the order must stem from an act of Congress or from the Constitution itself.").

Mariano-Florentino Cuéllar has dubbed the problem of "cyberdelegation."[125]

1. Cyberdelegation

The answer to the constitutional question of cyberdelegation should not be difficult—and not merely because the nondelegation doctrine has been widely recognized to have had only "one good year."[126] A congressional authorization of rulemaking by robot should not offend the nondelegation doctrine because presumably such an authorization would be made with the understanding that algorithmic policy determinations depend inherently on well-specified objective functions. It seems unlikely that any Congress that would expressly contemplate the use of artificial intelligence by agencies would not also include in legislation authorizing such use a sufficiently intelligible principle that would satisfy the demands of the nondelegation doctrine. Under this doctrine, rulemaking authority can be delegated to agencies only if Congress provides in its authorizing legislation "an intelligible principle" channeling and constraining the exercise of that delegated authority.[127] As every lawyer today knows, the level of intelligibility demanded by the courts has hardly been substantial. The courts have recognized that agencies can be delegated power "under broad general directives"[128] that facilitate governmental efficacy "in our increasingly complex society, replete with ever changing and more technical problems."[129] As a result, Congress must delegate in ways "delineat[ing] the general policy, the public agency which is to apply it, and the boundaries of this delegated authority."[130] Courts have even upheld delegations of authority directing, explicitly, only that agencies must act in the "public interest."[131]

To say that an intelligible principle must guide agency actions is to say that those actions must be aligned with a goal or, in mathematical terms, a type of objective function. An objective function is an essential prerequisite for the use of machine learning, and it will by necessity provide sufficient intelligibility to withstand the test embedded in the nondelegation doctrine. If a goal as broad and qualitative as "acting in the public interest" is legally sufficient, then goals defined in precise, quantifiable, and measurable ways must be so as well. A

125. Cuéllar, *supra* note 25. This first scenario is important to focus on also because it turns out ultimately to be where the statutory interpretation question raised by the second scenario leads. If a court were to conclude that a statute either authorized, or at least did not preclude, agencies to subdelegate to machines, the question then would still arise whether the Constitution permits Congress to delegate authority in such a way that allows agencies to delegate further to machine-learning algorithms.

126. Cass R. Sunstein, *Nondelegation Canons*, 67 U. CHI. L. REV. 315, 322 (2000). The Supreme Court has struck down legislation under the nondelegation doctrine on only a couple of occasions—and only during the New Deal. *Id.*

127. J.W. Hampton, Jr., & Co. v. United States, 276 U.S. 394, 409 (1928).

128. Mistretta v. United States, 488 U.S. 361, 372 (1989).

129. *Id*. *See also* Opp Cotton Mills, Inc. v. Adm'r, 312 U.S. 126, 145 (1941).

130. Am. Power & Light Co. v. SEC, 329 U.S. 90, 105 (1946); *see also* Yakus v. United States, 321 U.S. 414, 424–25 (1944).

131. *See, e.g.*, Nat'l Broad. Co., Inc. v. United States, 319 U.S. 190, 225–26 (1943).

Congress that deliberately contemplated and authorized an agency to use machine learning would presumably also understand the need to provide guidance about the necessary objective function for algorithms to optimize, and it would be more likely than usual to articulate a sufficiently clear set of goals that would pass the intelligibility muster.

What about the *recipients* of the authority to whom Congress delegates? Typically, these recipients have been agency officials, but on occasion Congress has tried to delegate certain kinds of authority to private actors. When Congress has delegated lawmaking power to private entities, the Supreme Court has declared such arrangements to be "legislative delegation in its most obnoxious form."[132] The Court's disapproval of congressional authorization of private decision making with legal implications raises a key question about the possibility of congressional authorization of algorithmic decision making. If delegating outside the government to the private sector is so disfavored, then presumably a delegation of rulemaking authority still further—not even to a human being, but to a machine—would seem more extreme and objectionable.

We think courts would be unlikely to equate delegation to machines with delegation to private individuals or entities for three reasons. First, machine-learning algorithms lack the essence of what makes a delegation to private individuals so obnoxious—the self-interest of those private parties. Private parties have their own biases and interests that lead them to make privately optimal decisions that may not be socially optimal.[133] By contrast, what machine-learning algorithms will optimize are the objectives that those deploying them specify should be optimized. As long as this specification is done by authorized government officials, the concerns about bias or self-interest that animate the constitutional prohibition on delegations to the private sector should not exist.

Second, even when they rely on machine learning, human governmental officials will retain ultimate control over the specification of algorithms and the translation of their outputs to regulatory actions. That the government maintains such a level of direction and control has made it sufficient for courts to uphold delegations to private parties. As long as a private party's participation in rulemaking falls short of possessing the ultimate decisional control, courts have allowed delegations to the private sector. The Supreme Court has deemed constitutional, for example, legislation providing industry with a role in policy-making by a government commission because private firms only "function[ed]

132. Carter v. Carter Coal Co., 298 U.S. 238, 311 (1936); *see also* Dep't of Transp. v. Ass'n of Am. R.Rs., 135 S. Ct. 1225, 1231 (2015); A.L.A. Schechter Poultry Corp., 295 U.S. 495, 537 (1935).

133. *See, e.g.*, George J. Stigler, *The Theory of Economic Regulation*, 2 BELL J. ECON. & MGMT. SCI. 3, 3–7 (1971). Arguably, the constitutional prohibition on delegation of rulemaking power to private entities emanates as much from due process considerations as from strictly nondelegation concerns. *See, e.g.*, Alexander Volokh, *The Shadow Debate Over Private Nondelegation in* DOT v. Association of American Railroads, 2015 CATO SUP. CT. REV. 359, 393 (2015) (pointing to "the fundamental unfairness of putting the regulation of an industry in the hands of an entity that has a profit-making interest in the outcome of the regulation").

subordinately to the Commission."[134] More generally, the Court has permitted delegation to private parties when their actions are limited to advisory decisions rather than ones that are binding on others.[135]

The kind of rulemaking by robot that we described in section I.C—and even just the phrase "rulemaking by robot"—might at first glance suggest a role for machines that exceeds the limits established for delegations to private entities. Yet even though machine-learning systems are more autonomous from humans than are conventional statistical techniques, their autonomy is still far from the constitutionally impermissible kind. Even if rulemaking were to be fully automated, the underlying algorithms, just to function correctly, must still be so well specified that important discretion would remain with the human creators and overseers of the algorithms. As previously discussed, algorithms could not themselves craft regulatory content de novo; humans must specify the targets, commands, and consequences of potential rules from which an embedded machine-learning system might choose the best.[136] More importantly, humans can, at any time, choose to reject a machine-chosen rule, alter an algorithm's specifications, or even "pull the plug" on the system entirely. If nothing else, as long as rules need to be officially signed or approved by a human official, then humans will retain ultimate control.[137]

Finally, although algorithms can act faster than humans—which might imply a lack of control—ultimately algorithms are mere measurement tools, which the courts widely accept as legally permissible. The objection, of course, would be that humans cannot, as a practical matter, exercise meaningful control over automated rulemaking systems in high-speed settings. In these (and perhaps other) situations, it may simply not be practical for human officials to monitor algorithms' output in real time and exercise their abilities to approve or reject rules or even "pull the plug" on the system. These officials would, of course, still retain the ability to disable or modify a rulemaking-by-robot system retroactively. Although it might be reasonable in some circumstances to demand more than retroactive review, in most instances automated artificial intelligence systems, once constructed by humans, will typically function as legally permissible measurement tools.

---

134. Sunshine Anthracite Coal Co. v. Adkins, 310 U.S. 381, 399 (1940); *see also, e.g.*, Currin v. Wallace, 306 U.S. 1, 15 (1939).

135. *See, e.g.*, Ass'n of Am. R.Rs. v. Dep't of Transp., 721 F.3d 666, 671 (D.C. Cir. 2013) ("Congress may formalize the role of private parties in proposing regulations so long as that role is merely 'as an aid' to a government agency that retains the discretion to 'approve[], disapprove[], or modif[y]' them." (quoting *Sunshine Anthracite Coal Co.*, 310 U.S. at 388)); Pittston Co. v. United States, 368 F.3d 385, 396 (4th Cir. 2004) ("These powers given to the Trustees are of an administrative or advisory nature, and delegation of them to the Trustees does not, we conclude, violate the nondelegation doctrine."); United States v. Frame, 885 F.2d 1119, 1129 (3d Cir. 1989) ("In essence, the Cattlemen's Board and the Operating Committee serve an advisory function . . . .").

136. *See supra* Section I.B.

137. *Cf.* Cary Coglianese, *Presidential Control of Administrative Agencies: A Debate over Law or Politics?*, 12 U. PA. J. CONST. L. 637, 646 (2010); Cary Coglianese, *The Emptiness of Decisional Limits: Reconceiving Presidential Control of the Administrative State*, 69 ADMIN. L. REV. 43 (2017).

In *Prometheus Radio Project v. FCC*, the FCC relied on a private company, Arbitron, to measure local radio station audiences, and these measurements provided critical inputs to rules on station ownership that were required to serve the public interest, in part, by ensuring competition.[138] The court ruled that reliance on Arbitron was not an unconstitutional delegation of legislative power, noting: "Arbitron will only provide a mechanism for measuring [market] concentration. Because the Commission remains the sole arbiter of whether a proposed radio station combination serves the public interest, no improper delegation will occur."[139] Thus, because the Commission specified the kind of rule—whether an ownership pattern is legal—and the goal to be served by the rule—a radio station combination ensuring diversity and competition sufficient to serve the public interest—private parties were allowed to control the methods used to measure diversity and competition and provide measurements that were then translated into rules.

Analogously, if an agency creates an embedded machine-learning system by supplying the possible rule options and the objective function, the implementation of an algorithm that maximizes that objective function and immediately promulgates the resulting rule should be sustained against nondelegation objections because it is functionally serving just as a measurement tool. From the standpoint of the nondelegation doctrine, the use of machine learning is not conceptually any different than the constitutional use of other machines or instruments. When FDA officials use measuring devices like thermometers to determine the optimal temperature at which to store fish, they are no more delegating impermissibly to those devices than if they were to rely on machine-learning algorithms.

For these reasons, not only would legislation authorizing machine-learning applications be unlikely to offend the intelligible principle requirement, it also should not constitute a prohibited delegation of authority to an entity outside of government. Cyberdelegation might well look somewhat novel, but it is not unlike government reliance on measurement devices and thus unlikely to exceed constitutional constraints on the delegation of governing authority.

2. Statutory Subdelegation

Having concluded that congressional authorization of rulemaking by robot is unlikely to offend the nondelegation doctrine, we return to the issue of subdelegation. We noted at the outset of this section that the subdelegation issue involves, in the first instance, answering a statutory interpretation question. Specifically, would an agency head's reliance on machine-learning algorithms exceed authority delegated by Congress via statutory language specifically authorizing "the

138. 373 F.3d 372, 387 (3d Cir. 2004). Specifically, Arbitron's measurements of radio station markets informed the legal limit on the number of radio stations within one market that may be owned by the same entity. *See id.*

139. *Id.* at 425.

Administrator" or "the Secretary" to act? It should be evident that, for reasons similar to those we have just presented about delegations to private entities, subdelegations to machines pose no categorical legal concerns. Administrators do not exceed their statutory authority by relying on thermometers or other measurement devices. Nor do they transgress their statutory authority by relying on aides and subordinate officials because ultimate decision-making authority in such cases is never completely or irrevocably subdelegated, given administrators' need to endorse, or their ability to override, what their subordinates do.[140] The same would be true of reliance on machine-learning algorithms.

Despite demonstrating that regulating by robot is unlikely to violate the nondelegation doctrine, our analysis should not, of course, be taken to imply that agency actions will never be struck down on grounds related to delegation. Delegations, for example, may be constrained by how much judicial deference courts give to agencies' interpretations of ambiguous statutes. Admittedly, *Chevron*[141] deference would appear to be considerable, as courts are generally instructed to defer to agencies' reasonable interpretations of ambiguous statutes.[142] For this reason, we find it difficult to see how an agency's decision to use machine learning, even if central to an agency's statutory construction, could rise to the level of unreasonableness typically required for courts to reject agency interpretations. Nonetheless, given that judges have yet to face robotic rulemaking processes like those we described in section I.C, there remains a possibility that they may, in certain situations, view administrative algorithms with a particular aversion. Furthermore, given what seems to be a growing currency of arguments that courts have been overly or inappropriately deferen-

140. Now-judge David Barron and now-Justice Elena Kagan have argued that the presence of an intra-agency subdelegation should affect the level of deference given to agency actions. David J. Barron & Elena Kagan, *Chevron's Nondelegation Doctrin*e, 2001 SUP. CT. REV. 201 (2001). Specifically, they proposed that interpretations authorized by agency leaders who also take responsibility for rules they personally authorize should be given *Chevron* deference, whereas interpretations made by lower-level employees should receive only more limited *Skidmore* deference. They argued that political accountability and thorough decision making are more likely to arise when high-level, appointed agency members have a significant hand in the rulemaking process. If courts followed their approach in cases challenging a subdelegation to a machine-learning system, perhaps such a system should receive additional scrutiny when only low-level agency members specify the objective functions and other normative choices embedded in the algorithms. Furthermore, although such an approach would calibrate deference based on a hierarchy of the human decision makers specifying the algorithms, an alternative application of Barron and Kagan's proposal might not start with the assumption that the decision makers are indeed humans. Rather, as Associate Justice Cuéllar has written, courts might view the algorithms themselves as the decision makers and give deference accordingly. *See* Cuéllar, *supra* note 25, at 17–19. He also rightfully acknowledges, though, that such "policing the line between automated decision and decisions support is difficult." *Id*. at 18.

141. Chevron U.S.A., Inc. v. NRDC, 467 U.S. 837 (1984).

142. *See* United States v. Mead Corp., 533 U.S. 218, 227 (2001) ("When Congress has 'explicitly left a gap for an agency to fill, there is an express delegation of authority to the agency to elucidate a specific provision of the statute by regulation,' and any ensuing regulation is binding in the courts unless procedurally defective, arbitrary or capricious in substance, or manifestly contrary to the statute." (quoting *Chevron*, 467 U.S. at 843–44) (citations omitted)).

tial to agencies,[143] the possibility of future changes in deference standards cannot be overlooked. It is also possible, of course, for Congress in the future simply to prohibit an agency—or perhaps all agencies—from subdelegating to automated systems based on machine-learning algorithms. Despite these future possibilities arising on statutory grounds or potential changes to the *Chevron* doctrine, when it comes to the core constitutional question, we do not foresee delegation considerations posing any substantial legal hurdle to regulatory applications of machine learning.

B. DUE PROCESS

If the prospect of rulemaking by robot might animate concerns about the nondelegation doctrine, the prospect of adjudicating by algorithm will no doubt raise due process concerns. Administrative agencies, after all, are expected to provide adequate procedural due process when taking actions that could deprive individuals or entities of protected liberty interests or property rights and entitlements.[144] These deprivations can occur whenever agencies exercise their adjudicatory authority, such as when the EPA orders businesses to engage in the costly cleanup of hazardous waste for which they are deemed responsible[145] or the Social Security Administration terminates financial payments of governmental benefits to individuals deemed to be no longer eligible.[146] In the machine-learning era, would an agency's reliance on an automated machine-learning system to make such adjudicatory decisions constitute a violation of an individual's constitutional right to due process?

As we noted at the outset of this Part, much will turn on how machine learning is used and how adjudications by algorithm are conducted. Machine-learning systems could be used, as discussed in section I.C, simply to generate inferences that constitute but one of several inputs into an independent judgment made by human officials. Or they could be used to make automatic decisions about deprivations of protected rights and interests. This latter approach, involving automated, algorithmic decisions in adjudicatory matters, would seem the most likely to deny individuals their due process right to a

143. *See, e.g.*, Ronald A. Cass, *Vive la Deference?: Rethinking the Balance Between Administrative and Judicial Discretion*, 83 GEO. WASH. L. REV. 1294, 1319–26 (2015).

144. Bi-Metallic Inv. Co. v. State Bd. of Equalization, 239 U.S. 441, 445–46 (1915) (suggesting that a right to be heard is necessary when a "relatively small number of persons [is] concerned, who [are] exceptionally affected, in each case upon individual grounds," but not in instances of adoption of widely-applicable rules); Londoner v. Denver, 210 U.S. 373, 386 (1908) ("[S]omething more than [an opportunity to submit objections in writing], even in proceedings for taxation, is required by due process of law. . . . [E]ven here a hearing in its very essence demands that he who is entitled to it shall have the right to support his allegations by argument however brief, and, if need be, by proof, however informal."). *See generally* Edward L. Rubin, *Due Process and the Administrative State*, 72 CAL. L. REV. 1044 (1984).

145. *See, e.g.*, Gen. Elec. Co. v. Jackson, 595 F. Supp. 2d 8, 21–29 (D.D.C. 2009).

146. *See, e.g.*, Mathews v. Eldridge, 424 U.S. 319, 323–25 (1976).

hearing—one of the canonical tenets of procedural due process.[147] Even so, we will explain why even such machine-learning-based decisions are in no way uniquely barred from satisfying due process. After reaching this conclusion, though, we will highlight an area of tension in current case law that may make it difficult to for agencies to determine when procedural protections are needed in an age of algorithmic adjudication.

1. Tests for Due Process

The landmark case of *Goldberg v. Kelly* is usually recognized for establishing that due process protections extend to recipients of government welfare benefits, which previously had been considered privileges that did not afford the same protection as constitutional rights.[148] Although *Goldberg* is often cited for the Court's rejection of the right–privilege distinction, just as central to the Court's decision was the importance of welfare recipients having the ability to appear in person before the government official who makes the decision about whether to terminate individual benefits.[149] The Court rejected a process that was based on written documentation, finding that it was even insufficient for a human case-worker to present the case on behalf of the welfare recipient.[150]

Six years later, in *Mathews v. Eldridge*, the Supreme Court held that, in the context of Social Security disability benefits, the government could structure a process based solely on paperwork review and without offering the recipients of disability benefits a pre-termination hearing before an administrative official.[151] The Court did not overturn *Goldberg*, but instead it offered the now-canonical balancing test involving three factors that must be weighed in determining whether due process has been satisfied:

> First, the private interest that will be affected by the official action; second, the risk of an erroneous deprivation of such interest through the procedures used, and the probable value, if any, of additional or substitute procedural safeguards; and finally, the Government's interest, including the function involved and the fiscal and administrative burdens that the additional or substitute procedural requirement would entail.[152]

147. Although we do not take up the possibility that machine-learning algorithms could actually "conduct" adjudicatory hearings, we note the remote possibility that advances in artificial intelligence may one day permit such hearings. *Cf.* Judith Newman, *To Siri, With Love*, N.Y. TIMES (Oct. 17, 2014), http://www.nytimes.com/2014/10/19/fashion/how-apples-siri-became-one-autistic-boys-bff.html [https://perma.cc/BF65-BH5G] (describing developmentally beneficial "discussions" between Apple's digital assistant Siri and autistic children).

148. 397 U.S. 254, 264 (1970).

149. *Id.* at 268 (finding unconstitutional procedures that "do not permit recipients to appear personally with or without counsel before the official who finally determines continued eligibility" and that do not allow recipients "to present evidence to that official orally, or to confront or cross-examine adverse witnesses").

150. *Id.* at 269.

151. 424 U.S. 319, 348–49 (1976).

152. *Id.* at 335.

The *Mathews* Court explicitly noted that, "[d]ue process is flexible and calls for such procedural protections as the particular situation demands."[153]

With *Mathews* balancing now defining procedural due process, we conclude that algorithmic adjudication would appear to fare suitably well in most instances (even if *Goldberg* might still preclude reliance on machine learning in the context of state welfare benefits). Of course, given that the *Mathews* standard calls for balancing, it is impossible to make sweeping generalizations about how machine-learning algorithms would fare when used by agencies making specific kinds of decisions. But at least one thing should be clear: No agency should categorically rule out the use of machine-learning algorithms to support adjudicatory decisions. Among the areas of the law in which "it depends" would seem to be the best answer, this is definitely one.

Mathews's first factor—the private interest at stake—is something entirely exogenous to machine learning. But the last two factors could very well be affected, perhaps dramatically so in some cases, by the use of algorithms. When used to administer extremely large programs, machine learning could save the government a substantial amount of money by forgoing pre-deprivation hearings, thereby significantly increasing the weight of the third factor in favor of machine learning in any due process balancing. It is the second factor, then, that will likely prove critical in determining the constitutional validity of automated adjudications—namely, how well machine learning stacks up against alternative, non-machine-based procedures in terms of avoiding erroneous deprivations. Agencies will need to be justifiably confident that automatic algorithmic deprivations have sufficiently low error rates to weigh heavily against the need for pre-deprivation hearings before human officials. We cannot, of course, determine ex ante what will constitute sufficiently low rates for all uses. Given that the nature of a balancing test like that in *Mathews* requires case-by-case determinations, we see little reason to write off all adjudicatory algorithms as unable to conform to due process expectations. On the contrary, because algorithms can reduce the possible introduction of negative features of human judgment—such as prejudice, bias, and mistakes—we have reason to expect that the use of machine learning by government can advance due process values.[154]

2. Evaluating Adjudicatory Algorithms

When agencies seek to assess the superiority of their machine learning in terms of error avoidance, the second *Mathews* factor, two conditions must be met: (1) error rates truly representing the decision processes' accuracy must be

153. *Id.* at 334 (quoting Morrissey v. Brewer, 408 U.S. 471, 481 (1972)) (internal quotation marks omitted).

154. Cary Coglianese, *Robot Regulators Could Eliminate Human Error*, S.F. CHRON. (May 5, 2016), http://www.sfchronicle.com/opinion/article/Robot-regulators-could-eliminate-human-error-7396749.php [https://perma.cc/WCH2-KNAB].

available, and (2) those error rates must prove to be acceptable in the overall balance of factors. The first of these two conditions should easily be met by virtue of machine learning being a purely mathematical decision-making process. Unlike qualitative methods for determining the factual bases of deprivations, mathematical classification systems provide estimates of their error rates before they are actually implemented; one can obtain from machine-learning algorithms, for instance, a form of output called confusion tables, which show various error rates in a test data set.[155] Assuming some degree of similarity between the test data and the population in which the algorithms will be implemented, these error rates can be generalized to real-world adjudications.

The second condition, an error rate's acceptability, might be determined by reference to a body of case law that characterizes acceptable error rates for other types of procedures. Relying on comparisons of error rates associated with other procedures would be in line with courts' overall emphasis, whenever possible in applying the *Mathews* test, on quantitative estimates of error based on empirical evidence rather than on abstract or hypothetical concepts of risk.[156] Although due process determinations will still be flexible and require case-by-case balancing, agencies could reasonably look to precedents about other procedures to guide their decisions about algorithmic autonomy. If the courts typically look favorably on conventional procedural mechanisms that have a certain level of accuracy, then automatic, algorithm-based adjudications with similar error rates should be likely to pass constitutional muster.

Looking to precedents will reveal, on the one hand, low error rates that courts have deemed acceptable. In *Mathews*, the risk of erroneous deprivation, as measured by reversals upon appeal, was 3.3 percent,[157] and the Court noted that with such a low error rate "[t]he potential value of an evidentiary hearing, or even oral presentation to the decision maker, is substantially less in this context than in *Goldberg*."[158] Error rates lower than 1 percent have been similarly treated as acceptable.[159] In *General Electric Co. v. Jackson*, the district court even held that an error rate of 4.4 percent was deemed "an acceptable rate of error,"[160] notwithstanding the court's recognition that "the private[] interest[s] [were] significant."[161]

---

155. *See* BERK, *supra* note 33, at 108–10. When predicting a binary outcome variable, these error rates will represent false positives and false negatives, and error rates from regression applications of machine learning will represent overestimates or underestimates. *See id.*

156. *See, e.g.*, Gen. Elec. Co. v. Jackson, 595 F. Supp. 2d 8, 33 (D.D.C. 2009).

157. 424 U.S. 319, 346 n.29 (1976).

158. *Id.* at 344–45.

159. *See, e.g.*, Gray Panthers v. Califano, 466 F. Supp. 1317, 1323 (D.D.C. 1979) (noting that an error rate of less than 1 percent "is a small fraction of the 3.3% rate the Court in *Mathews* upheld as consistent with due process").

160. *Gen. Elec. Co.*, 595 F. Supp. 2d at 37 (quoting Shands v. Tull, 602 F.2d 1156, 1160 (3d Cir. 1979)).

161. *Id.* at 38.

At the other end of the continuum, courts have concluded that high error rates that represent predictions no better, or perhaps even worse, than guessing by chance are unacceptable and necessitate additional due process protections to reduce these rates. For example, error rates of 50 percent[162] and 51.6 percent[163] have been deemed a "substantial risk,"[164] and an error rate of 74.6 percent has been labeled "unacceptably high."[165]

Between these extremes, court decisions reveal less agreement over the risk that more moderate error rates pose, and, in some of these cases, the error rate is not treated as a particularly weighty or controlling balancing factor at all. An error rate of 30 percent, for example, was deemed to constitute "a very high error rate for purposes of *Mathews v. Eldridge*,"[166] although an 18.96 percent error rate was considered a "slight"[167] risk of error, despite, in a different case, a court treating an error rate of 11.67 percent as a "high risk of error."[168] As a result, despite the precision that judicial notice of error rates to two decimal places might imply, no fixed percentage can be found in the case law demarcating a clean line between acceptable and unacceptable levels of error.

In addition to a lack of a clear, fixed threshold of acceptability, past precedents present another challenge in that reversal rates are not necessarily a complete or accurate measure of error. In the particular context of *Mathews*, the Court noted that the administrative review system "operated on an open-file basis,"[169] meaning that those denied disability benefits could, at any point following their deprivations, submit new evidence that could result in additional medical examinations. Individuals who availed themselves of these procedures to have an initial denial changed were not included in the 3.3 percent reversal rate because their cases were never appealed.[170] Similarly, and relevant to measurements in all scenarios, reversal rates may fail to capture those who are erroneously denied benefits but who choose for various unrelated reasons not to file an appeal that would lead to an evidentiary hearing and reversal. For these reasons, the actual rate of "error" in administrative systems will presumably be higher than indicated by a rate of reversal.[171]

---

162. Humphries v. County of Los Angeles, 554 F.3d 1170, 1195 (9th Cir. 2008), *rev'd on other grounds sub nom.* Los Angeles County v. Humphries, 562 U.S. 29 (2010).

163. White v. Mathews, 434 F. Supp. 1252, 1255 n.12 (D. Conn. 1976).

164. *Humphries*, 554 F.3d at 1195.

165. 595 F. Supp. 2d at 37 (referencing Dupuy v. Samuels, 397 F.3d 493, 505, 512 (7th Cir. 2005)).

166. Markadonatos v. Village of Woodridge, 739 F.3d 984, 1001 (7th Cir. 2014) (Hamilton, J., dissenting).

167. Graves v. Meystrik, 425 F. Supp. 40, 48–49 (E.D. Mo. 1977).

168. Atkins v. Parker, 472 U.S. 115, 151 (1985) (Brennan, J. dissenting) (citing Foggs v. Block, 722 F.2d 933, 939 (1st Cir. 1983) (finding an error rate of 585 households out of 5,013 households, or 11.67 percent)).

169. Mathews v. Eldridge, 424 U.S. 319, 347 (1976).

170. *Id.* at 346 n.29.

171. Perhaps it is for this reason that the *Mathews* Court noted that, "[b]are statistics rarely provide a satisfactory measure of the fairness of a decisionmaking process." *Id.* at 346.

These deficiencies in measuring error risk based on reversal rates may pose a challenge to an agency's search for judicial guidance on automatic algorithmic deprivations. Algorithms' test-data error rates are mathematically computed in a way that makes it difficult to compare them directly to the reversal rates currently relied upon within the case law. Although the case law seems to indicate that reversal rates under 5 percent should be sufficient to avoid the need for pre-deprivation hearings, agencies will be unable to determine ex ante what any particular algorithm's mathematical error rate equates to in terms of a reversal rate. As a more accurate estimate of the total error in a system will likely be higher than any reversal rate, an algorithm's error rate might be higher too, even when reliance on the algorithm actually results in a less error-prone process. Still, if an algorithm's error rate itself fell below 5 percent, then agencies relying on the algorithm for adjudication could presumably predict that the courts would deem this rate to be acceptable.

If nothing else, agencies could structure algorithmic adjudication systems to produce data on reversal rates. By incorporating a delay period between the time when an algorithmic prediction is made and when any corresponding deprivation of rights occurs, affected individuals or entities would have the opportunity to request a hearing, thus generating over time some data on reversals. These delay periods could accompany a phase-in or piloting of an automated adjudicatory system, providing agencies with the data needed to establish how their test-data error rates correspond with reversal rates. As long as these error rates prove sufficiently low, we anticipate no categorical reason why courts would reject adjudication by algorithm merely due to its digitized and automated nature.

3. Cross-Examination of Adjudicatory Algorithms

Perhaps the more substantial due process question presented by automated adjudication stems from how such a system would affect an aggrieved party's right to cross-examination. Claimants presumably should have a right to understand and provide comment on the design of any algorithm, as well as the underlying data used.[172] Unfortunately, it will seldom if ever be readily discernable why any particular prediction resulted from machine learning, and, consequently, it will not be easy to determine what evidence would be needed to

172. An agency could conceivably bias an algorithm in deliberate ways that could be exposed on appeal. In many machine-learning algorithms, it is possible for an agency to set the algorithm's cost ratio, or the preferred ratio of false positives to false negatives, which in turn affects the corresponding error rates. *See* BERK, *supra* note 33, at 139–45. Thus, an agency could potentially bias an algorithm to, for example, have more error when predicting that someone should be deprived than when predicting that someone should not be denied if the agency views not depriving someone as less costly than the pre-deprivation hearing that would be necessary if that individual were deprived. Outside of the administrative context, previous scholarship has discussed the difficulties faced in uncovering similar hidden biases and value choices in algorithms applied to criminal justice. *See* Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245, 1269–76 (2016).

rectify errors.[173] Any of the variables could have been predictively important ones, and any of those variables could have been the one (or ones) that contained the error. Probably the only meaningful way to identify errors would be to conduct a proceeding in which an algorithm and its data are fully explored. This could be accomplished through agency rulemaking establishing an automated adjudicatory system in the first place and in specifying the underlying algorithm's parameters. It might also be raised in individual hearings or appeals seeking to analyze separately an algorithm's forecast in each case, probing for sources of error in the data and specifications underlying those individual predictions.[174]

Given that this thorough examination of algorithmic error will entail sophisticated deconstructions of data quality and methodology, few individual claimants will readily possess the mathematical and analytical ability to conduct such an investigation without assistance.[175] Even with the appointment of counsel, or with a requirement that a hearing examiner effectively act as counsel, claimants may not be able to receive a fair hearing about the choices embedded in and the data used by algorithmic adjudication. Most potential counsel or agency hearing examiners do not possess the necessary skills to interrogate machine learning systems. An expert must have a good degree of statistical knowledge to understand how, for example, variables' importance values are interpreted and what consequences those interpretations hold for the effects of possible erroneous variable measurements on predictions. Adjudicating by algorithm may thus necessitate the establishment of a body of neutral and independent statistical experts to provide oversight and review, or more likely a prior rulemaking process informed by an expert advisory committee or subjected to a peer review process. Fortunately, such procedural steps are well established and have been

173. One may contend that the kind of evidence that must be provided to rectify errors relates to the underlying matter of fact in dispute and not to the method by which that underlying matter is estimated. As previously discussed, however, machine learning will often be applied to make predictions about risks not capable of being directly evidenced; if PHMSA orders the shutdown of a pipeline, for instance, the issue disputed at a hearing will be whether that pipeline indeed posed a risk for future failure, and the kind of physical evidence that could be presented at the hearing may not be capable of demonstrating the pipeline's lack of risk. Therefore, a pipeline operator in such a hearing will have to provide evidence instead about the potential sources of error in the algorithm that was used initially to predict the risk.

174. Outside the context of adjudicatory algorithms, previous scholarship has emphasized the need for individuals to be able to challenge automated decisions rendered by corporate entities' algorithms, such as those engaged in credit scoring. *See generally* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); Tom Baker & Benedict G.C. Dellaert, *Regulating Robo Advice Across the Financial Services Industry*, 103 IOWA L. REV. (forthcoming 2017).

175. Courts have held that counsel may be required for individuals who lack the abilities to present their case due to factors like low educational attainment or mental handicap. *See, e.g.*, Smith v. Sec'y of Health, Educ., & Welfare, 587 F.2d 857, 860–61 (7th Cir. 1978) (per curiam); Alamo v. Richardson, 355 F. Supp. 314, 316–17 (D.P.R. 1972). An alternative to providing counsel can be requiring that the body administering the hearing effectively serve as counsel. *See* Gold v. Sec'y of Health, Educ., & Welfare, 463 F.2d 38, 43 (2d Cir. 1972).

long used in other complex administrative matters.[176]

In the end, whether any particular algorithmic system will satisfy the standards of due process will depend on how well that system works and on the adequate validation of its performance. But nothing in principle stands in the way of satisfying the demands of due process, even for automated adjudication by algorithm, especially given the flexibility courts have long granted agencies in crafting adjudicatory procedures. Furthermore, given the well-established biases and errors that can creep into human judgment, and recognizing that machine-learning algorithms have demonstrated superiority over human decisions in other contexts,[177] it is reasonable to conclude that agencies will be able to satisfy the demands of due process even in the machine-learning era.[178]

C. ANTIDISCRIMINATION

As agencies face choices about how to design machine-learning algorithms in the service of adjudication and rulemaking, they will need to decide what kinds of variables should be included in the data these algorithms use to make predictions. They may have at their disposal data on demographic variables representing membership in legally protected classes, and inclusion of these variables in algorithms could increase their predictions' accuracy and potentially reduce disparities in accuracy across classes.[179] But including them, and

176. *See, e.g.*, U.S. FED. AVIATION ADMIN., UNMANNED AIRCRAFT SYSTEMS AVIATION RULEMAKING COMMITTEE (2011), https://www.faa.gov/regulations_policies/rulemaking/committees/documents/media/UASARC-6172011.PDF [https://perma.cc/UQ4L-DTE5]. General procedures for advisory committees are provided by the Federal Advisory Committee Act §§ 1–16, 5 U.S.C. app. 2 (2012). In addition, the Office of Management and Budget has established guidelines for peer review. Final Information Quality Bulletin for Peer Review, 70 Fed. Reg. 2,664 (Jan. 14, 2005).

177. *See, e.g.*, Berk, *Forecasting Domestic Violence*, *supra* note 104; Kriegler & Berk, *supra* note 51.

178. *See supra* note 104.

179. Imagine, for instance, that for a given protected attribute a data set contains information about two classes of individuals. One class is larger and has, on average, more "advantageous" outcomes in terms of the output variable, however defined. The other class is smaller and has, on average, less advantageous outcomes. If an algorithm were ignorant of or blind to the class identities of individuals in this data set, the algorithm would generate one classification rule that would generally perform more accurately for the majority class (simply because it is larger), causing less accurate predictions for the minority class. If, however, this algorithm explicitly took into account individuals' class memberships, it could generate two different classification rules, one for each class, allowing for more accurate predictions overall and, most pertinently, for the minority class. There may, of course, be instances in which the increases in overall accuracy and decreases in accuracy disparities are negligible. Particularly, nearly equally accurate predictions may result in data sets lacking class-related variables but containing many other variables that are highly correlated with class membership, provided that the data set is sufficiently large. *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671, 695 (2016). This finding may naturally prompt the question of whether including such proxy variables capable of recapitulating class membership raises equal protection concerns, and, indeed, this question has been considered outside the context of algorithmic decision making. *See generally* Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803 (2014). We do not address proxy variables in any depth in this Article because, as will be seen, the reasoning underlying our conclusion about class-related variables necessarily yields the same conclusion for sets of proxy variables.

then basing action on the resulting predictions, raises serious concerns about illegal discrimination, especially if those predictions lead to differential effects on members of a protected class. As a White House report on big data has warned, "[p]owerful algorithms can unlock value in the vast troves of information available . . . but also raise the potential of encoding discrimination in automated decisions."[180] In this section, we take up the question of whether inclusion of class-related input variables in federal agencies' algorithms would render them unconstitutional for equal protection reasons. We find that, although consideration of suspect class membership is almost universally struck down for contravening equal protection in traditional decision-making processes, quantitative or otherwise, some of the unique attributes of how machine learning operates may shield agencies from a finding of unconstitutionality. This is not to say that the use of machine-learning algorithms by federal agencies will always be unproblematic under equal protection; on the contrary, the use of algorithms could be driven by manifest, easily-discoverable animus or discriminatory intent as much as any other decision process. Our suggestion is simply that potential equal protection challenges to agency use of machine learning will likely face a high bar if officials in these agencies employ algorithms responsibly and in good faith, building a suitable record to demonstrate the propriety of their design and use.

Under the Fifth Amendment, agencies must respect the equal protection rights of those affected by their actions.[181] Federal agencies, though, do not violate those rights "*solely* because [an action] has a . . . disproportionate impact" on a protected class, such as race or religion.[182] Rather, to show a violation of equal protection by the federal government, "[a] purpose to discriminate must be present."[183] In other words, agency action that purposefully discriminates based on a protected class—that is, action that engages in dispa-

180. U.S. EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 45 (2014), http://www.whitehouse.gov/sites/default/files/docs/big_data_ privacy_report_may_1_2014.pdf [https://perma.cc/T7AN-BU8P]. Similar concerns over the potential for discrimination when algorithms are applied to big data have also been voiced in the press. *See, e.g.*, Katherine Noyes, *Will Big Data Help End Discrimination—or Make It Worse?*, FORTUNE (Jan. 15, 2015, 3:16 PM), http://fortune.com/2015/01/15/will-big-data-help-end-discrimination-or-make-it-worse [https://perma.cc/8PPY-7Y3A]; Michael Schrage, *Big Data's Dangerous New Era of Discrimination*, HARV. BUS. REV. (Jan. 29, 2014), https://hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination [https://perma.cc/Y7ME-HYLN].

181. Although the Fourteenth Amendment's Equal Protection Clause prohibits discrimination by the states, the Warren Court held that the concepts of due process and equal protection "are not mutually exclusive" and that "discrimination may be so unjustifiable as to be violative of" the Fifth Amendment's Due Process clause. Bolling v. Sharpe, 347 U.S. 497, 499 (1954). Accordingly, analysis of discrimination claims under the Fifth Amendment mirrors that of equal protection analysis under the Fourteenth Amendment, and much of the case law we discuss in this Part arises under the latter provision of the Constitution.

182. Washington v. Davis, 426 U.S. 229, 239 (1976).

183. *Id.* (quoting Akins v. Texas, 325 U.S. 398, 403–04 (1945)).

rate treatment—will be legally suspect and subject to heightened scrutiny.[184] At the end of this section, we will briefly suggest that regulatory decisions made by machine learning could in many instances withstand heightened scrutiny, but of course we recognize that evaluation under such a heightened standard of review will ultimately be highly fact-dependent. Our principal argument in this section is that algorithms that include variables indicating protected class membership will seldom if ever trigger heightened scrutiny, at least in the absence of any explicit showing of discriminatory intent or animus. Equal protection challenges to machine learning will, in short, likely fail at the first step of analysis that demands a finding that algorithms that include or analyze class-related variables are intentionally discriminatory.

1. Suspect Classifications

Because we assume that there will rarely exist direct evidence that a decision maker—in this case, an algorithm's programmers or their superiors—consciously intended to discriminate on the basis of a protected class, evidence of any such intent would need to be found indirectly and circumstantially. We will address this kind of search for inferences of discrimination in the next section, where we explain why machine learning is unlikely to give rise to such inferences. But first we take up the typical way that claimants alleging disparate treatment succeed in showing that governmental action deserves heightened scrutiny: namely, by demonstrating that an agency action involves a suspect *classification*. Such classifications are automatically subject to heightened scrutiny.[185]

Courts will likely find it difficult to conclude that agency algorithms that analyze class-related variables entail suspect classifications.[186] For one thing, the Supreme Court has never clearly defined what constitutes a suspect classification.[187] Its attitude has been relatively consistent when the *only* variable

184. Racial discrimination, for instance, is subject to strict scrutiny review. *See, e.g.*, Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 213 (1995). Disparate treatment on the basis of sex, however, is generally considered subject to intermediate scrutiny. *See, e.g.*, Miss. Univ. for Women v. Hogan, 458 U.S. 718, 724 (1982); Craig v. Boren, 429 U.S. 190, 218 (1976) (Rehnquist, J., dissenting).

185. *See, e.g.*, *Adarand Constructors, Inc.*, 515 U.S. at 235; *Craig*, 429 U.S. at 197; United States v. Carolene Prods. Co., 304 U.S. 144, 152 n.4 (1938).

186. It should be made clear that the use of "classification" to refer to the goal of a machine-learning algorithm must be disentangled from its use to refer to how the race-related information is treated within an algorithm. Scholars have suggested that, because a machine-learning algorithm has classification as its goal, the inclusion of class-related variables in such an algorithm necessarily renders it a legally harmful suspect classification. Barocas & Selbst, *supra* note 179, at 695. As our subsequent discussions will show, however, what matters for determining the level of judicial scrutiny is not just whether the goal of a government decision-making process is classificatory in overall nature, but whether within that process individuals are classified *by* their class memberships, such that being a member of one class is advantageous or disadvantageous when compared to being a member of another class.

187. *See* Stephen Menendian, *What Constitutes a "Racial Classification"?: Equal Protection Doctrine Scrutinized*, 24 TEMP. POL. & C.R. L. REV. 81, 82 (2014) ("The Supreme Court has, to date, not provided a definition of the term 'racial classification.' Increasingly a term of art, the Court has not

contributing to a governmental decision is membership in a protected class and where such membership universally leads to disadvantaged outcomes; such decision schemes are clearly suspect classifications and must be subjected to heightened scrutiny.[188] Beyond such manifestly problematic scenarios, the Court's criteria for suspect classifications have been murkier. Still, the Court has subjected to heightened scrutiny decision processes in which membership in a protected class was but one of many variables considered by government decision makers.[189] It might therefore be assumed that *any* inclusion of class membership, or even the mere consideration of a protected class, would suffice to conclude that a decision-making process depends on a suspect classification.[190] If this were the case, then any machine-learning algorithm that analyzes data containing variables for membership in a protected class would surely receive heightened scrutiny. Yet, as others have suggested, governmental actions that are class-conscious but do not *classify* on the basis of group membership may be exempt from heightened scrutiny.[191] In other words, government might consider a suspect *class* without necessarily resulting in a *classification*. How class membership is considered, beyond merely the fact that it is, can matter greatly when it comes to determining if a government action rests on a suspect classification.

Previous scholarship has put forward five such aspects of a classification: "(1) an official government label (2) proclaiming or identifying the [class] of (3) a particular individual, (4) which is then the basis for allocating benefits or imposing burdens (5) on the person classified."[192] The machine-learning algorithms we envision would certainly contain an official government label indicating particular individuals' class memberships. It is less clear, however, that these algorithms would allocate benefits or burdens on the basis of class membership in the same way as have almost every decision scheme that to date has been subjected to heightened scrutiny by courts. Decision schemes previously reviewed by the courts have universally exhibited what we label "categori-

---

always been careful about its usage or consistent in its application." (citations omitted)); Reva B. Siegel, *Equality Talk: Antisubordination and Anticlassification Values in Constitutional Struggles Over* Brown, 117 HARV. L. REV. 1470, 1542 (2004) ("American antidiscrimination law has no determinate criteria for deciding what practices are group-based classifications, and while courts sometimes articulate such criteria, they often apply them inconsistently."); Marcy Strauss, *Reevaluating Suspect Classifications*, 35 SEATTLE U. L. REV. 135, 138 (2011) ("The Supreme Court has not provided a coherent explanation for precisely what factors trigger heightened scrutiny.").

188. *See, e.g.*, Bolling v. Sharpe, 347 U.S. 497, 499 (1954).

189. *See, e.g.*, Gratz v. Bollinger, 539 U.S. 244, 251–57 (2003); Grutter v. Bollinger, 539 U.S. 306, 312–16 (2003).

190. Indeed, scholars have sometimes made this claim. For example, Sonja Starr stated that, in the 2013 iteration of *Fisher v. University of Texas at Austin*, "the Supreme Court . . . applied heightened constitutional scrutiny to the mere *consideration* of constitutionally suspect factors" in the university's undergraduate admissions process. Starr, *supra* note 179, at 864 (discussing Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411 (2013)).

191. *See* Menendian, *supra* note 187, at 91–95; Stephen M. Rich, *Inferred Classifications*, 99 VA. L. REV. 1525, 1579–86 (2013).

192. Menendian, *supra* note 187, at 102 (citations omitted).

cally different treatment" based on class membership. In the past, when an individual's race has been factored into a decision, being a member of a given race has caused individuals of that race to have their predicted outcomes affected in a consistently and categorically different way than individuals of another race. In these situations, being a member of a given race has always been, across individuals of that race, either advantageous or disadvantageous when compared to being a member of a different race.[193] Yet with machine learning, many, if not most, algorithms will not lead to categorically different treatment, even when they include consideration of variables for race and other protected classes.

Concern about categorically different treatment on the basis of class membership underlies many Supreme Court decisions under the Equal Protection Clause. In *Gratz v. Bollinger*, white undergraduate applicants were categorically denied twenty points awarded to students belonging to underrepresented minority groups in a point-based admissions system, resulting in a suspect classification.[194] In *Grutter v. Bollinger*, a law school used a more holistic admissions process that did not distribute points, but this process was still deemed a suspect classification; when race did factor into an admissions decision, membership in an underrepresented minority group was universally given a degree of preference for admission over lack of such membership.[195] The Court has similarly held that heightened scrutiny is warranted due to categorically different treatment arising from a public school board's method for determining teacher layoffs that gave preferential treatment to minority teachers.[196] Another school district's process of assigning students to schools triggered heightened scrutiny because, for each school, it gave preference to students of certain races whose assignment would place the school's racial makeup closer to that of the school's

193. Going forward, we discuss categorically different treatment of class membership in contexts where one class is advantaged—it is awarded benefits or allowed to avoid burdens—and another is disadvantaged. But the underlying conceptual formulation of categorically different treatment also applies to instances in which the attached benefits or burdens of governmental treatment are less evident. Take, for instance, *Johnson v. California*, which concerned the assignment of prisoners only to cells with cell mates of the same race; no prisoners of particular races were burdened any more than others, but the Court explicitly stated that allowing this action to escape strict scrutiny "ignores our repeated command that 'racial classifications receive close scrutiny even when they may be said to burden or benefit the races equally.'" Johnson v. California, 543 U.S. 499, 506 (2005) (quoting Shaw v. Reno, 509 U.S. 630, 651 (1993)). It should be clear, though, that this assignment of prisoners by race nonetheless involved categorically different treatment, even if not categorically different allocation of burdens or benefits; all prisoners of a given race were assigned to certain cells, whereas all prisoners of another race were assigned to different cells. *Id*. at 502 ("In fact, the CDC has admitted that the chances of an inmate being assigned a cellmate of another race are '[p]retty close' to zero percent." (citations omitted)).

194. *Gratz*, 539 U.S. at 266.

195. *See* Grutter v. Bollinger, 539 U.S. 306, 318–20 (2003). A similarly holistic admissions process was the subject of *Fisher v. University of Texas at Austin*; when considered, underrepresented minority status was categorically given preference over non-minority status to remedy the university's lack of "a 'critical mass' of minority students." *Fisher*, 133 S. Ct. at 2416.

196. Wygant v. Jackson Bd. of Educ., 476 U.S. 267, 273–74 (1986) (plurality opinion).

district.[197] In each of these cases, a public decision process treated membership in certain races as advantageous, while treating membership in others as disadvantageous.

These are just a few examples, but they demonstrate the pervasiveness of categorically different treatment of class membership. Traditionally, taking race or other protected class characteristics into account meant drawing a distinction that leads to one class being advantaged to some degree over another class by being more likely to receive benefits or avoid burdens. The Court's reasoning in its equal protection cases underscores how doctrinally important this categorically different treatment of class membership is to finding a reliance on a suspect classification. In *Fullilove v. Klutznick*, the Supreme Court applied strict scrutiny precisely because the legislative scheme gave and withheld "preference based on" ethnic membership.[198] Similarly, in *Wygant v. Jackson Board of Education*, the Court stated that the challenged teacher layoffs "operate[d] against whites and in favor of certain minorities, and therefore constitute[d] a classification based on race."[199] More recently, the Court emphasized its concern in *Fisher v. University of Texas at Austin* about a state "considering racial minority status as a positive or favorable factor in a university's admissions process."[200] The Court reasoned that "[t]he principle of equal protection admits no 'artificial line of a "two-class theory"' that 'permits the recognition of special wards entitled to a degree of protection greater than that accorded others.'"[201] In other words, equal protection law singles out for heightened scrutiny those classifications that arise when government draws lines that categorically favor some classes and disfavor others.[202]

What makes explicit consideration of class membership odious and offensive to equal protection principles is not just the government officials' consideration of race and other protected classes, but their use of class status as part of a *classification*, when a membership in one class is consistently treated as advantageous or disadvantageous compared to membership in another class. When it comes to the use of machine-learning algorithms, however, consideration of class membership will not necessarily, or even often, give rise to categorically different treatment. We can expect that most machine-learning applications will be used to forecast complex phenomena—such as tax fraud or disability eligibility—that are not easily predicted by standard, less powerful, statistical techniques. Such complex phenomena are complex precisely because their causes are not uniform and homogenous across the population. When machine learning is applied to predict such phenomena, its nonparametric nature and incorpora-

197. Parents Involved in Cmty. Schs. v. Seattle Sch. Dist. No. 1, 551 U.S. 701, 709–17 (2007).

198. 448 U.S. 448, 491 (1980) (plurality opinion).

199. 476 U.S. at 273–74.

200. 133 S. Ct. at 2417.

201. *Id.* (quoting Regents of Univ. of Cal. v. Bakke, 438 U.S. 265, 295 (1978) (plurality opinion)).

202. GEOFFREY R. STONE ET AL., CONSTITUTIONAL LAW 498 (7th ed. 2013) (stating suspect classifications are those where "the government has drawn [a] line between . . . favored and disfavored groups").

tion of multidimensional interaction effects allow this heterogeneity to be manifested—and more fully exploited for predictive benefit.[203] What this means for equal protection analysis is that if machine-learning algorithms incorporate an input variable representing individual membership in a particular class, it is highly unlikely that the effect of that variable on predicted outcomes will be consistently advantageous or disadvantageous for the multitude of individuals in that class. Such algorithms will be designed, after all, to optimize for accuracy in making predictions of fraud or benefits eligibility, not for giving preferential treatment based on race or other class characteristics. A learning algorithm will generally be able to improve its accuracy through more data and additional variables of all types. The forecasts an algorithm generates will likely support favorable treatment for some members of a class and unfavorable treatment for other members of that same class, as these forecasts depend on complex, non-linear interactions of multiple variables.[204]

In sum, determining that machine-learning algorithms create a suspect classification will hardly be a clear-cut task, even when government agencies include class-related variables in their data sets. Until now, explicit consideration of individuals' memberships in protected classes has virtually always been labeled as a classification because of the categorically different treatment that accompanies such consideration. We recognize that, for this reason, some courts may instinctually incline toward treating as a suspect classification the mere consideration of class-related variables by learning algorithms. Yet what truly makes such consideration a classification, and thus constitutionally suspect, is the attendant attachment of advantage to one class, and of disadvantage to another, by the officials who are collecting and analyzing data in traditional ways—an attachment that will neither necessarily nor usually be present with the use of machine learning. The courts' heightened scrutiny of suspect classifications exists not to prevent any and all governmental considerations of class characteristics. Rather, it exists to constrain governments from stereotyping and treating individuals as if they were "defined by" their status within a particular class.[205]

---

203. *See supra* Section I.A.

204. Furthermore, even if one nonetheless believed that consistent, homogenous effects of class membership had resulted, determining so would be impossible due to the inability to open the black box and see exactly what functional forms have resulted and exactly how variables have interacted. In other words, even if all members of a class were *ex post* subjected to the same outcome or treated in one manner, it would be difficult to conclude that this was because of the inclusion of class-related variables in the algorithmic analysis. If even an algorithm's designers cannot know how class-related variables might be affecting the outcomes, we find it hard to imagine a court could determine that a government agency that relied on such an algorithm to obtain more accurate predictions was intentionally discriminating on the basis of a protected class.

205. Parents Involved in Cmty. Schs. v. Seattle Sch. Dist. No. 1, 551 U.S. 701, 788–89 (2007) (Kennedy, J., concurring) ("[School authorities] are free to devise race-conscious measures to address the problem [of diversity] in a general way and without treating each student in different fashion solely on the basis of a systematic, individual typing by race. . . . These mechanisms are race conscious but do not lead to different treatment based on a classification that tells each student he or she is to be defined by race, so it is unlikely any of them would demand strict scrutiny to be found permissible.").

Traditional schemes, such as admitting minority students or retaining minority teachers, invariably do stereotype when they categorically assign a different weight in the government's decision analysis based on class membership. By contrast, machine-learning algorithms, especially those we have in mind for use in solving complex forecasting problems, are far less likely to do so. As courts begin to encounter governmental use of artificial intelligence, they should recognize that "classification" is not a synonym for explicit consideration of an individual's class membership[206]—at least not when applied to learning algorithms that analyze data that may contain class-related variables. Machine-learning algorithms do not operate by giving categorically different treatment based on class membership. As such, absent animus or other explicit intentionality about class in the selection of optimization parameters, machine learning's use should not contravene the foundational principles underlying the Supreme Court's invocation of suspect classifications.

2. Inferring Discriminatory Intent

Even if algorithms do not implicate suspect classifications, opponents of certain machine-learning applications might still argue that they should be subjected to heightened scrutiny by invoking circumstantial evidence to convince courts to infer discriminatory intent.[207] Such opponents might claim that the mere inclusion of a variable such as race in an algorithmic analysis, even if it does not result in consistently advantageous or disadvantageous predictions, still demonstrates an a priori intent to give disparate treatment or produce a disparate outcome. They might argue that conscious decisions by a regulatory agency to include such variables in a machine-learning analysis represent the kind of "specific . . . events leading up to the challenged decision"[208] that provide circumstantial evidence of discriminatory intent. Whatever surface appeal such arguments might hold, we think they will almost surely fail. In the context of machine learning, the black box nature of this analysis means that the act of including a variable does not by itself provide much insight into what an algorithm user might have hoped to gain by including that variable; additional direct or circumstantial evidence would be needed to connect the inclusion of a variable to discriminatory intent.

Yet perhaps machine-learning opponents will argue that government officials should reasonably foresee that the inclusion of variables related to a protected

206. The need for clearer recognition of the difference between classification and explicit consideration seems more than just theoretical. For example, as others have noted, in his opinion in *Parents Involved*, Chief Justice Roberts appears to use "race-conscious" synonymously with "racial classification." Rich, *supra* note 191, at 1571 n.239.

207. Such circumstantial evidence may include "historical background of the decision," the sequence of events leading up to the decision, the extent to which there is a departure from normal procedures, and the legislative history tracking deliberations by decision makers. Village of Arlington Heights v. Metro. Hous. Dev. Corp., 429 U.S. 252, 267–68 (1977).

208. *Id.* at 267.

class could create a disparate impact, and thus officials' persistence in including those variables evidences intent to discriminate.[209] Such a claim would be dubious. Given how machine-learning analysis works on a black-box basis, it is virtually impossible for anyone to know a priori what a given variable's likely importance in the algorithm will be or what its ultimate effects will be on any disparities of predictions.[210] Many machine-learning algorithms incorporate stochastic processes that randomly decide if and when a given variable factors into the classifications, and there are often complex and unpredictable interaction effects between different variables that may cause the inclusion of a protected-trait-related variable to decrease, rather than increase, the disparity of outcomes. Even after an algorithm is implemented, its black-box nature prevents anyone from determining the effects of that variable on any resulting disparity of the outcomes—whether it may have worsened or lessened those disparities.[211]

Machine-learning opponents might attempt, alternatively, to argue that, although disparate impacts cannot be expected a priori, a decision maker with discriminatory intent could run an algorithm multiple times with different variables included, determine ex post whether including a variable representing a protected trait produces greater disparity of outcomes, and then choose to implement an algorithm including such a variable if it produces relatively greater disparity.[212] The legitimacy of such a claim, however, is questionable due to the unknowable generalizability of the outcomes of any single machine-learning analysis. It is common practice to build an algorithm using a random sample of an entire data set (called the training data), and to then evaluate the algorithm's predictions in a different random sample of that data set (called the test data).[213] This allows an analyst to obtain some sense of how the algorithm may perform in new realizations of the data, as changes may result whenever an algorithm is used in novel predictive endeavors.

---

209. The Court has also allowed such foreseeability of disparate impact to serve as circumstantial evidence of discriminatory intent. *See, e.g.*, Columbus Bd. of Educ. v. Penick, 443 U.S. 449, 464 (1979); Pers. Adm'r of Mass. v. Feeney, 442 U.S. 256, 278–79 (1979).

210. *See supra* Part I.

211. Some machine-learning algorithms are capable of producing a type of output called partial dependence plots, which show the average functional (albeit not causal) relationship between a given predictor variable and the outcome holding all other variables constant. *See, e.g.*, D. Richard Cutler et al., *Random Forests for Classification in Ecology*, 88 ECOLOGY 2783 (2007). These plots do not, however, give one any insight into the effect of a given variable on the disparity of the ultimate classifications.

212. Even in such a scenario, one could not know whether it was indeed the inclusion of a given variable that produced the differences in disparity due to the stochastic nature of many machine-learning algorithms. That being said, if one's data set is sufficiently large and does not contain much noise, then one should not expect significantly different outcomes between iterations of an algorithm. Thus, if one were to observe that algorithms including a given variable resulted in significantly disparate impacts, then it could perhaps be assumed that the variable was responsible. However, this is data-dependent, making its likelihood difficult to quantify.

213. *See* BERK, *supra* note 33, at 31.

One problem common to any statistical analysis is that the analyst cannot know how close the test data are to new realizations of data and thus cannot predict how a model or algorithm will perform in the "real world." If the entire data set from which the test data were randomly sampled was itself a random sample from the real world population, then the test data and novel realizations may well be similar. But it is unlikely, even if an agency is working with randomly sampled data, for the population from which data are sampled to be the same population for which generalizations are desired to be made. Take, for instance, an IRS algorithm used to predict tax fraud. This algorithm's underlying data may have come from prior years' audits and their outcomes, but that population is not the same as the population of interest, which would presumably comprise all individuals filing tax returns—not just those audited—in the present year. Temporal differences between test data and real-world data can matter, as differences in the economy between prior years and the current year may impact the likelihood of individuals to engage in fraudulent tax practices. Thus, even if an agency analyst were to choose an algorithm that included a variable representing a protected class, such as race, a disparate impact is not foreseeable due to the inability of knowing the likelihood that disparate outcomes observed in test data would be borne out in real-world data. For these reasons, plaintiffs attempting to show circumstantial evidence of discriminatory intent in the use of machine learning will surely face a steep uphill climb.

3. Withstanding Standards of Review

Thus far, we have explained why administrative agencies' use of machine-learning algorithms that analyze class-related variables are unlikely to give rise to heightened scrutiny. Despite a lack of clear, settled judicial guidance in this area, it seems safe to conclude that learning algorithms' mathematical properties likely will preclude the courts from labeling the use of protected-class data in these algorithms as a suspect classification or inferring any circumstantial discriminatory intent. Yet if a court nevertheless did subject agencies' machine-learning tools to heightened scrutiny, this does not automatically mean that courts will find that the agencies have denied individuals of the equal protection of the law. Whether a decision process deemed to be discriminatory could survive heightened scrutiny will depend to a great degree on factors exogenous to the nature of machine learning—specifically, the nature of the government objective and the state interest motivating the process.[214] Presumably, regulatory agencies that use machine learning in the service of policy objectives such as public health protection or market stability would have strong arguments that they are seeking to advance compelling state interests.

214. To survive intermediate scrutiny, intentionally discriminatory processes must serve "important governmental objectives." Craig v. Boren, 429 U.S. 190, 197 (1976). Under strict scrutiny, "classification[s] 'must be justified by a compelling governmental interest.'" Wygant v. Jackson Bd. of Educ., 476 U.S. 267, 274 (1986) (plurality opinion) (quoting Palmore v. Sidoti, 466 U.S. 429, 432 (1984)).

Although a full consideration of how applications of machine learning would fare under heightened scrutiny will be highly fact dependent, government should be able to overcome at least one aspect of heightened scrutiny analysis—the prohibition on "overbroad generalizations" about different classes.[215] This specific equal protection concern has been raised in the context of evidence-based sentencing in which checklists used to inform criminal sentencing include gender-coding variables weighted on the basis of past regression analyses.[216] For example, Sonja Starr has put forward a compelling argument for why, in that context, inclusion of class-related variables unconstitutionally capitalizes on overbroad generalizations about class membership; checklists take gender into account only by assigning uniform weights to one gender, thereby subjecting all individuals to group-based treatment instead of treating them truly as individuals, "a core value embodied by the Equal Protection Clause."[217] This criticism, though, is unlikely to carry the same weight in the context of administrative agencies' use of machine-learning algorithms. If gender, for instance, were included in an algorithm, it would not be the case that, say, all females would have their predictions affected in the same way, simply because they were female. Rather, the effects of gender on the predictions would vary across individuals and would depend on interactions between the gender variable and all other input variables, effects that would again be unpredictable and heterogeneous across individuals.[218]

Opponents might argue that machine learning nonetheless creates a problematic statistical generalization of its own because algorithms generate predictions about what individuals will do based on what similar individuals have done in the past.[219] Admittedly, machine learning does rely on analysis of training data sets; however, such reliance on the past behavior of similar individuals occurs at a fundamentally different and much more individualized level in the algorithmic context than in the context of checklists. In the latter, generalizations are made about what females as a broad group have historically done. In the former, generalizations will be made about the past behaviors of, say, females of a certain age, with a certain educational background, living in a particular area, paying a given amount of taxes, driving a certain type of car, and so forth. Thus, even though Starr has provided compelling reasons to believe that checklist generalizations may be unconstitutionally "overbroad," the same assertion of excessive breadth will be much less tenable with respect to machine-learning algorithms.

We reiterate that challenges to agencies' algorithms will not likely reach the stage where these considerations will be relevant. The usual rationales for heightened scrutiny are not likely to apply to machine learning. Accordingly, the

215. United States v. Virginia, 518 U.S. 515, 533 (1996).
216. *See* Starr, *supra* note 179.
217. Starr, *supra* note 179, at 827.
218. *See supra* Section I.A.
219. *See supra* Section I.A.

applicable standard for equal protection purposes will likely be rational basis review, which "generally results in the validation of state action."[220] Rational basis review merely requires that a governmental action be "rationally related to furthering a legitimate state interest."[221] Assuming that the governmental interest to which an agency's machine-learning algorithm is applied is legitimate, it should be quite difficult for that algorithm to be struck down on equal protection grounds. An algorithm's objective function will itself provide a concrete, mathematical embodiment of the agency's goal, sufficient to demonstrate a rational relationship between the algorithm and the governmental interest at stake.

One of the few ways in which government actions can be struck down under rational basis review is through the finding of a motivating animus toward a certain group.[222] The Supreme Court has yet to offer clear guidance for determining the presence of animus or how exactly such a presence should affect judicial review.[223] Nevertheless, we can surmise that animus can be found either through direct evidence or via inference.[224] Direct evidence typically takes the form of statements, whether by legislators[225] or private individuals whose views are given effect by challenged actions,[226] showing sentiments of hostility, stereotype, bias, or bigotry. Such manifest indications of animus might well surround specific uses of machine-learning algorithms, but nothing unique about machine learning makes it any more vulnerable to explicit animus than any other decision-making process or governmental action.

Claims of indirect inferences of animus might conceivably be advanced by calling attention to disparate outcomes. Yet, at least with respect to machine learning, any such inferences would be hard to justify. Although animus has not been unambiguously defined, inferences of it would likely entail making a situation-dependent search for some kind of "logical connection" between the consideration of class and the government interest being served—the absence of which might plausibly support an inference of animus.[227] If an algorithm's supplementary, diagnostic output somehow indicated that a class-related variable was relatively important for the accuracy of the algorithm's predictions,

---

220. Kenji Yoshino, *The New Equal Protection*, 124 HARV. L. REV. 747, 755–56 (2011). By contrast, strict scrutiny has been famously described as being "'strict' in theory and fatal in fact." Gerald Gunther, *The Supreme Court, 1971 Term—Foreword: In Search of Evolving Doctrine on a Changing Court: A Model for a Newer Equal Protection*, 86 HARV. L. REV. 1, 8 (1972).

221. Mass. Bd. of Ret. v. Murgia, 427 U.S. 307, 312 (1976) (per curiam).

222. *See* Susannah W. Pollvogt, *Unconstitutional Animus*, 81 FORDHAM L. REV. 887, 889 (2012).

223. *See id.* at 924–30; Susannah W. Pollvogt, Windsor*, Animus, and the Future of Marriage Equality*, 113 COLUM. L. REV. SIDEBAR 204, 205–06 (2013).

224. Pollvogt, *supra* note 222, at 926–29.

225. *See, e.g.*, U.S. Dep't of Agric. v. Moreno, 413 U.S. 528, 534 (1973) (discussing statements by legislators expressing a desire to exclude "hippies" from food stamp programs).

226. *See, e.g.*, Palmore v. Sidoti, 466 U.S. 429, 433 (1984) (describing how a court's decision to take custody away from a mother who remarried a man of a different race gave effect to private biases against interracial marriages).

227. Pollvogt, *supra* note 222, at 927 (discussing City of Cleburne v. Cleburne Living Ctr., 473 U.S. 432 (1985)).

then it would appear easy to claim that that variable is connected to a valid governmental interest—an interest in taking action aimed at a legitimate purpose in the most efficient and accurate way possible. Alternatively, if a class-related variable could be shown to be not predictively important, then the consideration of class would likely not be driving the discrimination complained of, but rather it would be other, non-class-related variables that would underlie the outcomes. Opponents of machine learning might argue that these other variables form, in combination, a proxy for class, and this could be true, but then the inclusion of those variables would carry the same logical connection to the government's interest as would the inclusion of important class-related variables; the non-class-related variables would be contributing noticeably to the algorithm's accuracy. For these reasons, even when some evidence might be put forward as to the predictive power of certain variables, it is unlikely that agencies' algorithms will be particularly vulnerable to indirect inferences of animus that could lead them to be struck down under rational basis review.

Another potential argument opponents might make is that animus arises if an agency implements an algorithm and finds that it yields vastly disproportionate outcomes for individuals of a particular class, but then the agency fails to take any steps to alleviate these disparate outcomes.[228] After all, a series of novel statistical methods are emerging that can minimize the disparate impact of algorithms.[229] These methods, though, come at the expense of forecasting accuracy, which means that an agency's failure to adopt them may not necessarily imply animus. The agency may simply be seeking to ensure that accurate predictions underlie its actions. Furthermore, any steps an agency might take to rectify disparate impacts after an algorithm's implementation could itself constitute unconstitutional disparate treatment.[230]

---

228. It should also be noted, if the class in question is a constitutionally suspect one, then a similar argument could perhaps be advanced to show circumstantial evidence of discriminatory intent, as discussed earlier in this section.

229. *See, e.g.*, Toshihiro Kamishima et al., *Fairness-Aware Classifier with Prejudice Remover Regularizer*, *in* MACHINE LEARNING AND KNOWLEDGE DISCOVERY IN DATABASES 35–50 (Peter A. Flach, Tijl De Bie & Nello Cristianini eds., 2012) (describing an approach known as regularization, which is akin to mathematically punishing the algorithm for making vastly disparate classifications); Sorelle Friedler et al., Certifying and Removing Disparate Impact 10–11, Presented at the NIPS workshop on Fairness, Accountability, and Transparency in Machine Learning (2014), http://arxiv.org/pdf/1412.3756v1.pdf [https://perma.cc/Y967-BF56] (describing a data-altering approach to reduce disparate outcomes by what amounts, roughly speaking, to adding noise to the data).

230. *See* Ricci v. DeStefano, 557 U.S. 557 (2009). In *Ricci*, the Court held that, in the context of Title VII's prohibition of disparate impacts, decision processes may be modified ex post to minimize their disparate impacts only if the decision maker has a "strong basis in evidence to believe [the decision maker] will be subject to disparate-impact liability if it fails to take the race-conscious, discriminatory action." *Id*. at 585. Admittedly, this ruling does not directly address the scenario we have raised in this Article, in which the feared equal protection violation would result from a prohibition not of disparate impacts but of animus. Nonetheless, the Court's general hesitancy to allow ex post disparate treatment calls into question whether ex post rectifications of an algorithm's disparate impacts would be permissible.

Overall, our analysis indicates that agencies seeking to use machine learning to advance their public missions should not be put off by the potential for equal protection challenges.[231] We expect that the courts will permit these agencies to include class-related variables in their algorithms for the sake of accuracy and even for trying to mitigate disparate impacts. Although the inclusion of race and other class characteristics in decision making supported by machine learning may seem to run afoul of the Court's historical aversion to explicit considerations of protected classes, we have explained why the algorithmic consideration of class, unlike the consideration in traditional decision processes, is unlikely to yield categorically different treatment based on class membership. Agencies' algorithms are unlikely to be deemed suspect classifications triggering heightened scrutiny. Due to the distinctive mathematical properties of machine learning, agencies will also likely escape heightened scrutiny based on circumstantial evidence of intent. Absent direct evidence of animus, machine learning should be easily sustained under rational basis review. Even if subjected to heightened scrutiny, machine learning should, at least in some settings, also withstand challenge.

Our conclusions about agencies' ability to use machine learning in ways that avoid equal protection problems should hardly be construed as giving our blessing to the careless or indiscriminate use of machine learning. Principles of good government still necessitate, whenever feasible, the avoidance of possible discriminatory outcomes, even when such steps are not demanded by the

231. In reaching this conclusion, we have noted several uncertainties that exist, from a lack of clarity surrounding the meaning of "suspect classification" to the uncertain acceptability of *ex post* modifications of algorithms with disparate impacts. Such uncertainties are representative of the broader ferment that has characterized, and continues to characterize, equal protection doctrine, and it is in the context of this ferment that our conclusion should be understood. Perhaps these difficulties in obtaining clarity about the meaning of equal protection are not due to any easily rectifiable inadequacies in law, but simply to the problem of defining, in a more normative sense, what equality actually means. For example, many conceptions of equality rest on the notion that "persons similarly circumstanced shall be treated alike." F.S. Royster Guano Co. v. Virginia, 253 U.S. 412, 415 (1920). But some scholarship has argued that such a statement is circular and devoid of any real moral thrust; the only coherent way to define two individuals as "being alike" may not be to say that they are alike in any empirical way, but to say that they are morally alike, in that they should be treated alike. But then a determination of who should be treated alike is external to equality and must be derived from conceptions of rights or other moral principles. *See* Peter Westen, *The Empty Idea of Equality*, 95 HARV. L. REV. 537, 543, 547 (1982). This difficulty of defining equality has practical implications not only for understanding equal protection doctrine but also for the ways that agencies should implement machine learning to counter potential problems of unfairness. Mathematical definitions of equality must be grounded in a moral definition of equality to be responsive to moral concerns about discrimination. *Cf.* Sorelle A. Friedler et al., *On the (Im)possibility of Fairness*, ARXIV:1609.07236 [CS.CY] (2016), https://arxiv.org/pdf/1609.07236v1.pdf [https://perma.cc/225Y-LPAE] (discussing how multiple quantitative measures of algorithmic fairness exist but rely on unstated, and often incompatible, assumptions about fairness); Kory D. Johnson et al., *Impartial Predictive Modeling: Ensuring Fairness in Arbitrary Models*, ARXIV: 1608.00528 [STAT.ME] (2016), https://arxiv.org/pdf/1608.00528v2.pdf [https://perma.cc/C5QW-RY2U] ("The literature lacks a serious discussion of the philosophical components of fairness and how they should be operationalized in statistics. Doing so will require a spectrum of models to be defined, because fairness is a complicated philosophical topic.").

Constitution. Indeed, one important reason for agency experimentation with machine learning is to find ways to reduce or eliminate the errors and biases that creep into human decisions. Although agency officials should be fully cognizant of potential equal protection objections to their efforts to engage in algorithmic adjudication or robotic rulemaking, these legal risks appear quite manageable and should hardly stand in the way of responsible attempts to use machine learning, especially when doing so will reduce discrimination that would otherwise be introduced by human decision makers and their often inescapable, sometimes unconscious, prejudices.[232]

D. TRANSPARENCY

Up to this point, we have considered choices that agencies will face when deciding to implement algorithmic tools. Now we turn to the question of how could, and should, regulators ensure proper transparency about their choices. Some observers already fear that an era of artificial intelligence may render too much of society and its workings opaque.[233] The legitimacy of agencies' choices about algorithms will therefore depend on accountability and transparency.[234] Principles of transparency undergird much of administrative law, finding expression in statutes such as the Freedom of Information Act[235] and various requirements under the Administrative Procedure Act.[236] Taken together, these statutes provide a fundamental "means for citizens to know 'what their Government is up to.'"[237] The legitimacy of rules issued by administrative

232. *See* Coglianese, *supra* note 154; *see also* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. (forthcoming 2017) (describing how discriminatory effects of algorithms may result not from their black-box natures, which are often more neutral than potentially biased human decision makers, but from historical human biases baked into data sets). Human biases can manifest themselves in qualitative decision making and in traditional quantitative decision making, like regression analysis. Regression analysis is more susceptible to tacit bias because it is driven by theories about how individuals are likely to behave. A recent ProPublica report documents one form of disparate impacts from a traditional risk assessment algorithm. *See* Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/VAF3-XQEV]. Although the exact nature of the COMPAS algorithm at the center of the ProPublica report is proprietary and thus publicly unavailable, it does not seem to be a modern machine-learning algorithm but rather a "theory-guided instrument[]." Tim Brennan & William L. Oliver, *The Emergence of Machine Learning Techniques in Criminology: Implications of Complexity in Our Data and in Research Questions*, 12 CRIMINOLOGY & PUB. POL'Y 551, 559 (2013); *see also* Tim Brennan et al., *Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System*, 36 CRIM. JUST. & BEHAV. 21, 24 (2009) (noting how COMPAS models "use logistic regression, survival analysis, and bootstrap classification methods").

233. *See, e.g.*, John Danaher, *The Threat of Algocracy: Reality, Resistance and Accommodation*, 29 PHIL. & TECH. 245 (2016); Felten, *supra* note 20.

234. Cary Coglianese, Heather Kilmartin & Evan Mendelson, *Transparency and Public Participation in the Federal Rulemaking Process: Recommendations for the New Administration*, 77 GEO. WASH. L. REV. 924, 926 (2009).

235. 5 U.S.C. § 552 (2012).

236. *Id.* §§ 552–553.

237. Nat'l Archives and Records Admin. v. Favish, 541 U.S. 157, 171 (2004) (quoting Dep't of Justice v. Reporters Comm. for Freedom of the Press, 489 U.S. 749, 773 (1989)).

agencies requires that those being regulated have access to information showing how and why rules were adopted. Particularly, transparency requires that agencies ensure that their decisions are "clearly articulated" and "the rationales for these decisions are fully explained, and the evidence on which the decisions are based is publicly accessible."[238]

In the United States, transparency and reason-giving are embedded in a wide variety of statutory and court-constructed principles. Regulatory agencies must, for example, allow for public comment on proposed rules, and final rules must be published in the *Federal Register* along with a separate statement providing clear justification of agencies' policy choices.[239] Agencies also must generally disclose information to the public upon request,[240] and agencies headed by multiple commissioners must generally make their decision-making meetings open to the public.[241] All of these requirements have applied without much interfering with traditional administrative practices, including those in which qualitative decision processes by humans are supplemented with conventional quantitative techniques. By contrast, basing regulatory decisions on machine learning might be viewed by some observers as taking a step backwards and creating obfuscation because of learning algorithms' complexity and black-box nature.

How can the seemingly ineffable qualities of algorithmic analysis be squared with principles of open government? If an agency decides to adopt a rule because its machine-learning algorithm "said so," will that by itself constitute adequate justification when the functioning of the algorithm cannot be intuitively explained? This section takes up these and related questions. We conclude that well-informed, responsible use of machine learning can be compatible with long-established principles of accountable and transparent administrative decision making, even if some algorithmic components may be exempt from public disclosure.

Before delving into our main argument, though, let us briefly confront head-on any notion that the black-box nature of machine learning necessarily condemns it as being unlawfully opaque. Although such an initial conclusion might seem reasonable from the words "black box," such a label is a bit of a misnomer in this context. To say that machine learning has a black-box nature does not mean it is completely impenetrable to human examination. Rather, as discussed in section I.A, it means that machine-learning methods for transforming inputs to outputs are not as intuitively interpretable as more traditional forms of data analysis.[242] This is different than saying that no one can know at

238. Coglianese, Kilmartin & Mendelson, *supra* note 234, at 926.

239. 5 U.S.C. § 553(b)–(c).

240. *See id.* § 552.

241. *See id.* § 552(b).

242. As statistician Leo Breiman has noted, "My biostatistician friends tell me, 'Doctors can interpret logistic regression.' There is no way they can interpret a black box containing fifty [classification] trees hooked together." Breiman, *supra* note 42, at 209.

all how algorithms generate their predictions, which we would agree would undermine the transparency of any technique underlying administrative action. Machine learning can be understood and explained. Analysts can, and do, possess full knowledge of algorithms' inner workings, and they can mathematically explain how these algorithms optimize their objective functions. What they lack is simply an interpretive ability to describe this optimization in conventional, intuitive terms. They cannot say that a machine-learning analysis shows that X causes Y, and therefore a government agency aiming to reduce Y needs to regulate X.

1. Reason-Giving

Legitimacy demands that agencies give adequate reasons for their actions. With this core principle in mind, the shift in how easy it is to grasp intuitively the results of machine-learning algorithms might seem to mark an important qualitative change in how governmental decisions can be explained when they are based on artificial intelligence. Given the way courts often defer to agency expertise in cases involving complex scientific and mathematical analysis, however, decisions that rely on machine learning are unlikely to be rejected by the courts as insufficiently reasoned.

To be sure, administrative law has long compelled reason-giving by agency officials, especially for rulemaking, rendering reasons a cornerstone of an accountable and transparent government.[243] The Administrative Procedure Act's (APA) arbitrary and capricious standard requires an agency to "articulate a satisfactory explanation for its action, including a 'rational connection between the facts found and the choice made.'"[244] In *State Farm*, the Court offered what is now the canonical test for arbitrary and capricious review:

> Normally, an agency rule would be arbitrary and capricious if the agency has relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in view or the product of agency expertise.[245]

To satisfy the standard of reasonableness embedded within the APA, agencies need to explain clearly "the assumptions and methodology used in preparing the

243. *See* Cary Coglianese, *The Transparency President? The Obama Administration and Open Government*, 22 GOVERNANCE 529, 537 (2009) (discussing the importance of reasoned transparency that, unlike fishbowl transparency, demands explicit explanations for agency action); Jerry L. Mashaw, *Small Things Like Reasons Are Put in a Jar: Reason and Legitimacy in the Administrative State*, 70 FORDHAM L. REV. 17, 19 (2001) (discussing how the legitimacy of administrative actions, unlike legislative or judicial actions, must be grounded in reason).

244. Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983) (quoting Burlington Truck Lines, Inc. v. United States, 371 U.S. 156, 168 (1962)).

245. *Id.*

model" upon which they rely.[246] When computers implement a model, courts expect a reasoned explanation to ensure that "ultimate responsibility for the policy decision remains with the agency rather than the computer."[247] Thus, in administrative applications of machine learning, agencies will need to disclose algorithmic specifications, including the objective function being optimized, the method used for that optimization, and the algorithm's input variables.

Under the arbitrary and capricious standard, courts purport to give agencies' explanations a "substantial inquiry," one that is "thorough" and "probing."[248] Consequently, some uncertainty, especially in complex rulemakings, may always exist over whether a court will ultimately find an agency's explanations satisfactory, and such uncertainty may exist as well when agencies engage in rulemaking by robot or adjudication by algorithm. Still, courts are usually deferential to agencies' explanations, especially when they involve complex expert judgments. A year after *State Farm*, for example, the Court reviewed a highly complex rule developed by the Nuclear Regulatory Commission (NRC) in which the agency effectively assumed that long-term storage of radioactive wastes would pose no environmental impact. The Court upheld the NRC action, reasoning that "a reviewing court must remember that the Commission is making predictions, within its area of special expertise, at the frontiers of science. When examining this kind of scientific determination, as opposed to simple findings of fact, a reviewing court must generally be at its most deferential."[249]

Subsequent lower court decisions have exemplified this degree of deference to agencies' scientific and mathematical analyses. For example, a district court recently noted that "[t]he most important thing to remember is that even if plaintiffs can poke some holes in the agency's models, that does not necessarily preclude a conclusion that these models are the best available science. Some degree of predictive error is inherent in the nature of mathematical modeling."[250] Similarly, another district court involved in a recent review of an

246. Sierra Club v. Costle, 657 F.2d 298, 333 (D.C. Cir. 1981); *see also* Am. Pub. Gas Ass'n v. Fed. Power Comm'n, 567 F.2d 1016, 1039 (D.C. Cir. 1977) (noting that "the Commission is obligated to provide a complete analytical defense of its model").

247. *Sierra Club*, 657 F.2d at 334–35 ("The safety valves in the use of such sophisticated methodology are the requirement of public exposure of the assumptions and data incorporated into the analysis . . . and the insistence that ultimate responsibility for the policy decision remains with the agency rather than the computer. With these precautions the tools of econometric computer analysis can intelligently broaden rather than constrain the policymaker's options and avoid the 'artificial narrowing of options that [can be] arbitrary and capricious.'" (quoting Pillai v. Civil Aeronautics Bd., 485 F.2d 1018, 1027 (D.C. Cir. 1973)) (footnote omitted)).

248. *See, e.g.*, Citizens to Preserve Overton Park, Inc. v. Volpe, 401 U.S. 402, 415 (1971).

249. Balt. Gas & Elec. Co. v. Nat'l Res. Def. Council, Inc., 462 U.S. 87, 103 (1983). As Adrian Vermeule has noted, "*Baltimore Gas* review is in fact more consistent with Supreme Court practice in the past three decades than is *State Farm* (at least in its inflated form, as 'hard look review')." Adrian Vermeule, LAW'S ABNEGATION: FROM LAW'S EMPIRE TO THE ADMINISTRATIVE STATE 190 (2016); *id.* (noting that "arbitrary and capricious review is thin.").

250. Alaska v. Lubchenco, 825 F. Supp. 2d 209, 223 (D.D.C. 2011).

agency's conventional mathematical model noted that, "[w]hile there seems to be a reasonable difference of opinion regarding whether the model accurately predicted concentrations of [the chemical at issue], it is not within the purview of this Court to weigh the evidence supporting these extremely divergent scientific opinions and decide which of them is correct."[251] In these and other cases of dueling arguments over complex models, agency expertise usually wins out—and will likely continue to do so even as machine learning becomes the more common source of agency analysis.

2. Disclosure

In addition to the need to give adequate reasons to withstand arbitrary and capricious review, agencies confront other norms that reinforce disclosure of the analysis underlying their decisions. Executive orders on regulatory decision making impose principles of analysis as well as a process for the review of significant regulations by the Office of Information and Regulatory Affairs (OIRA) within the Office of Management and Budget (OMB).[252] The Information Quality Act authorizes the OMB Director to issue guidelines[253] "for ensuring and maximizing the quality, objectivity, utility, and integrity of information including statistical information disseminated by Federal agencies."[254] OMB Circular A-4 offers guidelines for conducting regulatory impact analysis that request that agencies "clearly document all of the assumptions and methods used in the analysis, discuss the uncertainties associate[d] with estimates, and publicly provide the supporting data and underlying analysis."[255] Similarly, courts have noted that, "to allow for useful criticism" in the rulemaking process, "it is especially important for the agency to identify and make available technical studies and data that it has employed in reaching the decisions to propose particular rules."[256]

---

251. Dow AgroSciences LLC v. Nat'l Marine Fisheries Serv., 821 F. Supp. 2d 792, 804 (D. Md. 2011) (footnote omitted).

252. *See* Exec. Order No. 12,866, Regulatory Planning and Review, 58 Fed. Reg. 51,735 (Oct. 4, 1993); Exec. Order No. 13,563, Improving Regulation and Regulatory Review, 76 Fed. Reg. 3821 (Jan. 21, 2011) ("[E]ach agency shall ensure the objectivity of any scientific and technological information and processes used to support the agency's regulatory actions.").

253. 44 U.S.C. § 3516 (2012).

254. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Republication, 67 Fed. Reg. 8,452, 8,458 (Feb. 22, 2002). These guidelines, though, are not legally binding. *See* Harkonen v. U.S. Dep't of Justice, 800 F.3d 1143, 1145 (9th Cir. 2015) (holding that the Information Quality Act does not create any legal right to agency information or to its quality).

255. OFFICE OF INFO. AND REGULATORY AFFAIRS, REGULATORY IMPACT ANALYSIS: A PRIMER 3 (2011), https://www.whitehouse.gov/sites/default/files/omb/inforeg/regpol/circular-a-4_regulatory-impact-analysis-a-primer.pdf [https://perma.cc/YGZ4-JG3K]. *See also* 67 Fed. Reg. at 8459 ("Where appropriate, data should have full, accurate, and transparent documentation, and error sources affecting data quality should be identified and disclosed to users.").

256. Conn. Light & Power Co. v. Nuclear Reg. Comm'n, 673 F.2d 525, 530 (D.C. Cir. 1982); *see also* Chamber of Commerce v. SEC, 443 F.3d 890, 899 (D.C. Cir. 2006) (stating agencies must reveal "technical studies and data" to public); Solite Corp. v. EPA, 952 F.2d 473, 484 (D.C. Cir. 1991) (same);

These standards for disclosure of underlying data and analysis are hardly absolute. OMB guidelines, for example, recognize that "it may often be impractical or even impermissible or unethical to apply the reproducibility standard to [certain] data."[257] Similarly, the Freedom of Information Act (FOIA) permits agencies to withhold data that constitute trade secrets or confidential business information[258] or that are collected or used for law enforcement purposes.[259] These exemptions have been successfully invoked to withhold data underlying rules,[260] and agencies employing machine learning to regulate various industries could presumably also lawfully rely on them to withhold protected data.[261]

FOIA's exemption for "information compiled for law enforcement purposes" applies "only to the extent that the production of such law enforcement records or information . . . would disclose guidelines for law enforcement investigations or prosecutions if such disclosure could reasonably be expected to risk circumvention of the law."[262] Law enforcement in this context includes regulatory inspections and audits—that is, detection of administrative violations as well as criminal ones. For agencies engaging in rulemaking via embedded machine learning, it would be reasonable to imagine they may design an algorithm to select a regulatory option based in part on the likelihood that regulated entities will violate it. Accordingly, they may mathematically build into the machine-learning decision processes of simulated regulated entities a model of the choice of whether to comply with different regulatory options. Such a mathematical model would presumably be similar to the enforcement algorithms that an agency would deploy after a rule's adoption to predict noncompliance with it. Due to this possibility of subsequently reusing, for a law enforcement purpose, a component of an embedded machine-learning algorithm that helped design a rule, agencies may legally seek to exempt from disclosure certain information about relevant algorithmic specifications used in developing regulations.[263]

The FOIA exemption for trade secrets and confidential business information might also apply when an agency contracts with a private company to imple-

---

Portland Cement Ass'n v. Ruckelshaus, 486 F.2d 375, 393 (D.C. Cir. 1973) (stating rulemaking process cannot function with inadequate data).

257. 67 Fed. Reg. at 8456.

258. *See* 5 U.S.C. § 552(b)(4) (2012).

259. *See id.* § 552(b)(7).

260. *See, e.g.*, Average Fuel Economy Standards Passenger Cars and Light Trucks Model Year 2011, 74 Fed. Reg. 14,196, 14,219 (Mar. 30, 2009) ("NHTSA cannot make public the entire contents of the product plans. The submitted product plans contain confidential business information, which the agency is prohibited by federal law from disclosing.").

261. For a general discussion of the kinds of data that typically fall under FOIA's exemptions, especially the exemption for trade secrets and commercial information, see Critical Mass Energy Project v. Nuclear Regulatory Comm'n, 975 F.2d 871, 872–74 (D.C. Cir. 1992).

262. 5 U.S.C. § 552(b)(7)(E) (2012).

263. Even though embedded machine learning is not initially compiled for a law enforcement purpose, these algorithms could be exempted due to the reasonable expectation that enforcement will ultimately ensue from collection of that information. *See* Gen. Elec. Co. v. EPA, 18 F. Supp. 2d 138, 144 (D. Mass. 1998).

ment machine-learning analysis.[264] Some of the specific methods that the private contractor uses for optimizing agents' objective functions in an embedded machine-learning analysis could constitute a trade secret, reflecting the mathematical innovation produced by the firm. Of course, the objective functions themselves could still be disclosed, as they would reflect the goals of the agency and perhaps the agency's assumption of the goals of regulated entities. What the private contractor working for the agency may possess expertise in, however, is the derivation of novel mathematical methods of optimizing those objective functions. Indeed, what made the machine-learning procedure known as "random forests" truly innovative is not what it maximizes, but how it does so. When this procedure is applied to predict the mean of a quantitative variable, its objective (or loss) function is mathematically identical to that of an ordinary least squares regression.[265] But the unique methods for optimizing that loss function result from years of research. Similar methods developed by private firms and applied to particular governmental loss functions can easily be repurposed to optimize other loss functions in a way that furthers the production of a trade commodity, thereby rendering those optimization processes "information relating to the 'productive process' itself,"[266] and thus protected from disclosure under FOIA. If these optimization processes are withheld but the objective functions being optimized are disclosed, it will be possible for the public to know the "why" but not necessarily or fully the "how" of the agencies' underlying analyses. Rules could be adopted because the analysis shows that their modeled consequences result in maximization of an agency's goals, but how exactly that maximization was computed could remain undisclosed.

In these ways, when agencies attempt to exploit the predictive benefits of machine learning, they may rely on methods that can protect some information from disclosure. But these practices and the resulting limits on disclosure associated with them are hardly unique or distinctive to machine learning, even if arguably they might become more prevalent with the routine use of machine learning.

Agencies would do well when using algorithms to follow good supplementary disclosure practices, even if they are not compelled to do so. One such practice would be the disclosure of all iterations of an algorithm or alternative algorithms that were considered, their predictions, and their corresponding specifications (assuming that those specifications are not exempt from disclosure). The end goal of an algorithm is accurate prediction via optimization of an objective function; determining which algorithm and specifications yield the most accurate predictions is often a matter of trial and error. Various algorithms

264. *See* 5 U.S.C. § 552(b)(4) (2012).

265. *See* BERK, *supra* note 33, at 193–203.

266. Ctr. for Auto Safety v. Nat'l Highway Traffic Safety Admin., 244 F.3d 144, 150–51 (D.C. Cir. 2001) (quoting Pub. Citizen Health Research Grp. v. FDA, 704 F.2d 1280, 1288 (D.C. Cir. 1983)); *see also* Appleton v. FDA, 451 F. Supp. 2d 129, 142–44 (D.D.C. 2006).

perform differently in data sets, and a whole host of factors can affect their accuracy, including how many times an algorithm is run before compiling results from all runs, what variables are included in the algorithm, and other tuning parameters that vary depending on the algorithm. In other words, the road to a final algorithm is paved with many others whose results ultimately inform properties of the final algorithm. Disclosing the earlier runs would be advisable because they would indicate that the ultimate methodology was chosen wisely and with consideration of possible alternatives.[267] Volunteering how those previous algorithms were specified and tuned should not reveal information that needs to be exempted from disclosure, and yet it would give some insights into agencies' deliberative processes. Perhaps more importantly, it would signal the adoption of "a presumption in favor of disclosure" that exemplifies good governmental practice.[268]

Agencies could also generate and release additional forms of algorithmic output that reveal information about the structural properties of data sets and the relationships among variables without disclosing exempted information. Such outputs might include, for example, "partial dependence plots," which reveal the functional forms of relationships (albeit with no implication of causality) between predictor variables and the outcome variable, as well as "predictor importance plots," which indicate, roughly speaking, how important each predictor variable is for resulting predictions.[269] Such plots are tangential to the metric that is key to policymaking based on algorithms—predictive accuracy—and thus they are not likely to factor into rules in any dispositive way that would make them legally subject to disclosure. But they can offer supplementary commentary that would further transparency by allowing interested parties to accrue some knowledge of information contained in withheld data sets or algorithms. Perhaps more importantly, output such as partial dependence plots[270] might provide some more intuitive explanations as to why an algorithm's predictions resulted, which may be valuable to courts or members of the public accustomed to decisions based on causal reasoning.[271]

---

267. There may be instances where this developmental information must be disclosed. *See, e.g.*, Am. Radio Relay League, Inc. v. FCC, 524 F.3d 227, 237 (D.C. Cir. 2008) (holding that redacted studies on which the FCC relied must be disclosed because "there is no APA precedent allowing an agency to cherry-pick a study on which it has chosen to rely in part"). If, on the other hand, supplementary information from which conclusions are not drawn merely "clarif[ies], expand[s], or amend[s] other data that has been offered for comment," then such information may be withheld. Chamber of Commerce v. SEC, 443 F.3d 890, 900, 903 (D.C. Cir. 2006). Thus, whether previous algorithm runs would legally need to be disclosed will depend on their role in a particular proceeding.

268. *See* Presidential Memorandum for the Heads of Executive Departments and Agencies Concerning the Freedom of Information Act, 74 Fed. Reg. 4683, 4683 (Jan. 26, 2009).

269. *See* BERK, *supra* note 33, at 236–49.

270. *See* Cutler et al., *supra* note 211, and accompanying text.

271. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 8, at 65 ("In our daily lives, we think so often in causal terms that we may believe causality can easily be shown. The truth is much less comfortable.").

Ultimately, transparency in administrative rulemaking is essential to avoid the creation of "secret law"[272] and to "ensure an informed citizenry, vital to the functioning of a democratic society, needed to check against corruption and to hold the governors accountable to the governed."[273] When regulations are made based on algorithmic predictions, transparency of decision making requires as much disclosure about the algorithm's specifications and its underlying data as possible.[274] However, as noted, FOIA may often allow the withholding of some data on commercial confidentiality or related grounds, rendering complete reproducibility of results difficult. But, if algorithmic specifications can be disclosed so as to make results at least theoretically reproducible, then the fundamental goal of knowing how and why decisions resulted might be satisfied. Of course, those specifications themselves, in particular the mathematical optimization processes used in a machine-learning analysis, could be exempted from disclosure as well if they support law enforcement or are developed by private contractors. But these are all legally permissible exemptions that have been invoked by agencies time and again outside the context of algorithmic regulation. As agencies move into the era of machine learning, they may continue to avail themselves of these same exemptions but would still be required or at least advised to release an appropriate amount of information to illuminate what may initially seem to be an opaque analytical process and satisfy prevailing norms of open government. In these ways, agencies will also be able to justify their actions with sufficient reasons to withstand charges that agencies have acted in an arbitrary or capricious manner.

III. THE MERITS OF MACHINE LEARNING IN THE REGULATORY STATE

The legal analysis presented in the previous Part indicates that when federal agencies use artificial intelligence to automate regulatory and adjudicatory decisions, they will likely face little difficulty in making machine-learning practices fit within existing administrative and constitutional constraints.[275] This analysis has taken place against a backdrop of growing concern over the proliferation of machine learning and artificial intelligence, which may make our conclusions surprising, if not a bit disquieting, to some readers. After all, as we noted at the outset of this Article, technologists and sociologists alike worry

---

272. NLRB v. Sears, Roebuck & Co., 421 U.S. 132, 138 (1975).

273. NLRB v. Robbins Tire & Rubber Co., 437 U.S. 214, 242 (1978).

274. As discussed *infra* Section I.C, these algorithms may be embedded or tied together in overarching computer programs that translate predictions into administrative actions. Forthcoming research suggests that optimal transparency and accountability may also require that agencies publish cryptographic commitments of these programs to demonstrate the regularity of their decision making. *See* Kroll et al., *supra* note 1, at 18–21.

275. In other words, we suggest that the administrative use of machine learning will be unexceptional from a legal perspective; it will not necessitate any "systemic change to laws or legal institutions in order to preserve or rebalance established values." Calo, *supra* note 120, at 553. Interestingly, Calo describes how robotics may possess a moderate level of exceptionalism in other legal domains, such as criminal and tort law. *Id*. at 552–55.

about the consequences of algorithmic automation spreading throughout society, bringing with it the possibility that highly-skilled jobs will be taken over by machines[276] as super-intelligent computers surpass the capacities of humans.[277] These anxieties will undoubtedly grow only more fevered when artificial intelligence falls into the hands of government officials, such as when it is used in predictive policing[278] or, in what has been called "the crossing of a moral Rubicon," in target-selecting military drones.[279] If the significant governmental power wielded by administrative agencies could also be taken over by autonomous algorithms, then surely society would face a grave specter of lives and livelihoods being regulated by robots.

Notwithstanding these ominous warnings, actual technological capabilities are hardly so threatening. For much the same reason that science alone can never "make" policy decisions,[280] machine-learning algorithms need humans to specify their objective functions and construct the mathematical processes that will maximize them.[281] Although machine learning could replace or supplement many routine governmental tasks, the oversight and direction of the government will remain in human hands even in the machine-learning era. Society's most consequential regulatory decisions are not routine and therefore will almost surely prove to be unsuitable candidates for automation; these most significant regulatory policy decisions present complexities, uncertainties, and value judgments that will resist the kind of specification needed to embed them in mathematical objective functions. Machine-learning analysis will be able to assist only by informing the most significant of regulatory choices, not by determining them.[282] The more routine decisions that algorithms will be able to

276. *See, e.g.*, Timothy Aeppel, *What Clever Robots Mean for Jobs*, WALL ST. J. (Feb. 24, 2015, 10:30 PM), http://www.wsj.com/articles/what-clever-robots-mean-for-jobs-1424835002 [https://perma.cc/E8JJ-VB6Y]; Claire Cain Miller, *Can an Algorithm Hire Better Than a Human?*, N.Y. TIMES (June 25, 2015), http://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html [https://perma.cc/N6QZ-AD55]; Zeynep Tufekci, *The Machines Are Coming*, N.Y. TIMES (Apr. 18, 2015), http://www.nytimes.com/2015/04/19/opinion/sunday/the-machines-are-coming.html [https://perma.cc/7WY5-ENV6].

277. *See, e.g.*, Peter Holley, *Apple Co-Founder on Artificial Intelligence: 'The Future Is Scary and Very Bad for People*,*'* WASH. POST (Mar. 24, 2015), https://www.washingtonpost.com/blogs/the-switch/wp/2015/03/24/apple-co-founder-on-artificial-intelligence-the-future-is-scary-and-very-bad-for-people [https://perma.cc/8ATY-TZK3].

278. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 8, at 158–62; *see also* Barocas & Selbst, *supra* note 179, at 673; Crawford & Schultz, *supra* note 119, at 103–05; Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 329–30 (2015); Ferguson, *supra* note 16 (manuscript at 3).

279. Robert H. Latiff & Patrick J. McCloskey, *With Drone Warfare, America Approaches the Robo-Rubicon*, WALL ST. J. (Mar. 14, 2013, 7:37 PM), https://www.wsj.com/articles/SB10001424127887324128504578346333246145590 [https://perma.cc/MV88-F9BQ].

280. *See* Cary Coglianese & Gary E. Marchant, *Shifting Sands: The Limits of Science in Setting Risk Standards*, 152 U. PA. L. REV. 1255, 1257–58 (2004).

281. *See supra* Section I.B.

282. Sometimes the use of benefit–cost analysis by administrative agencies has been resisted out of concern that it will substitute mechanistically for human judgment. *See, e.g.*, Lisa Heinzerling, *Regulatory Costs of Mythic Proportions*, 107 YALE L.J. 1981, 2070 (1998) (arguing that numbers "[a]t

make will surely affect people's lives and livelihoods, but in these cases machine-learning applications will be designed to meet human officials' specifications and will ultimately remain under human control. Autonomous robots will not be wielding regulatory power on their own.

For the reasons we have explained, nothing about algorithms makes them uniquely or automatically unsuitable for use by administrative agencies in terms of their ability to comport with core legal principles. Although machine-learning systems can fit quite comfortably within existing constitutional and administrative law, this does not mean that their use will always be warranted.[283] It especially does not mean that agencies should use machine learning in a haphazard or irresponsible manner. Agencies will have to act with care and, in designing and implementing their algorithms, consider potential pitfalls and areas of concern. In the following two sections, we briefly discuss how several key policy issues and good government principles apply to machine learning applications. The policy issues relate to the four legal doctrines discussed earlier: nondelegation, due process, antidiscrimination, and transparency. The other principles of good government apply more generally to the exploitation of algorithms by those in positions of power. Only by engaging in a thorough, case-by-case evaluation of such non-binding but vital considerations will agencies ensure that their use of machine learning conforms not just to the law but also to foundational principles of sound and legitimate public policy.

A. RELATED PUBLIC POLICY CONCERNS

In the previous Part, we demonstrated that, in principle, agencies' machine-learning algorithms should withstand legal challenges based on doctrines of nondelegation, due process, antidiscrimination, and transparency. Of course, an escape from judicial censure does not necessarily guarantee that the algorithms will be fully consonant with the public policy principles underlying those doctrines. Some of the policy concerns animating the legal doctrines discussed

---

worst" can "derail thoughtful discussion by offering the illusion of objective accuracy"). But even economists recognize that benefit–cost analysis can never fully determine a policy decision. *See, e.g.*, Kenneth J. Arrow et al., *Is There a Role for Benefit-Cost Analysis in Environmental, Health, and Safety Regulation?*, 272 SCI. 221, 221 (1996) (acknowledging that "benefit–cost analysis has a potentially important role to play in helping inform regulatory decision-making, although it should not be the sole basis for such decision-making"); John J. Donohue III, *Why We Should Discount the Views of Those Who Discount Discounting*, 108 YALE L.J. 1901, 1910 (1999) (noting that no one should "be a slave to such an analysis").

283. Nor will the use of machine learning always be easy. Although we have discussed how the need for specified objective functions will be more likely to produce legislation that assuages legal concerns over nondelegation, and how the same need for goal precision can mitigate concerns based on antidiscrimination and transparency principles, it may often be more difficult in practice for decision makers to agree on how to specify the objectives of administrative algorithms. As Justice Cuéllar has noted, these difficulties may arise frequently "because agreement at a high level of generality rarely translates into consensus on how to implement policies through administrative agencies." Cuéllar, *supra* note 25, at 16.

in Part II still deserve attention if agencies are to use machine learning in a responsible manner.

For instance, we noted that one of the nondelegation doctrine's motivating goals—ensuring political accountability—has prompted some scholars to advocate that the courts should give more deference when an agency head makes the key decision than when decisions are made by subordinate officials.[284] The courts have not embraced such a graduated approach, but its suggestion should prompt agency officials to consider how machine learning may affect the locus of decision making inside their agencies. When rules are intimately tied to the outputs of an algorithm, the programming of that algorithm will be a consequential task—one that presumably should not be assigned exclusively to a lower-level analyst, as traditional statistical analyses may be today. Assigning responsibility to lower-level analysts without adequate input and oversight could run the risk that higher-level officials—those who are more directly accountable to the political branches of government and to the public—will not fully understand critical details about an already ostensibly opaque rulemaking process. Perhaps worse, the lower-level analysts could make choices about an algorithm's specifications and tuning without realizing potentially far-reaching impacts of their decisions.[285] For these reasons, agencies should be mindful of who within an agency actually wields algorithm-specifying power and how well high-level officials understand the methods of regulating by robot or adjudicating by algorithm.

Our discussion of due process similarly suggests that there is no reason to view adjudicatory algorithms as uniquely threatening to fair hearings. Still, agencies will need to proceed thoughtfully and with care.[286] Agencies may need to delay the deployment of their algorithms to get a sense of how well the test-data error rates correspond with decision reversal rates—the metric underlying the existing body of relevant case law. Furthermore, agencies should increasingly seek out and engage neutral statistical experts to provide dispassionate assessments of consequential uses of algorithms. Individuals challenging an agency's deprivation of their rights or property cannot be expected to mount a sufficiently probing search for an algorithm's potential inadequacies and biases, either on their own or with the help of skilled, but mathematically naïve, counsel. Even if not legally required, agencies should still undertake the kind of probing inquiry needed to minimize possible errors and biases associated with any algorithms they use.

---

284. *See supra* note 140 and accompanying text.

285. In calling for increased artificial intelligence expertise in government, a recent report from Stanford notes that "insufficiently trained officials may simply take the word of industry technologists and green light a sensitive application [of artificial intelligence] that has not been adequately vetted." PETER STONE ET AL., ARTIFICIAL INTELLIGENCE AND LIFE IN 2030, at 43 (2016), https://ai100.stanford.edu/sites/default/files/ai_100_report_0901fnlb.pdf [https://perma.cc/9HPP-72XZ].

286. *See supra* Section II.B.

Agencies must also take note of the potential for their algorithms to cause a disproportionate impact on members of certain classes or groups, despite the probable constitutionality under the Fifth Amendment of the use of such algorithms.[287] Of all the policy concerns related to the legal doctrines discussed in this Article, the possibility of disproportionate impacts may be the most acute. In the popular and academic media, commentators have noted the ways in which algorithmic bias can manifest itself.[288] A series of violent interactions between police and African-Americans over the last few years has prompted public worry of pervasive discrimination by governmental authorities.[289] Agencies should thus seek ways of mitigating algorithms' disparate impact. But they should do so not just as a response to this public alarm, but also as a matter of ethical governance. If discrimination, no matter how unintentional, can be avoided, it should be. Recent advances in statistics, as noted earlier, may provide agencies with powerful mitigating tools, but agencies will still need to balance tradeoffs with forecasting accuracy.[290]

Finally, we have shown that agencies should have no problem disclosing sufficient information to meet the reason-giving and transparency requirements of the APA and FOIA.[291] Nevertheless, disclosing the bare minimum—probably just the objective functions and limited aspects of algorithms' specifications—should not be the ultimate goal of an open, forthcoming administrative state. Agencies should begin developing practices for documenting, retaining, and disclosing developmental algorithm specifications and final algorithm supplemental output to increase transparency and facilitate peer review.[292]

B. OTHER GOOD GOVERNMENT PRINCIPLES

In addition to policy concerns related to the four legal doctrines analyzed in Part II, several other good government principles should be considered by agencies when using machine learning. These principles relate to challenges associated with quantification, permissible error, lack of empathy, job losses, and privacy.

---

287. *See supra* Section II.C.

288. *See e.g.*, Barocas & Selbst, *supra* note 179; Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. (forthcoming 2017); Angwin et al., *supra* note 232; Kate Crawford, Opinion, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES (June 25, 2016), http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html [https://perma.cc/RCH6-MUKF]; Noyes, *supra* note 180; Schrage, *supra* note 180.

289. *See generally* Niraj Chokshi, *How #BlackLivesMatter Came to Define a Movement*, N.Y. TIMES (Aug. 22, 2016), http://www.nytimes.com/2016/08/23/us/how-blacklivesmatter-came-to-define-a-movement.html [https://perma.cc/34W8-GNRK] (chronicling the development of the Black Lives Matter movement in response to police violence).

290. *See supra* note 229 and accompanying text.

291. *See supra* Section II.D.

292. *See supra* Section II.D. Note that OMB guidelines have stated that technical methodology subjected to peer review can be presumed to be sufficiently objective. *See* Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies, 67 Fed. Reg. 8,452 (Feb. 22, 2002).

First, agencies should recognize that the use of algorithms will often compel agency decision makers to engage in quantitative coding of value judgments that have typically been made qualitatively. For example, machine-learning algorithms often permit the specification of their "cost" ratios—the ratio of false positives to false negatives.[293] This is a concrete, unambiguous quantification of agencies' relative normative values of their errors. For agencies not accustomed to making moral valuations through any kind of formal process, let alone one that assigns them numbers, machine-learning algorithms will necessitate addressing questions of organizational and democratic decision making. Who should have the power to transform qualitative moral judgments into a cost ratio? Should agencies involve the public or at least make interested parties aware of a transformation from qualitative to a quantitative assessment of errors and algorithmic tradeoffs? Is such a transformation even possible, or are human deliberations over morality too nuanced to be reduced to a matrix of error rates?[294] Agencies will have to confront not only the challenges of making accurate quantifications of error rates but also contend with a host of relevant normative questions about quantification as they proceed into the machine-learning era.

Second, agencies will have to decide what constitutes "acceptable" algorithmic error rates. This will go beyond the due process-related assessment of whether courts will consider algorithm error rates acceptable. Instead, it requires agencies to ask how large gains in accuracy must be to offset other difficulties in implementing algorithmic systems. Such issues would likely arise in the context of benefit–cost analyses that agencies conduct prior to taking important actions. For example, how much would, say, a five percent increase in accuracy when predicting hazardous waste pipeline incidents save an agency in inspection costs? Would these savings be worth the necessary investments in human capital and data infrastructure? And, beyond the internal difficulties faced by agencies, would the public demand particularly large increases in accuracy to compensate for the novelty, and potentially alarming nature, of robotic systems of regulation or adjudication?[295] Is a marginal improvement over the status quo acceptable, or should agencies strive for something more?

---

293. *See supra* note 172.

294. The difficulties of programming morality into artificially intelligent systems has been discussed previously, such as in the context of the choice of autonomous cars about which humans to harm when faced with scenarios inevitably resulting in some human casualties. *See, e.g.*, John Markoff, *Should Your Driverless Car Hit a Pedestrian to Save Your Life?*, N.Y. TIMES (June 23, 2016), http://www.nytimes.com/2016/06/24/technology/should-your-driverless-car-hit-a-pedestrian-to-save-your-life.html [https://perma.cc/PL2N-R7GR].

295. Governmental use of machine learning could create a "Tesla effect," of sorts. The Tesla effect refers to the principle that, even though driverless cars may be statistically far safer than manned ones, widespread alarm results when even a few individuals become victims of accidents involving a Tesla operating autonomously. *Cf.* Larry Greenemeier, *Deadly Tesla Crash Exposes Confusion over Automated Driving*, SCI. AM. (July 8, 2016), http://www.scientificamerican.com/article/deadly-tesla-crash-exposes-confusion-over-automated-driving [https://perma.cc/72C4-T33P] (discussing both self-

Third, we should consider what happens to human contact in a world of robotic regulators and algorithmic adjudicators. Citizens expect their government to be competent and efficient, but they also tend to view governmental institutions as more legitimate when these institutions operate with understanding and empathy.[296] The idea of the government reducing individuals to data points that are then fed into an algorithm will seem disconcertingly impersonal—even if ultimately more accurate and efficient. Administrative officials that use algorithms should seek to listen to interested members of the public as they design systems for rulemaking by robot or adjudicating by algorithm. These officials should also encourage participation through interactive methods that treat beneficiaries, targets of regulation, and all other affected parties with respect and dignity.[297]

Fourth, government should not turn a blind eye to the possibility that widespread use of algorithms will put many government workers out of their jobs.[298] The extent to which agencies' workforces will shrink as a result of adopting machine learning is unclear; as we have noted, human input will still be necessary at many steps of regulatory and adjudicatory processes, even when automated.[299] But there can be little doubt that the structures of agencies' workforces will change if day-to-day operations begin to take drastically different forms due to automated decision making. With over 1.8 million full-time

---

driving's safety features and the collective apprehension after a fatal accident involving a self-driving car).

296. *See, e.g.*, Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 30 CRIME & JUSTICE 283 (2003).

297. *See* CARY COGLIANESE, LISTENING, LEARNING, LEADING: A FRAMEWORK FOR REGULATORY EXCELLENCE 9 (2015), https://www.law.upenn.edu/live/files/4946-pprfinalconvenersreport.pdf [https://perma.cc/F89X-SGHB] (describing "empathic engagement" as a core attribute of regulatory excellence); Cary Coglianese, *Regulatory Excellence as "People Excellence,"* REGBLOG (Oct. 23, 2015), http://www.regblog.org/2015/10/23/coglianese-people-excellence/ [https://perma.cc/7AND-FB8K] ("[R]egulatory excellence demands the consistent achievement of three fundamental attributes: *utmost integrity, empathic engagement,* and *stellar competence*." (emphasis in original)). It may even be worthwhile for agencies to consider ways of developing online, anthropomorphic representations as part of a human–computer interface to encourage more empathic and emotionally positive interactions between citizens and their increasingly robotic regulators. *Cf.* Kate Darling, *"Who's Johnny?": Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy*, *in* ROBOT ETHICS 2.0 (forthcoming 2017), http://ssrn.com/abstract=2588669 [https://perma.cc/66PQ-6U85] (recognizing concerns around framing robotic technology in human terms but noting the benefits of anthropomorphizing robots); Will Davies, *Robot Amelia—A Glimpse of the Future for Local Government*, GUARDIAN (July 4, 2016, 2:10 AM), https://www.theguardian.com/public-leaders-network/2016/jul/04/robot-amelia-future-local-government-enfield-council [https://perma.cc/EZG3-GXCF] (describing how a London borough has begun responding to citizen requests, such as for permits, using a voice response system that employs natural language processing to interpret emotions in citizens' voices and respond appropriately and empathetically); Adriana Hamacher et al., Believing in BERT: Using Expressive Communication to Enhance Trust and Counteract Operational Error in Physical Human-Robot Interaction, Presented at the IEEE International Symposium on Robot and Human Interactive Communication (2016), http://arxiv.org/abs/1605.08817 [https://perma.cc/PV72-UZFM] (describing how robots with affective interaction styles recover more of humans' trust after they make errors than robots that are more efficient yet impersonal).

298. *See supra* note 9 and accompanying text.

299. *See supra* Section I.C.

employees in the executive branch of the federal government,[300] and many more working for government contractors, large-scale automation efforts may be appropriately combined with proactive job training for workers who might seek positions in the future in algorithm-support responsibilities.

Finally, although throughout this Article we have self-consciously eschewed a discussion of privacy concerns, for reasons noted earlier,[301] agencies still must never overlook these concerns. Administrative agencies are poised to collect more data on individuals with each passing day. Agencies must properly and securely store these data to minimize threats to privacy intrusions, especially when many administrative applications of machine learning will require inter-agency sharing through the cloud.[302] Agencies may also increasingly seek individuals' data from sources outside of the United States, which will require careful consideration of jurisdictional questions.[303]

C. A PATH FORWARD

The issues we have raised in the preceding two sections should be addressed as part of agencies' case-by-case assessments and benefit–cost analyses of specific applications of machine learning. Agency and cross-agency decisions about when and how to implement machine learning will benefit from clear guidance on how to assess machine learning's merits.[304] The Administrative Conference of the United States, the National Academy of Public Administration, or the National Academy of Sciences might be able to help in developing guidelines. When agencies ultimately apply such guidelines and conduct benefit–cost analysis of specific machine learning applications, we can expect administrative algorithms will prove extremely promising in some instances and less advantageous in others.[305] Agencies should be suitably discerning and pursue

300. U.S. Office of Pers. Mgmt., SIZING UP THE EXECUTIVE BRANCH: FISCAL YEAR 2015, at 5 (2016), https://www.opm.gov/policy-data-oversight/data-analysis-documentation/federal-employment-reports/reports-publications/sizing-up-the-executive-branch-2015.pdf [https://perma.cc/4FGX-2B4H].

301. *See supra* note 119 and accompanying text.

302. *Cf.* Paul M. Schwartz, *Information Privacy in the Cloud*, 161 U. PA. L. REV. 1623, 1661–62 (2013) (considering United States and European Union definitions of "personal information" and suggesting regulatory reforms for cloud storage to ensure "strong and effective protections for information privacy").

303. *Cf.* Andrew Keane Woods, *Against Data Exceptionalism*, 68 STAN. L. REV. 729, 731 (2016) (discussing competing conceptions of the territoriality of personal data).

304. This kind of ethical oversight has also been called for in recent considerations of a future medical profession driven substantially by machine learning. *See* Alison M. Darcy et al., Opinion, *Machine Learning and the Profession of Medicine*, 315 J. AM. MED. ASS'N. 551, 551 (2016) ("The profession of medicine has a tremendous opportunity and an obligation to oversee the application of this [machine-learning] technology to patient care.").

305. There may also be instances in which pursuing machine learning is advantageous and worthwhile only when human intuition can be incorporated into algorithms. *Cf.* Jens Jakob W. H. Sørensen et al., *Exploring the Quantum Speed Limit with Computer Games*, 532 NATURE 210, 213 (2016) (describing how addressing key technical issues facing quantum computing may be facilitated by incorporating the results of human intuition, gleaned from computer games simulating atomic movement, into machine-learning algorithms).

machine learning applications when they can lead to meaningful improvements in procedural and substantive outcomes.

Recognition of the need for careful consideration and reasonable safeguards when agencies use algorithms should not lead to any presumption against using machine learning entirely. These techniques are becoming ubiquitous in private industry for good reason; their ability to make accurate predictions of complex phenomena can render decision making vastly more effective and efficient, and it would be wise for administrative agencies also to seek these benefits.[306] Still, accountable managers must carefully oversee their use of algorithms, even in the private sector,[307] and they must take possible unintended consequences into account.[308] The same can be said of all innovative practices. The need to weigh the pros and cons of algorithms is in no way qualitatively different than the weighing of benefits and costs needed to inform other administrative choices,[309] including those prompted by other technological advancements.[310] Thoughtful implementation is always advisable for the adoption of any new administrative technology or process.

Deciding how and when to use machine learning may not come easily, which may make agency efforts to facilitate public participation in decisions about the use of machine learning even more important. It would certainly be better to take additional time to engage in robust public consultation and make thoughtful decisions about machine learning's use than to dismiss algorithms out of hand over exaggerated fears about unleashing artificial intelligence "demons."[311] Technological and analytical advances are continually invented and adopted because, despite their potential limitations, they offer the possibility to transform society for the better. The implementation of machine-learning algo-

306. Coglianese, *supra* note 22.

307. *See* Michael Luca, Jon Kleinberg & Sendhil Mullainathan, *Algorithms Need Managers, Too*, 94 HARV. BUS. REV. 96 (Jan.–Feb. 2016), https://hbr.org/2016/01/algorithms-need-managers-too [https://perma.cc/7YAB-MK4S].

308. *Cf.* Gökçe Sargut & Rita McGrath, *Learning to Live with Complexity*, HARV. BUS. REV. (Sept. 2011), https://hbr.org/2011/09/learning-to-live-with-complexity [https://perma.cc/94HV-65PW] (describing the management techniques required to oversee complex, data-driven systems); Latanya Sweeney, *Discrimination in Online Ad Delivery*, 56 COMMC'NS ACM 44, 53 (2013) (describing Google's need to contend with differential delivery of advertisements for arrest records when individuals search for names typically associated with different races). Algorithms applied in administrative contexts will face the same possible risks of bounded cognition that can accompany well-accepted, performance-based regulatory standards because objective functions will be defined in terms of desired outcomes. *See* Cary Coglianese, *Performance-Based Regulation: Concepts and Challenges*, *in* COMPARATIVE LAW AND REGULATION: UNDERSTANDING THE GLOBAL REGULATORY PROCESS 403 (Francesca Bignami & David Zaring eds., 2016).

309. *Cf.* Charles E. Lindblom, *The Science of "Muddling Through,"* 19 PUB. ADMIN. REV. 79 (1959) (discussing the complexity of the mosaic nature of traditional administrative decision making).

310. The Internet, for example, has vastly expanded the ways in which administrative agencies can communicate with the public in rulemaking, but taking advantage of these opportunities still requires careful consideration of factors such as how to make information equally accessible to all members of the public. *See* Cary Coglianese, *Enhancing Public Access to Online Rulemaking Information*, 2 MICH. J. ENVTL. & ADMIN. L. 1, 39–40 (2012).

311. *Cf.* Gibbs, *supra* note 11.

rithms by administrative agencies certainly possesses such potential. With proper forethought and care guided by the kind of analysis we have provided throughout this Article, an administrative state on the cutting edge of statistical innovation can be legally and responsibly realized. Moving in this direction can deliver marked improvements in overall well-being, especially as the government faces new challenges in overseeing a private sector that has already entered the machine-learning era.

CONCLUSION

We have considered a future American administrative state increasingly driven by machine-learning algorithms. Such a future may not look like the glamorously—or ominously—automated picture painted by popular media coverage of artificial intelligence, but it nonetheless could entail significant changes in how the government functions, whether in making certain rules or in applying and enforcing them. Administrative agencies have begun to take steps toward such a future, and our analysis offers support to agencies that seek to venture further into this new era of administration. Agencies that rely on machine-learning algorithms are unlikely to face unique, insurmountable legal barriers, at least not under the key legal principles of the administrative state that we have assessed here. By establishing the legality of the use of machine learning, we believe our analysis clears away one of the more significant hurdles to its further implementation. Of course, although our legal analysis supports agency use of machine learning, we recognize that agencies will need, as with any management or policy choice, to exercise care in how and when they deploy new technology and will need to build the human capital and information technology infrastructure to support rulemaking by robot and adjudication by algorithm.[312] To admit the need for such diligence in execution, however, is no reason for preemptively shying away from the benefits from government agencies' deployment of the latest statistical advancements.

If agencies encounter public hesitancy about governmental use of artificial intelligence, perhaps that anxiety will stem from a perception of machine-learning algorithms as fundamentally different in kind from existing decision-making processes. Something about the phrase "black box"—a common description of machine learning techniques—may make machine learning sound incompatible with notions of accountable government. But it would be more accurate to view machine-learning algorithms, or any other statistical procedures, not as complete black boxes, but rather as extensions of existing human decision making. After all, human cognition and computerized algorithms may process information in certain foundationally similar ways.[313] The latter are

312. Coglianese, *supra* note 22.

313. *See* Brenden M. Lake et al., *Human-Level Concept Learning Through Probabilistic Program Induction*, 350 SCIENCE 1332 (2015) (suggesting that the brain–computer analogy could profitably guide research); Gary Marcus, *Face It, Your Brain Is a Computer*, N.Y. TIMES (June 27, 2015), http://www.

black box in nature simply because they are, qualitatively speaking, more complex and less intuitively understandable than conventional techniques, perhaps also much like human decision making. Democratic government itself, in a collective sense, is decision making according to the algorithm of majority rule. The Constitution's establishment of a bicameral legislature and a system of checks and balances established a complex decision-making algorithm involving hundreds of officials elected by millions of voters. With the passage of time, Congress, presidents, and agencies have imposed their own layers of procedures for their work that are arguably as complex and counterintuitive as any machine-learning algorithm.[314] Just as some messiness and ineffability in collective decision making may be the price of the benefits of democracy, difficulty intuiting machine-learning results may be a relatively small price to pay for more accurate and effective decisions in the administrative state.

With the advent of artificial intelligence's widespread application throughout all other realms of society, it will not be long before the administrative state moves headlong to confront the questions we have analyzed in this Article. For administrative agencies, what will distinguish the machine-learning era is not a substitution of human judgment with some foreign and unfamiliar methodology, but rather an evolution of human judgment to incorporate fundamentally similar—albeit more accurate and often practically more useful—processes of decision making made possible by advances in statistical knowledge, data storage, and digital computing. The analysis offered here provides an antidote to visceral reactions against the use of artificial intelligence in the public sector, reactions we hope will give way to a measured optimism capable of guiding and improving the future of the administrative state.

nytimes.com/2015/06/28/opinion/sunday/face-it-your-brain-is-a-computer.html [https://perma.cc/6S4B-PENT].

314. Governmental decision-making processes have grown neither more straightforward nor less opaque since Bismarck's famous quip about the similarities between lawmaking and sausage making.