2019

Transparency and Algorithmic Governance

Cary Coglianese
University of Pennsylvania Carey Law School

David Lehr
University of Pennsylvania

Author ORCID Identifier:

Cary Coglianese 0000-0002-5496-2104

ARTICLES

TRANSPARENCY AND ALGORITHMIC GOVERNANCE

CARY COGLIANESE* & DAVID LEHR**

Machine-learning algorithms are improving and automating important functions in medicine, transportation, and business. Government officials have also started to take notice of the accuracy and speed that such algorithms provide, increasingly relying on them to aid with consequential public-sector functions, including tax administration, regulatory oversight, and benefits administration. Despite machine-learning algorithms' superior predictive power over conventional analytic tools, algorithmic forecasts are difficult to understand and explain. Machine learning's "black box" nature has thus raised concern: Can algorithmic governance be squared with legal principles of governmental transparency? We analyze this question and conclude that machine-learning algorithms' relative inscrutability does not pose a legal barrier to their responsible use by governmental authorities. We distinguish between principles of "fishbowl transparency" and "reasoned transparency," explaining how both are implicated by algorithmic governance but also showing that neither conception compels anything close to total transparency. Although machine learning's black-box features distinctively implicate notions of reasoned transparency, legal demands for reason-giving can

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

** Research Affiliate, Penn Program on Regulation; J.D. Candidate, 2020, Yale Law School. We thank Lavi Ben Dor, Harrison Gunn, Alexandra Johnson, and Jessica Zuo for their helpful research and editorial assistance, as well as Alissa Kalinowski, Caroline Raschbaum, and their colleagues on this journal for their careful editorial guidance. We are grateful for constructive substantive comments provided by Stuart Benjamin, Richard Berk, Harrison Gunn, Richard Pierce, and Arti Rai. We also acknowledge appreciatively a spirited discussion at Duke Law School's 2018 conference on artificial intelligence in the administrative state, which sparked our interest in developing this extended analysis of transparency issues.

be satisfied by explaining an algorithm's purpose, design, and basic functioning. Furthermore, new technical advances will only make machine-learning algorithms increasingly more explainable. Algorithmic governance can meet both legal and public demands for transparency while still enhancing accuracy, efficiency, and even potentially legitimacy in government.

INTRODUCTION

When Abraham Lincoln declared in 1863 that government "of the people, by the people, for the people, shall not perish from the earth,"[1] he spoke to enduring values of liberty and democracy. Today, these values appear to face an emerging threat from technology. Specifically, advances in machine-learning technology—or artificial intelligence[2]—portend a future in which

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

2. By "artificial intelligence" and "machine learning," we refer in this Article to a broad approach to predictive analytics captured under various umbrella terms, including "big data analytics," "deep learning," "reinforcement learning," "smart machines," "neural networks," "natural language processing," and "learning algorithms." For our purposes, we need not parse differences in the meaning of these terms, nor will we delve deeply into specific techniques within machine learning. It is sufficient to note that "[m]achine learning is not a monolith." David Lehr & Paul Ohm, Playing with the Data: What Legal Scholars Should Learn

many governmental decisions will no longer be made by people, but by computer-processed algorithms. What such a future will mean for liberty and democracy will depend to a significant degree on the extent to which these algorithms and their functioning can be made transparent to the public.

The government's use of machine-learning algorithms will follow from the great strides these algorithms have made in the private sector, where they are improving and automating important decisions, such as those in diagnosing medical conditions, operating motor vehicles, and detecting credit card fraud.[3] Public-sector institutions have started to take note. At both local and national levels, governments are beginning to rely on machine-learning algorithms to aid consequential decisionmaking.[4] Scholars and policy officials alike see increasing promise for the use of machine-learning algorithms by administrative agencies in a range of domestic policy areas.[5] Indeed, loom-

About *Machine Learning*, 51 U.C. DAVIS L. REV. 653, 669 (2017). Part I of the present Article provides a brief discussion of the basic properties of the computational tools we have in mind.

3. *See, e.g.*, CHRISTOPHER STEINER, AUTOMATE THIS: HOW ALGORITHMS CAME TO RULE OUR WORLD 4–7 (2012) (providing examples of how algorithms have "displaced humans in a growing number of industries"); Darrell M. West & John R. Allen, *How Artificial Intelligence Is Transforming the World*, BROOKINGS INSTITUTION (Apr. 24, 2018), https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/ ("There are numerous examples where [artificial intelligence (AI)] already is making an impact on the world and augmenting human capabilities in significant ways."). *See generally* EXEC. OFFICE OF THE PRESIDENT, ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY (2016), https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF (discussing implications of artificial intelligence for the economy).

4. *See, e.g.*, EXEC. OFFICE OF THE PRESIDENT NAT'L SCI. & TECH. COUNCIL COMM. ON TECH., PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (describing opportunities and challenges associated with the use of artificial intelligence in the private and public sectors); P'SHIP FOR PUB. SERV. & IBM CTR. FOR THE BUS. OF GOV'T, THE FUTURE HAS BEGUN: USING ARTIFICIAL INTELLIGENCE TO TRANSFORM GOVERNMENT (2018), http://www.businessofgovernment.org/report/using-artificial-intelligence-transform-government [hereinafter IBM CTR., THE FUTURE HAS BEGUN] (presenting case studies of government agencies' use of artificial intelligence).

5. *See, e.g.*, EXEC. OFFICE OF THE PRESIDENT NAT'L SCI. & TECH. COUNCIL COMM. ON TECH., *supra* note 4, at 1 (observing that "[t]he effectiveness of government itself is being increased as agencies build their capacity to use AI to carry out their missions more quickly, responsively, and efficiently"); STEPHEN GOLDSMITH & SUSAN CRAWFORD, THE RESPONSIVE CITY: ENGAGING COMMUNITIES THROUGH DATA-SMART GOVERNANCE 107–08 (2014) (discussing the use of machine learning and other data intensive strategies at the local level of

ing just over the horizon, agencies could soon develop systems that use algorithms to make key decisions automatically, raising the increasingly realistic prospect of robotically created regulations and algorithmically resolved adjudications.

Existing and future applications of machine learning in governmental settings present important new questions about the proper scope for and design of algorithmic governance. One of the most salient questions centers on transparency and arises from the relatively inscrutable nature of these new techniques.[6] Unlike the traditional statistical analysis on which governmental decisionmakers have long relied—an analysis in which humans specify models relating input variables to output variables—machine-learning techniques have a decidedly "black box" character to them.[7] This makes it difficult to understand and put into intuitive prose how learning algorithms reach

government); Joel Tito, *Destination Unknown: Exploring the Impact of Artificial Intelligence on Government* 6 (Sept. 2017) (unpublished working paper) (on file with the Centre for Public Impact), https://publicimpact.blob.core.windows.net/production/2017/09/Destination-Unknown-AI-and-government.pdf (noting that "the impact on governments of AI adoption will be enormous").

6. *See, e.g.*, VIKTOR MAYER-SHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 179 (2013) (pointing to "the risk that big-data predictions, and the algorithms and datasets behind them, will become black boxes that offer us no accountability, traceability, or confidence"); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 8 (2015) (raising alarm over "authority increasingly expressed algorithmically" because "[t]he values and prerogatives that the encoded rules enact are hidden within black boxes"); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1254 n.23 (2008) (decrying the "opacity of automated systems" and recommending steps for agencies to take to ensure they provide "meaningful notice"); Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, REG. & GOVERNANCE 12–13 (2017), https://onlinelibrary.wiley.com/doi/pdf/10.1111/rego.12158 (arguing that, because they are "opaque, inscrutable 'black boxes,'" algorithms present a fundamental challenge to the aspiration of a "liberal society . . . to be a *transparent order*") (emphasis in original) (internal quotation marks omitted).

7. Michael Luca, Jon Kleinberg & Sendhil Mullainathan, *Algorithms Need Managers, Too*, HARV. BUS. REV. (2016) ("Algorithms are black boxes . . . . [They] often can predict the future with great accuracy but tell you neither what will cause an event nor why."). Moreover, a government official relying on these advanced analytic techniques to address a public problem will not obtain from them any clear understanding of what is *causing* the problem the official seeks to solve. This is because machine-learning algorithms are predictive tools that do not directly support the drawing of causal inferences, which means, for the purpose of governmental transparency, that a government official will not find from the use of these algorithms a causal reason for adopting a particular policy. *See infra* Section I.B.

the results they do.[8] It may be one thing for private-sector organizations to rely on inscrutable algorithms, but governmental decisionmakers have long been bound by principles of transparency.[9]

Can algorithmic governance be squared with legal demands for transparency? In this Article, we consider this question in depth and offer a comprehensive assessment of issues of transparency implicated by methods of algorithmic governance.[10] We begin in Part I by highlighting current and prospective uses for machine learning by governmental entities, explaining what makes machine learning different from other types of analysis, and drawing a key distinction between machine-learning applications that support human decisions versus those that substitute for human decisions. In Part II, we articulate the principles of transparency applicable to government in the United States to show what current legal standards demand. We distinguish between "fishbowl transparency" and "reasoned transparency,"[11]

8. *See, e.g.*, JUDEA PEARL & DANA MACKENZIE, THE BOOK OF WHY: THE NEW SCIENCE OF CAUSE AND EFFECT 359 (2018) (noting that, with machine-learning techniques, "the programmer has no idea what computations [the algorithm] is performing or why they work"); Cliff Kuang, *Can AI Be Taught to Explain Itself?*, N.Y. TIMES MAG. (Nov. 21, 2017), https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html (observing "that artificial intelligences often excel by developing whole new ways of seeing, or even thinking, that are inscrutable to us").

9. On the rationale for and principles of transparency as applied to governmental entities, see Cary Coglianese et al., *Transparency and Public Participation in the Federal Rulemaking Process: Recommendations for the New Administration*, 77 GEO. WASH. L. REV. 924, 926, 961 (2009), and Seth F. Kreimer, *The Freedom of Information Act and the Ecology of Transparency*, 10 U. PA. J. CONST. L. 1011 (2008). With respect to private uses of machine learning, it should be noted that the European Union (EU) has imposed transparency-related regulatory obligations on private use of personalized data, with liability extending to private entities beyond Europe. Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter EU General Data Protection Regulation]. Although our analysis here focuses on the transparency of governmental uses of machine-learning algorithms in the United States, the EU's so-called right to explanation does bear affinities with parts of American administrative law applicable to federal agencies.

10. In an earlier article, we raised transparency as one of several issues implicated by governmental use of machine learning, but we could only provide a brief sketch there of the open-government legal issues presented by algorithmic governance. *See* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1205–13 (2017). Our present Article tackles the important issue of transparency head on, providing the comprehensive legal analysis that our earlier work lacked.

11. *See infra* Section II.A (defining and discussing these terms).

explaining that both are implicated by, and relevant to, algorithmic governance—but also noting that under neither conception do current legal standards demand anything close to total transparency. In Part III, we assess whether machine learning's ostensibly black-box features will prevent governments that use this technology from meeting legal standards of transparency. We conclude that, when governments use algorithms responsibly, machine learning can pass muster under prevailing norms. Moreover, we point to a widening panoply of techniques that data scientists are developing to make learning algorithms more explainable. Overall, we find reason to be optimistic that, notwithstanding machine learning's black-box qualities, responsible governments can provide sufficient transparency about their use of algorithms to supplement, and possibly even replace, human judgments.

I. TOWARD A BLACK-BOX GOVERNMENT?

Private industry has turned to machine learning because it offers unparalleled accuracy, surpassing not only other statistical methods but also human judgment.[12] Today, uses of machine learning abound in the private sector, where its ability to make extraordinarily accurate predictions in complex decision spaces has made it integral to consumer recommendation systems, marketing campaigns, supply chain optimization, self-driving cars, and much more.[13] Machine learning's value derives from its ability to learn for itself how to detect useful patterns in massive data sets and put together information in ways that yield remarkably accurate predictions or estimations. Given these algorithms' advantages, governmental authorities have many opportunities to take advantage of them as well. In this Part, we survey several existing governmental applications and, more strikingly, sketch some possibilities for how governmental use of machine learning could develop in the future. We then turn to what makes machine learning distinctive, explaining briefly how it works and showing why, despite its advantages in delivering accuracy, its results can be harder to explain.

A. Methods of Algorithmic Governance

When considering governmental use of machine learning, applications can be distinguished by the extent to which they are outcome determinative. By this we mean the extent to which the output of an algorithm corresponds

12. *See* Coglianese & Lehr, *supra* note 10, at 1158 n.40 and accompanying text.

13. *See, e.g.*, West & Allen, *supra* note 3.

directly to governmental action eventually taken: Does the output of an algorithm directly determine what governmental action is taken?[14] An algorithm could be outcome determinative if, either by the design of a governmental procedure or in conjunction with other computer programs, it directly initiates an action or makes a decision in a way that effectively leaves humans "out of the loop."[15] Alternatively, an algorithm's output could merely be passed along as one factor for consideration by a human official who possesses complete control over what action is ultimately taken.

Today, most governmental applications of machine learning are not determinative of final actions. For instance, machine-learning algorithms have been applied to direct police officers toward potentially high-crime areas, but not to determine by themselves whom to arrest. They have been used by local officials to direct restaurant inspectors toward establishments that are likely violating food safety standards, but not to impose penalties.[16] They have been used in similar ways by federal agencies to identify individual tax returns for auditing,[17] predict toxicities of chemicals that could potentially be regulated,[18] identify fishing boats to inspect for compliance with by-catch

14. *See* Coglianese & Lehr, *supra* note 10, at 1167–76 (discussing how machine learning might be used in determining government actions).

15. *See* Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1450–52 (2013) (applying the related concept of an "algorithm-based decision"). In using the term "out of the loop," we do not mean to suggest that humans are not involved in the process at all. Even with an outcome-determinative system, humans design the system and make the decision to use it—and can make the decision to discontinue its use. In other words, machine-learning algorithms do not possess lives of their own. *See* PAUL SCHARRE, ARMY OF NONE: AUTONOMOUS WEAPONS AND THE FUTURE OF WAR 30, 32 (2018) ("Fully autonomous systems sense, decide, and act entirely without human intervention" but "[a]utonomy doesn't mean the system is exhibiting free will or disobeying its programming"). For an accessible discussion of "loop-related" terminology, see *id.* at 28–32.

16. *See* 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.

17. *See* JANE MARTIN & RICK STEPHENSON, INTERNAL REVENUE SERV., RISK-BASED COLLECTION MODEL DEVELOPMENT AND TESTING 142–58 (2005), http://www.irs.gov/pub/irs-soi/05stephenson.pdf.

18. *See* U.S. ENVTL. PROT. AGENCY OFFICE OF RESEARCH & DEV., TOXICITY FORECASTER (TOXCAST™) (2016), https://www.epa.gov/sites/production/files/2016-12/documents/tox_cast_fact_sheet_dec2016.pdf; Robert Kavlock et al., *Update on EPA's ToxCast Program: Providing High Throughput Decision Support Tools for Chemical Risk Management*, 25 CHEMICAL RES. TOXICOLOGY 1287, 1295 (2012).

rules,[19] discern patterns in vaccine adverse-event reports,[20] and assist in conducting quality control reviews of Social Security disability claims processing.[21] In all of these instances, humans retain complete say over any final governmental action taken. Furthermore, although these examples involve rather consequential decisions and policies, machine learning has also been applied to automate a host of more routine tasks that are less salient, such as sorting mail[22] and sifting through survey responses about workplace injuries.[23]

But this currently limited nature of machine learning will not last long. With the advancement of machine-learning techniques and the proliferation of supporting back-end data infrastructures,[24] the role for algorithms in government is likely to expand.[25] Not only could machine learning soon be employed in more determinative ways, but it could do so, broadly speaking, to yield two different kinds of determinations: adjudications and regulations.

For adjudication by algorithm, no longer might algorithms merely *inform* adjudicatory decisions, such as by targeting inspectors to certain facilities or flagging tax returns for a full review by human auditors. Rather, machine learning, in conjunction with other computer systems, might directly and automatically conduct an audit or inspection, deem a tax return fraudulent,

19. Richard Berk, *Forecasting Consumer Safety Violations and Violators*, *in* IMPORT SAFETY: REGULATORY GOVERNANCE IN THE GLOBAL ECONOMY 131, 136 (Cary Coglianese et al. eds., 2009)

20. *See* HESHA J. DUGGIRALA ET AL., DATA MINING AT FDA (2015), https://www.fda.gov/downloads/ScienceResearch/DataMiningatFDA/UCM443675.pdf; Taxiarchis Botsis et al., *Novel Algorithms for Improved Pattern Recognition Using the US FDA Adverse Event Network Analyzer*, 205 STUDENT HEALTH TECH. INFO. 1178–82 (2014).

21. FELIX F. BAJANDAS & GERALD K. RAY, IMPLEMENTATION AND USE OF ELECTRONIC CASE MANAGEMENT SYSTEMS IN FEDERAL AGENCY ADJUDICATION 49–51 (May 23, 2018), https://www.acus.gov/sites/default/files/documents/2018.05.23%20eCMS%20Final%20report_2.pdf.

22. *See* Ofer Matan et al., Handwritten Character Recognition Using Neural Network Architectures (Nov. 1990) (unpublished paper presented at Proceedings of the 4th USPS Advanced Technology Conference), http://yann.lecun.com/exdb/publis/pdf/matan-90.pdf.

23. *See* IBM CTR., THE FUTURE HAS BEGUN, *supra* note 4, at 17.

24. By "back-end infrastructures," we mean to refer to various computing capabilities often needed to make machine learning operational. This includes sufficient data storage needed to support the use of the large data sets on which machine learning operates best. Furthermore, the deployment of machine-learning algorithms—putting them into practice and allowing them to, say, dictate what adjudicatory action is taken—requires developing ancillary computer programs that turn a machine-learning prediction into an action. *See* Lehr & Ohm, *supra* note 2, at 701 n.173 and accompanying text.

25. *See* Coglianese & Lehr, *supra* note 10, at 1167–76.

decide whether an individual should receive an airplane pilot's license, award or withhold disability benefits, or assign prisoners to cells based on predictions of their propensity for future violence.[26] It takes little technical imagination to see how these applications could materialize; they would be relatively straightforward applications of machine learning. As suggested by existing private-sector uses of machine learning, the quintessential tasks to which learning algorithms customarily apply are individual-level predictions—such as whether a consumer will buy a product, whether an email is spam, and so forth. Adjudicating by algorithm relies on the same kind of predictions—whether a worksite has a safety violation, whether a tax return is fraudulent, or whether an individual meets benefits eligibility criteria. To be sure, employing machine learning in a way that fully determines adjudicatory outcomes will require significant technical investments in back-end data infrastructures. But at base, the statistical tools that will facilitate adjudicating by algorithm already exist and are already being employed in analogous endeavors.[27]

A bit more technical imagination and advancement may be required for machine learning to usher in automatic regulation—that is, the making of rules by robot. In part this is because what automated rules mean might take several different forms. Perhaps the simplest form would be a regulatory authority mandating the use of a particular machine-learning system in lieu of stating a rule in canonical text.[28] In other words, the regulator would mandate the use of algorithmic adjudication, with the algorithm constituting the

26. *See, e.g.*, BAJANDAS & RAY, *supra* note 21, at 7 (noting that although currently the Social Security Administration uses an algorithmic system to support quality checks on how disability claims are handled by humans, "eventually deep learning algorithms may be able to process some claims to final resolution").

27. For a discussion of a successful private-sector analogue—the fully automated dispute resolution system developed and already used by eBay to settle tens of millions of disputes each year—see Colin Rule, *Resolving Disputes in the World's Largest Marketplace*, ACRESOLUTION, Fall 2018, at 8–11 (2008) and BENJAMIN H. BARTON & STEPHANOS BIBAS, REBOOTING JUSTICE: MORE TECHNOLOGY, FEWER LAWYERS, AND THE FUTURE OF LAW 111–15 (2017). In June 2018, the Administrative Conference of the United States adopted a recommendation that agencies that have electronic case records "should consider how to analyze and leverage" these sources of data "to improve their adjudicative processes, including through the use of natural language processing, machine learning, and predictive algorithms." Admin. Conf. of the U.S., Recommendation 2018-3, *Electronic Case Management in Federal Administrative Adjudication*, 83 Fed. Reg. 30,686, 30,687 (June 29, 2018).

28. *Cf.* Cary Coglianese, *E-Rulemaking: Information Technology and the Regulatory Process*, 56 ADMIN. L. REV. 353, 370–71 (2004) (describing how information technology might "transform[] rules from text contained in the *Code of Federal Regulations* to software packages akin to

applicable rule according to which adjudications should be made. For example, consider an illustrative case of an individual applying for a governmental license—say, a prospective pilot asking the Federal Aviation Administration (FAA) to grant her flight certification. Currently, the FAA has established a set of fixed rules for when a commercial pilot certificate can be awarded.[29] The rules specify requirements for age, hours of flight training and experience, and performance on written and in-flight tests.[30] These rules may work well, but it is not beyond the imagination to think that one day the FAA might instead rely on a machine-learning algorithm to improve the process of determining when a pilot's license should be granted. Under such an algorithmic government model, the "rule" would be the algorithm. The application for a license to be issued under the rule might comprise the signing of a consent form to allow the FAA to run its authorized algorithm through all of the available data about the applicant—say, school records, medical records, social media postings, and fine-grained data from the flight recorders from previous training flights flown by the applicant. The FAA could award a pilot's license when the machine-learning algorithm forecasts the applicant's risk to be below a specified threshold.

In such a case of mandatory algorithmic adjudication, with the algorithm substituting for the rule, humans would still be instrumental in designing that algorithm and specifying the level at which forecasted risk would deny a license to an applicant. Humans could also conceivably go further to design systems that would not merely substitute for rules but that could actually craft or select rules. Such fully automated rulemaking would contemplate removing humans from the selection of an administrative rule. Consider, for example, whether algorithms might eventually be able to replace the current process by which the Occupational Safety and Health Administration (OSHA) establishes permissible exposure limits (PELs) for chemicals in workplaces. The process today depends on research and analysis by humans—so much so that OSHA cannot realistically establish a PEL for every chemical to which workers might be exposed. If chemicals' health risks could be forecasted by the use of a machine-learning algorithm, perhaps OSHA in the future could create an algorithmic system that would automatically establish PELs.

Automated rulemaking will be a more challenging scenario to realize. As with an adjudication, the relevant questions to be answered in a rulemaking go beyond the information processing needed to make an individualized

the popular TurboTax® or other commercially available compliance software").

29. *See* Certification: Pilots, Flights Instructors, and Ground Instructors, 14 C.F.R. § 61 (2018).

30. *Id.*

forecast. Unlike adjudicating, however, regulating typically does not rely on simple factual predicates of the kind long predicted by machine learning. For one, making a rule fundamentally demands identifying relevant normative values or policy goals—e.g., how safe is safe enough?—and these value choices will be ones that humans must make and then use to inform the parameters specified in an algorithm's objective function.[31] Even with respect to the factual forecasts that algorithms can make to inform rulemaking decisions, almost any major rulemaking will be multi-faceted, with tradeoffs to be made across multiple factual outcomes. Resolving such tradeoffs will be another choice calling for human judgment. For example, even a rule as seemingly straightforward as a PEL cannot be accomplished by a single algorithm predicting, say, how many cases of a specific disease would occur if a very specific amount of that pollutant were released. Setting a workplace chemical exposure standard demands understanding the effects that exposure to varying levels of a chemical will have on multiple diseases and other consequences, as well as attending to other factors such as the costs of emissions control or the likelihood that regulated entities would comply with different standards. A single machine-learning algorithm cannot by itself make multi-faceted tradeoffs.

But a machine-learning algorithm can be built into a larger automatic rulemaking system where normative choices and tradeoffs have been specified in advance. If the policy options can be clearly conceived in advance, and if the basis for choosing between them depends simply on a forecast that a machine-learning algorithm can make, a system could be designed in which the algorithm in essence automatically "chooses" the rule. Such a possibility already exists in an algorithmic system the city of Los Angeles has created to operate its traffic signals.[32] For any given driver approaching an intersection at any given time, the machine-learning system automatically determines the "rule" confronting the driver about whether to stop or go. The rule choices in the system are exceedingly simple and well specified: the rule can be "red," "yellow," or "green." The value choice reflected in the algorithm's objective

31. For a discussion of the inherent normativity involved in setting regulatory standards, see Cary Coglianese & Gary E. Marchant, *Shifting Sands: The Limits of Science in Setting Risk Standards*, 152 U. PA. L. REV. 1255, 1262, 1277 (2004). In suggesting that normative choices necessarily underlie rulemaking, we certainly do not mean to suggest that such choices have no bearing on adjudication. With adjudication, however, a learning algorithm's objective function is based on underlying rules, while ultimately rules themselves must be based on normative or policy choices.

32. Ian Lovett, *To Fight Gridlock, Los Angeles Synchronizes Every Red Light*, N.Y. TIMES (Apr. 1, 2013), https://www.nytimes.com/2013/04/02/us/to-fight-gridlock-los-angeles-synchronizes-every-red-light.html.

function centers on minimizing traffic congestion. Drawing on data gathered from sensors implanted in streets throughout the city, the system automatically chooses the rule (i.e., light color) at each intersection, and at each moment, that will minimize congestion based on the algorithm's forecasts.

Los Angeles's traffic system provides a concrete and accessible example of the use of a machine-learning algorithm in a system that automatically generates rules. Although a city's use of machine learning to operate traffic lights may seem somewhat banal, a traffic signaling system provides a point of reference for seeing how automated rulemaking systems could be developed in other contexts. Even if the regulatory choices in other settings were more complex than just "stop" or "go," as long as a regulator could predetermine the array of policy options, specify in advance the overall objective to be maximized, and spell out any tradeoffs, a system could in principle be developed that embeds machine-learning algorithms into an automated decisionmaking structure. Based on the forecasts produced by these algorithms, such a system would generate a final outcome by "choosing" from among pre-specified options the one that would maximize the objective given the tradeoffs and the resulting algorithm-generated forecasts. For more complex regulatory problems, developing the decisionmaking structure for such a system could itself be a considerable challenge—one which will obviously depend on humans to create—but, once created, the system might be able to produce and modify rules quickly, a virtue in some settings within a world of increasing digital forms of economic and social interaction.

Still more sophisticated automated rulemaking systems could be based on a set of non-learning algorithms called agent-based models (ABM) or multi-agent systems (MAS), which can have machine-learning algorithms embedded within them. ABM or MAS algorithms could be developed to model mathematically the regulated environment. In the PEL example, this environment comprises the set of all regulated workplaces and the various aspects of the workplace setting that could be affected by pollutants. The modeled environment would also contain one of many different possible variants of a proposed rule—that is, possible variants of a PEL—and then, in the context of the model, simulated regulated "entities" could choose to take actions in accordance with, or in contravention to, the rule under analysis. Machine-learning algorithms would be embedded within the larger ABM/MAS algorithm and would forecast whether regulated entities would comply with a regulation. The system would then, based on this forecasted compliance, determine the net effect of the rule under analysis on the modeled environment. The ultimate ability of a regulatory agency to use this fusion of ABM/MAS and machine learning to make rules would stem from its ability to run multiple iterations of the overall system, with each iteration using a different variant of a possible rule. The system would be designed to select

automatically, as the promulgated rule, the variant that yields the best outcomes in the modeled environment, however "best" is defined in advance by the system designers at the regulatory agency.

Yet, as we have already indicated, even in the most sophisticated cases of robotic rulemaking, human involvement will not be entirely absent. The objective function incorporated into any ABM/MAS must reflect normative choices that must be made by humans. Furthermore, the system could not itself come up with the multiple possible permutations of the rule from which it selects the best. Some standard templates for different rules would need to be identified in advance by humans. Still, this process could be substantially automated; a human could program such a system, for instance, to try every possible regulatory variant within a very large range. In such a case, it would be quite reasonable to say that the system, not the human, has in effect chosen the rule autonomously. Of course, because of resource constraints on running complicated algorithms repeatedly, in practice it might be more likely the case that humans would first whittle down the possible rule permutations to a more manageable number, from which the system then would choose the optimal one.

Although this vision of regulating by robot has not yet fully materialized for workplace health standards or any other federal rulemaking, the fact that machine learning is already making rules in slightly different, but fundamentally analogous, ways to govern traffic in the city of Los Angeles shows the potential for robotic rulemaking in the future. Perhaps at some point in the future the U.S. Department of Transportation's Pipeline and Hazardous Materials Safety Administration (PHMSA) could mandate the installation of sensors throughout the nation's pipelines—much like Los Angeles has installed sensors in all of its roads—and a PHMSA-mandated algorithmic control system could automatically block off certain sections whenever an algorithm detects unsafe conditions.[33] Similar automated rulemaking systems might be helpful for regulating high-speed trading on securities markets—or in any other setting where government needs to set and adapt rules rapidly. The need for rapid, automatic rulemaking systems seems increasingly plausible, if not inevitable, as a response to the private economy's growing reliance on machine learning, especially as critical infrastructures and economic activities themselves increasingly operate automatically by algorithms. It is not at all far-fetched to think that, to regulate a growing data-driven algorithmic private economy, government will need to develop its own automatic regulatory systems in response.[34]

33. *See* Coglianese & Lehr, *supra* note 10, at 1167–69 (elaborating on this possibility).

34. *See generally* Cary Coglianese, *Optimizing Regulation for an Optimizing Economy*, 4 U. Pa. J.

B. What Makes Machine Learning Distinctive?

The prospect of novel applications of machine-learning algorithms that can automatically determine adjudicatory or regulatory outcomes raises important legal and policy questions. Among the most significant of these questions is one related to governmental transparency. Machine-learning algorithms possess what is often described as a "black box" nature—that is, they discern patterns and make predictions in a way that cannot be intuitively understood or explained in the same way as conventional analysis can be. To understand why machine-learning algorithms are considered so opaque, it is necessary to understand what machine learning is, how it works, and what distinguishes it from traditional statistical analysis.[35]

At its most general level, an algorithm is little more than a set of steps stated in mathematical or logical language.[36] From this view, the use of algorithms by government is hardly problematic at all; virtually any decisionmaking process can be characterized as an algorithm. But taking a closer look, we can compare machine learning with more traditional statistical techniques to understand better what makes machine learning distinctive. Traditional techniques and machine-learning algorithms do, of course, have much in common: they both operate by attempting to achieve a mathematical goal. In an ordinary least squares regression, for instance, the goal is to minimize the sum of the squared residuals, where the residuals are the differences between actual and predicted values. Machine-learning algorithms similarly attempt to achieve mathematical goals, typically referred to as "objectives" or "objective functions," and some machine-learning algorithms even share the same goals as traditional techniques like ordinary least squares regression.[37] Furthermore, both techniques attempt to achieve their goals by analyzing historical data that have been collected from a population of interest—referred to in the machine-learning literature as the "training data."[38]

But where traditional and machine-learning techniques diverge is in how the analysis occurs—that is, how the objective function is met or "opti-

L. & PUB. AFF. 1, 1–13 (2018) (discussing the need for the government to make use of machine learning and other digital advances in order to keep better pace with innovations in the private sector).

35. Our account here is highly stylized and general. We make no claim to describe all the myriad ways that machine-learning algorithms can be structured, nor do we provide a comprehensive introduction to machine learning.

36. *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 640 n.14 (2017) (defining an algorithm as "a well-defined set of steps for accomplishing a certain goal").

37. *See* Lehr & Ohm, *supra* note 2, at 671–72.

38. *See* PATANJALI KASHYAP, MACHINE LEARNING FOR DECISION MAKERS: COGNITIVE COMPUTING FUNDAMENTALS FOR BETTER DECISION MAKING 8, 40, 45 (2017).

mized." In traditional statistical analysis, humans play the key role in setting up the analysis, leaving little to the algorithm. For example, in a traditional regression analysis, humans specify what input variables in the data set the regression should consider and how they should be put together to yield a prediction, or estimate, of the outcome variable—that is, whether the input variables should be added together, multiplied by each other, or take on other functional forms. All that is left for the regression to learn is how the variables—or combinations of variables, as specified by the humans—are weighted. In contrast, with machine learning humans do not specify how input variables should be put together to yield predictions. The algorithm itself tries many possible combinations of variables, figuring out how to put them together to optimize the objective function.[39] In other words, the algorithm "learns" how to make accurate predictions or estimates.

This is not to say that humans have no role in the machine-learning context. Human analysts exert substantial control over myriad aspects of an algorithm's functioning. They are still required to provide the algorithm with its data, select a particular kind (or family) of algorithm to implement, and "tune" details of the algorithm's optimization process.[40] Humans also play a major role in evaluating a learning algorithm's performance. Although the algorithm learns how to make useful predictions on a set of training data, it is not possible to make any meaningful assessment of the algorithm's accuracy with this same historical data set; for a variety of reasons, the real-world data onto which the "trained" algorithm is eventually applied in order to make predictions will likely differ from the historical data. Human analysts thus purposely separate out some historical data from the training data set before they use it. These separate historical data—called the "test data"—are excluded from the training data and, thus, not considered by the algorithm during training. After training, human analysts assess the algorithm's accuracy by using the separate test data, making adjustments to the algorithm as necessary to increase accuracy as measured in the test data.[41] This is very much a process of trial and error—one in which complex, powerful algorithms make predictions using methods repeatedly guided and nudged, but not dictated, by humans in the establishment and refinement of the algorithm.

39. *See* Lehr & Ohm, *supra* note 2, at 671–72.

40. For a more detailed discussion of residual human involvement in machine learning, see *id.* at 672–702.

41. Both the training data and test data include measures of the historical outcomes of interest—that is, the outcomes the analyst seeks to predict—as well as a variety of other variables that the algorithms use to identify patterns associated with and predictive of the outcomes.

Using machine-learning algorithms that are constructed in this way yields numerous benefits. Most significantly, the algorithms' flexible learning process and ability to discern predictively useful patterns in large data sets can make them extremely accurate.[42] They can outperform traditional statistical techniques and, for many tasks, can surpass the abilities of human decisionmakers.[43]

Machine-learning algorithms also can facilitate large gains in efficiency. Because of their ability to "see" complex patterns, they allow for the automation of tasks previously thought outside the realm of algorithmic control. Machine-learning algorithms can also be continuously refined as new data become available; they can be continuously fed new data, which are incorporated into new training data sets on which the algorithm can be automatically retrained.[44]

But these advantages come at a cost to transparency. Machine-learning algorithms are deemed "black boxes" because it is difficult to put into intuitive language how they function.[45] With a traditional regression analysis, an analyst can say exactly how the algorithm's predictions result: input variable values are put together according to the human analyst's specifications, with just the weights, or coefficients, determined by the regression. Such a simple, intuitive explanation is not available for machine-learning algorithms because the combinations of variables and mathematical relationships between them that the algorithms "learn" can be difficult to uncover and, when uncovered, are often extremely complex.

Even machine-learning methods that are considered relatively simple evade intuitive explanation. For example, one class of algorithms known as "random forests" operates essentially by constructing hundreds or even thousands of classification or regression "trees" and deciding on final predictions by averaging or combining predictions from each individual tree.[46] An analyst can examine the structure of a given tree and the predictive rules it displays, but that information tells the analyst nothing about how predictions are made in the forest as a whole. For still more complex techniques, like

42. *See* Coglianese & Lehr, *supra* note 10, at 1158 n.40.

43. One of the more widely known examples to date has been the success of Google's deep learning system in beating champion human players at the game of Go. *See generally* David Silver et al., *Mastering the Game of Go Without Human Knowledge*, 550 NATURE 354 (2017).

44. *See* Lehr & Ohm, *supra* note 2, at 702.

45. *See* L. Jason Anastasopoulos & Andrew B. Whitford, *Machine Learning for Public Administration Research, with Application to Organizational Reputation*, J. PUB. ADMIN. RES. & THEORY, Nov. 5, 2018, at 16, OXFORD (advance article), https://doi.org/10.1093/jopart/muy060 ("Good predictions often require a tradeoff between accuracy and interpretability.").

46. *See* Leo Breiman, *Random Forests*, 45 MACHINE LEARNING 5, 5–6 (2001).

"deep learning" techniques, the inner workings of the algorithms are even more difficult to divine and translate into prose.[47]

The difficulty of generating intuitive explanations from machine-learning algorithms is further exacerbated by the kind of "big data" on which they often operate. Although in principle both machine-learning algorithms and more traditional techniques can operate on the same data sets, the predictive power of machine learning manifests itself in large, complex data sets, so it tends to be these data sets on which learning algorithms are applied. But complex data sets necessarily contain complex inter-variable relationships, making it even more difficult to put into intuitive prose how a machine-learning algorithm makes the predictions it does.

Even if analysts could discover the inter-variable relationships that a machine-learning algorithm keys in on, they cannot overlay any causal inferences onto those relationships.[48] In other words, they cannot say that a relationship between an input variable and the output variable is causal in nature. In fact, some of the patterns that are predictively useful might not be causal at all, and some may be so non-intuitive as to have never occurred to humans—perhaps, say, if the third letter of a tax filer's last name helps in forecasting cases of tax fraud.

To put machine learning's advantages and disadvantages into focus, let us return to our hypothetical example of an automated system that the FAA might use to award commercial pilot licenses. Such a system could analyze the data from all of an applicant's training flights and other records and use that analysis to forecast the applicant's risk. Such an algorithm might well do a better job of selecting safe pilots than the FAA's current system—a prospect that would be consistent with learning algorithms' generally superior performance in other contexts and would presumably counsel in favor of the FAA giving serious consideration to the use of an algorithmic system. But when it comes to transparency, such an algorithmic system would not enable the FAA to provide a conventional, intuitive, prosaic explanation for exactly why any specific individual was—or was not—selected to receive a license.[49]

As a result, the use of so-called black-box algorithms would seem, at least at first glance, to run up against the law's general requirement that government provide adequate reasons for its actions. Their use might seem to undermine basic good-government principles designed to promote accountability and build trust. The question becomes whether, notwithstanding

47. *See* Lehr & Ohm, *supra* note 2, at 693 n.135.

48. *See* RICHARD A. BERK, STATISTICAL LEARNING FROM A REGRESSION PERSPECTIVE 331 (2d ed. 2016). Of course, much traditional statistical analysis, especially absent an overarching experimental research design, can also fail to sustain causal claims.

49. *Id.*

learning algorithms' advantages, their ostensibly inscrutable nature will keep government officials from providing sufficient reasons or explanations for why certain decisions were made. In other words, can government make algorithmic outputs sufficiently transparent? To assess whether the use of machine learning by government agencies can be squared with legal expectations of transparency that ordinarily apply to administrative actions, the next Part explores in greater depth the relevant legal principles.

II. LEGAL PRINCIPLES OF OPEN GOVERNMENT

Transparency is integral to a legitimate government and a fair society. When government is open, officials can be expected to do their jobs better because public accountability presumably inhibits them from advancing their own self-interests at the expense of their duty to produce public value.[50] As Louis Brandeis famously quipped, "Sunlight is … the best of disinfectants."[51]

Transparency not only deters officials from taking shortcuts and bribes, it also gives other governmental actors—the courts, for example—a basis for their oversight roles.[52] The visibility of governmental institutions also gives other nongovernmental organizations—the media, non-profit advocacy organizations, academic researchers, law firms, and businesses—the opportunity to monitor what the government is doing. When private individuals and organizations can learn about what government is doing, they can do a better job of organizing their own affairs by anticipating the establishment of new laws or understanding changes in government programs. Significantly, in a democracy, transparency can also help build an informed citizenry and provide a basis for more meaningful public participation in all facets of governmental decisionmaking.[53]

50. *See, e.g.*, Adriana S. Cordis & Patrick L. Warren, *Sunshine as Disinfectant: The Effect of State Freedom of Information Act Laws on Public Corruption*, 115 J. PUB. ECON. 18, 35–36 (2014) (finding a decrease in indicators of corruption following the passage of state open records laws). On the importance of public value creation by government officials, see generally MARK H. MOORE, CREATING PUBLIC VALUE: STRATEGIC MANAGEMENT IN GOVERNMENT (1995).

51. LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 92 (1914).

52. For example, when it comes to overseeing government agencies' use of machine learning under a range of legal doctrines, such as those combatting discrimination, the courts will require transparency about algorithms and their use. *See* Coglianese & Lehr, *supra* note 10, at 1195–1205; Farhad Manjoo, *Here's the Conversation We Really Need to Have About Bias at Google*, N.Y. TIMES (Aug. 30, 2018), https://www.nytimes.com/2018/08/30/technology/bias-google-trump.html.

53. *See* Coglianese et al., *supra* note 9, at 926–30 (discussing benefits of transparency in government).

Still, despite its many virtues, governmental transparency can be difficult to assess in practice because what "open government" means can vary. The concept of governmental transparency has been invoked in different ways by different legal scholars, political theorists, and public officials.[54] To provide conceptual clarity and a foundation for our analysis of algorithmic governance, we distinguish in this Part between two types of governmental transparency: "fishbowl transparency" and "reasoned transparency."[55] The former prioritizes the disclosure of information about *what* government is doing, while the latter aims to promote an understanding of *why* government does what it does. Both find support in different ways in U.S. administrative law. As we will explain more fully in the next Part, both types of transparency are also implicated, in different ways, by governmental use of machine learning.

Due to its ostensibly black-box character, the most distinctive questions about machine learning, especially if it is used to replace human decisionmaking, will center on how it may affect the government's ability to give reasons for its actions. In this Part, we lay the groundwork for an analysis of reason-giving as it applies to algorithmic governance. We begin by elaborating on fishbowl and reasoned transparency before turning to a review of the

54. *See, e.g.*, Cary Coglianese, *Open Government and Its Impact*, REG. REV. (May 8, 2011), https://www.theregreview.org/2011/05/08/open-government-and-its-impact/ (observing that "little clarity currently exists over what open government means").

55. Cary Coglianese, *The Transparency President? The Obama Administration and Open Government*, 22 GOVERNANCE 529, 530, 537 (2009). The distinction between fishbowl transparency and reasoned transparency that we rely on here parallels to a degree the distinction sometimes made by computer scientists and other scholars between the *transparency* of algorithms (i.e., access to their parameters and the underlying data they analyze) and the *explainability* of algorithms (i.e., the reasons why they reach certain results). *See, e.g.*, AARON RIEKE, MIRANDA BOGEN & DAVID G. ROBINSON, PUBLIC SCRUTINY OF AUTOMATED DECISIONS: EARLY LESSONS AND EMERGING METHODS 24 (2018), http://www.omidyar.com/sites/default/files/file_archive/Public%20Scrutiny%20of%20Automated%20Decisions.pdf (distinguishing "transparency" from "explanation"); Finale Doshi-Velez & Mason Kortz, Accountability of AI Under the Law: The Role of Explanation 6 (2017) (unpublished working paper) (on file with the Ethics and Governance of Artificial Intelligence Initiative), http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584 ("[E]xplanation is *distinct* from transparency."). However, we eschew the latter terms and rely on the former for two main reasons. First, we want to make clear that we are assessing algorithmic governance against the legal principles of governmental transparency that generally apply to all administrative decisions. Second, even in the context of algorithmic governance, the transparency required of an agency's decision to use an algorithmic tool could, at least in principle, demand more than just the "explainability" of the algorithm itself and its functioning; it could also demand openness about other factors surrounding the government's decision to use an algorithmic tool in the first place.

major principles of governmental reason-giving. In doing so, we aim to explicate the core transparency doctrines applicable to the U.S. federal government. As we note in concluding this Part, these doctrines are far from absolute. At its core, transparency law is pragmatic.

A. Types of Transparency

Fishbowl transparency, as its name suggests, refers to the public's ability to peer inside government and acquire information about what officials are doing.[56] It focuses on public access to information the government holds and information about what the government does. It includes public access to government hearings, records stored in filing cabinets, and materials available on government computers.

A series of federal statutes demand that the government provide fishbowl transparency. The Government in the Sunshine Act and the Federal Advisory Committee Act require meetings of multi-member commissions and advisory bodies, respectively, to be open to the public.[57] The Freedom of Information Act (FOIA) requires agencies to provide government documents to members of the public on request.[58] The Administrative Procedure Act (APA)[59] requires agencies to publish notices of new regulations they propose to adopt.[60] The E-Government Act calls upon agencies to create websites and make information accessible via the Internet.[61] These and other legal requirements help ensure that those who are affected by governmental decisions can monitor what officials are doing and respond on an informed basis.

By contrast to fishbowl transparency's emphasis on public *access* to information about what government is doing, reasoned transparency emphasizes the usefulness of that information—that is, whether government reveals *why*

56. A concrete manifestation of this type of transparency, as well as this terminology, can be found in the principles of public access to information under which the U.S. Environmental Protection Agency has long operated and that were originally outlined in what is known as the "fishbowl memo" issued in 1983 by then-Administrator William Ruckelshaus. *See* Press Release, EPA, Ruckelshaus Takes Steps to Improve Flow of Agency Information [Fishbowl Policy] (May 19, 1983), https://archive.epa.gov/epa/aboutepa/ruckelshaus-takes-steps-improve-flow-agency-information-fishbowl-policy.html#memo.

57. Government in the Sunshine Act, 5 U.S.C. § 552(b) (2018); Federal Advisory Committee Act, 5 U.S.C. app. § 10(a)(1).

58. Freedom of Information Act, 5 U.S.C. § 552.

59. Administrative Procedure Act, 5 U.S.C. §§ 551–559, 561–570(a), 701–706.

60. *Id.* § 553(b).

61. E–Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899 (2002) (codified as amended in scattered sections of 5, 10, 13, 31, 40, and 44 U.S.C.)

it took action. Reasoned transparency stresses the importance of government explaining its actions by giving reasons.[62] As one of us has written elsewhere, under principles of reasoned transparency, the government must provide explanations that are

> based on application of normative principles to the facts and evidence accumulated by decisionmakers—and [that] show why other alternative courses of action were rejected. Sound policy explanations are not the same as the kind of account that journalists, historians, or social scientists would give if they were trying to explain, as an empirical matter, why a policy was in fact adopted, an account that would clearly be aided by an expansion of fishbowl transparency. Instead, reasoned transparency depends on making a substantive evaluation of the soundness of an official's reasoning, not on knowing whether that official might have met with one interest group or another.[63]

Administrative lawyers will readily recognize that the basis for reasoned transparency can be found in the due process clauses of the Fifth and Fourteenth Amendments[64] as well as in the APA's procedural requirements.[65]

Both fishbowl transparency and reasoned transparency are important for good government—and both bring with them legal obligations that government must honor whenever it relies on machine learning. These two types of transparency are related in another way too: reasoned transparency inherently depends on some level of fishbowl transparency. After all, for the government to offer a public explanation of *why* it took a specific action, it must, if nothing else, disclose *what* action it took. Reasoning also necessitates the disclosure of facts that the government has collected and any analyses it has conducted that justify its actions.[66]

Still, despite this minimal connection, fishbowl and reasoned transparency are different kinds of transparency, grounded in different sources of law. These differences matter for algorithmic governance because machine learning presents its most distinctive challenge to reasoned transparency, not fishbowl transparency. The ostensibly black-box nature of machine-learning algorithms raises the question of whether government will be able to explain

62. *See generally* Martin Shapiro, *The Giving Reasons Requirement*, 1992 U. CHI. LEGAL F. 179 (1992) (discussing federal law's requirement to give reasons).

63. Coglianese, *supra* note 55, at 537.

64. U.S. CONST. amends. V, XIV. The elements of procedural due process include a requirement for some kind of "statement of reasons for the decision" made by the government official. Mathews v. Eldridge, 424 U.S. 319, 325 n.4 (1976) (citing Goldberg v. Kelly, 397 U.S. 254 (1970)). *See generally* Henry Friendly, *Some Kind of Hearing*, 123 U. PA. L. REV. 1267, 1279–95 (1975); Citron, *supra* note 6, at 1281–88.

65. *See, e.g.*, 5 U.S.C. § 706(2)(A).

66. *See generally* Cary Coglianese et al., *Seeking Truth for Power: Informational Strategy and Regulatory Policymaking*, 89 MINN. L. REV. 277 (2004) (discussing the centrality of information to governmental decisionmaking).

sufficiently why learning algorithms reach the predictions they do. This is not to say that fishbowl transparency does not matter when governments use machine learning; rather, that algorithmic governance presents few if any truly distinctive questions in terms of fishbowl transparency.

B. Legal Demands for Reason-Giving

To determine whether governmental use of machine-learning algorithms is compatible with reasoned transparency, it is necessary to understand the major sources of the law's demand for governmental reason-giving. These sources include the doctrines of substantive and procedural due process, rules of administrative procedure, and standards for arbitrary and capricious review.

When governmental adjudication adversely affects private interests protected by the Fifth and Fourteenth Amendments—"life," "liberty," and "property"—both substantive and procedural due process will be implicated.[67] To comport with the demands of substantive due process, the government's action must generally be found justifiable as a rational means to achieve a legitimate government purpose.[68] Separately, procedural due process, which also applies, demands that government provide an individual or

67. As Jerry Mashaw has noted, the courts' application of the due process clauses in the Constitution "has dramatically increased the demand for transparently rational administrative adjudication." Jerry L. Mashaw, *Small Things Like Reasons Are Put in a Jar: Reason and Legitimacy in the Administrative State*, 70 FORDHAM L. REV. 17, 26 (2001).

68. The same rational basis test also forms the legal analysis under the Equal Protection Clause of the Fourteenth Amendment, which has been incorporated into the Fifth Amendment's due process demands on the federal government. United States v. Carolene Prods. Co., 304 U.S. 144, 152 (1938) (holding that substantive due process analysis of "regulatory legislation" begins with "the assumption that it rests upon some rational basis"); Bolling v. Sharpe, 347 U.S. 497, 499 (1954) (basing in the Fifth Amendment's Due Process Clause the application of equal protection constraints on the federal government); *see also* Minnesota v. Clover Leaf Creamery Co., 449 U.S. 456, 470 n. 12 (1981) (reasoning that, from a "conclusion under equal protection" that a law meets the rational basis test, "it follows a fortiori that the Act does not violate the Fourteenth Amendment's Due Process Clause"). A more demanding set of reasons is required under the "strict scrutiny" test when fundamental rights (for due process) or suspect classifications (equal protection) are implicated by governmental actions. 16B AM. JUR. 2D *Constitutional Law* § 862 (2018). Given the focus of this Article on general principles of transparency and how they are implicated by machine learning's black-box nature, we only discuss the rational basis test here. We recognize, of course, that government could use machine-learning algorithms to target fundamental rights. But if heightened scrutiny under substantive due process were demanded of a machine-learning application, it would not be due to the algorithm's black-box nature per se, but to the fundamental rights

other legal entity deprived of protected interests with notice of its action and an opportunity to be heard.[69] The precise procedural steps required can vary markedly across different adjudicatory settings, but the affected individual must be provided enough information about the government's decisionmaking to permit the individual a fair opportunity to correct for potential errors.[70] At its core, procedural due process requires impartial decisionmaking based on accurate evidence and relevant legal rules. As the Supreme Court noted in *Goldberg v. Kelly*,[71] "[t]o demonstrate compliance with this elementary requirement, the decision maker should state the reasons for his determination and indicate the evidence he relied on."[72]

Even beyond these constitutional requirements for due process, statutory rules of administrative procedure direct federal agencies to provide reasoned transparency. When Congress has dictated in other statutes that agencies provide an on-the-record hearing[73]—that is, engage in so-called formal adjudication or rulemaking—the APA requires that agencies provide "a statement of findings and conclusions, and the reasons or basis therefor, on all the material issues of fact, law, or discretion."[74] The APA further provides that such actions subject to formal procedures must be justified on the basis of substantial evidence.[75] Of course, most agency actions are not subject to the APA's formal requirements. Still, even for certain informal actions, agencies must provide reasons. The APA requires that in informal—or "notice-and-comment"—rulemaking, agencies must provide a "concise statement of the basis and purpose" of the rule ultimately adopted.[76] In addition, under the Unfunded Mandates Reform Act, the most significant rules adopted by Executive Branch agencies must be accompanied by a "written statement" that

implicated by the application of the algorithm.

69. For a helpful overview of procedural due process requirements, see Sidney A. Shapiro & Richard E. Levy, *Heightened Scrutiny of the Fourth Branch: Separation of Powers and the Requirement of Adequate Reasons for Agency Decisions*, 1987 DUKE L.J. 387, 405–06, 418 n.144, 431 n.213 (1987).

70. Greene v. McElroy, 360 U.S. 474, 496–97 (1959) (describing as an "immutable" principle that whenever governmental action affecting a protected interest "depends on fact findings, the evidence used to prove the Government's case must be disclosed to the individual so that he has an opportunity to show that it is untrue").

71. 397 U.S. 254 (1970).

72. *Id.* at 271; *see also* Friendly, *supra* note 64, at 1268, 1291–92 (describing a "statement of reasons" as one of the most important requirements of procedural due process).

73. *See generally* United States v. Fla. E. Coast Ry. Co., 410 U.S. 224 (1973).

74. 5 U.S.C. § 557(c)(3)(A) (2018).

75. *Id.* § 706(2)(E).

76. *Id.* § 553(c).

includes a "qualitative and quantitative assessment" of the estimated costs and benefits of the rule.[77]

Moreover, agency action subject to judicial review can be struck down if a court it finds it to be "arbitrary" and "capricious."[78] The Supreme Court has interpreted the APA to require that agencies "examine the relevant data and articulate a satisfactory explanation" for each of their actions, especially with respect to rulemaking.[79] The Court's interpretation of the arbitrary and capricious standard imposes an affirmative obligation on an agency to provide some kind of administrative record containing evidence and reasons for new actions at the time it undertakes them.[80] On judicial review, the courts will then seek "to determine whether the [agency] has considered the relevant factors and articulated a rational connection between the facts found and the choice made."[81]

The requirements imposed by the APA, together with due process requirements for adequate notice and explanation, push administrative agencies to provide reasoned justifications for their actions. As Jerry Mashaw has noted, "[t]he path of American administrative law has been the path of the progressive submission of power to reason."[82] Reasoning requires going beyond mere expediency, whim, or self-interest.[83] Agencies need to be able to explain their choices as based on evidence and justified in terms of the criteria

77. Unfunded Mandates Reform Act of 1995, Pub. L. No. 104-4, 109 Stat. 48 (codified as amended in scattered sections of 2 U.S.C.). Similar requirements for the most significant rulemakings have been imposed by executive order. *See* Exec. Order No. 12,866, 58 Fed. Reg. 51,735 (1993).

78. 5 U.S.C. § 706(2)(A); *see also id.* § 702 (providing the right of judicial review for any person "adversely affected" by final agency action).

79. Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983).

80. Citizens to Pres. Overton Park, Inc. v. Volpe, 401 U.S. 402, 420 (1971) (reasoning that although agencies are not necessarily required to make formal findings, courts must examine the "full administrative record that was before the [agency] at the time [of its] decision"); *see also* Encino Motorcars, L.L.C. v. Navarro, 136 S. Ct. 2117, 2127 (2016); *State Farm*, 463 U.S. at 42; SEC v. Chenery Corp., 318 U.S. 80, 94–95 (1943).

81. Balt. Gas & Elec. Co. v. Nat. Res. Def. Council, Inc., 462 U.S. 87, 105 (1983) (internal citations omitted).

82. Mashaw, *supra* note 67, at 26.

83. Not even the change in political ideology of an administration's leadership, occasioned by a presidential election, constitutes a sufficient reason for agency action. *State Farm*, 463 U.S. at 42; *see also* Frederick Schauer, *Giving Reasons*, 47 STAN. L. REV. 633, 657–58 (1995) (observing that "[a] reason-giving mandate will . . . drive out illegitimate reasons when they are the only plausible explanation for particular outcomes").

or purposes contained within the statutes governing their authority.[84]

Satisfying the law's reason-giving requirement also necessarily entails some degree of fishbowl disclosure. At a minimum, the government must disclose the information in its possession that provided the factual basis for its decision. In the rulemaking context, administrative law has generally pushed agencies to make publicly available all of the information that is before the agency at the time of its decision, including information that might run counter to the agency's decision.[85] To justify a rulemaking decision, an agency must confront adverse information and contrary arguments presented in public comments, and it must explain why such information and arguments do not undermine the basis for the agency's rule.[86]

C. Law's Pragmatism About Open Government

Notwithstanding these various legal demands for open government, the law does not require that the government provide total transparency. Fishbowl transparency necessitates an openness that would be better described as *translucent*, not transparent. FOIA contains nine major exemptions from disclosure, including those related to personnel information, law enforcement protocols, and confidential business information and trade secrets.[87] Furthermore, the APA does not require that agencies provide public notice for all new rules. In fact, the APA's "good cause" exemption[88] is wide enough that, according to Connor Raso's empirical research, many federal agencies' regulatory issuances occur under that exemption and thus are issued without any advance notice at all.[89]

Another exemption in the APA allows certain agency actions to escape judicial oversight. Although the APA establishes a presumption in favor of judicial review of all final agency actions—and thus subjects such actions to

84. *Overton Park*, 401 U.S. at 416 (noting that under the arbitrary and capricious standard, the agency must show that it "acted within the scope of [its] statutory authority" and its decision "was based on a consideration of the relevant factors").

85. Some statutes require agencies to maintain a publicly accessible rulemaking docket with all such information. *E.g.*, Clean Air Act, 42 U.S.C. § 7607 (2018). The advent of the federal web portal, Regulations.gov, now makes it relatively easy for agencies to provide such docket information online.

86. *See, e.g.*, United States v. N.S. Food Prods. Corp., 568 F.2d 240, 252 (2d Cir. 1977) ("It is not in keeping with the rational process to leave vital questions, raised by comments which are of cogent materiality, completely unanswered.").

87. Freedom of Information Act, 5 U.S.C. § 552(b)(1)–(9) (2018).

88. 5 U.S.C. § 553(b)(3)(B).

89. Connor Raso, *Agency Avoidance of Rulemaking Procedures*, 67 ADMIN. L. REV. 65 (2015).

the potential application of the arbitrary and capricious test—this presumption does not apply to agency actions that are "committed to agency discretion by law"[90] or for which there is "no law to apply."[91] Courts treat these categories as narrowly defined, but the Supreme Court has held that government decisions about whom to inspect or target for enforcement are committed to agency discretion—meaning that an agency need not explain why it chooses to investigate certain individuals or facilities but not others.[92] Agencies effectively have complete discretion over whom they choose to target.[93]

Even for actions that are subject to judicial review, reasoned transparency usually does not require exhaustive or extensive explanations. The rational basis test for substantive due process, for instance, is hardly any test at all. The relationship between a governmental action and a proper purpose need only be "debatably rational"[94] and the government's purpose can be merely "conceivably legitimate."[95] Indeed, to withstand the rational basis test, the government never actually needs to give any reason at all; the test merely requires that a reason *could* be found linking a government action to a legitimate end.[96]

90. 5 U.S.C. § 701(a)(2).

91. Citizens to Pres. Overton Park, Inc. v. Volpe, 401 U.S. 402, 410 (1971).

92. Heckler v. Chaney, 470 U.S. 821, 832 (1985). Another category excluded from judicial review comprises "non-actions," such as decisions to avoid initiating a rulemaking, at least if no statutory obligation or deadline compels taking action. *See generally* Cary Coglianese & Daniel Walters, *Agenda-Setting in the Regulatory State: Theory and Evidence*, 68 ADMIN. L. REV. 93 (2016) (discussing agenda-setting at administrative agencies).

93. Of course, if agencies' targeting practices are based on unconstitutionally discriminatory grounds, the courts may certainly step in. *Cf.* Floyd v. City of New York, 959 F. Supp. 2d 540, 562 (S.D.N.Y. 2013) (finding that New York City had "adopted a policy of indirect racial profiling by targeting racially defined groups for stops based on local crime suspect data" with the result of "disproportionate and discriminatory stopping of blacks and Hispanics in violation of the Equal Protection Clause"). Furthermore, for inspections of business facilities, once a specific facility has been targeted for inspection, the government may be required to obtain a search warrant to gain access to the facility to conduct the inspection and thus would need to provide a reason for its requested search. But that reason could be simply that the facility has been identified for inspection "on the basis of a general administrative plan for the enforcement of the [law] derived from neutral sources." Marshall v. Barlow's, Inc., 436 U.S. 307, 331–32 (1978). It may perhaps seem unnecessary to note that, once the government conducts an inspection and seeks to impose a fine for noncompliance, it must then provide reasons, grounded in law and in fact, for why penalties should be imposed.

94. Reid v. Rolling Fork Pub. Util. Dist., 854 F.2d 751, 753 (5th Cir. 1988) (explicating the test for rational basis review under the Equal Protection Clause). The rational basis test is the same for substantive due process and equal protection. *See supra* note 68.

95. *Reid*, 854 F.2d at 753.

96. United States v. Carolene Prods. Co., 304 U.S. 144 (1938).

Procedural due process might seem to hold more bite. But even there, the Court has made it clear that the reasons agencies must provide in the adjudicatory context "need not amount to a full opinion, or even formal findings of fact and conclusions of law."[97] More generally, the demands of procedural due process are now encapsulated under a balancing test that calls upon courts to make pragmatic judgments about how strict procedural demands should be, expressly taking into account the impact of potential procedural demands on governmental resources.[98]

With respect to the APA, although the arbitrary and capricious test is often referred to as "hard look" review, the Supreme Court has made clear that "the scope of review under the 'arbitrary and capricious' standard is narrow, and a court is not to substitute its judgment for that of the agency."[99] The Court has instructed judges to "uphold a decision of less than ideal clarity if the agency's path may reasonably be discerned."[100]

The Supreme Court has also made clear that judicial inquiry into an agency's reasoning does not entail anything like the extensive discovery process that typically takes place in civil litigation. Rather, review is limited to what the agency had available in its record at the time of its decision.[101] Courts cannot compel the cross-examination of evidence supplied by the agency.[102] Courts also ordinarily must not "probe the mental processes of the Secretary in reaching his conclusions"[103]—they merely can look to the reasons that agency officials have provided.[104]

Furthermore, when the reasons that an agency offers depend on highly technical matters, such as those demanding statistical and scientific expertise, courts give the agency considerable deference. In its decision in *Baltimore Gas*

97. Goldberg v. Kelly, 397 U.S. 254, 271 (1970).

98. Matthews v. Eldridge, 424 U.S. 319 (1976); *see also* Coglianese & Lehr, *supra* note 10, at 1184–91.

99. Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983).

100. *Id.* (quoting Bowman Transp., Inc. v. Arkansas-Best Freight Sys., Inc., 419 U.S. 281, 286 (1974)).

101. Citizens to Pres. Overton Park, Inc. v. Volpe, 401 U.S. 402 (1971); *see also* SEC v. Chenery Corp., 318 U.S. 80 (1943).

102. Vt. Yankee Nuclear Power Corp. v. Nat. Res. Def. Council, Inc., 435 U.S. 519, 541 (1978).

103. Morgan v. United States, 304 U.S. 1, 18 (1938).

104. The requirement is not to offer an empirical or psychological description of the origins of a decision, but rather to provide a legal or policy justification for it. As Frederick Schauer has noted, "[d]escription is not justification." Schauer, *supra* note 83, at 651.

& Electric Co. v. Natural Resources Defense Council,[105] the Supreme Court cautioned that courts should not second-guess a government agency when it is "making predictions, within its area of special expertise, at the frontiers of science."[106] The Court emphasized that, "[w]hen examining this kind of scientific determination, as opposed to simple findings of fact, a reviewing court must generally be at its most deferential."[107]

Judges by and large do not hold agencies to extremely high standards of rationality under the arbitrary and capricious standard. Rather, as Adrian Vermeule has noted, they are simply looking to rule out "clear and indefensible legal error or irrationality."[108] Vermeule has reported that, of the sixty-four cases since 1982 in which the Supreme Court has considered an agency action under the arbitrary and capricious test, the Court has upheld the agency decision over ninety percent of the time.[109]

It is clear that, although administrative law does demand both fishbowl and reasoned transparency, the demands of each are far from onerous.[110] Transparency law has been largely pragmatic. Fishbowl transparency requirements include exemptions that take into account administrative imperatives, and

105. 462 U.S. 87, 103 (1983).

106. *Id.*

107. *Id. See also* FERC v. Elec. Power Supply Ass'n, 136 S. Ct. 760, 782 (2016) (noting the narrowness of arbitrary and capricious review and stating that "nowhere is that more true than in a technical area"); Mobil Oil Expl. & Producing Se. v. United Distrib. Cos., 498 U.S. 211, 231 (1991) (explaining that "we are neither inclined nor prepared to second-guess the agency's reasoned determination in this complex area").

108. ADRIAN VERMEULE, LAW'S ABNEGATION 34 (2016). *See also* Thomas O. McGarity & Wendy E. Wagner, *Legal Aspects of the Regulatory Use of Environmental Modeling*, 33 ENVTL. L. REP. 10,751, 10,757 (2003) ("Despite the high level of judicial deference, EPA's models are frequently subject to tedious, technical nitpicking . . . . Courts often consider these challenges in detail, but if after this analysis the courts discover that the disagreement concerns a battle of the experts, they typically defer to the Agency's judgment.").

109. VERMEULE, *supra* note 108, at 158.

110. Some administrative law scholars have asserted that judicial review under the arbitrary and capricious standard does impose a heavy burden on agencies. Richard J. Pierce, Jr., *Rulemaking Ossification Is Real: A Response to* Testing the Ossification Thesis, 80 GEO. WASH. L. REV. 1493 (2012). However, researchers who have investigated these claims empirically and systematically have found them wanting. Cary Coglianese, *Empirical Analysis and Administrative Law*, 2002 U. ILL. L. REV. 1111 (2002); William S. Jordan III, *Ossification Revisited: Does Arbitrary and Capricious Review Significantly Interfere with Agency Ability to Achieve Regulatory Goals Through Informal Rulemaking?*, 94 NW. U. L. REV. 393 (2000); Anne J. O'Connell, *Political Cycles of Rulemaking: An Empirical Portrait of the Modern Administrative State*, 94 VA. L. REV. 889, 896 (2008); Jason W. Yackee & Susan W. Yackee, *Testing the Ossification Thesis: An Empirical Examination of Federal Regulatory Volume and Speed, 1950-1990*, 80 GEO. WASH. L. REV. 1414 (2012).

agencies have accordingly managed to adapt their operations to provide routinized disclosure of government information.[111] When it comes to reasoned transparency, courts do expect agencies to explain themselves, at least when they take actions not committed to agency discretion; however, they also give agencies considerable deference as to what counts as sufficient reasoning.[112]

III. REASON-GIVING IN ALGORITHMIC GOVERNANCE

Having established what transparency law demands, we now turn to the question of whether—or how—governmental reliance on machine learning might meet those demands. As for any legal question, the answer will ultimately depend on how government actually uses machine learning—and even on what kind of machine-learning algorithm it uses. After all, machine learning itself refers to a broad range of techniques which apply to varied types of data, perform many different functions, and vary in their mathematical complexity. Any general analysis must thus proceed with appropriate caveats.

We also must acknowledge that there will always be ways that government could conceivably use machine learning that would clearly violate both fishbowl and reasoned transparency requirements. It would be obviously irresponsible and unlawful for a governmental authority to rely on machine-learning algorithms deceptively by both concealing the automation of otherwise reviewable actions and offering no valid independent reasons for those actions. More generally, the government could always act maliciously and deceptively to bias machine learning's results to achieve nefarious policy ends. We are mindful of these possibilities, but we also recognize that nefar-

111. This is not to say that agency practices under FOIA have been optimal—or even sufficient. Some observers have lamented what they perceive to be the high costs of FOIA compliance. *See, e.g.*, Antonin Scalia, *The Freedom of Information Act Has No Clothes*, REGULATION, Mar.–Apr. 1982, at 15, 16. Others have noted that government agencies are far from prompt or forthcoming in responding to requests for information. *See, e.g.*, SEAN MOULTON & GAVIN BAKER, CTR. FOR EFFECTIVE GOV'T, MAKING THE GRADE: ACCESS TO INFORMATION SCORECARD 2015 (Mar. 2015), https://www.foreffectivegov.org/ sites/default/files/info/access-to-information-scorecard-2015.pdf; *Delayed, Denied, Dismissed: Failures on the FOIA Front*, PROPUBLICA (July 21, 2016, 8:01 AM), https://www.propublica. org/article/delayed-denied-dismissed-failures-on-the-foia-front (noting that government agencies are far from prompt or forthcoming in responding to requests for information). For findings from a government audit of FOIA practices, see U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-18-452T, FREEDOM OF INFORMATION ACT: AGENCIES ARE IMPLEMENTING REQUIREMENTS BUT NEED TO TAKE ADDITIONAL ACTIONS (2018).

112. *See supra* notes 108–109 and accompanying text.

ious governmental action can take place entirely independently of any application of machine learning. The possibility of dishonest, oppressive, or even evil government should never be dismissed—but it is a concern separate from our focus here.

Arguably equally concerning, though, would be the possibility that a government agency uses machine learning merely irresponsibly or in a technically uninformed manner. As mentioned, the development of machine-learning algorithms, especially for the kinds of specialized applications to which they would be applied by government officials, is a challenging endeavor. It requires expertise in statistics and computer or information science. It also requires knowledge of how policy choices can be embedded in mathematical choices made while designing the algorithm. It is certainly conceivable—and perhaps likely under current conditions—that many government agencies will not have staff with the requisite expertise to make these choices in an informed way.[113] Indeed, perhaps this is why governmental authorities that have so far looked to machine learning have tended to contract with private companies to obtain the relevant expertise to develop their algorithmic applications. To oversee adequately the private contracting of algorithmic design, though, governments still need sufficient in-house expertise to ask the right questions and demand that their contractors act responsibly and with the appropriate degree of transparency.

Of course, we also must recognize that in many instances government officials will use machine learning in myriad ways that will hardly even implicate transparency law at all. When the U.S. Postal Service uses machine-learning algorithms in its mail-sorting equipment to "read" handwritten zip codes, and when the National Weather Service uses machine-learning algorithms to help generate weather forecasts, no serious transparency concerns arise.[114] Even less banal uses, such as the use of machine learning to identify targets for subsequent human inspection or auditing, may also easily satisfy transparency law—if for no reason other than that the legal demands in some of these circumstances will be minimal. Agency use of algorithms for law

113. On the need for adequate resources and human capital to take advantage of algorithmic governance, see Coglianese, *supra* note 34; Robert L. Glicksman, David L. Markell & Claire Monteleoni, *Technological Innovation, Data Analytics, and Environmental Enforcement*, 44 ECOLOGY L.Q. 41, 47 (2017) ("Optimal use of big data . . . will require [government] to hire experts in data analytics and make significant investments in computer systems capable of collecting, transporting, storing, and analyzing the data.").

114. For background on the U.S. Postal Service's use of hand-writing learning algorithms and the National Weather Service's use of machine learning in meteorological forecasting, see Coglianese & Lehr, *supra* note 10, at 1162.

enforcement targeting, after all, would presumably fall outside of both FOIA's requirements and the scope of judicial review under the APA.[115] Similarly, whenever governmental agencies use machine-learning analysis as but one factor in a larger human decisionmaking process or as a mere supplement to human decisionmaking, they should experience relatively few difficulties in satisfying transparency laws' demands—as long as the human officials have sufficient independent reasons to justify their actions.[116]

The more interesting, and presumably more difficult, cases will arise when the government uses machine learning in outcome-determinative ways to make consequential and nondiscretionary decisions. We referred earlier to cases like these as ones of either *regulating by robot* or *adjudicating by algorithm*, denoting situations where automated, algorithmic systems essentially make governmental decisions by either establishing policies or resolving individual disputes or claims. We draw here on these outcome-determinative uses of machine learning to create something of a test case for the purpose of analyzing whether government can meet transparency standards. If government agencies that use machine learning can satisfy the law's demands for transparency when used in such determinative ways, then their use of such algorithms in other, less determinative applications should by extension pass muster.[117]

115. *See supra* notes 87–89 and accompanying text.

116. If an agency cannot provide any independent reasons for its decision, then machine learning cannot really be serving as a mere supplement to human decisionmaking—because, by definition, if the algorithm is a mere supplement then there must be some reason other than the algorithm's output. On the other hand, it is possible for a human official to have another reason but still to have the human judgment depend on an algorithm's output. A human decisionmaker might say, "I decide X whenever my human judgment is to do X, based on Y reason, *and* when the algorithm's output is Z." In such a case, where the human decisionmaker depends on some affirmation from the algorithm to support a decision, the human-generated reason may be separate from the algorithm but it is still contingent on and interactive with the algorithm's output. Such a case is one that we will treat here, for our purposes, as conceptually indistinguishable from the complete substitution of the algorithm for human decisionmaking, as the human decision is not entirely independent of the algorithm.

117. As the discussion in the previous note ought to indicate, by "determinative" or "outcome determinative," we mean that the justification for a particular governmental action lies purely with the algorithm and its output. It does not matter whether the government's action is itself entirely executed by a machine or whether the machine's output needs to be ratified pro forma by a human official, provided that the sole justificatory basis for the action results from the algorithm. Although we use terms like "determinative" to distinguish qualitatively different degrees of reliance on algorithms, it is not essential to delve into the metaphysics of decisionmaking to claim that the government's decision was in fact "made" by an algorithm.

We begin this Part by reiterating the distinction between fishbowl and reasoned transparency, both of which could be implicated by specific governmental uses of machine learning. Much criticism of machine learning emphasizes fishbowl transparency, but any problems with fishbowl transparency created by the government's application of machine learning are neither new nor particularly difficult to resolve. It is reasoned transparency that would appear to be most distinctively implicated by machine learning's black-box nature. Consequently, after initially distinguishing fishbowl and reasoned transparency, we next turn to the kind of reasons that analysts can offer about how algorithms work—that is, how they can peer inside the black box. We conclude that government officials should be able quite easily to satisfy the demands of reasoned transparency. A high-level explanation of an algorithm's functioning is both possible and legally sufficient. Beyond this, we also highlight in the final section of this Part the many technical advances that machine-learning researchers are making to improve methods for extracting reasons from algorithms, enabling officials to go beyond what is legally required when explaining their algorithms. Given the current ability of agencies to meet prevailing legal demands, combined with the promise of ongoing technical advances in the explainability of machine learning, whatever unease may remain about algorithmic governance should only wane with time.

A. Situating Fishbowl and Reasoned Transparency

Fishbowl transparency figures prominently in contemporary discussion of machine learning. Critics worry about a world "increasingly controlled by secret models."[118] Investigative reporters emphasize "the largely hidden effect of algorithms," raising worries about the "proprietary" nature of algorithms

As one of us has suggested in another context, in any governmental setting where there may be multiple potential sources of input, it may be difficult to single out any one source as *the* decisionmaker. *Cf.* Cary Coglianese, *The Emptiness of Decisional Limits: Reconceiving Presidential Control of the Administrative State*, 69 ADMIN. L. REV. 43 (2017); Cary Coglianese & Kristin Firth, *Separation of Powers Legitimacy: An Empirical Inquiry into Norms About Executive Power*, 164 U. PA. L. REV. 1869 (2016). For another, when human officials have designed an algorithmic system and chosen to use it as the basis for making decisions, those officials still have some meaningful claim to having shaped even a particular decision by having created and used the algorithmic system. Our purpose in this Article is simply to use terms like "determinative" in a conventional sense to connote the government's reliance on the output of an algorithm as the pivotal factor that justifies a particular decision.

118. CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 13 (2016). O'Neil also worries about what she characterizes as the "arbitrary" nature of learning algorithms—a reasoned transparency concern. *Id.*

developed for courts by private contractors who do "not publicly disclose the calculations used."[119] Researchers study algorithmic transparency by filing FOIA-type requests with state and local governments, finding "a big gap between the importance of algorithmic processes for governance and public access to those algorithms."[120]

Clearly, algorithmic governance presents real concerns about fishbowl transparency.[121] Governmental use of machine learning generates a broad range of potentially disclosable information, including the algorithm's source code, its objective function, its specifications and tuning parameters, its training and test data sets, and the programming details of any ancillary computer programs that translate its predictions into actions.[122] The desire for public access to some or all of this information is understandable, especially if disclosure of at least some information is needed to provide a satisfactory reasoned explanation of actions determined by algorithms.

Not surprisingly, concerns about fishbowl transparency have also found their way into litigation over algorithmic governance. In Wisconsin, a criminal defendant challenged a state trial court's use of a risk assessment algorithm in determining his sentence.[123] Among other claims, the defendant argued that his due process rights were effectively violated by the trial court

119. *See* Julia Angwin et al., *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks.*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. For a discussion of the statistical errors in the ProPublica report, see Jennifer Doleac & Megan Stevenson, *Are Criminal Risk Assessment Scores Racist?*, BROOKINGS INSTITUTION (Aug. 22, 2016), https://www.brookings.edu/blog/up-front/2016/08/22/are-criminal-risk-assessment-scores-racist/.

120. Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103, 132–33 (2018).

121. Others have raised similar concerns. *See, e.g.*, Citron, *supra* note 6, at 1291–93 (discussing tension between FOIA and the limited information governments have provided about automated decision systems used by governments); Joshua A. Kroll et al., *supra* note 36, at 638 (noting that machine learning "is particularly ill-suited to source code analysis"—which the authors describe as "the most obvious approach" to providing transparency about automated decision tools).

122. *See generally* Lehr & Ohm, *supra* note 2 (describing various features of a machine-learning algorithm).

123. State v. Loomis, 881 N.W.2d 749 (Wis. 2016). The case centered on the court's reliance on the "COMPAS" algorithm that figured in the investigative report cited in Angwin, *supra* note 119. COMPAS does not appear to be a *machine-learning* algorithm. *See* Coglianese & Lehr, *supra* note 10, at 1205 n.232. However, the legal issues presented in the case are still relevant here to our analysis of machine-learning algorithms.

because of "the proprietary nature" of the algorithm on which the trial court had relied.[124] The algorithm had been developed by a private firm that considered it "a trade secret" and thus would "not disclose how the risk scores are determined or how the factors are weighed."[125] The defendant argued that this secrecy "denied [him] information which the [trial] court considered at sentencing," constituting a violation of his rights to procedural due process.[126] Ultimately, the Wisconsin Supreme Court rejected the defendant's argument, pointing to the fact that the private firm had released a full list of variables used in calculating risk scores and then noting that the risk analysis was entirely "based upon [the defendant's own] answers to questions and publicly available data about his criminal history."[127] The court held that the availability of this information afforded the defendant the requisite "opportunity to verify that the questions and answers listed on the [risk assessment] report were accurate."[128]

The outcome in the Wisconsin case suggests that governmental authorities can lawfully use algorithms without divulging all the potentially disclosable information about them. Although under some circumstances some information about algorithms might well need to be released under FOIA, transparency law certainly does not require full disclosure of everything.[129] A government agency that uses machine learning to identify facilities to inspect, for example, would presumably not need to disclose information about how its algorithm operates, as that information would be covered by the exemption for law enforcement-related information.[130] Moreover, as the Wisconsin case confirms, if a government agency contracts with a private company to design and deploy an algorithm, that company can lawfully claim that various pieces of algorithmic information are trade secrets that must be protected from disclosure.[131] In addition, some of the underlying data used in particular machine-learning applications might be subject to various legal privacy

124. *Loomis*, 881 N.W.2d at 757.

125. *Id.* at 761.

126. *Id.*

127. *Id.*

128. *Id.* (rejecting the defendant's other arguments as well, although it was important to the court's rulings on those other ground—but not at all mentioned in its ruling on procedural due process—that the risk assessment scores were not used in a determinative way but were merely one factor in sentencing decisions).

129. For a discussion of the circumstances under which algorithmic information might need to be disclosed, see Coglianese & Lehr, *supra* note 10, at 1209–12.

130. 5 U.S.C. § 552(b)(7) (2018).

131. *Id.* § 552(b)(4).

protections, such as where data are drawn from individual medical, educational, credit, or census records.[132]

Admittedly, from the standpoint of anyone concerned about robust fishbowl transparency, the fact that information can be lawfully withheld due to trade secret and other FOIA exemptions might do little more than restate the problem. Legal or not, the information is still withheld from the public. In some circumstances, a reasonable case might well be made on policy grounds against the withholding of information—and for greater disclosure than the law requires.[133] On the other hand, valid and important policy reasons do exist to protect the confidentiality of certain types of information, so the law's exemptions from disclosure might still be justified on policy grounds.[134] It is not our purpose here to engage in a policy debate over how much fishbowl transparency the government should provide, nor is it to analyze the proper scope of data privacy laws on policy grounds. Rather, in highlighting how, under current law, governmental entities that rely on algorithmic systems need not disclose all potentially disclosable information about these systems, we seek to draw two main implications about the role for fishbowl transparency in analyzing the legal viability of machine learning.

The first implication is that any questions about the optimal level of fishbowl transparency run orthogonal to an analysis of machine learning and its black-box character. For example, consider concerns about government agencies outsourcing algorithmic development to private contractors. Nothing about machine learning raises distinctive concerns about the role of private actors in supporting government functions or about the protection of confidential business information used by government officials. Debates over the relationship between public and private actors extend much more broadly than the current debate about machine learning. Some policymakers and scholars advocate greater privatization of government services, while others oppose extensive outsourcing and urge greater reliance on government bureaucrats.[135] In much the same vein, consider claims about the need

132. *See, e.g.*, Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936; The Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g; The Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681(a)-(x) (2018).

133. On related issues, one of us has put forward recommendations for best practices in transparency that go beyond disclosure that is legally mandated. Coglianese et al., *supra* note 9, at 934–46. Our legal analysis here does not necessarily mean that compliance with the law defines the entirety of good government practices.

134. Coglianese et al., *supra* note 66, at 330–31.

135. *See generally* JOHN J. DIIULIO, JR., BRING BACK THE BUREAUCRATS: WHY MORE FEDERAL WORKERS WILL LEAD TO BETTER (AND SMALLER!) GOVERNMENT (2014); PAUL R. VERKUIL, VALUING BUREAUCRACY: THE CASE FOR PROFESSIONAL GOVERNMENT (2017);

for secrecy to protect the intellectual property rights of algorithm developers. Some proponents of intellectual property rights urge stronger protections to promote innovation, while others favor weaker rights to discourage market abuses or combat inequities—but this debate rages on even outside the context of machine learning.[136] The point is simply that many of the concerns about machine learning and fishbowl transparency implicate much larger debates. The policy considerations relevant in other contexts will be basically identical to what is needed to address the debate over the optimal level of fishbowl transparency in a world of algorithmic governance.

Moreover, even if the optimal level of fishbowl transparency does call for expanded disclosure in connection with algorithmic governance, the solutions will have little or nothing to do with machine learning per se. If the concern is with the proprietary nature of algorithms developed by private contractors, government agencies could expand fishbowl transparency simply by changing the terms of their contracts to clarify or even waive trade-secret protections.[137] Or perhaps, instead of contracting with private firms, agencies could create open-source competitions for the design of algorithms[138] or build their own in-house expertise so they do not need to rely on private contractors.[139] Irrespective of the actual role played by private contractors, whenever government agencies develop their own algorithmic systems, they can always consult with independent reviewers and advisory committees or solicit public input through a notice-and-comment process.[140] These are all readily available solutions. None are novel, and none have anything to do with machine learning's distinctive "black-box" character. In the end, the challenges in identifying and providing an optimal level of fishbowl transparency in a world of algorithmic governance are simply not unique to machine learning.

The second implication we wish to draw about machine learning and fishbowl transparency is that this kind of transparency can indirectly but inte-

Mildred E. Warner, *Reversing Privatization, Rebalancing Government Reform: Markets, Deliberation and Planning*, 27 POL'Y & SOC'Y 163 (2008).

136. MEIR PEREZ PUGATCH, THE INTELLECTUAL PROPERTY DEBATE (2006).

137. This is Brauneis and Goodman's chief recommendation. Brauneis & Goodman, *supra* note 120, at 164–66.

138. Edward L. Glaeser et al., *Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy*, 106 AM. ECON. REV. PAPERS & PROC. 1 (2016).

139. Coglianese, *supra* note 34.

140. *See, e.g.*, Coglianese & Lehr, *supra* note 10, at 1190–91 (discussing use of advisory committees and rulemaking proceedings). For further discussion of the role of advisory committees and notice-and-comment rulemaking, see *supra* notes 57 and 60 and accompanying text.

grally affect government's ability to meet the demands of reasoned transparency. It is reasoned transparency, after all, that does appear to pose a distinctive challenge for algorithmic governance. As we noted earlier, because reasoned transparency is all about explanation and meaning—the ability of government to say why and how an action was taken—the black-box nature of algorithms seems, at least at first glance, to threaten this form of transparency in potentially novel ways. But it is important to keep in mind that government agencies' ability to satisfy the demands of reasoned transparency could very much be affected by the extent of their fishbowl transparency.

A recent preliminary decision issued by a federal district court in Texas illustrates this connection between fishbowl transparency and reasoned transparency. The court's decision arose in a case centered on a Houston school district's use of algorithms to rate their teachers' performance and to provide the district a basis for dismissing poorly rated teachers.[141] The school district relied on an outside vendor to develop and run its algorithmic system, and the vendor treated its "algorithms and software as trade secrets, refusing to divulge them to either [the school district] or the teachers themselves."[142] Nine teachers and the local teachers' union challenged the school district's reliance on the algorithms on multiple grounds, including the procedural due process claim that, without "access to the computer algorithms and data necessary to verify the accuracy of their scores," they were denied their constitutional rights.[143]

The district court rejected the school district's motion for summary judgment on the procedural due process claim, finding that the teachers had put forward enough evidence to take their case to the jury.[144] The court asserted that, "without access to . . . proprietary information—the value-added equations, computer source codes, decision rules, and assumptions—[the teachers'] scores will remain a mysterious 'black box,' impervious to challenge."[145]

The court did not question trade secret protection itself. It accepted the school district's argument that "'the Due Process Clause does not empower Plaintiffs to put [the vendor] out of business' by requiring the disclosure of its

141. Hous. Fed'n of Teachers, Local 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168 (S.D. Tex. 2017). It is not altogether clear from the court's opinion whether the algorithm at issue was a machine-learning algorithm, but the relevant legal issues we discuss here are indistinguishable from those that would be implicated by learning algorithms.

142. *Id.* at 1177.

143. *Id.* at 1176.

144. *Id.* at 1180.

145. *Id.* at 1179.

trade secrets."[146] Rather, the court said that the question for the jury was simply whether "a policy of making high stakes employment decisions based on secret algorithms [is] incompatible with minimum due process."[147] If it is incompatible, the remedy would be for the school district to find some other policy that does not rely on secret algorithms, not to repudiate a firm's trade secrets.[148] The upshot is that reasoned transparency can become affected by limits on fishbowl transparency whenever sufficient reasons for decisions based on machine learning require the disclosure of confidential source code or other trade secrets.

Importantly, though, the preliminary nature of the district court's decision in the Texas case means that the court did not rule out the possibility that, with further evidence presented at a trial, a jury could still properly find for the school district.[149] It might be possible to protect trade secrets while still providing teachers with sufficient information to satisfy the demands of procedural due process. Just as in the Wisconsin Supreme Court's decision, circumstances may exist where the government can rely on a private contractor but still put forward enough non-proprietary information to provide adequate reasons for machine learning decisions. We thus turn next to what it means for government to offer adequate reasons in the context of algorithmic governance.

B. The Adequacy of Reason-Giving in Algorithmic Governance

To discern what reasons government must give to support determinative uses of machine learning, let us begin by assuming, for the sake of analysis, a setting in which there exist no limitations whatsoever on public disclosure of information related to a given algorithmic system. Any interested person can be given access to that algorithm's objective function, its specifications (e.g., the kind of algorithm selected, a list of what the input variables were, and the algorithm's tuning parameters), the training and testing data, and even the full source code. As a result, all interested persons will be able to access everything about that algorithm and even test it out for themselves. In principle, then, all members of the public and all affected persons could conceivably understand how the algorithm works as well as does anyone in government. Still, even with this assumed full disclosure, there will remain some irreducible inscrutability, at least compared to what can be obtained when using traditional statistics. As we have outlined, it is difficult to put into intuitive prose

146. *Id.*

147. *Id.*

148. *Id.*

149. *Id.* at 1173 (explaining the summary judgment standard).

how exactly learning algorithms operate—that is, exactly what relationships they are keying in on and how those relationships affect the predictions.[150] Would this inherent level of inscrutability, even assuming a best-case scenario of full fishbowl transparency, bar federal agencies from relying on machine-learning algorithms?[151]

The answer would almost certainly be "no." Machine-learning algorithms' irreducible inscrutability should not form a bar to their use by government officials, even to substitute for human decisionmaking in circumstances where adequate reasons must be given. What will count as an adequate reason will vary depending on the legal source of the reason-giving requirement, but in general, as we explained in Part II, the law's demand for reason-giving is neither absolute nor total.[152] Ultimately it is pragmatic. Transparency law can assuredly accommodate a government agency's choice to employ a state-of-the-art tool that can improve accurate decisionmaking in the service of a valid governmental purpose, even if the nature of that tool is such that it does not readily permit the government to offer intuitive accounts of its inner workings in the same way as traditional but often less accurate statistical tools.

1. Substantive Due Process

Consider first the reason-giving needed to meet the rational basis test of substantive due process. As we indicated earlier, it may be a stretch even to call this a "test."[153] To withstand rational basis scrutiny, the government

150. *See* discussion *supra* Section I.B.

151. As we delve into our discussion here of inscrutability, what the law requires, and what methods make algorithms more scrutable, it may be worthwhile to situate our use of the term "inscrutability" in the context of work by Andrew Selbst and Solon Barocas. They distinguish between inscrutability—an inability to understand what the complex predictive rules that a machine-learning algorithm discovers are—and non-intuitiveness—an inability to understand how the input information that the algorithm considers are relevant to the outcome. *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1091, 1094 (2018). We use "inscrutability" largely as they do. Also, we do not take up in depth non-intuitiveness because, as we discuss, an understanding of whether input features are intuitively relevant to the outcome—such as whether there is an intuitive causal relationship between them—is not at all required to justify administrative decisions. Furthermore, as Selbst and Barocas rightly point out, causal inferences cannot be obtained from machine learning, and, even if causal relationships could be discovered, they are likely to be just as complex as correlational ones. *Id.* at 1098. But government does not need to establish causality to satisfy principles of reasoned transparency.

152. *See* discussion *supra* Section II.C.

153. *See supra* notes 94–96 and accompanying text.

does not need to give a reason for its action; rather, the action simply must be one that is capable of being supported by a reason.[154] When translated into a technical demand with respect to machine learning, substantive rationality will be satisfied by disclosing merely the outcome variable used in the machine-learning analysis and the objective function whose optimization yields predictions for the outcome.[155] These two pieces of information will reveal the goal of the decisionmaking process and that it was indeed that very goal that the algorithm in fact attempted to achieve, which is all that is needed to satisfy the rational basis test.[156]

2. Procedural Due Process

The strictures of procedural due process, which apply when government makes individual adjudicatory decisions, are more demanding in that they necessitate that the government offer some statement of reasons. Yet, as we explained in Part II, the Supreme Court has made clear that these reasons need not take the form of any formal opinion.[157] Providing the outcome variable and objective function will form part of an adequate statement of reasons. Procedural due process will also require that the government provide information about the accuracy of the algorithm in satisfying the government's objective. This is because procedural due process not only aims to ensure that the government treats individuals procedurally fairly, but also that government procedures are not prone to serious error.[158]

---

154. *See id.*

155. Note that this description applies more to adjudicating by algorithm than regulating by robot. As mentioned earlier, it is only for adjudicating by algorithm that a single algorithm can accomplish the administrative goal. When considering the more complex algorithms that presumably will be needed in most instances of regulating by robot, slightly different disclosures would meet the demands of substantive due process. There, the goal of the analytic system would be the criteria by which the "best" rule is defined and selected. Mathematically, this is likely to be more complex than simply an outcome variable predicted and the objective function optimized to generate predictions, but the underlying point is the same: the only information that has to be revealed to meet substantive due process relates to the goal of the algorithm, not every detail of how it meets that goal.

156. The federal district court in Texas discussed in Part III.A rejected a substantive due process claim even where the government relied on a "highly secretive" algorithm that was "impossible to replicate." Hous. Fed'n of Teachers, Local 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1180–83 (S.D. Tex. 2017) (rejecting plaintiffs' arguments that school district's reliance on secret algorithm violated their right to substantive due process and granting defendant's motion for summary judgment).

157. *See id.*

158. Mathews v. Eldridge, 424 U.S. 319, 346 n.29 (1976) (noting that "in order fully to

With machine learning, the procedural demand for error-correcting information can likely be met by the government disclosing three pieces of information. First, the government should provide affected individuals challenging an algorithmic adjudication with the data—input variables—collected about them to verify that they are accurate.[159]

Second, the government should provide information about how accurate the algorithm is across individuals when evaluated in a test data set. If an algorithm is predicting, for example, whether an individual is disabled, how often does the algorithm get such a prediction wrong? Even if the algorithm is not using any incorrect input information, an algorithm that is simply bad at its job—that is, one that is unacceptably prone to error—could violate procedural due process. That said, because machine learning is used precisely for its ability to make predictions with greater accuracy than other techniques and even humans, this is unlikely to be a barrier in practice.[160]

Finally, the government should disclose the results from verification procedures that certify to individuals their adjudications were the outcome of algorithms being implemented correctly and without glitches. In recent work, Joshua Kroll and his coauthors have outlined a series of verification methods, like cryptographic commitments and zero-knowledge proofs, that can ensure "procedural regularity"—namely, that algorithmic decisions "were made under an announced set of rules consistently applied in each case."[161] By using methods such as these, government officials can verify that predictions resulting from their algorithmic systems are working as intended.[162]

assess the reliability and fairness of a system of procedure, one must also consider the overall rate of error"); Fuentes v. Shevin, 407 U.S. 67, 81 (1972) (describing the purpose of due process as extending beyond procedural fairness but including the desire "to minimize substantively unfair or mistaken deprivations" of protected interests).

159. This is similar to the Fair Credit Reporting Act's reliance on revealing input information as a way to correct potentially erroneous credit scoring. Fair Credit Reporting Act, 15 U.S.C. § 1681 (2018).

160. *See* Coglianese & Lehr, *supra* note 10, at 1158 n.40 and accompanying text.

161. Kroll et al., *supra* note 36, at 637.

162. When the government relies on a physical rather than an algorithmic device—say, a thermometer—it would generally suffice to justify penalizing a person who failed to meet a temperature requirement (e.g., for storing food safely) that a thermometer measures temperature levels accurately and that the specific use of that thermometer to show noncompliance followed proper techniques. When comparable showings can be made for machine learning as for thermometers or other physical machines, there should be no question that they can satisfy the reason-giving demanded of the government under constitutional and statutory law.

Under prevailing principles of procedural due process, adjudication by algorithm will likely satisfy legal requirements without much difficulty.[163] The overarching framework for procedural due process analysis calls for a pragmatic balancing of three factors: (1) the private interests at stake in a governmental decision; (2) the "risk of an erroneous deprivation" of those interests from the decisionmaking process; and (3) the "fiscal and administrative burdens" associated with a particular procedural arrangement.[164] Of these three factors, the first will be obviously exogenous to machine learning, but the second two are undoubtedly affected by machine learning—and almost certainly for the better. The appeal that machine learning holds for government agencies, after all, is precisely that it can facilitate more accurate, automated decisionmaking, resulting in both a reduction in the risk of errors as well as a savings in the time and expense of governmental decisionmaking.[165] Agencies that deploy machine learning in the adjudicatory context will obviously need to validate their use of algorithmic tools to demonstrate their ability to achieve these advantages, and they will need to afford individuals or entities subject to adjudication access to information to ensure that the algorithms were correctly applied.[166] But nothing about machine learning's inscrutability should prevent agencies from making such demonstrations. In the end, "recognizing that machine-learning algorithms have demonstrated superiority over human decisions in other contexts, it is reasonable to conclude that agencies will be able to satisfy the demands of due process even in the machine-learning era."[167]

3. Arbitrary and Capricious Review

Turning to what government agencies must show to satisfy the APA's arbitrary and capricious standard, we can see that here too agencies should be able to meet courts' demand for reason-giving, notwithstanding the ostensibly black-box nature of machine-learning algorithms. It may be referred to

Presumably any responsible agency using a machine-learning algorithm would be able to justify that use by explaining an algorithm's objective function, releasing data used in the individual case, and disclosing the results of validation tests.

163. Coglianese & Lehr, *supra* note 10, at 1184–91.

164. Mathews v. Eldridge, 424 U.S. 319, 335 (1976); *see also* Schweiker v. McClure, 456 U.S. 188, 200 (1982) ("Due process is flexible and calls for such procedural protections as the particular situation demands.").

165. Coglianese & Lehr, *supra* note 10, at 1185–86.

166. *See id.* at 1186–91 (discussing validation and "cross-examination" of adjudicatory algorithms).

167. *Id.* at 1191.

as "hard look" review, but the arbitrary and capricious standard has never required a full explanation of the kind that psychologists, historians, or political scientists might demand if they wanted to understand exactly why government officials reached a decision.[168] If the courts have no need to peer into the minds of government administrators,[169] they presumably should have little worry that they cannot peer into the "minds" of machine-learning algorithms either.

Arbitrary and capricious review applies to any agency action, but it is especially salient in the context of judicial review of rulemaking. What will matter to the courts is that the agency has sufficiently justified its design of and reliance on a particular algorithmic tool. The agency will need to reveal and justify its choice of an outcome variable and objective function. As the selection of an objective function and design of an algorithm necessarily entail making tradeoffs that call for policy judgment,[170] an agency will need to explain its choices about these tradeoffs in terms of factors that are consistent with the agency's statutory authority, and it will need to respond to meaningful public comments submitted during a rulemaking.[171] Agencies will also need to validate that the algorithm performs as intended and that it achieves the justified objectives. The courts will scrutinize agencies' reasoning about these choices and their validation efforts, but, in the end, the legal test is supposed to be "whether there has been a clear error of judgment" in designing and validating an algorithm to achieve a valid purpose—not whether the specific results of a machine-learning algorithm will be intuitively explainable.[172]

168. *See supra* notes 108–109 and accompanying text; *see also* Ethyl Corp. v. EPA, 541 F.2d 1, 97–98 (D.C. Cir. 1976) (explaining that although the courts should educate themselves about the evidence the agency considered in making its decision, in the final analysis, a judge is to "look at the decision not as the chemist, biologist or statistician that [they] are qualified neither by training nor experience to be, but as a reviewing court exercising [their] narrowly defined duty of holding agencies to certain minimal standards of rationality").

169. *See supra* note 103 and accompanying text.

170. *See, e.g.*, Richard Berk et al., Fairness in Criminal Justice Risk Assessments: The State of the Art (May 30, 2017) (unpublished manuscript) (on file with Cornell University Library), https://arxiv.org/pdf/1703.09207.pdf.

171. Citizens to Pres. Overton Park, Inc. v. Volpe, 401 U.S. 402, 416 (1971) (explaining that "the court must consider whether the decision was based on a consideration of the relevant factors"); Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983) (describing as arbitrary and capricious an agency decision "so implausible that it could not be ascribed to a difference in view or the product of agency expertise").

172. *Overton Park*, 401 U.S. at 416.

Federal courts have long demonstrated a tendency to defer to administrative agencies' judgments in cases involving complex mathematical modeling and scientific analysis.[173] Although the arbitrary and capricious test calls for judges to ensure that agencies have carefully considered relevant factors, the Supreme Court has also indicated that, when a government decision "'requires a high degree of technical expertise,' we must defer to 'the informed discretion of the responsible agencies.'"[174] Designing and validating a machine-learning algorithm will certainly require a high level of technical expertise. Furthermore, as the Ninth Circuit Court of Appeals has noted, when a governmental action "involves a great deal of predictive judgment," these "judgments are entitled to particularly deferential review."[175] Enhancing the government's ability to make predictive judgments constitutes the main purpose of designing and validating machine-learning algorithms, so we should expect that judicial deference would be afforded in cases where agencies rely on algorithmic governance.[176]

173. *See State Farm*, 463 U.S. at 43; Balt. Gas & Elec. Co. v. Nat. Res. Def. Council, 462 U.S. 87, 103 (1983).

174. Marsh v. Or. Nat. Res. Council, 490 U.S. 360, 371 (2011) (quoting Kleppe v. Sierra Club, 427 U.S. 390, 412 (1976)).

175. Trout Unlimited v. Lohn, 559 F.3d 946, 959 (9th Cir. 2009).

176. We acknowledge that, in the past, courts have deferred to human decisionmaking by expert agencies, whereas arguably such deference might be less suited to decisions made by outcome-determinative algorithmic systems, especially in cases of rules generated by automated rulemaking systems. We grant that some aspects of current law could well impose certain impediments if agencies rely on automated rulemaking by robot. For example, an agency would need to demonstrate "good cause" if the use of such a system were to bypass the normal notice-and-comment process. *See supra* note 88. Yet, for purposes of arbitrary and capricious review, we see no fundamental impediment when the agency is able to justify adequately the way it has designed and operated its algorithmic system. An algorithmic system is simply a tool selected by the agency to aid agency officials in fulfilling their statutory responsibilities. Even when that tool is an automated rulemaking system, such a system can only be designed in such a manner that comports with human policy and design choices embedded within it. *Cf.* Coglianese & Lehr, *supra* note 10, at 1180–84. The agency rule will simply be one that has nested within it the possibility for a series of subsidiary, automated rule "decisions" made contingent on algorithmic predictions. Granted, the agency rule will thus be much more complicated than otherwise, but conceptually, it will not be fundamentally different than when an agency embeds within its rules other contingencies or requires the use of other tools to measure those contingencies. Just as with the selection of any other tool (e.g., an air quality monitoring device or even a thermometer), agencies will need to justify their choices, but when these choices call for technical expertise, as they necessarily will with machine learning, the applicable legal standard remains a deferential one.

The district court's reasoning in *Alaska v. Lubchenco*[177] is instructive. Although that case involved traditional statistical techniques, the court nevertheless confronted charges that the agency failed to provide adequate reasons to support its forecasting analysis.[178] That analysis predicted a sustained decline in the population of beluga whales in the Cook Inlet of Alaska.[179] On the basis of this forecast, the National Marine Fisheries Service (the Service) designated the whale population in Alaska as an endangered species and therefore subjected Alaska to a whaling ban.[180] The state of Alaska challenged the Service's decision in federal court in Washington, D.C. The court reviewed the agency's decision under the APA's arbitrary and capricious standard—but also against an additional requirement contained in the Endangered Species Act (ESA) that the Service make its determinations "solely on the basis of the best scientific and commercial data available."[181]

The court described in some detail the nature of the statistical model the Service used "to determine the probability of extinction" that then constituted the agency's reason for declaring the beluga whale to be endangered:[182]

> The Service performed extensive testing on the model's sensitivity to these variables by running more than ten thousand individual trials for further analysis. Using statistical methods, the Service then compared models with these different effects to the observed population trend from 1994 to 2008 in order to determine which model best matched the existing data. The model was also peer-reviewed by independent scientists, including researchers from Alaska's own Department of Fish and Game. On the basis of this sensitivity analysis, the Service selected a model that most closely fit the observed population trends. The "most realistic" model predicted a 1 percent risk of extinction in 50 years, a 26 percent risk of extinction in 100 years, and a 70 percent risk of extinction in 300 years. But even under the "base case scenario" or "healthy population" model, there was 29 percent risk of extinction in 300 years. As a measure of confidence in this negative trend, the model estimated that there is only a 5 percent probability that the population growth rate is above 2 percent, while there is at least a 62 percent probability that the population will decline further.[183]

Alaska claimed that the Service's "population model was arbitrarily chosen from among the thousands of trial runs produced by the Service's population

177. 825 F. Supp. 2d 209 (D.D.C. 2011).

178. *See id.* at 213 (describing the modeling technique as a time-series econometric model).

179. *See id.* (explaining the results of the National Marine Fisheries Service's (the Service's) time-series model).

180. *Id.* at 212.

181. 16 U.S.C. § 1533(b)(1)(A) (2018).

182. *Lubchenco*, 825 F. Supp. 2d at 221.

183. *Id.*

viability analysis" and "that the Service gave no explanation for relying on one model—the 'most realistic' one—out of the thirty-one possible models that could result from mixing and matching the independent variables."[184]

The court rejected Alaska's arguments, concluding that the Service had satisfied the relevant legal tests for reason-giving. The court observed that:

> There is no "better" way to assess a species' likelihood of extinction. Plaintiffs do not suggest a more accurate method for estimating the abundance of marine mammals, nor do they point to a superior method of projecting the observed population trend into the future. "If no one propose[s] anything better, then what is available is the best."[185]

Of course, today or in the near future, the use of machine-learning algorithms might well provide that better, or more accurate, way to predict species extinction.[186]

It is telling that the court's assessment of the agency's analysis depended on factors that could easily be met if agencies rely on inscrutable machine-learning algorithms: process (e.g., peer review) and predictive performance (i.e., accuracy). The court did not delve into the inner workings of the statistical models or ask *why* the models performed well. It did not demand any replication or even the submission of the underlying data to the court. Ultimately, the court rejected Alaska's arguments entirely on pragmatic grounds:

> The 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. The standard under the ESA is that "the Service must utilize the 'best scientific . . . data available,' not the best scientific data possible." In this case, plaintiffs do not point to any superior data that the Service should have considered. And the State's own peer reviewer concluded that although the model assumptions "could have been more detailed" or "better discussed," "the assumptions made considering what is known about beluga biology and life history were reasonable." Thus, it ultimately makes no difference that plaintiffs can point to a few shortcomings here and there in the Service's modeling. The agency's population via-

184. *Id.* at 223.

185. *Id.* at 221 (quoting Massachusetts *ex rel.* Div. of Marine Fisheries v. Daley, 170 F.3d 23, 30 (1st Cir. 1999)).

186. *See, e.g.*, Morteza Mashayekhi et al., *A Machine Learning Approach to Investigate the Reasons Behind Species Extinction*, 20 ECOLOGICAL INFORMATICS 58, 66 (2014) (arguing that a machine learning approach to species population analysis "may prove to be beneficial for conservation biologists from the point of view of being able to detect early signals of extinction"); Julian D. Olden et al., *Machine Learning Methods Without Tears: A Primer for Ecologists*, 83 Q. REV. BIOL. 171, 172 (2008) (noting that "a number of [machine learning] techniques have been promoted in ecology as powerful alternatives to traditional modeling approaches").

> bility analysis represents the best available science and is therefore entitled to deference.[187]

Even if Alaska had claimed that another statistical approach would have been superior, the court still probably would have deferred to the agency. When litigation turns into a "'battle of the experts' . . . the courts traditionally reject" the challenger's claims and declare the agency the winner.[188]

The district court's approach in *Alaska v. Lubchenco* is emblematic of courts' more general deferential posture toward agency reason-giving under the APA. Although many courts will scrutinize agencies' reasoning, even when it is based on mathematical or other technical analysis, it is the outlier court that demands much more than the *Lubchenco* court.[189] In most cases, it will likely be enough for government officials to satisfy the arbitrary and capricious test if they can show that (a) an algorithmic system was constructed to advance a legally valid purpose by revealing the goal of an algorithm, (b) it is functioning correctly to advance that purpose (i.e., the program is not malfunctioning and it is producing validated results), and (c) it is being used as intended.[190] Demanding much more would go far beyond any notion of reason-giving demanded of government officials today.

4. Reasoned Transparency Under Conditions of Limited Fishbowl Transparency

Up to this point, we have proceeded under an assumption of full disclosure about all elements of an algorithmic system. Doing so has allowed us to show that, at least in principle, the ostensibly black-box nature of machine-learning algorithms should not by itself impede agencies from providing entirely sufficient explanations that will meet the tests of due process or arbitrary and capricious review. The remaining question now is whether government can meet these tests with anything less than a best-case assumption of total fishbowl transparency.

This question arises in practice because the law does not demand total fishbowl transparency and, in many cases, total transparency will not be able to be provided for justifiable reasons.[191] But in most cases, all the information

187. *Lubchenco*, 825 F. Supp. 2d at 223.

188. McGarity & Wagner, *supra* note 109, at 10,769.

189. *See id.* at 10,759 (noting "several outlier cases . . . evincing very little deference to agencies").

190. Courts may even accept less, such as by failing to insist on validation efforts. *See id.* at 10,768 (noting that "an agency's decision to forego the validation or calibration of models is usually, but not always, respected by the courts").

191. *See* Freedom of Information Act, 5 U.S.C. § 552(b)(1)–(9) (2018) (identifying nine major exemptions from FOIA disclosure).

that must be disclosed to satisfy legal demands of reasoned transparency will be able to be disclosed. Admittedly, as in the Wisconsin and Texas legal actions, which challenged algorithms developed by private firms,[192] whenever government relies on contractors to develop algorithms, some of the contractors' information, such as the underlying source code, may well be subject to protection as a trade secret or confidential business information.[193] But for several reasons, even trade secret protection is unlikely to constitute a major limiting factor in most cases. First, the objective functions and outcome variables that should be disclosed will likely not be properly classified as trade secrets because they represent the *government's* goal; the mathematical form of the objective function will often be dictated, or created, by the government, not by the private party.[194] Second, other essential elements of reason-giving—such as results from testing and validation procedures, and the claimants' own data to ensure accuracy—will in most cases be fully releasable

192. As noted earlier, the Wisconsin Supreme Court upheld the challenged use of an algorithmic risk assessment on procedural due process grounds because the appellant had access to sufficient non-proprietary information. State v. Loomis, 881 N.W.2d 749, 753–54 (Wis. 2016). The federal district court in Texas, on the other hand, held that the proprietary nature of the algorithm, in that case at least, raised a question for a jury as to whether the plaintiffs had available enough information to satisfy demands of procedural due process. Hous. Fed'n of Teachers, Local 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1175 (S.D. Tex. 2017).

193. *See generally* Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018) (discussing how the trade secret status of algorithmic components may affect criminal proceedings).

194. To be clear, there may certainly be instances in which an objective function does contain information that could be considered a trade secret and that goes beyond revealing merely the predictive goal of the algorithm. Particularly, "regularization" methods used to prevent overfitting to, in very rough and broad terms, push the algorithm to make predictions in particular ways can involve adding additional mathematical terms to the objective function. In other words, the objective function may contain more than simply an indication of how it is assessing "accuracy," such as by reference to residuals or errors. *Cf.* Lehr & Ohm, *supra* note 2, at 704–05 (discussing the use of regularization to reduce an algorithm's disparate impacts). The mathematical additions to objective functions could perhaps be deemed trade secrets if they were developed by private companies. But this is unlikely to be problematic for two reasons. First, those additions could always be excerpted before disclosure, leaving only the non-confidential parts of the objective function to be disclosed; this would still indicate whether the goal pursued by the algorithm is legitimate—the purpose of disclosing the objective function in the first place. Second, as we discuss, in camera review is always an option.

without needing to divulge any protected source code or other trade secrets.[195] Third, even if the proprietary nature of a private contractor's underlying source code does bar the government from disclosing necessary information, that information can always be reviewed by a court in camera, thus protecting any trade secrets or confidential business information. In the rulemaking context, independent peer reviews could be conducted under non-disclosure agreements. Finally, as noted in Section III.A, agencies could proactively work to avoid any conflict between trade secret protection and reasoned transparency simply by drafting contracts with their private consultants to clarify what information must be treated confidentially and what information may be disclosed.[196]

In most instances, we would expect the fishbowl transparency issues to be worked out without much difficulty because of the pragmatic nature of transparency law's demands. We note as well that we have been assuming all along that the agency will be using machine-learning systems in outcome-determinative ways. Where this is not the case, the agency will presumably be able to provide an alternative explanation that does not require the disclosure of any confidential information. The important takeaway is that in most cases the demands of reasoned transparency will be able to be met while still respecting authorized limitations on fishbowl transparency.[197]

C. Technical Advances in Algorithmic Transparency

The previous Section showed that meeting the demands of reasoned transparency, even under circumstances of limited fishbowl transparency, should not be difficult for responsible government officials to meet. This alone

195. Furthermore, as Joshua Kroll and his coauthors have pointed out, the source code may not reveal anything more about the functioning of a machine-learning algorithm. *See* Kroll et al., *supra* note 36, at 638.

196. *See supra* notes 120–137 and accompanying text.

197. Our focus throughout this article is on federal transparency law in the United States as it applies to the actions of government agencies. We note that, at a more general level, the analysis we provide here might accord with how the EU's "right to explanation" may eventually be understood under the General Data Protection Regulation (GDPR). *See generally* EU General Data Protection Regulation, *supra* note 9. At least one early assessment suggests that private organizations using machine learning may simply need to provide "a basic explanation of how the model is working" to satisfy the GDPR's requirement of "meaningful information" about the "logic" and "significance" of the automated algorithmic system. Andrew Burt, *Is There a 'Right to Explanation' for Machine Learning in the GDPR?*, IAPP: PRIVACY TECH (June 1, 2017), https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/#.

should assuage any concern that fundamental principles of transparency enshrined in law might be threatened by governmental use of so-called "black box" algorithms. But we now take our analysis one step further, showing how machine-learning analyses are growing less inscrutable due to technical advances. Over the last few years, there has been expanding interest—among members of both the legal and technical communities—in improving the interpretability of algorithms in ways that go well beyond what is legally demanded.[198] Research is accumulating that details technical methods for improving the ability to explain the inner workings of machine learning in more intuitive ways.[199] For the reasons we have discussed above, these technical developments are not essential for satisfying the legal requirement of

198. To the extent that legal considerations have helped motivate some of this interest, we suspect that a driving force has been the EU's GDPR with its ambiguous right to explanation. *See generally* EU General Data Protection Regulation, *supra* note 9, § 71. The GDPR applies to private-sector uses of machine learning, and uncertainty over what counts as a sufficient explanation for purposes of the new EU regulation presumably has produced significant enough stakes to attract the attention of both researchers and market actors. Of course, researchers have long attempted to get reasons, of sorts, out of machine-learning algorithms. Indeed, some of the basic methods that we describe in this section were developed earlier in the 2000s. *See, e.g.*, RICHARD A. BERK, STATISTICAL LEARNING FROM A REGRESSION PERSPECTIVE 226–29, 277–92 (1st ed. 2008) (describing partial dependence plots). But these early methods are limited in two ways. First, as a technical matter, they are limited in applicability; as we will discuss, for some particularly advanced machine-learning techniques in use today, analogues to these basic methods either have not yet been developed or have been developed only more recently. Second, these methods do not serve the same purpose, and were not developed in response to the same pressures, as the cutting-edge techniques we discuss later in this Section. Some early machine-learning methods, like random forests, were initially applied to social science problems, and it is social scientists' objective to attempt to model phenomena. As a result, the goal of many reason-giving methods was to give social scientists tools for telling descriptive stories about what processes could be generating the data they observed. That goal contrasts with the goal motivating development of many cutting-edge techniques we discuss later—reducing the opacity of algorithms whose applications in sensitive contexts mandates a certain level of reason-giving. This goal emerged from a host of primarily legal and policy scholars who critiqued applications of machine learning for being too opaque. *See* Lehr & Ohm, *supra* note 2, at 658–64. This concern has in turn sparked innovation from technical scholars, often working in tandem with the legal scholars. *See, e.g.*, *ACM Conference on Fairness, Accountability, and Transparency (ACM FAT)*, ACM FAT* CONF., https://fatconference.org/ (last visited Jan. 22, 2019); *Fairness, Accountability, and Transparency in Machine Learning*, FAT/ML, http://www.fatml.org/ (last visited Jan. 22, 2019).

199. *See* Andrew D. Selbst, *A Mild Defense of Our New Machine Overlords*, 70 VAND. L. REV. EN BANC 87 ("Black boxes can generally be tested, and the relationship between inputs and outputs is often knowable, even if one cannot describe succinctly how all potential inputs map

reasoned transparency; however, their emergence reinforces optimism that algorithmic governance methods can be deployed in ways that are sufficiently transparent, both to satisfy existing legal demands as well as to fulfill potentially broader aspirations of good government.

Our concern with transparency here centers on humans' analytic ability to understand how an algorithm's internal processes ostensibly produce predictions. This contrasts with understanding whether those processes are in fact what are producing the predictions—that is, information that verifies its functioning. The former describes how the internal math of an algorithm operates to yield predictions, while the latter—verification—refers to issues already discussed in Section III.B concerning whether it is indeed that math that yields predictions instead of computer glitches or faulty programming.

To explain what reason-giving methods exist and how they elucidate algorithmic functioning, let us return to an example we first presented in Section I.A of an algorithm hypothetically deployed by the FAA to determine pilot certification. Suppose that a candidate pilot was predicted to be unworthy of certification. What kinds of reasons could be given for why the pilot was so predicted? We can distinguish between the reasons about an algorithm's operations at an *individual* level and at a *group* level.[200] The former refers to being able to understand why a particular prediction or estimation resulted. In this case, an individual-level explanation would provide understanding of what aspects about the specific candidate led to the prediction in

to outputs. To say that something is a black box is not to say we can understand nothing about it.") Note that the methods we discuss in this Section are what Selbst and Barocas would refer to in their work as post hoc methods—those that do not place any constraints on how the algorithm is initially specified. *See* Selbst & Barocas, *supra* note 151, at 34–35. By contrast, they also contemplate opportunities to increase the scrutability of algorithms by restricting the algorithm's complexity—by, for example, limiting the number of input variables or choosing (and appropriately tuning) algorithms that are, by their nature, less complex than others. We do not address these because, as we mention at the outset, many analysts as well as government officials will properly welcome complexity; it is this complexity that enables machine learning's prowess. Furthermore, given our discussion of the level of reasoned transparency necessitated by the law, administrative uses of machine learning should face no significant legal demands to be less complex. If anything, the courts have indicated that they will give greater deference to agencies under the Administrative Procedure Act when issues are complex. *See* Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983); Balt. Gas & Elec. Co. v. Nat. Res. Def. Council, 462 U.S. 87, 103 (1983).

200. Individual-level explanations are also sometimes referred to in the technical literature as *local* explanations, and group-level explanations are referred to as *global* explanations. *See, e.g.*, Riccardo Guidotti et al., *A Survey of Methods for Explaining Black Box Models*, 105 ACM COMPUTING SURVS. 93:1, 93:6 (2018).

his case. A group-level explanation, on the other hand, does not try to reveal why a particular entity's prediction resulted; rather, it reveals patterns or trends in the factors affecting predictions across entities. It might reveal, say, what features of pilots tend to lead to predictions of flight worthiness (or non-worthiness) across all pilots examined by the algorithm.

Within both individual- and group-level explanations, we can also distinguish between methods explaining the importance of input variables and those explaining the functional forms of input variables.[201] The former attempts to reveal, roughly speaking, the magnitude of the effect of a given input variable—say, the applicant's age—on the ability of the algorithm to make successful forecasts. A method aimed at explaining importance will essentially seek to measure how much that variable matters relative either to that variable not being considered at all by the algorithm (i.e., the algorithm dropping age from the analysis) or to having the values the variable takes on randomized (i.e., stripping the age variable of any predictively useful information it may contain).[202] When implemented on a group level, measures of importance are often interpreted as reductions in overall accuracy across all individuals examined—for example, if the variable about the applicant's age were removed from consideration or randomly shuffled, the algorithm would make a certain percentage more errors across all candidates when predicting that they are not worthy of certification.[203] Technical methods for achieving this kind of group-level meaning about variable importance have existed for a while for less complex machine-learning algorithms, but have recently started to be developed or refined for more complex ones.[204]

When importance methods are implemented on an individual level, they take on a slightly different interpretation. Because they do not operate across multiple individuals, they cannot be interpreted as percentage reductions in accuracy when a variable is dropped or randomly shuffled. Rather, they are often interpreted as indicating how "pivotal" an input variable is to an individual's prediction. In other words, they indicate how likely it is that, if an

201. In the regression literature, input variables are also referred to as independent variables, while output variables are referred to as dependent variables.

202. *See* Lehr & Ohm, *supra* note 2, at 679–81.

203. *See id.*

204. *See, e.g.*, Anupam Datta et al., *Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems*, 2016 IEEE SYMP. ON SECURITY & PRIVACY 598, 601, 608-09; Marina M.-C. Vidovic et al., Feature Importance Measure for Non-Linear Learning Algorithms (Nov. 22, 2016) (unpublished conference paper) (on file with Cornell University Library), https://arxiv.org/pdf/1611.07567.pdf (analyzing different methods for teaching complex machine-learning algorithms).

input variable, like the age of the candidate, were dropped or randomly shuffled, a candidate's predicted flight worthiness would change. If dropping that variable or randomly shuffling its value has no effect on a candidate's ultimate prediction, then one could say that the variable is not "important" or "pivotal." Analytic methods that accomplish this kind of individual-level explanation of importance are actively being developed.[205]

In addition to methods aimed at explaining the importance of an input variable, technical methods are being refined to explain input variables' *functional forms*, adding a different kind of meaning. The importance methods discussed above do not give any indication of the direction or manner in which the input variable about an applicant's age affects the predicted outcome; they do not say, for instance, that an increase in a pilot's age tends to be associated with an increase in the predicted outcome variable (or a higher probability, if the outcome variable is a binary prediction). That is what methods explaining functional forms attempt to explain.[206] To explain functional form, however, these methods produce, as a practical matter, a different form of output. The importance of an input variable can be indicated by a number indicating, say, the percent increase in error (for a group-level explanation) or the probability of a different prediction (for an individual-level explanation). By contrast, the functional form of a variable is often revealed on a graphical plot. Such plots indicate, roughly speaking, the effect that increases or decreases in a given input variable, like the applicant's age, have on the outcome variable, holding the values of all other input variables constant. Notably, unlike importance measures, explanations of functional forms are available only on the group level, not the individual level. Furthermore, while group-level explanations of functional form have been available for some time for less complex machine-learning methods, they have only started to be developed for more complex algorithms, like various forms of deep learning.[207]

205. *See, e.g.*, Grégoire Montavon et al., *Explaining NonLinear Classification Decisions with Deep Taylor Decomposition*, 65 PATTERN RECOGNITION 211 (2017) (focusing data analysis on individual data points); Wojciech Samek et al., *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*, ITU J., Oct. 2017 (analyzing individual predictions in machine leaning and artificial intelligence); Wojciech Samek et al., *Interpreting the Predictions of Complex ML Models by Layer-wise Relevance Propagation*, *in* 9887 LNCS: ARTIFICIAL NEURAL NETWORKS AND MACHINE LEARNING – ICANN 2016 (Alessandro E.P. Villa, Paolo Masulli, and Antonio Javier Pons Rivero, eds., 2016) (summarizing a technique that explains predictions in machine learning); Vidovic et al., *supra* note 204.

206. *See* Coglianese & Lehr, *supra* note 10, at 1212; Lehr & Ohm, *supra* note 2, at 709–10.

207. *See* Lehr & Ohm, *supra* note 2, at 709–10; *see, e.g.*, Marco Tulio Ribeiro et al., Model-

In our example, a group-level explanation of importance might reveal to the candidate that, if an input variable indicating the applicant's age were omitted from the algorithm, the algorithm would make thirty percent more errors across individuals when predicting that candidates are not worthy of certification. Taken alone, and despite it being a group-level explanation, this could suggest to the candidate that she was predicted as not being flight-worthy because of her age. But an individual-level explanation of importance could reveal a contradictory story. It might reveal, for instance, that when the relevant input variables were randomly shuffled, there was only a five percent chance of the particular candidate's negative prediction changing. In other words, the candidate's age probably did not have a bearing on her ultimate prediction. It should be clear that, in such a circumstance, the individual-level explanation will typically be of far greater value for understanding how the algorithm functioned.

Turning to an explanation of functional form, suppose that the individual-level explanation of importance indicated that the age variable was important to the candidate's negative prediction. That alone does not reveal how the candidate's age affected her outcome—just that it did. A group-level plot of functional form could elucidate this. It could reveal that, holding all other things constant, pilots who are older are more likely to be deemed worthy of certification. If the candidate were younger, this group-level explanation could suggest that the applicant's denial was negatively affected by her age. But, of course, this may not necessarily be true because this is a group-level explanation. It could have been the case that, for this particular candidate's prediction, given other attributes about her, the algorithm actually found that the younger age made her more worthy of certification.

Admittedly, there is much work still underway in designing, using, and understanding machine-learning algorithms. Much work that has been completed to date on importance and functional form has yet to undergo rigorous testing or evaluation by the statistical community. But the rapid pace at which these methods are being refined in response to growing interest in algorithmic transparency provides a basis for even greater optimism for the future reliance on algorithmic governance. The sophistication and refinement of these methods will only continue to grow, and they are already sufficiently developed to support confidence in the explainability of machine-learning algorithms. Even when algorithms are applied to make individual-level predictions—often, at least at first blush, some of the most visceral of

Agnostic Interpretability of Machine Learning (Jun. 16, 2016) (unpublished conference paper) (on file with Cornell University Library), https://arxiv.org/pdf/1606.05386.pdf (arguing for explaining machine learning predictions using model-agnostic approaches).

applications, such as predictive policing—government agencies will likely have strategies available to them to provide rather detailed individual-level explanations.

CONCLUSION

If machine-learning applications are designed and managed well, governmental reliance on them should be able to withstand legal challenges based on principles of reason-giving. The fact that the use of such algorithms can satisfy legal demands for transparency does not mean that algorithms should immediately be deployed in any particularly administrative domain. But it does mean that scholars, policymakers, and the public ought to be receptive to the use of machine-learning algorithms where they can improve public administration. The responsible use of algorithms—even in outcome-determinative ways—will not contravene legal principles of transparency. Although the potential for outcome-determinative uses of machine learning by governments loom on the horizon, algorithms will likely be applied more often to assist, rather than supplant, human judgment. If, as we have shown, even outcome-determinative applications of machine learning can meet the law's demands, then there should be even less concern over the less determinative uses.

Of course, there is always something to be said for promoting transparency even beyond what agencies must do to withstand judicial scrutiny of their reasoning.[208] After all, to have trust and confidence in their government, citizens may well hold more demanding expectations for meaningful information than do the courts. We are thus heartened that data scientists are already finding ways to do more than is required to coax explanatory information out of ostensibly black-box algorithms. We are not only confident that governments will be able to meet demands for explainability in their use of algorithms, especially under prevailing legal standards, but we are cautiously optimistic that algorithmic governance might in important circumstances even enhance public trust and legitimacy in government. In some cases, well-designed algorithms may increase public trust by achieving

208. *See* CARY COGLIANESE, LISTENING, LEARNING, LEADING: A FRAMEWORK FOR REGULATORY EXCELLENCE 5–6 (2015). We have emphasized the transparency of algorithms in this Article because it is a critically important consideration in a decision to use them in public-sector applications. But it is certainly not the only consideration. Algorithmic governance also implicates other values—fairness, equality, privacy, efficiency—that should be considered on a case-by-case basis when officials contemplate a move to machine learning.

faster and fairer outcomes for those individuals who interact with government agencies and are subject to their decisions.[209]

In the future, a government that makes use of so-called black-box algorithms need not be a black-box government. With responsible practices, government officials can take advantage of machine learning's predictive prowess while remaining faithful to principles of open government. Algorithmic governance can meet the law's demands for transparency while still enhancing efficacy, efficiency, and even legitimacy in government.

209. Earlier we noted that eBay has had a successful experience in relying on an automated dispute resolution process to settle tens of millions of disputes each year. Strikingly, eBay has found that "parties who engaged in the process were more likely to return and purchase other items through eBay, a pretty remarkable result from settling a dispute." BARTON & BIBAS, *supra* note 27, at 113.