PROCUREMENT AS POLICY: ADMINISTRATIVE PROCESS FOR MACHINE LEARNING

Deirdre K. Mulligan[†] *& Kenneth A. Bamberger*[††]

ABSTRACT

At every level of government, officials contract for technical systems that employ machine learning—systems that perform tasks without using explicit instructions, relying on patterns and inference instead. These systems frequently displace discretion previously exercised by policymakers or individual front-end government employees with an opaque logic that bears no resemblance to the reasoning processes of agency personnel. However, because agencies acquire these systems through government procurement processes, they and the public have little input into—or even knowledge about—their design or how well that design aligns with public goals and values.

This Article explains the ways that the decisions about goals, values, risk, and certainty, along with the elimination of case-by-case discretion, inherent in machine-learning system design create policies—not just once when they are designed, but over time as they adapt and change. When the adoption of these systems is governed by procurement, the policies they embed receive little or no agency or outside expertise beyond that provided by the vendor. Design decisions are left to private third-party developers. There is no public participation, no reasoned deliberation, and no factual record, which abdicates Government responsibility for policymaking.

This Article then argues for a move from a procurement mindset to policymaking mindset. When policy decisions are made through system design, processes suitable for substantive administrative determinations should be used: processes that foster deliberation reflecting both technocratic demands for reason and rationality informed by expertise, and democratic demands for public participation and political accountability. Specifically, the Article proposes administrative law as the framework to guide the adoption of machine learning governance, describing specific ways that the policy choices embedded in machine-learning system design fail the prohibition against arbitrary and capricious agency actions

DOI: https://doi.org/10.15779/Z38RN30793

[†] Associate Professor, School of Information, University of California, Berkeley; Faculty Director, Berkeley Center for Law and Technology.

[††] The Rosalinde and Arthur Gilbert Foundation Professor of Law, University of California, Berkeley; Faculty Director, Berkeley Center for Law and Technology. Much appreciation to Nitin Kohli for his expert input and close reading, Margot Kaminski for her detailed comments on an earlier draft, participants at the ACM FAT* conference, members of the UC Berkeley Algorithmic Fairness and Opacity working group, and participants in the Simons Institute for the Theory of Computer Science Summer 2019 Cluster on Fairness for insightful and helpful comments and discussions; to the Berkeley Center for Law and Technology for its support of this project; and to Sanjana Parikh and Miranda Rutherford for their superb editing and research contributions to this Article. Research for this Article has been supported by generous funding from the US NSF INSPIRE SES1537324.

absent a reasoned decision-making process that both enlists the expertise necessary for reasoned deliberation about, and justification for, such choices, and makes visible the political choices being made.

Finally, this Article sketches models for machine-learning adoption processes that satisfy the prohibition against arbitrary and capricious agency actions. It explores processes by which agencies might garner technical expertise and overcome problems of system opacity, satisfying administrative law's technocratic demand for reasoned expert deliberation. It further proposes both institutional and engineering design solutions to the challenge of policymaking opacity, offering process paradigms to ensure the "political visibility" required for public input and political oversight. In doing so, it also proposes the importance of using "contestable design"—design that exposes value-laden features and parameters and provides for iterative human involvement in system evolution and deployment. Together, these institutional and design approaches further both administrative law's technocratic and democratic mandates.

I. INTRODUCTION

The U.S. Solicitor General's 2017 arguments opposing Supreme Court review of *Loomis v. Wisconsin*,[1] a case presenting the constitutionality of the use of risk assessment software—software that uses statistical models to predict the likelihood of an individual failing to appear at trial or engaging in future criminal activity—in sentencing, may have prevailed in convincing the Justices to deny the petition for certiorari.[2] The Solicitor General conceded that one of the issues raised in the case—"the extent to which actuarial assessments considered at sentencing" may take gender into account—"is a serious constitutional question."[3] Yet he argued that Mr. Loomis's challenge to the use of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system used by the State of Wisconsin in

[1] *See* Brief for the United States as Amicus Curiae, Loomis v. Wisconsin, 138 S. Ct. 2290 (2017) (No. 16-6387), https://www.scotusblog.com/wp-content/uploads/2017/05/16-6387-CVSG-Loomis-AC-Pet.pdf [https://perma.cc/L98E-8AVH]; *see also* State v. Loomis, 881 N.W.2d 749 (Wis. 2016). The Wisconsin Supreme Court case generated petition for *certiorari*. *Id.*

[2] *See Order List: 582 U.S.*, SUP. CT. U.S. 5 (June 26, 2017), https://www.supremecourt.gov/orders/courtorders/062617zor_8759.pdf [https://perma.cc/X85J-PGRK].

[3] Brief for the United States, *supra* note 1, at 19.

sentencing was "not a suitable vehicle" for Supreme Court review because "it is unclear *how* COMPAS accounts for gender."[4]

Yet, however persuasive this argument might have been in the context of Supreme Court case management, the implications of this concession are shocking as a matter of policy. At no time during the challenge, which was appealed all the way to the Wisconsin Supreme Court, could the courts even determine how constitutionally relevant variables were used in the system's analysis.[5] More significantly, it is unclear whether the government ever deliberated about—or was even fully aware of—the way gender was used during the procurement of this system, or its application in the sentencing over thousands of cases.[6] The state asserted that it used "the same COMPAS risk assessment on both men and women, but then compares each offender to a 'norming' group of his or her own gender."[7] In the end, however, all evidence suggests that the State of Wisconsin left the decision of how gender was to be used at the discretion of the software vendor.

[4] *Id.*

[5] This is particularly striking because regardless of how gender is used, the decision would not constitute a trivial detail, as under the Due Process Clause, a sentencing court may not consider as "aggravating" factors characteristics of the defendant "that are constitutionally impermissible or totally irrelevant to the sentencing process, such as for example race, religion, or political affiliation." Zant v. Stephens, 462 U.S. 862, 885 (1983). The Supreme Court of Wisconsin too prohibits the use of gender as a sentencing factor. *See* State v. Harris, 786 N.W.2d 409, 416 (Wis. 2010).

[6] The court record does not document any evidence of such deliberation, and we could find no evidence of such deliberation elsewhere. In fact, there are indications that the state had not even adopted high level guidelines for the design of tools. SUZANNE TALLARICO ET AL., NAT'L CTR. FOR STATE COURTS, EFFECTIVE JUSTICE STRATEGIES IN WISCONSIN: A REPORT OF FINDINGS AND RECOMMENDATIONS, 122 (2012), https://www.wicourts.gov/courts/programs/docs/ejsreport.pdf [https://perma.cc/L78K-VSRT] (suggesting that draft standards developed by a national coordinating network, which require risk tools to be "equivalently predictive for racial, ethnic and gender sub-groups represented in the Drug Court population," "*could* serve as a model for standards *should the state of Wisconsin wish to develop them*") (emphasis added). It is, moreover, difficult to assess what courts are doing to consider the embedded policies in these tools, even with substantial effort. *See generally* Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103, 137–38 (2018) (reporting that only one of sixteen courts provided any information about a risk assessment tool (not COMPAS) in response to public records acts, with most claiming to be exempt).

[7] *Loomis*, 881 N.W.2d at 765. The Practitioner's Guide provided by Northpointe does not mention norming. Wisconsin may be referring to either what Northpointe calls "normative subgroups," which include (1) male prison/parole, (2) male jail, (3) male probation, (4) male composite, (5) female prison/parole, (6) female jail, (7) female probation, and (8) female composite. *Practitioner's Guide to COMPAS Core*, NORTHPOINTE 1, 11–12 (Mar. 19, 2015), http://www.northpointeinc.com/files/technical_documents/Practitioners-Guide-COMPAS-Core-_031915.pdf [https://perma.cc/775A-6GMH].

While deeply troubling, this phenomenon is widespread. At every level of government, officials purchase, or contract for use of, technology systems that employ machine learning—systems that perform tasks without using explicit instructions, relying on patterns and inference instead. These systems frequently displace discretion previously held by either policymakers charged with ordering that discretion, or individual front-end government employees on whose judgment governments previously relied, with an opaque logic that bears no resemblance to the bounded and rational reasoning processes of agency personnel, but rather by patterns that machines induce by observing human actions.[8]

However, research reveals that government agencies purchasing and using these systems most often have no input into—or even knowledge about—their design or how well that design aligns with public goals and values. They know nothing about the ways that the system models the phenomena it seeks to predict, the selection and curation of training data, or the use of that data—including (as in the *Loomis* case) whether and how to use data that relate to membership in a protected class. And agencies have no input into the system's analytic technique, treatment of risk or uncertainty, preferences for false positives or false negatives, or confidence thresholds. In short, governments play no role in setting important policy.

Indeed, in a recent study by Robert Brauneis and Ellen Goodman involving open records requests seeking information about six algorithmic programs used by forty-two different agencies in twenty-three states, only one jurisdiction provided the algorithm and details about its development.[9] In most instances, by contrast, agency documents revealed that they did not have access to the algorithm, the model's design, or the processes through which the algorithm was generated or adjusted.[10] Indeed, most government bodies did not even have a "record of what problems the models were supposed to address, and what the metrics of success were."[11]

Algorithmic systems generally, and those that design and sell them, are increasingly subject to criticism for inattention to context and culture, the values baked into their design, and the biases they embed.[12] Yet government

[8] *See infra* Section III.B.1 (discussing decision making by machine learning systems).

[9] Brauneis & Goodman, *supra* note 6, at 137 ("[O]nly one of the jurisdictions, Allegheny County, was able to furnish both the actual predictive algorithms it used (including a complete list of factors and the weight each factor is given) and substantial detail about how they were developed.").

[10] *Id.*

[11] *Id.* at 152.

[12] *See* Mary Flanagan et al., *Embodying Values in Technology: Theory and Practice*, *in* INFORMATION TECHNOLOGY AND MORAL PHILOSOPHY 322, 322–47 (Jeroen van den Hoven & John Weckert eds., 2008) (arguing that technology can embody values by design

agencies seeking to automate tasks left to their discretion seem persistently tone deaf to the need for greater agency and public participation in shaping technology systems. Across the country there is a smattering of public efforts to assess the policies embedded in algorithmic systems, but these are exceptions. A January 2019 Request for Proposal (RFP) issued by the Program Support Center of the U.S. Department of Health and Human Services sought a contractor who could in turn coordinate the procurement of Intelligent Automation/Artificial Intelligence (IAAI) on behalf of a range of agencies.[13] In the words of the proposal, "[t]his contract is the next logical step to integrating IAAI technologies into all phases of government operations."[14] This RFP reflects the dominant mindset of agencies: It positions machine learning systems as machinery used to support some well-defined function, rather than new methods of arranging how an institution makes sense of and executes on its mission, which is often tied to an empiricist epistemology where prediction, rather than causation, is a sufficient justification for action.[15]

The marked absence of a public sector culture of algorithmic responsibility reflects a "procurement" mindset that is deeply embedded in the law of public administration. Technology systems are acquired from third-party vendors with whom government agencies enter into contracts for goods or services. Public procurement is governed by an extensive body of regulation intended to promote certain bureaucratic values—including price,

and developing a framework for identifying moral and political values in such technology); Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-By-Design*, 106 CALIF. L. REV. 697, 708–13 (2018) (discussing the science and technology studies as well as computer science and legal literatures on "Values in Design"); Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169, 169–85 (2000) (discussing biases in the creation of search indexes and search results); James H. Moor, *What is Computer Ethics?*, 16 METAPHILOSOPHY 266, 266–75 (1985) (discussing the ethical implications of invisible abuse, emergent bias due to designers' values, and bias rooted in complexity within computer systems).

[13] *See* Aaron Boyd, *HHS Contract Will Offer AI Tech, Support to All of Government*, NEXTGOV.COM (Jan. 10, 2019), https://www.nextgov.com/emerging-tech/2019/01/hhs-contract-will-offer-ai-tech-support-all-government/154078/ [https://perma.cc/W8NH-CYHY].

[14] *Solicitation/Contract/Order for Commercial Items: Solicitation Number 19-233-SOL-00098*, U.S. DEP'T HEALTH & HUM. SERVS. 9 (Jan. 10, 2019) https://www.fbo.gov/utils/view?id=39d0a0ce8bfe09391b9fee07833274de [https://perma.cc/6DEC-L5WQ] [hereinafter *Solicitation Number 19-233-SOL-00098*].

[15] Rob Kitchin, *Big Data, New Epistemologies and Paradigm Shifts*, BIG DATA & SOC'Y 3–5 (2014), https://doi.org/10.1177/2053951714528481 [https://perma.cc/3N7Q-3LYG] (describing and critiquing Big Data "empiricism, wherein the volume of data, accompanied by techniques that can reveal their inherent truth, enables data to speak for themselves free of theory").

fairness in the bidding process, innovation, and competition[16]—and elaborates methods of challenging contracting decisions on these elements. This body of regulation generally limits standing to challenge contracting decisions to jilted commercial competitors. Both public contracting and decision making about agency management are largely exempted from administrative procedures that govern decisions of policy[17]—procedures intended to promote a different set of public values: substantive expertise, transparency, participation and political oversight, and reasoned decision making. Thus, current agency perception and practice leave the policies that algorithms embed obscured, unarticulated, and unvetted.

This Article makes the case that because choices in the design, adoption, and use of machine learning systems often make substantive policy, design, adoption, and use should be approached with a different mindset—a "policymaking" mindset—and should reflect the frameworks for legitimate policymaking embodied in administrative law.

Designing algorithmic and machine learning systems involves decisions about goals, values, risk and certainty, and a choice to place constraints on future agency discretion. If these systems employ adaptive machine learning capabilities, their design choices make policy—not just once when they are designed, but over time as they adapt and change. When the adoption of those systems is governed by procurement, the policies they embed receive little or no agency or outside expertise beyond that provided by the vendor: no public participation, no reasoned deliberation, and no factual record. Design decisions are left to private third-party developers. Government responsibility for policymaking is abdicated.

An important body of scholarship has explored the possibilities and shortcomings inherent in algorithmic systems,[18] suggested ways in which

[16] *See generally* Steven L. Schooner, *Desiderata: Objectives for a System of Government Contract Law*, 11 PUB. PROCUREMENT L. REV. 103 (2002) (summarizing nine goals identified for government procurement systems: competition, integrity, transparency, efficiency, customer satisfaction, best value, wealth distribution, risk avoidance, and uniformity).

[17] *See, e.g.*, 5 U.S.C. §§ 553(a)(2)–(3) (2012) (containing the Administrative Procedure Act's exemption of matters relating to "agency management" or to "public property, loans, grants, benefits, or contracts" from the section's general requirements of notice-and-comment rulemaking).

[18] *See, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015) [hereinafter BLACK BOX]; Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1 (2018); Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 724 (2010); Peter A. Winn, *Judicial Information Management in an Electronic Age: Old Standards, New Challenges*, 3 FED. CTS. L. REV. 135 (2009); Guy Stuart, *Databases, Felons, and Voting: Bias and Partisanship of the Florida Felons List in the 2000 Elections*, 119 POL. SCI. Q. 453 (2004); Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. (Apr. 1, 2013), https://hbr.org/

individual government determinations based on algorithmic systems might be challenged,[19] and proposed methods for increasing transparency and accountability.[20] Fewer researchers have extended these insights to accommodate the pressing challenges of machine learning,[21] and even fewer have explored what moving technology systems acquisition and design from a "procurement" mindset to a "policymaking" mindset would mean in terms of technical design, administrative process, participation, and deliberation.[22]

This Article begins to fill that gap. It argues that, in contexts in which policy decisions are likely to be made through procurement, process suitable

2013/04/the-hidden-biases-in-big-data [https://perma.cc/MH6U-28M2]; Julia Angwin et al., *Machine Bias: There's Software Used Across the County 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 [https://perma.cc/WK73-BW9S].

[19] Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1252 (2008).

[20] *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 633 (2017) (suggesting a "technological toolkit to verify that automated decisions comply with key standards of legal fairness"); Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. TELECOMM. & HIGH TECH. L. 235, 235–36 (2011); Katherine Fink, *Opening the Government's Black Boxes: Freedom of Information and Algorithmic Accountability*, INFO., COMM. & SOC'Y 1–19 (May 30, 2017), https://doi.org/10.1080/1369118X.2017.1330418 [https://perma.cc/ATP4-KRZ8] (reviewing current state of law and practice with respect to whether algorithms would be considered "records" under the Freedom of Information Act and reviewing agency bases for withholding algorithms and source code under FOIA requests); *see also* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 121 (2019) (arguing that recently introduced provisions protecting employees against trade secret actions could immunize whistleblowers policing algorithms from within firms).

[21] *See, e.g.*, Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1 (2019); Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147 (2017); Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399 (2017).

[22] Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 6, 26–30 (2019) (proposing a regulatory toolkit to govern the use of algorithms in the private sector, including "substantive rulemaking mechanisms, such as the use of safe harbors and private sector codes of conduct, and accountability mechanisms, such as the use of oversight boards and audits"); Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 110 (2017) (calling for "criminal justice expertise and political process accountability" to be brought into the design of recidivism risk tools); Andrew D. Selbst, *Disparate impact in big data policing*, 52 GA. L. REV. 109, 109 (2017) (recommending police be required to complete "algorithmic impact statements" before adopting predictive policing technology); Catherine Crump, *Surveillance Policy Making By Procurement*, 90 WASH. L. REV. 1595 (2016) (proposing steps to strengthen democratic input); Dillon Reisman et al., *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, AI NOW INST. (Apr. 2018), https://ainowinstitute.org/aiareport2018.pdf [https://perma.cc/N6W6-JRHQ]. Danielle Citron's work examining a prior generation of expert systems has provided foundational analysis for thinking about ways that administrative law concerns about delegation and process might be translated to the technological context. Citron, *supra* note 19, at 1252.

for substantive administrative determinations should be used: process ensuring the type of deliberation that safeguards fundamental administrative law values. Such processes must satisfy administrative law's *technocratic* demands that policy decisions be the product of reasoned justifications informed by expertise—elements grounded in the rule of law.[23] And they must reflect *democratic* requirements of public involvement and political accountability. The Article thus makes the case that the policies designed into machine learning systems adopted by government agencies must be surfaced and deliberated about through new processes and brought fully within an administrative law mindset. Governance through technology cannot be allowed to quietly route around the processes that ground agency action's legitimacy.

Part II describes the ways that the integration of machine learning into governance has been viewed as a matter of procurement and the failures of that approach. Government agencies have relied on private vendors for the design of algorithmic systems, largely exacerbating the challenges of governing through technology by abdicating government's role in shaping important design choices. It then explores five examples of ways in which system design embeds policy decisions to make the case that machine-learning system adoption should often instead be understood as policymaking.

Part III examines administrative law as an alternative framework for the adoption of machine learning in governance. Describing the specific ways in which machine learning systems displace administrative discretion and human logic, this Part argues that the policy choices embedded in system design fail the prohibition against arbitrary and capricious agency actions absent a reasoned decision-making process that enlists the expertise necessary for reasoned deliberation, provides justifications for such choices, makes visible the political choices being made, and permits iterative human oversight and input. This Part focuses on changing the system-adoption process, arguing that design choices should occur through a decision-making process that reflects the technocratic and democratic goals of administrative law.

[23] Kevin M. Stack, *An Administrative Jurisprudence: The Rule of Law in the Administrative State*, 115 COLUM. L. REV. 1985, 1989 (2015) (grounding reasoned justification as a rule-of-law requirement).

III. BRINGING MACHINE-LEARNING SYSTEM DESIGN WITHIN ADMINISTRATIVE LAW

A. ADMINISTRATIVE PROCESS FOR MACHINE LEARNING DESIGN

Identifying the ways that the design of machine learning systems can embed value decisions reveals the ways that the adoption of machine learning systems through procurement can render policymaking invisible. Design choices set policy without input from agency employees, stakeholders, or other experts. The models, assumptions, metrics, and, at times, even the data that drive such systems, are largely opaque and unknown to government officials who acquire them and the public they govern.

When such systems embed policies, the current method of adoption lacks all hallmarks of legitimate governance. Administrative actors are excused from reasoning, analysis, and the requirement that they justify policy choices. They bring no expertise to bear. They elicit no public participation or input. Their decisions evade judicial review and political oversight. Scholarship has largely failed to address this phenomenon of lawless governance. To be sure, a robust literature has focused on the challenge of system opacity, proposing algorithmic "transparency" as a means to address the ways opacity can obscure bias, error, and outcomes that diverge from public goals.[102] Proposals for transparency have focused on open sourcing a

[102] *See generally* Charles Vincent & Jean Camp, *Looking to the Internet for Models of Governance*, 6 ETHICS & INFO. TECH. 161, 161 (2004) (explaining that automated processes remove transparency); Paul Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Computer*, 43 HASTINGS L.J. 1321, 1343–74 (1992) (setting forth an influential paradigm for addressing data-driven governance, which includes making data processing systems transparent; granting limited procedural and substantive rights to the data subject; and creating independent governmental monitoring of data processing systems).

given system's software code and releasing it to the public for inspection; mandating disclosure of system methodology;[103] disclosing the sources of any data used;[104] requiring audit trails that record the facts and rules supporting administrative decisions when they are based on automated systems; mandating that hearing officials explain in detail their reliance on an automated system's decision;[105] and notifying those affected when algorithmic systems are used.[106]

Such transparency mechanisms, in turn, are intended to facilitate accountability.[107] Openness about the algorithms that drive technological systems government agencies use permits public analysis and critique,[108] and an assessment of the fairness of their use. It allows software audits[109] that identify correct, and incorrect, inputs and outputs, back-testing of those input and outputs to assure the system is executing its intended goals,[110] and testing of software on specific scenarios with pre-determined outcomes. It can allow individuals to contest, inspect, and adjudicate problems with data or decisions made by a system, facilitating challenges to government determinations based on algorithmic systems. Such measures facilitate mechanisms to "vindicate the norms of due process" and administrative

[103] PASQUALE, *supra* note 18, at 14–15; Giovanni Buttarelli, *Towards A New Digital Ethics: Data, Dignity, And Technology*, EUR. DATA PROTECTION SUPERVISOR 2 (Sept. 11, 2015); Rob Kitchin, *Thinking Critically About and Researching Algorithms*, 20 INFO. COMM. & SOC'Y 14 (2017); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. U. L. REV. 1, 21 (2014); Nicholas Thompson et al., *Emmanuel Macron Talks to WIRED About France's AI Strategy*, WIRED (Mar. 31, 2018, 06:00 AM), https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/ [https://perma.cc/X7HH-4K6K].

[104] PASQUALE, *supra* note 18, at 14.

[105] Citron, *supra* note 19, at 1310–12.

[106] *See* Citron & Pasquale, *supra* note 103, at 21; *see also* Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 125–28 (2014) (advocating a right to "procedural data due process" to address the harms of predictive systems).

[107] Kroll et al., *supra* note 20, at 657 (describing transparency as "[a] native solution to the problem of verifying procedural regularity" and describing its utility and limits); Fink, *supra* note 20, at 1453–56 (explaining limits of transparency due to current state of law and practice with respect to whether algorithms would be considered "records" under the Freedom of Information Act (FOIA) and agency bases for withholding algorithms and source code under FOIA requests); Pasquale, *supra* note 20, at 235–36.

[108] Citron, *supra* note 19, at 1311–12.

[109] *See* Citron & Pasquale, *supra* note 103, at 20–22 (advocating for transparency requirements for data and calculations and placing scoring systems used in the context of employment, insurance, and health care under licensing and audit requirements); *see also* Crawford & Schultz, *supra* note 106, at 122–23.

[110] PASQUALE, *supra* note 18, at 14–15; Diakopoulos, *supra* note 44, at 399–402; Citron & Pasquale, *supra* note 103102, at 21–22.

decision making even when decisions are automated.[111] This allows individuals to plead extenuating circumstances that software cannot anticipate[112] and accords the subjects of automated decisions the right to inspect, correct, and dispute inaccurate data.[113]

Yet, while critics have debated the limits of these approaches,[114] the debate has focused largely on the use and effectiveness of transparency, whistleblowers, ex post challenges, and oversight. The ex post focus positions accountability after critical design decisions have been made. And while new scholarship has begun to focus on the process of machine learning system design,[115] this literature has not explored the full potential of administrative law to remedy the abdication of government agencies' involvement in design questions, even when they implicate issues that we usually regard as involving traditional substantive policy questions.

Administrative law maps another direction. It suggests that, when the design of machine learning systems embeds policy, policymakers should be required to engage in reasoned decision making. To be meaningful, given the character of the decisions involved in machine learning design, that

[111] Citron, *supra* note 19, at 1301.

[112] *Id.* at 1304.

[113] PASQUALE, *supra* note 18, at 145.

[114] Kroll et al., *supra* note 20, at 657–58 (explaining that while "full or partial transparency can be a helpful tool for governance in many cases . . . transparency alone is not sufficient to provide accountability in all cases"); *see generally* Katherine Noyes, *The FTC Is Worried About Algorithmic Transparency, and You Should Be Too*, PC WORLD (Apr. 9, 2015, 08:36 AM), https://www.pcworld.com/article/2908372/the-ftc-is-worried-about-algorithmic-transparency-and-you-should-be-too.html [https://perma.cc/7KHT-GHZ7]; Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*, UNIV. MICH. (May 22, 2014), http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf [https://perma.cc/TJ3Y-2UZK] (presenting at "Data and Discrimination: Converting Critical Concerns into Productive Inquiry," a preconference at the 64th Annual Meeting of the International Communication Association). Critiques include the fact that open sourcing a given machine learning system's neural network does not necessarily mean an outside third party will verify how the system determined a given output. *See* Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability*, NEW MEDIA & SOC'Y 973, 983–84 (2016); Jakko Kemper & Daan Koklman, *Transparent to Whom? No Algorithmic Accountability Without a Critical Audience*, INFO. COMM. & SOC'Y (2018); Brauneis & Goodman, *supra* note 6, at 137–38 (pointing out the difficulty of understanding complex AI systems and the shortcomings of knowing inputs and outputs of a given system as the basis for adequate oversight); Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181, 194–96 (2016). For the impediment posed to transparency by trade secret law, see Brauneis & Goodman, *supra* note 6, at 153–57; David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135, 180 (2007).

[115] *See supra* note 22; *see also* Katyal, *supra* note 20, at 54.

deliberation must address an understanding, informed by both technical and domain expertise, of the methodologies adopted and the value choices behind them, and provide justifications for those choices' resolution. Administrative law, moreover, provides guidance about what types of concerns should trigger such requirements, and how, given the characteristics of machine learning, those concerns translate to the particular context of system design.

B. A FRAMEWORK FOR REASONED DECISION MAKING ABOUT MACHINE LEARNING DESIGN

The administrative state's legitimacy is premised on the foundational principle that decisions of substance must not be arbitrary or capricious.[116] Rather, those decisions must be the product of a contemporaneous process of reasoned decision making.[117] Requiring such process vindicates core public law values: it ensures, on the one hand, that technical expertise has been brought to bear on a decision; and on the other, that the decisional visibility necessary to permit public accountability exists.[118] Together, a transparent reasoning process prohibits an agency from "simply asserting its preference."[119]

Specifically, an agency must produce a record that enables courts "to see what major issues of policy were ventilated," and "why the agency reacted to them as it did."[120] Thus the agency must have engaged in reasoned analysis about relevant factors consistent with the record before it, and they may not have considered irrelevant factors or decided without sufficient evidence. An agency falls short where there is no record of "examin[ing] the relevant data" or "articulat[ing] a satisfactory explanation for its action including a 'rational connection between the facts found and the choice made.' "[121]

By the terms of this standard, the complete abdication of any agency role in considering the important policy choices inherent in a machine learning system's design would be an abject failure. This Section explores the alternative, using the arbitrary and capricious paradigm to identify the types

[116] Courts may "hold unlawful and set aside [an] agency action" they deem to be "arbitrary [or] capricious." 5 U.S.C. § 706(2)(A).

[117] SEC v. Chenery Corp. (Chenery II), 332 U.S. 194, 196 (1947) (holding that courts may uphold an agency's action only for reasons on which the agency relied when it acted); *see generally* Kevin M. Stack, *The Constitutional Foundations of Chenery*, 116 YALE L.J. 952 (2007) (grounding the *Chenery* norm in the Constitution).

[118] Cass R. Sunstein, *From Technocrat to Democrat*, 128 HARV. L. REV. 488 (2014) (discussing the technocratic and democratic directions in administrative law).

[119] *Id.* at 496.

[120] Auto. Parts & Accessories Ass'n v. Boyd, 407 F.2d 330, 338 (D.C. Cir. 1968).

[121] Motor Vehicle Mfrs. Assn. v. State Farm Mut., 463 U.S. 29, 43 (1983) (quoting Burlington Truck Lines, Inc. v. United States, 371 U.S. 156, 168 (1962)).

of machine learning systems, and system elements, whose design should be guided by reasoned and transparent decision making, and what such decision making would require in the machine learning context to survive legal challenge.

1. *Determining What System Choices Should Require Reasoned Decision Making*

Government agencies increasingly rely on artificial intelligence across their operations. Many functions—from monitoring IT system security to managing government supply lines and procurement—involve largely management support, and therefore may not implicate the types of policy decisions that should trigger the type of decisional record discussed above. This raises a threshold challenge in distinguishing systems that are inward-facing from those that create public-facing policy of the type that agencies should deliberate about and ventilate in a public manner.

Administrative law has dealt with comparable distinctions in a range of contexts and offers some insights into where and how we might draw lines about when a machine learning system is engaged in policymaking of concern to us, and when it is not. Specifically, jurisprudence has identified important indicia of contexts in which administrative choices trigger concerns necessitating a reasoned and transparent decision making process, and the creation of a record sufficient for judicial review: whether the agency action in question limits future agency discretion in deciding issues of legal consequence, and whether the action reflects a normative choice about implementation. Each of these inquiries offer useful insight for the question of which machine learning systems' design, and which system elements' design, should be treated as making policy.

a) Design Choices that Limit Future Agency Discretion

In a variety of contexts, courts have identified the constraining effect of an administrative decision on the future substantive discretion of agencies or their staff as a baseline determinant of whether agency decisions will be subject to judicial review, and therefore to analysis under the arbitrary and capricious standard. When current decisions hem in choices about the law's application going forward, they reflect binding policy choices and thus may be reached openly, explicitly, and through reasoned analysis.

Even in contexts in which executive discretion is broad—as it is in internal agency management—such factors argue for requiring reasoned decision making. Thus, while agencies have largely unreviewable discretion regarding enforcement decisions,[122] judicial oversight is appropriate when an

[122] *See* Heckler v. Chaney, 470 U.S. 821, 832 (1985). The court opinion stated:

agency adopts a "general enforcement policy" that "delineat[es] the boundary between enforcement and non-enforcement."[123] Such actions limit agency discretion going forward, with implications for "a broad class of parties."[124] In such contexts, in contrast to individual decisions to forgo enforcement, an agency is expected to present a clearer and more easily reviewable statement of its reasons for acting.[125]

Related concerns govern the determination of whether an agency action is "final," which is a second Administrative Procedure Act (APA) prerequisite for judicial review.[126] To satisfy this requirement, an agency action must not simply mark the "consummation" of an agency's "decision-making process"—a standard satisfied by many nonbinding or advisory decisions, even when they are made informally.[127] The decision must also "be one by which rights or obligations have been determined, or from which legal consequences will flow."[128] The Supreme Court has recently counseled a "pragmatic" approach to the interpretation of this standard, focusing on the prospective limits it places on agency discretion as a key component of the "legal consequences" test.[129] Lower courts have already taken such a pragmatic approach—looking at whether, as a practical matter, a purportedly non-binding agency decision effectively guides future agency decisions and constrains agency discretion, such as if "an agency act[s] 'as if a document issued at headquarters is controlling in the field.' "[130]

The distinctions drawn with respect to finality track those governing whether agency actions must satisfy the notice and comment procedures prescribed by § 553 of the APA. While reasoned decision making sufficient

> [W]e recognize that an agency's refusal to institute proceedings shares to some extent the characteristics of the decision of a prosecutor in the Executive Branch not to indict—a decision which has long been regarded as the special province of the Executive Branch, inasmuch as it is the Executive who is charged by the Constitution to "take Care that the Laws be faithfully executed."

Id.; APA excludes from review "agency action . . . committed to agency discretion by law." 5 U.S.C. § 701; *see also* Citizens to Preserve Overton Park, Inc. v. Volpe, 401 U.S. 402, 410 (1971) (holding that the "committed to agency discretion" exception to judicial review is "very narrow" and "is applicable in those rare instances where 'statutes are drawn in such broad terms that in a given case there is no law to apply' ").

[123] Crowley Caribbean Trans. v. Pena, 37 F.3d 671, 676–77 (D.C. Cir. 1994).

[124] *Id.*

[125] *Id.*

[126] The APA extends judicial review only to "final agency action." 5 U.S.C. § 704.

[127] Bennett v. Spear, 520 U.S. 154, 177–78 (1997).

[128] *Id.*

[129] *U.S. Army Corps of Eng'rs v. Hawkes Co.*, 136 S. Ct. 1807, 1814–15 (2016); *see* William Funk, *Final Agency Action After Hawkes*, 11 N.Y.U. J.L. & LIBERTY 285 (2017).

[130] Appalachian Power Co. v. EPA, 208 F.3d 1015, 1021 (D.C. Cir. 2000).

for system design and adoption decisions to survive arbitrary and capricious review can certainly occur through a range of administrative processes beyond informal rulemaking, this jurisprudence offers an informative framework in which courts have thought carefully about which agency actions should trigger more robust process, reflecting reasoned deliberation, participation, expertise, and judicial review.

In this context, courts have developed extensive doctrine regarding what types of agency actions are "non-legislative" and therefore exempt from such process requirements, as compared to those that are "substantive" and therefore must satisfy them. Such exempt actions (involving, for example, internal agency procedure, agency management, or guidance to regulated parties) do not carry the "force of law" in that they do not make substantive changes to the legal rights and obligations of regulated individuals. As understood by case law, agency guidance statements are those "issued by an agency to advise the public prospectively of the manner in which the agency proposes to exercise a discretionary power."[131] These statements provide agencies with the opportunity to announce their "tentative intentions for the future" in a non-binding manner. An agency articulation, then, that "genuinely leaves the agency and its decision makers free to exercise discretion" raises few process concerns.[132] The agency may adopt it with little process, and it is not, in and of itself, reviewable by courts.

By contrast, courts are also sensitive to the concern that agencies are circumventing the need for decision-making process when they make substantive policy in a manner purported to govern only internal agency procedure or provide only informal guidance. As a result, courts sometimes find that notice-and-comment is necessary, even when the agency statement in question does not seem in and of itself to have any binding legal effect on regulated entities. This seems especially so when the relevant statutes and legislative rules give the agency wide discretion, but the challenged agency statement indicates that agency personnel will in reality exercise that discretion only in narrowly defined circumstances.[133] In those situations, courts have found that the agency action is "practically" (although not formally) binding. Because of the severe constraints that the agency's "informal" action imposed on agency discretion, the agency should have engaged in the full notice-and-comment rulemaking procedure.

[131] Am. Bus. Ass'n v. United States, 627 F.2d 525, 529 (D.C. Cir. 1980) (internal citation omitted).

[132] *Id.*

[133] Gen. Elec. v. EPA, 360 F.3d 188 (D.C. Cir. 2004); Cmty. Nutrition Inst. v. Young, 818 F.2d 943 (D.C. Cir. 1987).

Tracking these standards, existing jurisprudence regarding the setting of formulae and numerical cutoffs, and the choices regarding underlying methodology, provides useful guidance for identifying aspects of machine-learning systems that set discretion-constraining policy.

Pickus v. United States Board of Parole,[134] a case arising in the challenge to an agency's decision to adopt a formula informally (without a notice and comment process), describes well the ways in which the such adoption can set future policy by limiting agency discretion going forward. In *Pickus*, the D.C. Circuit considered a challenge to two rounds of Parole Board "guidelines" that set formulae by which parole would be determined. The court rejected the Board's contention that, under the APA, the issuance of such guidelines lacked legal force because they were merely "general statements of policy, interpretative rules," or "rules relating to agency organization, practice or procedure."[135]

In so doing, the court focused on the practical implications on agency decision-making discretion, and the subsequent legal consequences. As the court described, the first set of guidelines "consist of nine general categories of factors, broken down into a total of thirty-two sub-categories, often fairly specific." Therefore,

> [a]lthough they provide no formula for parole determination, they cannot help but focus the decisionmaker's attention on the Board-approved criteria. They thus narrow his field of vision, minimizing the influence of other factors and encouraging decisive reliance upon factors whose significance might have been differently articulated had [more formal decision-making processes] been followed.[136]

Because of this narrowing of decision-making focus, the court held, the guidelines "were of a kind calculated to have a substantial effect on ultimate parole decisions."

The second agency action, styled an "announcement," consisted of a "complex, detailed table which purport[ed] to state the range of months

[134] Pickus v. United States Board of Parole, 507 F.2d 1107 (D.C. Cir. 1974). In a later case, Prows v. United States Dep't of Justice, 704 F. Supp. 272 (D.C. Cir. 1988), a Program Statement from the Federal Bureau of Prisons declaring that inmates had to deposit at least 50% of their payment from prison jobs to "legitimate financial obligations" was struck down. Analogizing the rule to the guidelines in *Pickus*, the court found the Statement "has been interpreted by defendants in a 'formula like' manner," without any discretion and therefore wasn't an interpretative rule nor a policy statement and should have proceeded through notice and comment. *Prows*, 704 F. Supp. at 277.

[135] *Pickus*, 507 F.2d at 1112 (D.C. Cir. 1974) (citing 5 U.S.C. § 553(a)(2) and providing exemptions).

[136] *Id.* at 1111–13.

which the Board [would] require an inmate to serve depending upon the severity of his offense (six classifications) and his 'salient factor score' (four classifications)."[137] The score, the court continued,

> is computed using only those criteria, and the quantitative input of each is specified as well. Computation of the score is a purely mechanical operation. Third, the chart sets a narrow range of months of imprisonment that will be required for a given category of offense and a given salient factor score. This is not to suggest that these determinants are either unfair or undesirable, but merely that they have significant consequences.[138]

Thus, the court concluded, both policies defining parole selection criteria "are substantive agency action," and "the interested public should have an opportunity to participate, and the agency should be fully informed, before rules having such substantial impact are promulgated."[139]

Moreover, in *Community Nutrition Institute v. Young*,[140] the D.C. Circuit determined that FDA "action levels"—the allowable levels of unavoidable contaminants in food, and again a precise number—while purportedly without the "force of law," practically bound third parties and should have gone through the notice-and-comment procedure required for legislative rules. Pursuant to its statutory mandate to limit the amount of "poisonous or deleterious substances" in food,[141] the FDA established "action levels"—which the FDA characterized as guidance statements—that set permissible levels of unavoidable contaminants such as aflatoxins in food. Producers who exceed action levels are subject to enforcement proceedings. The FDA claimed that action levels were "nonbinding statements of agency enforcement policy," but the court found that setting precise numerical limits cabined the FDA's enforcement discretion, effectively binding the FDA and therefore affecting the rights of regulated parties.[142]

b) Normative Choices Between "Methods of Implementation"

Judge Richard Posner, writing for the Seventh Circuit in *Hoctor v. U.S. Department of Agricriculture*,[143] has articulated the way that numerically-based

[137] *Id.* at 1110–11.

[138] *Id.* at 1113.

[139] Id.

[140] Cmty. Nutrition Inst. v. Young, 818 F.2d 943 (D.C. Cir. 1987).

[141] 21 U.S.C. § 346.

[142] *Cmty. Nutrition Inst.*, 818 F.2d at 946–48 ("[T]his type of cabining of an agency's prosecutorial discretion can in fact rise to the level of a substantive, legislative rule."). That is exactly what has happened here.

[143] Hoctor v. U.S. Dep't of Agric., 82 F.3d 165, 171 (7th Cir. 1996).

line-drawing can often reflect a particularly unconstrained form of normative policymaking—which, when is does, enhances the need for more robust process. *Hoctor* involved a challenge to an informal USDA internal memorandum fixing a specific requirement for the height of perimeter fences used to contain "dangerous animals." While the background regulation in force for a number of years had required fencing "appropriate" for the animals involved, the memorandum sought uniformly to require eight-foot fences. The court, however rejected the agency's attempt to arrive at a numerical standard with little decisionmaking process, which, in the court's mind, undermined the decision's democratic legitimacy.[144]

Generally, Judge Posner emphasizes the policy-making nature of administrative decisions that "translate[ ] a general norm into a number"[145]—a phenomenon, he notes, that arises "especially in scientific and other technical areas, where quantitative criteria are common."[146] Moreover, he describes, the "flatter" (or more specific) the ultimate line drawn by the agency, "the harder it is to conceive of it as merely spelling out what is in some sense latent in a statute or regulation"[147] and the more it represents a choice among "methods of implementation."[148] Such choices are legislative in nature, and should be treated as such.

Jurisprudence reviewing agency decision making under the arbitrary and capricious standard reflects these insights about numerical or formula-based agency implementation choices, and provides important foundations for identifying which elements of machine learning systems must satisfy the arbitrary-and-capricious metric in their adoption. Indeed, courts have explicitly held that agencies must engage in reasoned analysis in choosing methods of implementation reflecting many of the very type of decisions inherent in machine learning design described in Part II.

In assessing risk, courts have held, agency decision makers must actively consider the decision whether to err in the direction of false negatives or false positives, and provide reasons for their choice.[149] Similarly, agencies must justify the assumptions behind their use of specific models when

[144] SUSAN ROSE-ACKERMAN, STEFANIE EGIDY & JAMES FOWKES, Due Process of Lawmaking: The United States, South Africa, Germany and the European Union 91 (2015).

[145] *Hoctor*, 82 F.3d at 171.

[146] *Id.*

[147] *Id.*

[148] *Id.* at 170.

[149] *See* Int'l Union, United Mine Workers of Am. v. Fed. Mine Safety & Health Admin., 920 F.2d 960, 962–66 (D.C. Cir. 1990) (remanding to the agency for "more reasoned decision making" on the issue of whether carbon monoxide monitors provide enough protection for workers, after it engaged only in an analysis of false negatives, without discussion of false positives, "ignor[ing] this problem altogether").

determining costs and designing impact statements[150] and provide information to justify the methodology behind models that they use for risk prediction.[151] They must take steps to confirm the validity of their chosen models[152] and, in deciding whether to use a particular scientific methodology, both demonstrate its reliability and transparently discuss its shortcomings. With respect to data, agencies must provide information on its source.[153]

What reasoned deliberation entails is set out across a range of procedural contexts—from cost-benefit analysis to environmental impact assessments—and in a range of substantive policy contexts. The failure to identify, disclose, engage with, and justify the consequent policy choices within models closely correlated to machine learning systems—statistical and economic models, for example—has been held to constitute a "complete lack of explanation for an important step in the agency's analysis."[154] And absent efforts to surface and affirmatively explain the assumptions underlying decision-making models, they may remain "fatally unexplained" and unappreciated.[155]

c) Application to Machine Learning Systems

Decisions about the design of a machine learning system—particularly one modeling fairness—constrain agency discretion much like the formulae in *Pickus*, and the action levels in *CNI*. These cases underscore the ways in which precise numerical limits or formulae have anchoring effects that constrain agency action, and the consequent importance of robust process in their adoption. Machine learning systems are rife with similar issues, such as cutoffs determining who is high, medium, or low risk in recidivism risk systems, or the thresholds in the Amazon Rekognition service described in Part II.

[150] Nat. Res. Def. Council, Inc. v. Herrington, 768 F.2d 1355, 1412–19 (D.C. Cir. 1985) (analyzing the Department of Energy's use of a real annual discount rate of 10% when determining life cycle costs and the net present value of savings from appliance energy efficiency standards).

[151] Owner-Operator Indep. Drivers Ass'n, Inc. v. Fed. Motor Carrier Safety Admin., 494 F.3d 188, 199–204 (D.C. Cir. 2007) (holding that an agency must disclose the methodology of the agency's operator-fatigue model, a crash-risk analysis that was a central component of the justification for the final rule).

[152] Ecology Ctr. v. Austin, 430 F.3d 1057 (9th Cir. 2005). The Forest Service used a model to conclude that treating old-growth forest through salvage logging was beneficial to dependent species but did not confirm their hypothesis by any on-the-ground analysis.

[153] *Nat. Res. Def. Council*, 768 F.2d at 1412–19.

[154] *Owner-Operator Indep. Drivers Ass'n, Inc.*, 494 F.3d at 204.

[155] *Natural Res. Def. Council*, 768 F.2d at 1414–19 (analyzing the Department of Energy's use of a real annual discount rate of 10% when determining life cycle costs and the net present value of savings from appliance energy efficiency standards).

There is, moreover, often no unique connection between the cutoffs or thresholds chosen and a statutory mandate or technological requirement, as in *Hocter*. Rather—like the agency actions in the arbitrary-and-capricious cases involving which risk models to adopt, whether to prefer false negatives and positives, what data to use, and which scientific methodology to employ—those decisions reflect normative choices between methods of implementation. In the machine learning context, these might also include the unit of analysis (the algorithm, the algorithmic system, or the overall system of justice);[156] how fairness is measured—whether it is by group-level demographic parity, equal positive predictive values, equal negative predictive values, accuracy equity, individual fairness metrics such as equal thresholds, or a devised similarity metric—and the related question of whether and how to use attributes related to protected classes such as in the *Loomis* case.

When the design of a machine learning system deprives an agency and its staff of future substantive discretion,[157] especially through numerical or methodological choices that reflect normative judgments on implementation rather than ones required directly by statute or technical or scientific knowledge, the design choices embedded in machine learning systems should not be reached in an arbitrary or capricious manner. Thus if a record lacks evidence of agency deliberation or reveals deliberations that demonstrate one of the other indicia of arbitrariness, an agency's reliance on the system should be subject to legal challenge.

2. *Designing Agency Decision Making: Reflecting the Technocratic and Democratic Requirements of Administrative Law*

Where the policy decisions embedded in system design supplant administrative discretion, what would it mean, in the language of the arbitrary and capricious jurisprudence, that on the one hand "the agency should be fully informed"[158] and provide a justification for its choice based on a "rational connection" with the "facts found,"[159] and on the other, that decisions should be open to public engagement and political accountability?

[156] Andrew D. Selbst et al., *Fairness and Abstraction in Sociotechnical Systems*, *in* PROC. CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 59, 60–61 (2019) (describing the "framing trap"—the tendency to analyze fairness at the level of inputs and outputs of the model rather than at the level of socio-technical system in which the machine learning system is embedded).

[157] Am. Bus. Ass'n v. United States, 627 F.2d 525, 529 (D.C. Cir. 1980) (asking "whether a purported policy statement genuinely leaves the agency and its decision-makers free to exercise discretion").

[158] Pickus v. United States Board of Parole, 507 F.2d 1107, 1113 (D.C. Cir. 1974).

[159] Motor Vehicle Mfrs. Assn. v. State Farm Mut., 463 U.S. 29, 43 (1983) (quoting Burlington Truck Lines, Inc. v. United States, 371 U.S. 156, 168 (1962)).

These elements reflect dual (and sometimes competing) impulses—technocratic and democratic—animating the law of administrative process.[160]

As to the first, to engage in reasoned deliberation, agency staff must address their lack of technical knowledge, enlist additional expertise to "inform" themselves sufficiently, and provide reasons justifying the resolution of four questions specific to machine learning systems. Those questions include: (1) for what a system is optimizing; (2) what determinations are being made about the choice and treatment of data; (3) what assumptions and limitations are implied by the choice of model; and (4) what interfaces and policies structure agency staff's interactions with machine learning systems—the human-machine loop. Importantly, as to the second question, meaningful processes must address the opacity of value choices made through design by ensuring "political visibility,"[161] to surface the fact that technical choices involve a policy judgment. In this context, transparent decision making involves not simply making algorithms transparent, but making policy visible.[162]

a) Technocratic Elements in Reasoned Decision Making About Machine Learning Systems

A comparison of machine learning systems with prior automation—generally so-called "expert systems"—helps identify particular aspects of machine learning system design decisions that displace traditional modes of expert administrative decision making. Danielle Citron's 2008 work on *Technological Due Process*[163] provided a foundational analysis of the ways that administrative automation based on a prior generation of expert systems transformed the technological decision-making landscape in ways that matter for policymaking norms.[164] It is further instructive in highlighting the ways machine learning has both compounded and redirected the displacement of expert human judgment, a challenge with which agencies must grapple when adopting such systems.

[160] *See generally* Sunstein, *supra* note 118 (discussing the technocratic and democratic strains in administrative law).

[161] Mulligan & Bamberger, *supra* note 12, at 776–80.

[162] *See, e.g.*, Eaglin, *supra* note 22, at 88 (noting how recidivism risk tools make it "difficult to ascertain . . . policy decisions").

[163] Citron, *supra* note 19.

[164] For an excellent and accessible discussion of expert systems and what lessons from their development suggest about the discussion for explainability and interpretabiliy in machine learning, see generally David C. Brock, *Learning from Artificial Intelligence's Previous Awakenings: The History of Expert Systems*, 39 AI MAG. 3 (2018).

i) Citron's Concerns: Displacement of Expert Agency Judgment

Citron identified a related set of objections to earlier attempts to automate agency processes. First, she described how "[a]utomated systems inherently apply rules because software predetermines an outcome for a set of facts."[165] This, in turn, displaces the ongoing exercise of human judgment, which is better reflected in standards. She thus concludes that "[d]ecisions best addressed with standards should not be automated."[166] Citron further drew on the "rules versus standards" debate to emphasize the distinction between automated systems, which implement rules and favor consistency, and human decision-making systems, which favor "situation-specific discretion."[167]

Second, Citron raised the related question of *who* sets the rules that displace the standards-like exercise of human judgment. Her concern involved the displacement of expert agency decision making by the choices of engineers who design technical systems.[168] In particular, she was concerned that engineers' interpretations and biases, and their general preference for tractable binary questions, distort decision making.

Finally, Citron expressed concern regarding the lack of record-keeping and transparency about the rules automated systems apply. Absent such a digital trail, the ability to seek redress or accountability is limited. To enable individual due process and support overall accountability, Citron advocates that systems be built to produce "audit trails that record the facts and rules supporting their decisions."[169]

ii) Updating Concerns: How Machine Learning Displaces Rational Expert Agency Decision Making

While Citron's conception of what is inherent in automation may have been largely accurate with respect to the automated systems used by government at the time (prior to her 2008 publication date), the rote

[165] Citron, *supra* note 19, at 1303.

[166] *Id.* at 1304.

[167] *Id.* at 1303; *see* Bamberger, *supra* note 18, at 676 ("Computer code [in contrast to human judgment] operates by means of on-off rules, while the analytics it employs seek to quantify the immeasurable with great precision.") (internal quotation marks omitted).

[168] Citron, *supra* note 19, at 1261 ("Code writers also interpret policy when they translate it from human language to computer code. Distortions in policy have been attributed to the fact that programmers lack 'policy knowledge.' "); *id.* at 1262 ("Changes in policy made during its translation into code may also stem from the bias of the programmer. . . . Policy could be distorted by a code writer's preference for binary questions, which can be easily translated into code.").

[169] *Id.* at 1305.

application of predetermined rules she documents is an inapt description of the machine learning systems coming into government use today. Machine learning systems do not apply predetermined rules to sets of facts, but rather develop probabilistic models that optimize for a particular goal. They are then allowed to learn in the field, generate new rules on the fly, and iteratively update them.

In this way, like earlier expert systems, machine learning systems too displace agency reasoning and expertise, and constrain future agency discretion. However, the displacement takes new forms, stems from additional sources, and requires distinct responses. The risk of displacement no longer stems from the explicit reasoning of engineers translating agency rules into code, but rather arises from the "logic" the model machine learning systems derive from training data reflecting past agency actions. The assumptions and policy choices built into the machine learning model used to generate the predictive model, as well as policy choices in the application of the predictive model, rather than engineer-coded rules, are the key hidden constraints.

a. Element 1: Delegating "Logic-Making" to Machines

Today's machine learning systems, then, delegate "logic-making" to algorithms. Unlike expert systems that Citron rightly identified as displacing nuanced and fact-specific agency staff decisions with the rote application of predetermined rules as coded by engineers, machine learning systems *construct their own logics* from training data. Machine learning systems skip the process of codifying an agency's decision-making process, and instead rely on the machine learning model to learn a classifier—its own machine logic—from a set of training data that reflects past agency actions. While the machine logic captured in the classifier could be considered more analogous to the intuition and instinct associated with agency experts,[170] importantly, it in no way reflects the logic of agency decision makers. In fact, it answers without the causal reasoning associated with logic, or as one scholar notes, "they don't 'think' in any colloquial sense of the word—they only answer."[171]

While many consequential decisions are made by the engineers, the decision about how to model agency judgments is not explicitly constructed by engineers through rules—abstract or specific—but rather learned by

[170] Of course, human intuition is produced by neurological processes and machines' through computational processes. While machine learning abounds with terms that evoke the brain, only some machine learning systems attempt to mirror cognitive processes.

[171] Jonathan Zittrain, *The Hidden Costs of Automated Thinking*, NEW YORKER (July 23, 2019), https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking [https://perma.cc/D9YK-JYHZ].

algorithms through analysis of data traces reflecting agency decision making. In theory, one might surmise that because machine learning models are trained on data that represents the past decisions and related outcomes of the agency—or "similar" ones—they might naturally align more closely with the judgment of agency experts and, by design, provide less room for interference or usurpation of judgment by engineers. If that were so, perhaps machine learning systems should raise *less* concern about displacement of human expert judgment than earlier automated systems.

Unfortunately, this surmise breaks down under more careful scrutiny. The training data reflects patterns reflected in professional decisions, but not professionals' *decision-making processes*. This is an important distinction. It means that a machine learning model's "logic" may well reflect the actions and outcomes of professional decision making (the outputs) but bear little resemblance to the rationales and justifications behind those decisions.

Significantly, the "reasoning" of complex machine learning systems often bears no resemblance to human logic and is impossible to discern.[172] The divergence in intuition is intrinsic because machines and humans "see" in different ways. For example, machine learning systems can identify complex patterns and scan across massive data sets. Humans, by contrast, can identify things they've seen (such as faces) despite a wide range of subtle and relatively extreme perturbations (changes to hair style, plastic surgery, aging, etc.). The different intuitions developed by human and machine systems may therefore produce similar outputs in some instances, but not in others, or similar outputs but for very different reasons.

Thus, the machine has learned its own logic based on the training data: it has not learned to mirror the agency's logic, only to predict outcomes of it. Like two students producing correct answers to a math problem, unless they "show their work," it would be wrong to assume they used the same method, let alone that either used the right method or appropriately applied it. The design process of machine learning systems does not explicitly transfer expert reasoning and therefore does not create the pattern of displacement found in expert systems. Yet because machine learning classifiers are developed by studying the outcomes of agency logics rather than the logic itself, it creates potentially more troubling displacement effects.

[172] For example, machine learning image recognition systems are famous for appearing to perform well on a task but actually relying on a simplistic and poorly-chosen heuristic. In addition, as Kaminsky describes, algorithms lack the contextual understandings of acceptable bases for decision making and the common sense of humans. Kaminsky, *supra* note 22, at 14–15.

b. Element 2: Constraints on Policymaking Evolution

Machine learning systems develop *probabilistic models that optimize for a particular goal*—and then, where they are allowed, update them as they learn in the field. Rather than the automated rules that concerned Citron, the constraints imposed on agency discretion in machine learning systems are found in choices about what a system is optimizing for and how the goal is operationalized going forward within the system.

Once deployed, the logic of the model—whether fixed, or allowed to learn over time—remains constrained by the assumptions and choices made during design. In contrast, the judgment of agency professionals and staff may evolve over time, sometimes on a gentle slope, but at other times diverging swiftly in response to new research, new political winds, or other internal or external jolts. While a machine learning model may learn new ways to optimize for the goal established, it is tethered to the beliefs and biases that are fixed in the model, as well as the assumptions and ingested data used during development. As a result, machine learning systems can instill patterns of racism, debunked science, or other faulty or unjust reasoning that may be captured in the training data or optimization choices.

Even if the policies embedded in a model are fully aligned with agency decision making at the time of its initial deployment, if the system is not updated to reflect changing agency understanding of sound judgment and agency practice, machine learning systems can constrain agency discretion in particularly problematic ways.

Finally, the extent to which a model developed on one data set can be safely used on another is an immensely important policy question. It is well documented that models trained on one data set can perform catastrophically poorly on a data set that many might assume to be similar by some set of metrics.[173] For example, models trained on newswire copy perform poorly on texts from other domains.[174] Even for discrete Natural Language Processing (NLP) tasks, such as identifying words as nouns, verbs, adjectives, etc.—called part-of-speech or grammatical tagging or word-category disambiguation—which lay people might consider simple and transferable

[173] Selbst et al., *supra* note 156, at 4–5 (calling this the "portability trap" and tying it to the quest for abstractions and tools that can be reused across contexts).

[174] *See* David Bamman, *Natural Language Processing for the Long Tail*, DIGITAL HUMAN. 2 (2017) ("[T]he performance of an out-of-the-box part-of-speech tagger can, at worse, be half that of its performance on contemporary newswire. On average, differences in style amount to a drop in performance of approximately 10–20 absolute percentage points across tasks.").

across corpora, models trained on news articles perform quite poorly on literary works.[175]

iii) The Challenge: Reintroducing Expert Justification for Agency Decisions

The way in which machine learning systems generate decisions without decision-making processes challenges administrative law's fundamental mandate of reasoning. To be legitimate, reliance on machine learning in governance requires processes that reintroduce appropriate expertise in providing justifications for administrative choices.

The requirement of "justification" regarding a system's design and its subsequent choices is critical. Justification is distinct from two elements identified by computer scientists related to system accountability: interpretability—properties or qualities or techniques related to a system that help humans understand the relationship between inputs and outputs[176]—and explainability: the ability to explain the operation of a machine learning system in human terms.[177] Explainability provides the reasoning behind the relationship between inputs and outputs interpretability reveals.[178]

While both interpretability and explainability might be helpful, they are not sufficient to satisfy administrative legitimacy.[179] Explaining an algorithm's operation without providing informed justifications for the choices reflected in that operation fails the "arbitrary and capricious" threshold. Instead,

[175] *See id.* (summarizing research investigating the disparity between training data and test data for several NLP tasks).

[176] Finale Doshi-Velez & Been Kim, *Towards a Rigorous Science of Interpretable Machine Learning*, ARXIV (Mar. 2, 2017), https://arxiv.org/pdf/1702.08608.pdf [https://perma.cc/6CVL-DWRK].

[177] Explanations can describe the operation of a model in general (so-called "global" explanations) or for a particular mechanism in the model used to relate inputs and outputs (so-called "local" explanations). Upol Ehsan et al., *Rationalization: A Neural Machine Translation Approach to Generating Natural Language Explanations*, ARXIV (Dec. 19, 2017), https://arxiv.org/pdf/1702.07826.pdf [https://perma.cc/GK5X-MXC5].

[178] Both explainability and interpretability are areas of debate and research among computer scientists and the multiple disciplines within the broader "fairness, accountability and transparency" research community. For a discussion of these terms and others within and across relevant disciplines, see Nitin Kohli et al., *Translation Tutorial: A Shared Lexicon for Research and Practice in Human-Centered Software Systems*, CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY (2018).

[179] For a discussion of the relationship between explanations and justifications in criminal law, and probable cause in particular, see Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1288 (2017) ("Apart from safeguarding constitutional values, explanations also vindicate rule-of-law principles. A key tenet of legality, separating lawful authority from ultra vires conduct, is the idea that not all explanations qualify as justifications.").

design choices that embed policy choices must reflect reason, a rational connection to the facts, context, and the factors mandated by Congress in the relevant organic statute, while avoiding elements extraneous to the legislative command. And such justification, in turn, requires the application of a range of forms of expertise, including technical knowledge about machine learning and algorithm design, as well as statistics, domain expertise, and specialized fields such as those represented in the Fairness, Accountability, and Transparency in Machine Learning (FAT*) and ethics, law and sociology communities, whose members investigate the social and political consequences of algorithmic systems.

As an initial matter, when systems are adopted by governments, agencies must be able to enlist sufficient expertise at the design phase to permit knowledgeable exploration of technical design and data choices that embed policy. As discussed above, decisions about system goals (what is optimized for), how to operationalize the goal into a target variable for the system to optimize for, and what modeling frameworks to use, all require expert input because they are fundamental policy decisions. In addition, determinations about the data—its selection, curation, cleaning, and similarity to the data on which it will be used—and about the triggers for updating or replacing it are all essential policy questions with which agencies must grapple explicitly. Decisions about the use and inclusion of data about protected traits warrant particular scrutiny. Precise numerical limits such as cut-offs or thresholds—particularly those that cabin discretion—must be the product of reasoned agency decision making.

Additionally, consistent with the case law's emphasis on agency discretion, agencies must comprehend and address the impact of a system on future agency choices. Traditionally, agency staff are able to adjust to new informational inputs as a situation requires.[180] They can selectively pull data in and out of the decision-making frame based on case-specific, situational knowledge. Machine learning—like other automated systems—can constrain the ability to flexibly alter the data brought to bear on a decision in response to the particular problem or person presented.[181] While machine learning systems can process tens of thousands of data points, they can only consider the data predetermined to be relevant. Setting bounds on what can be considered—ensuring, for example, that information about race, gender, age, or other protected attributes does not infiltrate agency decision making—may align with a simplistic notion of fairness. But using such simple

[180] *Cf.* Kaminsky, *supra* note 22, at 13–14 (describing how moving from a human to an automated decision can eliminate "cultural knowledge about what is or is not an appropriate decisional heuristic in a particular case").

[181] *See* Citron, *supra* note 19, at 1304 (explaining that policies allowing "individuals to plead extenuating circumstances that software cannot anticipate" should not be automated).

categories has been found to frequently be at odds with justice, the goal it purportedly serves.

Even where systems are billed as "decision support," ostensibly allowing decision makers to consider other information, automation bias may lead to overreliance on machine outputs.[182] Without efforts—policy, system design, and accountability frameworks—to foster questioning, agency staff may come to defer to machine outputs, particularly over time. In doing so, systems may elevate ideals of procedural fairness at the cost of substantively just and right outcomes. Angele Christin's research documents that automation bias may not always result and suggests that this tension between different visions of fairness may be a point of resistance. She found different kinds of resistance and tinkering with recidivism risk tools in the justice system—some of which appears to be grounded in battles over competing conceptions of fairness, its relation to justice, and the role that discretion, rather than rigidity, plays in advancing the latter.[183] The risks posed by automation bias nevertheless loom large when relevant professional, regional, or site-specific experts are not consulted during system development,[184] or when the systems are acquired as commercial off-the-shelf products rather than collaboratively developed or tailored for the conditions and context of use.

Because of this limited input and the ways these systems constrain agency staff's ability to expand or narrow the data used to render a decision, and to shift their reasoning over time, machine learning systems risk upsetting context-specific, domain-specific, and evolving judgments—key rationales for agency existence. For these reasons, the interfaces and policies that structure agency staff's interactions with machine learning systems must be the subject of agency deliberation and involve reasoned application of

[182] *See* Kate Goddard et al., *Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators*, 19 J. AM. MED. INFORMATICS ASS'N 121 (2011) (reviewing literature on automation bias in health care clinical decision-support systems).

[183] Angèle Christin, *Algorithms in practice: Comparing web journalism and criminal justice*, BIG DATA & SOC'Y 1, 9 (July 16, 2017) (discussing a senior judge's perspective on recidivism risk tools: "I don't look at the numbers. There are things you can't quantify . . . [y]ou can take the same case, with the same defendant, the same criminal record, the same judge, the same attorney, the same prosecutor, and get two different decisions in different courts. Or you can take the same case, with the same defendant, the same judge, etc., at a two-week interval and have completely different decision. Is that justice? I think it is" and finding probation officers similarly resisting rigidity by tinkering with the criteria to obtain the score they thought adequate for a given defendant).

[184] For example, criminal justice risk-assessment tools, which have been around for decades and are often simply logistic regressions, are almost uniformly created outside of the jurisdictions in which they are deployed. There are fewer than sixty tools used across the entire United States. Angèle Christin et al., *Courts and predictive algorithms*, DATA & CIV. RTS.: A NEW ERA POLICING & JUST. (Oct. 27, 2015).

expertise about the human-machine loop. This includes agency policies of the type Citron recommends—such as training on automation bias and requiring explanations of the facts and findings produced by automated systems on which agency staff rely[185]—as well as decisions about system interfaces, such as whether to communicate uncertainty and, if so, how to do so.

b) Democratic Elements in Reasoned Decision Making About Machine Learning Systems

In addition to gathering the expertise necessary to understand, explain, and justify these design choices, the arbitrary and capricious jurisprudence points to deeper issues about what meaningful deliberation would require in the machine learning context. Specifically, its emphasis on the public disclosure of the decisions made and the assumptions behind them reflects the reality that "[m]odels and proxies are built on numerous assumptions, often based in scientific principles but also laden with value judgments."[186] As political scientist Shiela Jasanoff describes, "there is growing awareness that science cannot answer all of our questions about risk and that both scientific and value judgments are involved in the processes of risk assessment and risk management."[187]

Agencies cannot create a meaningful record of pertinent "issues of policy" involved in machine learning system design and "why the agency reacted to them as it did"[188]—indeed they cannot be transparent to the public, if they fail to disclose both information about the code and its underlying models, limits, defaults, assumptions, training data, and the very fact that they engaged in a policy judgment and how those judgments were resolved. Decisional transparency must involve not only openness about design but also publicity about the very existence and political nature of value questions being resolved through design processes.

Thus "political visibility,"[189] rather than algorithmic transparency, is the essential characteristic of legitimate processes for adopting complex algorithmic systems. Administrative legitimacy is predicated on the explicit public articulation of value choices under consideration and transparent

[185] Citron, *supra* note 19, at 1306–07.

[186] Sara A. Clark, *Taking a Hard Look at Agency Science: Can the Courts Ever Succeed*, 36 ECOLOGY L.Q. 317, 331 (2009).

[187] Sheila Jasanoff, *Cultural Aspects of Risk Assessment in Britain and the United States*, *in* THE SOCIAL AND CULTURAL CONSTRUCTION OF RISK 359, 359 (B. B. Johnson & V. T. Covello eds., 1987).

[188] Auto. Parts & Accessories Ass'n v. Boyd, 407 F.2d 330, 338 (1968).

[189] Mulligan & Bamberger, *supra* note 12, at 251.

deliberation about their resolution.[190] When values are embedded in design choices they are "less visible as law, not only because it can be surreptitiously embedded into settings or equipment but also because its enforcement is less public."[191] The regulative features of technology design can appear "constitutive"—non-normative and part of the natural state of things.[192] If they are not explicitly surfaced (as they often are not), the policy decisions built into machine learning systems "fade into the background and hide the political nature of [their] design."[193] Value trade-offs, unrecognized as governance, remain unaddressed at the design stage, hindering both robust consideration of substantive policy and ex post oversight.

[190] *See, e.g.*, *Boyd*, 407 F.2d at 338 (D.C. Cir. 1968) (noting that an agency rulemaking record must make visible "what major issues of policy were ventilated" and "why the agency reacted to them as it did").

[191] Lee Tien, *Architectural Regulation and the Evolution of Social Norms*, 7 YALE J.L. & TECH. 1, 22 (2004).

[192] Mireille Hildebrandt, *Legal and Technological Normativity: More (and Less) than Twin Sisters*, 12 TECHNE 169, 179 (2008).

[193] Mulligan & Bamberger, *supra* note 12, at 778.

[194] Cass R. Sunstein, *Constitutionalism After the New Deal*, 101 HARV. L. REV. 421, 442–44 (1987) (discussing the "New Deal belief in the importance of technical expertise" as a justification for according agencies "a large measure of autonomy").

[195] *Cf.* Jody Freeman & Adrian Vermeule, *Massachusetts v EPA: From Politics to Expertise*, SUP. CT. REV. 51 (2007) (discussing the Supreme Court's "expertise-forcing" jurisprudence ensuring that "agencies actually do exercise expert judgment"); Heckler v. Chaney, 470 U.S. 821, 833 n.4 (1985) (applying arbitrary and capricious review, even in enforcement contexts characterized by high executive discretion, when an agency's failure to exercise its discretion "amount[s] to an abdication of its statutory responsibilities").