DRAFT REPORT FOR THE
ADMINISTRATIVE CONFERENCE OF THE UNITED STATES

ARTIFICIAL INTELLIGENCE AND REGULATORY ENFORCEMENT

Michael Karanicolas

This report was prepared for the consideration of the Administrative Conference of the United States. It does not necessarily reflect the views of the Conference (including its Council, committees, or members).

Recommended Citation

Michael Karanicolas, Artificial Intelligence and Regulatory Enforcement (Sept. 27, 2024) (draft report to the Admin. Conf. of the U.S.).

Artificial Intelligence and Regulatory Enforcement

Michael Karanicolas

I. Executive Summary

[to come – based on review and feedback]

I. Introduction

Across the federal government, artificial intelligence (AI) has already taken root in a variety of administrative processes and procedures. While the public has been captivated by the possibilities, and dangers, of AI, agencies have been incorporating the technologies into a broad range of institutional functions. In 2020, a report commissioned by the Administrative Conference of the United States (ACUS) identified 157 use cases of AI across 64 federal agencies.[1] Just two years later, a survey by the Government Accountability Office (GAO) identified over 1,200 current and planned use cases, with NASA and the Department of Commerce leading the way.[2] Studies like the ones commissioned by ACUS or by the GAO are likely only presenting a partial picture of where AI is being used across the federal government. What is happening today is likely also just the tip of the iceberg when it comes to AI's longer-term impact on how government operates.

While creative and progressive approaches to the provision of public services are welcome, the expanding use of AI in regulatory enforcement carries natural tradeoffs, including against values like public trust, due process, and the expertise—and the essential human character—that underlies administrative agencies' place in America's constitutional system.[3] There is a need for careful institutional analysis of the pros and cons of incorporating AI into investigations and enforcement activities. These conversations should be public, and should go beyond the narrow, risk-based analysis that dominates current models of assessment and encourage longer term thinking about the agency's character, what it may be giving up through its growing reliance on AI, and the overall impacts on key values like public trust, legitimacy, and fairness. Agency considerations should also involve a careful and critical assessment of where these tools are likely to be effective, as opposed to assuming that technologically-driven solutions will inevitably improve operations.

This paper attempts to provide an initial framework for assessing the role of AI in regulatory enforcement and recommendations for agencies considering introducing or expanding the use of AI for these purposes.

- **Understanding Terms**

Part of the challenge in discussing appropriate impacts and safeguards for "artificial intelligence" lies in pervasive confusion and inconsistencies in how this term is understood and applied. The term is often found alongside references to "machine learning," and indeed, in many popular contexts, the terms are used interchangeably. But while "machine learning" systems can be defined as algorithms which have the capacity to improve themselves based on training data, "artificial intelligence" belies any such technical definition, since the term is generally used as shorthand for any machine-based system which performs tasks that are traditionally reliant on human intelligence.[4] While these two qualities—the ability to learn from data inputs and the ability to

[1] Ho et al., p. 16.
[2] https://www.gao.gov/assets/d24105980.pdf
[3] Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L. J. 798, 804 (2021).
[4] https://www.aaas.org/sites/default/files/2022-09/Paper%204_AI%20and%20Bias_NIST_FINAL.pdf?adobe_mc=MCMID%3D52000037841860946489208149054056839995%7CMCORGID%3D242B6472541199F70A4C98A6%2540AdobeOrg%7CTS%3D1688688000&ref=internet.exchangepoint.tech p. 11.

perform tasks generally associated with human intelligence—recur in most legal definitions of AI, there are also substantial differences in how AI is defined across different frameworks.

For example, the 2019 National Defense Authorization Act describes AI as a system that "performs tasks under varying and unpredictable circumstances without significant human oversight."[5] By contrast, the National Artificial Intelligence Initiative Act of 2020 defines AI as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments".[6] These definitions show considerable overlap, of course, but also differ from one another. Significant variance is further noted throughout the multitude of state-level, agency-level, and international definitions of AI.

When considering these questions from an institutional perspective, consistency is more important than precision. Governance questions are best suited to focus on impacts, which suggest that a relatively open-ended definition is preferable.

The National Institute of Standards Technology (NIST), in its widely cited Artificial Intelligence Risk Management Framework (AI RMF), refers to an AI system as an engineered or machine-based system with varying levels of autonomy that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments.[7] NIST's definition is largely drawn from the Organization for Economic Co-operation and Development (OECD) Recommendations on AI, which were published in 2019.[8] Because the AI RMF appears to be gaining steam as the dominant decision-making model for agencies considering new AI use cases, this paper will adopt the NIST definition for the sake of consistency, with the general understanding that the term should be understood inclusively.

- **About the Project and Methodology**

This paper was developed through a contract with the Administrative Conference of the United States (ACUS), which called for a study, with accompanying recommendations, to "examine the potential benefits and risks of using algorithmic tools to support agencies' regulatory enforcement efforts and identify policies, practices, and organizational structures agencies can put in place to ensure they enforce the law fairly, accurately, and efficiently." Although the project description calls for research into the use of "algorithmic tools", this paper frames the discussion on AI for two reasons.

First, "algorithmic tools" are an extremely broad category of programs, which can include everything from a simple handheld calculator to the operating system used to type this paper. Most of these functions are not subjects of concern, and have been employed across the federal government for decades, attracting little controversy. What is novel is the technology's sophistication and, more importantly, its function, which has begun to complement, and in some cases even supplant, the judgment of humans within the public service. The issue, in other words,

[5] Add cite
[6] https://www.congress.gov/bill/116th-congress/house-bill/6216/text#toc-H41B3DA72782B491EA6B81C74BB00E5C0
[7] Add cite
[8] Add cite

is automation, and the new functions that these tools are beginning to play relative to the regulatory enforcement process, rather than anything inherent in the technology itself.

Second, while, as noted in the previous section, AI is also an imperfect and imprecise term, it has become the lingua franca across the U.S. government, and around the world, to address the economic and social challenges that are coming as a result of the rise of sophisticated machine learning tools which are likely to automate a broad range of work-related functions over the coming decade. This is appropriate when one considers that the core challenge related to these products is their capacity to displace human workers, especially in a decision-making function. AI, as a concept, is fundamentally about displacement of humans by machines, which cuts to the core of the social, legal, and administrative challenges that are the focus of this research project. Because the NIST definition of AI, when interpreted inclusively, already encompasses virtually all of the algorithmic tools which are matters of controversy due to their potential to supplant or replace human workers, this paper defers to using the term AI for the sake of simplicity and clarity.

The focus of this paper is also on regulatory enforcement, which is defined in the project as including detecting, investigating, and prosecuting current and potential noncompliance with the laws that agencies administer.[9] Some examples of these activities include operations at the Securities and Exchange Commission, whose mandate includes enforcing federal securities laws,[10] the Internal Revenue Service, which investigates financial crimes such as tax fraud,[11] and the Environmental Protection Agency, which enforces a range of rules related to pollution and waste products.[12] However, while regulatory enforcement is often a particularly high impact use case for AI, it is relatively unusual for accountability structures targeted at AI, in the US or elsewhere, to consider these applications as categorically different from other uses of AI within the decision-making process. As a result, while this paper is specifically focused on regulatory enforcement, most of the recommendations and conclusions are equally applicable to other high impact uses of AI, and are framed more broadly where this is the case.

The paper was developed through a yearlong research process that included consultation with a wide range of experts in AI and administrative law, as well as discussions of initial findings with audiences at Yale, UCLA, and at the University of California Center Sacramento. Although the research process included engagement with public service employees at the state and federal level, this paper does not provide an exhaustive list of use cases related to regulatory enforcement, though a few illustrative examples are included in Part II. This is because a general mapping was already carried out through an earlier ACUS report, *Government by Algorithm*, and because a comprehensive use case database is currently under development at AI.gov, including over 700 use cases as of September 2024.[13]

The author offers his sincere thanks to all of the experts who agreed to speak with me, including Kevin De Liban, Andrew Selbst, Margot Kaminski, Daniel Ho, Elham Tabassi, Janet Haven, Sanford Williams, Barry Johnson, Reza Rashidi, Phil Lindenmuth, Melodi Dincer, Marc-Etienne

[9] https://www.acus.gov/sites/default/files/documents/Algorithmic-Tools-in-Enforcement-RFP.pdf
[10] https://www.sec.gov/enforcement-litigation
[11] https://www.irs.gov/about-irs/office-of-fraud-enforcement-at-a-glance
[12] https://www.epa.gov/enforcement/national-enforcement-and-compliance-initiatives
[13] https://ai.gov/ai-use-cases/

Ouimette, Benoit Deshaies, Jith Meganathan, Achuta Kadambi, Yuan Tian, Nanyun Peng, Jon Michaels, and Tom Speaker. Thanks as well to student research assistants who contributed research or background material, including Yuyang (Kate) Hu, Nicholas Wilson, Alyssa Stolmack, and to Kazia Nowacki, Adam Cline, Jeremy Graboyes and the rest of the team at the Administrative Conference of the United States, for their helpful feedback and support in preparing this document.

All errors remain the sole responsibility of the author, and opinions expressed do not necessarily reflect those of the Administrative Conference of the United States.

II. Background: AI, Public Perceptions, and Administrative Agencies

Despite the interest surrounding OpenAI's release of ChatGPT in 2022, excitement around the promise of these new technologies, such as generative AI, has been met by an equal measure of public skepticism, and even fear, about their broader impact on humanity.[14] At the extreme, visions of a "Terminator-style" apocalypse driven by autonomous weapons systems have propagated, alongside highly publicized concerns about the supposed "existential risk" that these technologies pose to humanity.[15] Many experts consider the worst-case fearmongering about AI to be more reflective of excitement around AI than reality, and there has been considerable room for error within the doomsaying. One widely shared story concerning an AI-enabled drone, which had apparently opted to attack its operator when it found the restrictions on its use of force to be too onerous, was ultimately debunked when it turned out the scenario had been a thought-experiment, rather than an actual exercise.[16]

However, not all reports of AI malfeasance are bogus. One commonly cited cautionary tale concerns the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), a proprietary algorithm built by a private contractor to assess the risk of recidivism for criminal defendants.[17] COMPAS works by developing a "risk score" based on a questionnaire which is meant to predict the likely danger from a person's release. COMPAS, or systems like it, have been widely incorporated into sentencing or bond hearings, including in Arizona, Colorado, Delaware, Florida, Kentucky, Louisiana, Oklahoma, Virginia, Washington, and Wisconsin. However, an investigation in 2016 found that the software was returning results which were biased against Black subjects. The company's audits failed to capture these discriminatory impacts through their own internal assessments because their key performance indicator focused on accuracy, which was roughly equivalent between the different racial groups. The audit failed to uncover that the system tended to err by placing Black defendants into a higher risk category, and white defendants into a lower risk one. In other words, there was a mechanism in place to catch the problem, but it failed because the bias manifested in a way which was different from what the auditing program was looking for.

Another high-profile failure concerned the automation of allocation decisions for Medicaid resources, which led to drastic cuts in services for housebound patients in Arkansas, Idaho, and elsewhere.[18] Again, early complaints from individuals subjected to the systems' decision making went unheeded until litigation by the ACLU of Idaho and Legal Aid Arkansas forced the problems into the light.

Every use case for AI is different and should be evaluated based on its specific context and risk profile. There is a world of difference between, for example, an automated tool which manages the allocation of parking spaces for a large agency and one which is setting bombing targets or

[14] https://www.politico.com/newsletters/digital-future-daily/2023/09/25/ai-vs-public-opinion-00118002
[15] https://futureoflife.org/open-letter/pause-giant-ai-experiments/
[16] https://www.bbc.com/news/technology-65789916
[17] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[18] https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy

deciding whether a person should have access to home care visits or bail. However, the preceding cases are a good illustration of why agencies may face an uphill public relations battle in developing and deploying AI for sensitive or high impact purposes, such as regulatory enforcement. There are particularly fraught consequences when considering the importance of trust and perceptions of legitimacy on regulatory enforcement functions. Agencies that wish to capitalize on the potential benefits of AI face a pressing challenge of how to maintain trust and legitimacy while pursuing greater automation.

- **Trust and Legitimacy in Regulatory Enforcement**

In parallel to the broader expansion of AI across the economy, administrative agencies have begun to pilot these technologies for a range of functions related to regulatory enforcement. At the SEC, for example, a number of AI-based tools have been developed to support investigations into financial crimes. These include the Corporate Issuer Risk Assessment tool (CIRA), which includes a machine-learning component that identifies potentially suspicious filings to predict misconduct based on historical datasets.[19] Another SEC tool, the Abnormal Trading and Link Analysis System (ATLAS), attempts to detect insider trading through examining differences in behavior between traders that lost money versus those who profited. At the EPA, a proof-of-concept developed by their Office of Compliance in partnership with the University of Chicago has been used to target facility inspections more accurately and efficiently, resulting in a 47% improvement of detecting violations of the Resource Conservation and Recovery Act.[20] The Animal and Plant Health Inspection Service under the US Department of Agriculture has piloted a similar program aimed at improving the efficiency of methods to detect invasive pest species at ports of entry.[21] Although each of these programs pursues important goals, and the use of AI within their respective enforcement processes offers significant potential in terms of boosting efficacy and efficiency, significant questions remain regarding the broader impact of the expansion of AI in regulatory enforcement on agencies' public trust.

Public trust is a fragile thing which may be built up over generations and destroyed virtually overnight. The importance of cultivating legitimacy through popular participation, regular testing of bureaucratic expertise, and normative reflection of policy choices is baked into the foundations of America's administrative apparatus, particularly through the Administrative Procedure Act.[22] At the core of these procedural protections is the fundamental principle that, in establishing legitimacy, it may often be necessary to accept tradeoffs relative to other institutional values, particularly efficiency of operations.[23]

In addition to legal requirements, there are practical reasons why administrative agencies should prioritize efforts to cultivate public trust and popular perceptions of their legitimacy. Where agencies issue rules that impact the public, and particularly where public compliance is necessary for the success of their mandate, perceptions of legitimacy are of paramount importance to

[19] Government by Algorithm, p. 23.
[20] https://www.epa.gov/data/epa-artificial-intelligence-inventory
[21] https://www.usda.gov/data/ai_inventory.csv
[22] Bruce Ackerman, *The New Separation of Powers*, 113 HARV. L. REV. 633, 697 (2000).
[23] Lisa Shultz Bressman, *Beyond Accountability: Arbitrariness and Legitimacy in the Administrative State*, 78 N.Y.U. L. REV. 461, 546 (2003).

encourage respect for agency authority and voluntary compliance.[24] While the threat of sanctions may loom large over individual decisions to file a tax return, the system, as a whole, depends on voluntary compliance. The sovereign citizen movement is a good illustration of the challenges and frustrations in trying to deal with a group that rejects the legitimacy of government agencies, even if the group's perceptions are based on spurious logic. An agency's sense of legitimacy must be resilient enough to survive even individual unpopular agency decisions or outright errors, which should ideally be seen as exceptions within a fundamentally valid structure even by those who are disadvantaged by them.[25] Public relations are an important operational consideration for modern executive branch structures for practical, as well as political, reasons.[26]

None of this is to argue that agencies should be overly defensive about how they act, or their place in the constitutional order. In *The Procedure Fetish*, Nicholas Bagley argues persuasively that, contrary to concerns about some democratic deficit that is inherent to the administrative state, agencies are natural outgrowths of America's democratic and constitutional structure, which includes balanced roles for the executive, legislative, and judicial branches in maintaining robust public accountability over their operations and decision-making.[27] Nonetheless, in considering their operational future, it would be shortsighted for agencies to ignore the mounting attacks on their legitimacy. While the sovereign citizens are a fringe movement, public skepticism of administrative agencies extends far more broadly.[28] It also appears to be reaching a crescendo in the present political moment, as calls for encouraging a significant reduction in the reach of administrative agencies have become firmly entrenched in the political mainstream.[29] While there is no question that politicians have been instrumental in driving this narrative, they are also responding to public sentiment which shows that Americans' trust in government is at historically low levels.[30] . It is also worth noting that public support for governments' use of AI tends to correlate with trust in government more generally, leading to the potential for vicious (or virtuous) cycles as adoption accelerates.[31] In this moment, it is critically important to scrutinize how automation in general, and the use of AI in particular, are likely to impact public, judicial, and political perceptions of administrative agencies' role in American governance and the legitimacy of their regulatory enforcement functions.

- **Expertise, Discretion, and Regulatory Enforcement**

A traditional justification for the administrative state, generally, is that the technological complexity of the modern world necessitates a level of regulatory expertise which is beyond the capacity of Congress to independently regulate.[32] This argument featured heavily in Justice

[24] TOM R. TYLER, WHY PEOPLE OBEY THE LAW 4 (2006).
[25] Adrian Vermeule, *Bureaucracy and Distrust: Landis, Jaffe, and Kagan on the Administrative State*, 130 HARV. L. REV. 2463, 2463 (2017).
[26] DANIEL CARPENTER, REPUTATION AND POWER: ORGANIZATIONAL IMAGE AND PHARMACEUTICAL REGULATION AT THE FDA (2010).
[27] Nicholas Bagley, *The Procedure Fetish*, 118 Mich. L. Rev. 345, 377 (2019).
[28] See, e.g., DAVID SCHOENBROD, POWER WITHOUT RESPONSIBILITY: HOW CONGRESS ABUSES THE PEOPLE THROUGH DELEGATION 14–18 (1993).
[29] See, e.g., https://www.cnn.com/2023/08/23/politics/republican-government-cuts-what-matters/index.html.
[30] https://www.pewresearch.org/politics/2024/06/24/public-trust-in-government-1958-2024/
[31] https://www.bcg.com/publications/2019/citizen-perspective-use-artificial-intelligence-government-digital-benchmarking
[32] Kenneth Culp Davis, *A New Approach to Delegation*, 36 U. CHI. L. REV. 713, 715 (1969).

Kagan's dissenting opinion in *Loper Bright*, which queried how district and appellate court judges were meant to know things like whether or not Washington's western grey squirrels were a distinct population or if alpha amino acid polymers are proteins.[33] Today, a similar argument undergirds some of the drive towards incorporating AI into regulatory enforcement, as continued technological progress means that even human experts are outmatched by the volume and complexity of the regulatory challenges they face. Part of this is a matter of scale. When the Securities and Exchange Commission was created in 1934, around 300,000–400,000 shares were traded every day on the New York Stock Exchange. Today, that number is in the billions, presenting a far more difficult challenge to track suspicious activity. For another example, healthcare devices have become increasingly complex and specialized, with some integrating AI into their features. This presents a unique challenge to the Food and Drug Administration's typical regulatory paradigm due to their tendency to degrade or drift after approval,[34] requiring new and data intensive modes of post-market surveillance.[35] In other words, the spread of AI-enabled devices across the healthcare system increases the need to equip regulators with AI-enabled oversight tools.

The corollary to this increasing focus on automation in regulatory enforcement is an erosion of the human elements in administrative processes, and potentially a hollowing out of the expertise which is a core pillar of legitimacy underlying administrative agencies. This is true not only in the legal sense, as noted by Justice Kagan above, but among public perceptions. Survey data shows a strong correlation between perceptions of administrative expertise and public perceptions of legitimacy.[36] Some of the most trusted federal agencies, such as the Federal Reserve, maintain this status despite being relatively light in terms of the procedural rigor underlying their decision-making.[37]

The spread of AI in regulatory enforcement presents a potential existential challenge to administrative agencies, since it outsources human expertise to a set of hard-coded rules interpreted and enforced by machines. This can obviously play into existing concerns about opacity in administrative operations, as well as perceptions of arbitrariness, due to the inscrutability of AI decision-making. The fact that administrative agency staff may not be capable of controlling or even explaining the outputs of their tools poses a challenge to the expertise and exercise of discretion which are a key underlying justification supporting the administrative state.[38]

Where AI systems are being developed by third-party contractors, accountability is further stymied by trade secrecy claims, which can serve as an additional shield against external accountability. The tension between transparency and effective law enforcement is not unique to the AI realm,

[33] Loper Bright Enters. v. Raimondo

[34] AI models tend to change, and often degrade in their performance, in subtle ways over time, which poses an oversight challenge to systems which are designed to track significant material changes. See https://www.ibm.com/topics/model-drift.

[35] Akshay Sreekumar & Peter Horton, *Liability Preemption in the New Regulatory Framework of Data Driven Healthcare*, UCLA INSTITUTE FOR TECHNOLOGY, LAW & POLICY (2022), https://www.dropbox.com/scl/fi/2gfq6gdx1dd7wsy4fxrvp/LIABILITY_AND_PREEMPTION.pdf?rlkey=bbe36xhus19qiw2pyumx2xy92&e=1&dl=0.

[36] https://bpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/2/3167/files/2022/01/bureaucratic_trust.pdf

[37] Nicholas Bagley, *The Procedure Fetish*, 118 Mich. L. Rev. 345, 382 (2019).

[38] Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L. J. 798, 818 (2021).

and it manifests in some form under virtually every freedom of information or right to information framework.[39] Police routinely complain that disclosures of information about their investigative techniques, whether in response to an information request or in a judicial context, will compromise the efficacy of their operations and help bad actors to get away.[40] However, opacity can also undermine efforts to cultivate public trust in favor of the use of AI in regulatory enforcement, frustrating the ability to track and monitor a system's performance and to obtain buy-in from impacted communities and other key external stakeholders. While increasing model complexity and adding randomness can make AI systems harder to game, it creates a separate tradeoff related to interpretability.

Similarly, while the democratic tensions flowing from the government's increasing reliance on private sector contractors are not unique to AI, they take on greater salience in the context of AI due to the increasing level of autonomy exercised by AI in the decision-making process, as well as the challenges that agency staff face in explaining their outputs.[41] In a 2021 article on the subject, Ryan Calo and Danielle Keats Citron pointed out that, if regulatory enforcement is essentially being delegated to AI tools supplied by third-party contractors, there is an argument to be made that administrative agencies are no longer necessary at all, since Congress could just as easily contract directly with the companies providing the enforcement tools in order to achieve their regulatory aims.[42]

While challenges around trust and legitimacy are paramount, since they cut to the core of administrative agencies' functions and mandates, there is a laundry list of other concerns related to AI's integration in regulatory enforcement. These include the fundamentally regressive nature of AI, since the systems necessarily rely on historical data to form their understanding of a particular challenge.[43] Like the proverbial general focused on fighting the last war rather than the next one, AI's dependence on data from earlier enforcement efforts may render it ill-equipped to handle the dynamic and adversarial nature of modern malfeasance. These challenges may be compounded if the data that they are trained on contain errors or undue amounts of noise, though there are active learning methods which can mitigate this problem.[44]

There are also concerns that AI-enabled regulatory enforcement tools may be subject to gaming, particularly given how well-resourced many enforcement targets are. While this challenge is not specific to machines, Daniel Ho, et al, raise particular concerns around the risk that third-party contractors who design these tools, or employees within the contracting agencies, may sell their

---

[39] *See, e.g.*, MICHAEL KARANICOLAS ET AL, INTERPRETATION OF EXCEPTIONS TO THE RIGHT TO INFORMATION: EXPERIENCES IN INDONESIA AND ELSEWHERE (2012), p. 75-85, https://www.law-democracy.org/wp-content/uploads/2010/07/Interpretation-of-Exceptions-To-the-Right-To-Information-Experiences-in-Indonesia-and-Elsewhere.pdf.

[40] Hannah Bloch-Wehba, *Visible Policing: Technology, Transparency, and Democratic Control*, 109 CAL. L. REV. 917, 964-965 (2021).

[41] GOVERNMENT BY CONTRACT: OUTSOURCING AND AMERICAN DEMOCRACY (Jody Freeman & Martha Minow eds., 2009).

[42] Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L. J. 798, 818 (2021).

[43] AIURELIEN GERON, HANDS-ON MACHINE LEARNING WITH SCIKIT-LEARN, KERAS, AND TENSORFLOW: CONCEPTS, TOOLS, AND TECHNIQUES TO BUILD INTELLIGENT SYSTEMS (Rachel Roumeliotis & Nicole Tache eds., 2nd ed. 2019) p 25-26.

[44] *Geron Id 27, 89-90.*

knowledge about the inner workings of the machines to enforcement targets or other parties with skin in the game.[45] One can easily imagine that a person with knowledge of what kinds of patterns are likely to be flagged by an AI tool tracking suspected insider trading for the SEC could be tempted to sell this knowledge or even exploit it themselves.[46] Potential manipulation can come in more subtle varieties, such as knowing what keywords are likely to route a patent application to an examiner with a particularly favorable rate of approval.[47]

AI also presents a challenge to the importance of discretion in agency enforcement decisions, since these systems typically deal poorly with edge cases, where an enforcement decision could conceivably go in either direction. Cascading failures are another concern, as are broader worries about the data and energy intensive models that these systems rely on. and the implications of their development and expanding use for privacy and for the environment. Kate Crawford has worked extensively to document the massive energy and extractive costs that flow from the rise in popularity of generative AI.[48]

Together, these concerns paint a picture of a need for administrative agencies to tread cautiously in adopting AI, particularly for contentious or sensitive applications such as regulatory enforcement. While it is understandable that agencies are keen to shed public perceptions of bureaucratic inefficiency and portray themselves as being on the leading edge of technological innovation, support for more technologically enhanced government is likely to evaporate if the public comes to believe that AI systems are not trustworthy in performing these functions.[49]

In the next section, the existing legal landscape in the United States and elsewhere is discussed to present a snapshot of the regulatory safeguards that currently exist, and the conceptual gaps in these models.

[45] Ho at 63.
[46] Ho at 87.
[47] Ho at 87.
[48] https://www.nature.com/articles/d41586-024-00478-x. See also Kate Crawford, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.
[49] ACUS p. 7.

III. The Existing Governance Landscape

Although successive U.S. administrations have taken an increasingly keen interest in the economic and sociological ramifications of AI, there has been relatively little specific focus on the technology's impact on regulatory enforcement, and only sporadic attention paid to its impact on government operations more generally. Legislative attention has been far more concerned with the private sector companies developing these new technologies, and in establishing an appropriate regulatory framework for their work, than on parallel developments across the administrative state.[50]

Even as Obama-era recommendations for studying AI's potential functions[51] gave way to more targeted sets of principles, such as the "Blueprint for an AI Bill of Rights,"[52] considerations of the appropriate safeguards and regulations for this technology remain relatively high-level and general. Policy development processes are likely to advance particularly slowly in the regulatory enforcement space, since guiding rules or principles are likely to be tested, refined, and, in some cases, nullified through successive waves of judicial review.

The two most significant moves to develop more concrete and actionable standards for how the administrative state should approach AI in a regulatory enforcement context have come from the National Institute of Standards and Technology (NIST) and the Office of Management and Budget (OMB).

In January 2023, NIST published its "Artificial Intelligence Risk Management Framework" (AI RMF), which provides a model assessment process for agencies to map potential risks, develop tracking mechanisms, and respond appropriately. The AI RMF establishes a taxonomy of potential risks flowing from the use of AI. Challenges defined under the AI RMF include reliability, accuracy, robustness, resilience, security, accountability, explainability, interpretability, privacy, fairness, and bias. The AI RMF provides a model risk assessment process for agencies to map potential risks, develop tracking mechanisms based on this mapping, measure risks as they emerge, and manage and respond appropriately.

Considered in the context of regulatory enforcement, it is particularly important for risks to be assessed institutionally and systematically, rather than purely from the perspective of harms that flow directly to the subjects of the decisions or other direct stakeholders. A robust assessment process should consider the risk that a system, even if it works perfectly, might nonetheless serve to undermine confidence in an agency and perceptions of legitimacy. Likewise, while human performance may provide a useful baseline for comparison, the fact that an AI program may return a lower level of erroneous decisions as compared to a traditional decision-making system, or lower levels of bias, should not be the end of the conversation in assessing whether it is fit for a purpose. If subjects of AI-driven enforcement decisions perceive that they are more unfair or arbitrary, then it is possible that the deployment of these systems will have a net negative impact on an agency's

[50] *See, e.g.*, https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047

[51] https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf p. 16.

[52] https://www.whitehouse.gov/ostp/ai-bill-of-rights/

legitimacy, even if the new system actually improves the accuracy of decision-making. Risk, in other words, should be understood holistically, and considered in the context of an entire organization, as opposed to limiting the assessment to a particular process. Similarly, while the AI RMF specifically mentions the importance of engaging with impacted communities, in the context of regulatory enforcement the perspectives of targets of enforcement, as well as communities that are otherwise impacted by decisions not to prosecute, should be complemented by considerations of the perspectives of the broader public.

Another important characteristic of the AI RMF is that it emphasizes ongoing evaluations throughout the AI lifecycle.[53] In other words, while early-stage assessments to inform decisions on whether to proceed with regulatory enforcement are vital, approval for a system to be developed or deployed should not mark the conclusion of the risk assessment process. Risks, impacts, and trade-offs should be mapped on an ongoing basis and include continuous assessment of whether the technology is delivering as promised or whether it is proving unfit for its purpose. A willingness to retire poor-performing systems, and to avoid falling victim to a sunk-cost fallacy, is vital. In the context of regulatory enforcement, allowing space for systems to be phased out presents a particular challenge insofar as appeals against adverse agency decisions may carry on for years. A post-hoc admission that an AI system—which was instrumental to previous enforcement decisions—is no longer fit for its purpose may undermine the agency over the course of appeals or reconsideration processes. This is an understandable disincentive to critical assessment. However, avoiding such a determination would only cause delay and make the inevitable decision to discontinue a problematic system even more difficult.

While the AI RMF provides a useful starting point, the format leaves significant discretion to implementing agencies. Risks are ultimately contextual determinations. They are resistant to centralized definition since they depend on the particular use case. There can even be significant variance within individual tools. A system may manifest a particular risk profile at the testing phase and introduce completely different problems in implementation.[54] The human element is also a significant factor to consider. Risk may depend on the individuals who are interacting with or using the tools, as well as their expectations and perceptions of its capabilities. The AI RMF relies on a sense of collective responsibility for managing the impacts of AI across the implementing agency, emphasizing the importance of diversity among the team considering potential risks.[55] It also requires a willingness to ask difficult and resource-intensive questions about the tradeoffs flowing from various use cases.

The second major regulatory development which is worth flagging is OMB's AI policy memo, which was published in March 2024.[56] The memo includes a number of requirements and recommendations for executive branch agencies, including the designation of a Chief AI Officer and the establishment of AI Governance committees at CFO Act agencies to guide and coordinate issues related to AI implementation, including managing risks.

[53] AI RMF p 11.
[54] NIST RMF P. 24
[55] NIST RMF P. 15
[56] https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

Probably the most noteworthy aspect of the memo is the requirement for agencies to track and publicly report all AI use cases, as well as to identify where AI uses are "rights-impacting" or "safety-impacting." The memo also includes substantial discussion of risk management and mitigation efforts regarding rights-impacting or safety-impacting uses of AI, including conducting impact assessments and real-world testing, and implementing measures to address discrimination. There is significant overlap between these requirements and the content of the AI RMF. However, the introduction of a centralized framework for collecting these assessments and monitoring how these technologies are being deployed is a vital addition. The lack of such tracking has been a significant impediment to efforts to craft an appropriate response to the use of AI across the federal government. It is difficult to come up with a coherent public policy response if not everyone has a comprehensive understanding of the implications associated with using these technologies.

OMB's policy memo serves as an initial attempt to corral federal agencies around a rough and general set of standards by developing a framing of how and where these technologies are being used and by encouraging agencies to construct a model of different types of risk and accompanying mitigation strategies. In response to this prompt, agencies have begun issuing their own internal guidance on the use of AI. One example is the IRS, which in May 2024 issued an interim guidance memorandum which, among other things, establishes a use case inventory and defines an approval and workflow approving new AI applications, as well as establishing minimum practices for safety-impacting or rights-impacting AI.[57] The IRS interim guidance memorandum also designates a cross-functional AI Assurance Team and AI Project Teams to review and execute key governance steps, including steps such as impact assessments and ongoing risk evaluations, which broadly follow the standards spelled out in the AI RMF.

While the OMB policy memo and the AI RMF are the most prominent frameworks, there are other institutional actors which are relevant to potential regulatory efforts. These include ACUS, which in 2020 published a set of standards and considerations for federal agencies using artificial intelligence, including related to transparency, bias, capacity, procurement, privacy, data management, security, oversight, and decisional authority.[58] In 2021, the Government Accountability Office (GAO) published its own "Guidance for Creating Agency Inventories" around federal agency uses of AI. It is worth noting, however, that a follow-up study documented widespread noncompliance: of the 19 agencies to whom GAO offered recommendations, only 10 fully agreed to comply.[59] Compliance challenges are a major issue underlying any attempt to create effective standards across the executive branch, and there is a long-running debate regarding the relative benefits of binding, sanctions-based systems versus more informal structures built around administrative support and capacity-building.[60] Without commenting on the substance of the GAO recommendations, it is worth noting that there are few things which can undermine the perceived legitimacy of an administrative oversight structure more than issuing a recommendation or requirement which is subsequently ignored.[61] This is not to argue against ambitious standards or the demand for compliance with robust best practices, but it does illustrate the importance of

[57] https://www.irs.gov/pub/foia/ig/spder/interim-guidance-raas-10-0524-0001-artificial-intelligence-governance-and-principles-redacted.pdf
[58] https://www.acus.gov/document/statement-20-agency-use-artificial-intelligence
[59] https://www.gao.gov/assets/d24105980.pdf
[60] Michael Karanicolas & Margaret B. Kwoka, *Overseeing Oversight*, 54 CONN. L. REV. 655 692-5 (2022).
[61] Nicholas Bagley, *The Procedure Fetish*, 118 Mich. L. Rev. 345, 392 (2019).

ensuring that the oversight bodies are equipped with the tools and resources to spur meaningful change across administrative agencies.

- **Gaps in the Regulatory Environment**

The emerging regulatory environment, in guiding how AI may be used in administrative enforcement, places a heavy emphasis on transparency. While transparency is an essential ingredient in any effective oversight structure, it does not by itself present a solution, or even a response, to the accountability and other structural challenges posed by AI's deployment.[62] The AI RMF notes that accountability presupposes transparency.[63] But while the latter is a precondition for the former, transparency is not itself sufficient to provide robust accountability.

Explainability, or the ability to characterize AI decisions in a way which renders their reasoning comprehensible to humans, is often cited as another key value, though it, too, is not an end in itself.[64] Rather, explainability is valuable because of its utility in facilitating meaningful review, supporting human autonomy, facilitating due process, strengthening perceptions of legitimacy, and providing guidance to future decision-makers.[65] The value of explainability, in other words, depends on complementary mechanisms to support these follow-on goals.

In dealing with high-risk applications, such as regulatory enforcement, placing a "human in the loop" is often emphasized as a mitigation tactic. This is unsatisfactory as a solution, in part because creating meaningful review over AI-generated decisions requires more than just human intervention. The oversight must be meaningful, and humans have a tendency to defer to automated recommendations. In addition, as due process rights related to high stakes decisions made by machines escalate, it begs the question as to whether the purported efficiency gains through the use of these systems may, in some instances, be illusory. The intensity of human review required to ensure that enforcement decisions have meaningful oversight may require just as many man-hours as having a human carry out the decision-making process independently. In such circumstances, it is worth asking whether a blanket prohibition against the use of AI as anything other than a research tool for human staff may be preferable in certain highly sensitive or contentious enforcement roles.

A related problem, which pervades much of the AI governance space, relates to challenges in connecting the general principles that usually ground high level guidance to more concrete and operational directions. As Cary Coglianese noted, it is one thing for governments to dictate that AI systems should be "fair", "safe", "explainable" and so forth: but determining what that means from an operational perspective is an entirely different matter.[66] The absence of clear performance standards, Coglianese observes, is what gives rise to a reliance on management-based regulation, which relies on process and protocol rather than attempting to achieve particular outcomes.[67] This, in turn, leads to new challenges as regulatory agencies grapple with how to ensure that risk

[62] See infra Section IV for a full discussion of transparency and best practices.
[63] AI RMF p. 15.
[64] https://www.ibm.com/topics/explainable-ai
[65] Lost in Translation p. 18. More details in notes.
[66] https://theregreview.org/2024/01/15/coglianese-how-to-regulate-artificial-intelligence/
[67] *Id.*

assessments and other related processes are actually meaningful, as opposed to becoming mere paperwork exercises.[68]

However, probably the most pervasive structural weakness in existing accountability frameworks is that they assume a continued organic expansion of the use of AI across administrative agencies. Government attitudes to the use of AI strongly emphasize the importance of ensuring that agencies have adequate space to experiment and pilot new applications, with challenges around accountability and legitimacy to be addressed reactively. The OMB policy memo specifically points to the need to remove barriers to the responsible use of AI and achieve enterprise-wide improvements in AI maturity.[69]

A sense of apprehension at how to structure specific binding standards is understandable, given the novelty of these technologies, their complexity, the rate at which they are evolving, and the incredible range of functions where they are being piloted. At the same time, commenters talk of an "avocado ripeness problem" in finding the opportune time to impose strict regulations: just as an avocado can seemingly transition directly from being underripe to overripe, there is a thin line between when a fast-moving technology is too novel for observers to see clearly and understand its inherent risks, and when its use has become so deeply ingrained in government or the economy as to make effective regulation impossible.[70]

Before moving on, it is useful to examine a few comparable frameworks from the state and international levels, to develop a broader sense of how regulators elsewhere are responding to the emerging challenges posed by AI in regulatory enforcement.

- **AI Regulation at the State Level**

There have been a number of state-level initiatives aimed at regulating AI across the public sector. Though relatively few of them apply specifically to regulatory enforcement, several would likely impact the use of AI by the relevant state governments in different ways.

In California, AB 302 requires the California Department of Technology to coordinate with other interagency bodies to compile a comprehensive inventory of "high-risk automated decision systems" that state agencies are using, developing, or procuring.[71] These systems are defined as those that assist or replace human decision-making and have significant legal impacts, such as around access to housing, education, employment, credit, healthcare, and criminal justice.[72] The inventory must detail the decisions these systems make, the data they use, and any measures to mitigate risks, including cybersecurity and bias.[73] The Department must submit this inventory in a report to the State Legislature by January 1, 2025, and annually thereafter.[74]

[68] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3598264#page=19
[69] Policy Memo p. 9.
[70] https://cyber.jotwell.com/what-sts-can-and-cant-do-for-law-and-technology/
[71] Cal. Gov. Code § 11546.45.5 (2023).
[72] Cal. Gov. Code § 11546.45.5, subd. (a)(4) (2023).
[73] Cal. Gov. Code § 11546.45.5, subd. (a)(4) (2023).
[74] Cal. Gov. Code § 11546.45.5, subd. (d)(1) (2023).

Although the AB 302 does not apply specifically to regulatory enforcement uses, these will likely be considered among the "high risk" uses being developed, and the requirements to document and evaluate the performance of these systems, and to implement safeguards against risks such as discrimination or cybersecurity threats, will likely have a substantial impact on how these technologies are deployed.

Another law under consideration in California is SB 896, the Generative Artificial Intelligence Accountability Act. Again, this is not specifically targeted at regulatory enforcement (or even public sector) functions, though it encourages agencies to engage in GenAI-focused rulemaking to clarify if and how existing regulations apply to GenAI or other automated decision-making systems.[75] The bill also mandates that state agencies using GenAI inform members of the public when they are interacting with GenAI regarding government services and benefits.[76]

The Maryland Artificial Intelligence Act, which was passed in May 2024, requires state agencies to develop a publicly available inventory of all systems using high-risk AI, including basic information about the AI systems such as their purpose and intended use.[77] A newly-created AI Subcabinet is tasked with defining "high-risk AI", though it will also have a broader mandate to support AI innovation across the state government.[78] Earlier versions of the bill contemplated additional responsibilities, like identifying best use cases across state government units and testing proofs of concept, though these were ultimately excluded from the final draft.

In Washington, SB 5356 has been under discussion, in various forms, since at least 2021, and would require public notice and accountability measures for automated decision-making tools used by state agencies to produce legal effects on natural persons.[79] These include requiring each agency to complete an algorithmic accountability report for each automated decision-making tool in use, as well as to require agencies to notify people impacted by the use of automated decision-making tools of the system's use, how to contest any decision involving an automated decision-making tool, and the degree to which human review resulted in the final decision, among other things.[80] The bill would also make any decision made or informed by an automated decision-making system subject to appeal "if a legal right, duty, or privilege is impacted by the decision".[81]

In Illinois, there are, as of August 2024, several bills relevant to the use of AI in regulatory enforcement under consideration. HB 5116, known as the Automated Decision Tools Act, applies to deployers, including in administrative agencies, that use an automated decision tool, including those powered by AI, to make consequential decisions that produce significant effects on a person's life and livelihood.[82] HB 5116 imposes several requirements, including that deployers conduct and submit annual impact assessments, and that they inform individuals when an automated decision tool is used to make or influence a consequential decision about them. HB

[75] Cal. Senate Bill 896 § 2, subd. (h) (Aug. 19, 2024).
[76] Cal. Senate Bill 896 (Aug. 19, 2024).
[77] Md. S.B. 818 (2024).
[78] 496 Md. 3.5–803, subd. (A)(1); 496 Md. 3.5–801, subd. (D)(1) (2024).
[79] Wash. S.B. 5356 (2024).
[80] Wash. S.B. 5356 § 4, subd. (8)(a), § 5, subds. (1)(f), (4) (2024).
[81] Wash. S.B. 5356 § 4, subd. (8)(c) (2024).
[82] H.B. 5116, 103rd Gen. Assemb. (Ill. 2024).

5116 would also require each deployer to develop and maintain a governance program to mitigate the risks of algorithmic discrimination.

HB 4705, the Artificial Intelligence Reporting Act, would require each state agency to designate a Chief Artificial Intelligence Officer from its existing staff to prepare an annual report detailing its use of covered algorithms for operations including enforcement.[83] These reports are to be published publicly by the Department of Innovation and Technology.

Finally, HB 4836 requires state agencies using AI systems, as well as entities deploying state-funded AI systems, to adhere to NIST standards for trustworthiness, equity, and transparency, and to submit algorithmic impact assessments based on the AI RMF to the Auditor General, and the Department of Innovation and Technology.[84]

- **International Case Studies: European Union**

Probably the best known and most influential international model is the European Union's *AI Act*, which imposes a sliding scale of requirements based on the purported risk of the use case, including obligations related to transparency, auditing, and oversight.[85] However, the *AI Act* also prohibits uses of AI which are deemed unacceptably risky. Although the *AI Act* is not specifically targeted at the public sector, the latter category includes some government applications, particularly the use of AI for predictive policing, to develop social credit scores, or for real-time biometric tracking in public spaces. The *AI Act* also includes some discussion of public sector uses that would be considered high risk, including any use of these systems to determine access to essential public services and benefits (such as healthcare), as well as all uses related to law enforcement, migration, border control, and the administration of justice and democratic processes. High risk systems are also required to be registered in a public database unless their uses are for law enforcement or migration. The *AI Act* also contains blanket exclusions for AI systems that are exclusively designed for military, defense, or national security purposes.

The core of the mitigation practices envisioned by the *AI Act* revolve around a conformity assessment, designed to ensure that the system complies with data quality, traceability, transparency, human oversight, accuracy, cybersecurity, and robustness standards. The assessment is meant to be repeated every time the system or its purposes is substantially modified, though defining a substantial modification may pose a conceptual challenge as a result of the tendency of some AI systems to change in steady but subtle ways after they enter the market.[86] The *AI Act* also requires the development of risk management systems that include testing and assessment at both the piloting and the post-market phases, with accompanying reporting requirements, as well as requirements related to transparency, accuracy, and data quality.

---

[83] H.B. 4705, 103rd Gen. Assemb. (Ill. 2024).
[84] H.B. 4836, 103rd Gen. Assemb. (Ill. 2024); U.S. DEP'T OF COM., NAT'L INST. STANDARDS AND TECH., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK: GENERATIVE ARTIFICIAL INTELLIGENCE PROFILE (2024), https://doi.org/10.6028/NIST.AI.600-1.
[85] *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* COM (2021) 206 final (Apr. 21, 2021), at 5.2.2 [https://perma.cc/NLS2-AY53].
[86] https://www.ibm.com/topics/model-drift

Key oversight functions from the *AI Act* are delegated to technical standard setting organizations, though the main enforcement responsibilities are in the hands of national authorities, as well as the European Commission's AI Office, which provides strategic guidance and governs general-purpose AI models. It is worth noting that the *AI Act* includes substantial sanctions, of up to €35M or 7% of the total worldwide annual turnover of the preceding financial year (whichever is higher) for infringements. The *AI Act* also envisions regular audits and post-market monitoring by these authorities. Together, the rules are meant to create a system which is highly adaptable and iterative, in line with the evolving nature of the underlying technologies.

- **International Case Studies: Canada**

In Canada, government uses of AI are governed by the *Directive on Automated Decision-Making* (the *Directive*), under the auspices of the Treasury Board of Canada Secretariat.[87] Canada was an early mover in the AI governance space, having passed the *Directive* in 2019, though it has been subject to regular updates since then. Although the law does not specifically distinguish between uses for regulatory enforcement and other functions, it is limited in its application to cases where the AI system is processing "client data," which in practical terms means that it is heavily focused on cases where a person or organization is seeking a government service or benefit, or is a target of enforcement. In other words, use cases like background policy research or personnel management functions are outside the purview of the rules. It is also worth noting that the *Directive* is forward-facing, only applying to applications subsequent to its entry into force, which allows for a gradual ramp up of oversight responsibilities.

The *Directive* relies on a combination of public accountability and risk-based impact assessments to support responsible use of AI. It introduces a requirement to carry out an Algorithmic Impact Assessment (AIA) prior to the production of an automated decision-making system, and to publish the results online.[88] The impact of the decision and the importance of the rights or interests engaged leads to a sliding scale of obligations. At lower levels, these include requirements for data bias testing and the provision of generalized explanations for common decision results. At the higher end, requirements include human intervention in the decision-making process, publication and peer-review, the provision of a "a meaningful explanation" for negative outcomes, and Treasury Board approval for the system to operate.[89]

Regarding public accountability, the *Directive* introduces a robust notification requirement, mandating institutions that utilize automated decision-making systems provide clear, prominent, and plain-language notices to the public of this fact on their website.[90] The *Directive* requires that AIAs must be published online, with the intention of spurring public engagement to ensure the process is meaningful. The *Directive* also includes recommendations for consultation and engagement with impacted communities, though cost and logistical concerns mean that these are not currently required.

[87] Treasury Board of Canada Secretariat, *Directive on Automated Decision-Making* (2019), http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592&section=html.
[88] https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html.
[89] *Directive on Automated Decision-Making*, *ibid* at Appendix C.
[90] *Directive on Automated Decision-Making*, *ibid* at s 6.2].

Like its American counterparts, the *Directive* places few hard and fast restrictions on the use of AI for various applications, though cases using higher levels of risk require approval from a senior political appointee. In practice, administrative agencies have been reticent to hand over direct authority to automated decision-makers, instead incorporating them as research or assessment tools to aid human decision-making. In Canada, as elsewhere, there is a tension between the need for robust centralized oversight and the requirement that impact assessments be delegated to those with the greatest contextual understanding of a particular use case. Although Treasury Board involvement in most assessments is not strictly mandatory, in practical terms agencies have been keen to draw on the expertise that TBS is able to offer in developing a robust assessment process.

- **International Case Studies: Singapore**

Singapore has been another early leader in developing AI governance structures, particularly through the launch of its *Model AI Governance Framework* and, more recently, the development of *A.I. Verify*, a testing framework toolkit to support self-assessment by those developing or deploying AI technologies.[91] These frameworks are framed as voluntary guidance, rather than strictly binding requirements.

The *Framework* leans heavily on the values of explainability, transparency, and fairness, as well as emphasizing that the technologies should be human-centric and focused on supporting human capabilities and the interests of human beings. It also emphasizes the importance of iteration, calling on relevant bodies to institute a documented review process which will "continually identify and review risks relevant to their technology solutions, mitigate those risks, and maintain a response plan should mitigation fail."[92]

In addition to emphasizing the importance of assessing data sets for inaccuracy or bias, including through maintaining robust records of data provenance and lineage, the *Framework* suggests differentiating the data sets used for training, testing and validation. The *Framework* also suggests expanding human oversight and human involvement in decision-making where risk is particularly heightened, with the latter circumstance being defined as a multiplier of the severity of harm by the probability of harm.

[91] *Singapore's Approach to AI Governance*, PERSONAL DATA PROTECTION COMM'N (May 2022), https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework [https://perma.cc/G45K-BAAL].
[92] *Model AI Governance Framework* at 29.

IV. Key Values Underpinning an Appropriate Framework for AI in Regulatory Enforcement

The use of AI in regulatory enforcement presents opportunities to shift traditional administrative paradigms in novel and valuable ways. Because these systems are more malleable than human decision-makers, they offer new possibilities for achieving regulatory objectives while combating bias. Aside from potential gains in efficiency and processing power, machines can be fine-tuned and pushed in desired directions in a way that human staff cannot. However, their effective use depends on their ability to be deployed in a manner which maintains public trust in the federal government. Trustworthiness is a challenging commodity, especially as it pertains to complicated institutional structures. Moreover, an agency or system is only as trustworthy as its weakest characteristics.[93]

This section considers the earmarks of a strong system of oversight for the use of AI in regulatory enforcement and provides recommendations to safeguard the legitimacy of the federal government in the context of expanding experimentation with AI.

A. Understanding Risk and Risk Assessments

In the United States and around the world, the dominant governance model for AI focuses on assessing and mitigating risk. This idea is central to NIST's AI RMF and virtually every other major guidance document published across the executive branch, as well as to parallel efforts in the European Union, Canada, and Singapore.

It is easy to understand the appeal of a risk-based framework since it builds on existing models of regulation that are applied to a range of other roughly analogous harms—from environmental pollution to privacy and human rights impacts.[94] A commonality between these categories of harm is that they are all diffuse and difficult to measure or establish strict causality for. Algorithmic impact assessments have emerged as a core component of responsible AI use, as a successor to established models for environmental impact assessments, privacy impact assessments, and human rights impact assessments.[95]

Much of the momentum in favor of risk-assessment models lies in this familiarity, which may be particularly valuable in attempting to build guardrails around a novel and fast-moving technology like AI. But the relatively long track record for this model of governance also demonstrates that, along with its strengths, there are weaknesses and blind spots.

---

[93] RMF p. 12.

[94] A. Michael Froomkin, *Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements*, 5 UNIVERSITY OF ILL. L. REV. 1713, 1757–58 (2015); *Report of the Special Representative of the Secretary General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie: Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, UNHRC, 17th Sess, UN Doc A/HRC/17/31 (2011), https://www.ohchr.org/Documents/Issues/Business/A-HRC-17-31_AEV.pdf.

[95] Andrew D. Selbst, "An Institutional View of Algorithmic Impact Assessments," Harvard Journal of Law & Technology, (2021).

For example, while risk regulation is well-adapted to mitigate certain structural harms, it is less effective at mitigating individualized harms. This point is key in considering the use of AI in regulatory enforcement, where the consequences of decision-making are particularly sharp for the individual or entity on the other end of the process. Where AI systems are being used for regulatory enforcement, implementing entities should understand the limitations of a risk-based approach. Responsible use of AI for regulatory enforcement may require that risk assessments be complemented by prohibitions on certain particularly sensitive use cases (such as where decisions have a significant impact on fundamental rights),[96] or even liability-based structures that aim to compensate individuals for specific harms incurred.[97]

Experience also suggests that risk regulation is better at addressing predictable and easily quantifiable harms as opposed to the sort of "unknown unknowns" that are prevalent in considerations of the impact of AI.[98] In a 2011 article assessing the efficacy of the Nuclear Regulatory Commission's treatment of uncertain risks, Daniel Farber describes a rulemaking process which assumed that certain wastes would have no impact on the environment since they would be in a sealed repository.[99] Although the agency eventually acknowledged that the risk of a leak was unknown, and was not zero, the perception that the danger was relatively remote led the agency to effectively round-down their assessment.[100] This approach proved misguided when a proposed disposal site was found to have fractures in its bedrock, which would have allowed for water percolation and potential leakage of nuclear materials.[101]

The challenge of "unknown unknowns" is particularly thorny in the context of AI given that many of the harms which are built into existing risk management frameworks, including the AI RMF, are fuzzy at best. There is little agreement on how terms like "fairness" should be applied, in practical terms, and even less consensus on how to understand these principles mathematically.[102] By contrast, audits that are assessing known flaws, particularly data-based ones, such as the impacts of biased or otherwise problematic data sets, are on more familiar ground. The result is that harms like a loss of public trust or legitimacy, which are more difficult to pin down, are likely to be obscured or devalued in a risk-based analysis, which naturally focuses on harms that are easier to measure such as error rates.[103]

As a result, it is important for risk assessment processes to be implemented with what Margot Kaminski has dubbed "epistemic humility," by acknowledging the tendency of AI systems to surprise us, sometimes in harmful or destructive ways, and to incorporate this understanding into the heart of the decision-making process.[104] Similarly, assessments should factor in the natural

[96] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62019CA0817.
[97] Kaminski, Risk Regulation Toolkit, p. 8.
[98] Kaminski, Risk Regulation Toolkit, p. 14.
[99] Daniel A. Farber, "Uncertainty," GEORGETOWN LAW JOURNAL 99 (2011): 901, 910–11.
[100] *Ibid.* at 911.
[101] *Ibid.* at 950.
[102] Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NORTHWESTERN UNIVERSITY LAW REVIEW 357, 420 (2022).
[103] Jacob Metcalf, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, & Madeleine Clare Elish, "Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts," *Proceedings of the 2021 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2021).
[104] Kaminski, Risk Regulation Toolkit, p. 16.

human tendency to perceive AI systems as working more effectively than they do, and as being applicable to settings where they are not fit for purpose.[105] As a consequence of the natural tendency of new technologies to spread, regulatory assessments of the impact of AI should also incorporate a presumption that systems will expand beyond their initial approved uses. Above all, risk assessments should ensure adequate space to consider worst case scenarios, rather than allowing the outcomes which are perceived as most likely to dominate the calculus.

Different uses of AI can present vastly different risk profiles for administrative agencies. Uses related to regulatory enforcement represent some of the areas of greatest concern, due to the direct proximity of these systems with agency outputs, and the fact that enforcement decisions nearly always engage with thorny procedural and civil rights questions that are core to perceptions of legitimacy in agency decision-making. While it is unlikely that the use of AI in a regulatory enforcement context requires a fundamentally different risk assessment process than other public sector applications, for these use cases a rough corollary may be drawn between the degree of risk and the impacts of the AI on agency outputs.[106] The latter determination may be connected to the level of autonomy that these systems enjoy, though with the caveat that human review does not necessarily mitigate the dominant role AI systems may play in determining outcomes. Nonetheless, the use of AI as a research support tool for human decision-makers is likely to raise fewer concerns than where the AI is directly making enforcement decisions or recommendations. This distinction may be difficult to pin down, given the heavy influence that even early-stage research support can have over the shape of the final decision, and the deference with which humans treat AI-powered recommendations. As a result, in a risk assessment process over the use of AI in regulatory enforcement, the influence of the system over the final decision should be understood as a spectrum, rather than a binary question of whether or not it is producing autonomous outcomes.

Ultimately, while risk-based processes are an important tool in promoting appropriate guardrails around the use of AI in regulatory enforcement, their efficacy will be heavily dependent on the spirit which implementing agencies apply these assessments. Risk regulation is an extremely broad concept, which can mean different things to different groups, guided by a range of legal, sociological, institutional, and historical factors. In order to ensure that assessments are meaningful, and guided by appropriate understandings of the risk landscape, a robust external consultation structure is essential.

B. Public Engagement

It is a core principle of the rule of law that legitimacy rests on adequately justifying laws to the public.[107] Public engagement fulfils several important functions in a robust governance framework, including helping to support democratic accountability and, in the context of regulatory enforcement processes which are targeted at individuals, to support a sense of dignity among those subject to administrative decisions, for better or for worse.

---

[105] AI RMF p. 4.
[106] ACUS p. 10.
[107] Mireille Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, 20 THEORETICAL INQUIRIES L. 83, 113 (2019).

In addition to core democratic reasons why any uses of AI within the public sector should be subject to robust public consultation, external engagement is a valuable tool to boost the quality of risk assessments, and avoid their capture by the private sector or government agencies which have a direct stake in the systems' approval. Representatives from communities which are impacted by algorithmic enforcement, either directly through being a subject of regulation or indirectly through secondary effects of regulatory decisions, may be helpful to quantify and enumerate harms and concerns which may be difficult to isolate as part of an internal risk assessment process.[108] Impacted communities may also be able to spot problems with AI systems' outputs that elude their operators or designers, and which may not even be captured by sophisticated auditing or assessment processes. For example, in the case of Arkansas' infamous experiment with algorithmic healthcare determinations, the patients realized that there were errors in the outcomes even while officials, and the system's designers, insisted that it was working exactly as intended.[109]

Not all public engagement processes are created equal. At the extreme, it is possible to distinguish between meaningful opportunities for the public to influence agency decision-making, and the mere opportunity to provide feedback. The latter may serve as a procedural smokescreen, masking capture or preset decisions behind formal procedural equality.[110]

Although there are aspects of public engagement which are unique to the context of AI in regulatory enforcement, the challenge of providing meaningful avenues for public policy consultation is common to democracies around the world, as well as to intergovernmental organizations, and virtually any other institution which seeks public legitimacy. As a result, there is a robust set of international standards to draw on in attempting to make these processes meaningful.

For example, the Council of Europe's Code of Good Practice for Civil Participation in the Decision-Making Process (Code) sets four levels of participation: information, consultation, dialogue, and partnership.[111] The latter category, which is the highest standard, involves close collaboration, including service provision activities, participatory forums, and the establishment of co-decision-making bodies.[112] The Code envisions an engagement process which spans the entire lifecycle of the decision-making process, beginning at the Agenda Setting phase and extending through implementation to Monitoring and Reformulation processes.

Consultation structures take a number of forms that may include engaging with the public to inform them of rule-making procedures, through stakeholder meetings, the designation of public representatives, structured briefings to air differing views on a controversial public policy question, advertising campaigns to solicit input, negotiated rulemaking, etc.[113]

[108] Kaminski, UCLA Discourse, 179
[109] https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy
[110] Nicholas Bagley, *The Procedure Fetish*, 118 Mich. L. Rev. 345, 382 (2019).
[111] https://www.coe.int/en/web/ingo/civil-participation p. 3.
[112] https://www.coe.int/en/web/ingo/civil-participation p. 3.
[113] Nicholas Bagley, *The Procedure Fetish*, 118 Mich. L. Rev. 345, 385 (2019).

While transparency is not a panacea to resolve all concerns related to the use of AI in regulatory enforcement, robust transparency is a precondition for effective stakeholder engagement. In a 2024 paper on the subject, Margot Kaminski presents a model of transparency which focuses on the two-way flow of information, including both the "voices in" [to government], through meaningful opportunities to provide feedback, and the "voices out" [from the government], through robust transparency practices which ensure the public is well-informed about the context in which these technologies are being developed and deployed.[114] A good first step to ensuring a robust flow of information to relevant stakeholders, which can inform their participation and responses, is the publication of risk assessment or AI impact assessment results. However, the most relevant community stakeholders may not have adequate subject-matter expertise to process the kinds of technical data released in standard audit reports.[115] Agencies seeking to cultivate relationships with external stakeholders from impacted communities will need to either invest the resources to translate these documents so that they are more generally accessible, or to sponsor training and upskilling for community representatives or organizations to the point where they can engage with advanced questions related to the use of AI in regulatory enforcement.[116]

Agencies interested in boosting public participation in external consultation processes will also need to be mindful of the timing and location of engagement opportunities. A strict adherence to 9-5 business hours may mean that working people are unable to join. Childcare and travel costs may also present an obstacle to in-person participation, which may require resources to help mitigate. These issues are less likely to manifest where a consultation is carried out using remote participation, though an online format may be less satisfying for participants, and less conducive to robust and candid conversations. Where impacted groups include persons with disabilities, or persons who may not speak English, as well as other historically underrepresented communities, there may be a need for additional measures to bridge these challenges. Community organizations can serve as a key liaison to support participation among such historically marginalized groups, though this requires agencies to devote resources to cultivating productive and meaningful relationships with civil society partners.

On the latter point, it is noteworthy that there are a small but growing number of civil society organizations which are specifically focused on AI-related issues. For example, the Algorithmic Justice League, a Cambridge, Massachusetts based non-profit, founded the Algorithmic Vulnerability Bounty Project as a mechanism for outsourcing the identification of AI-driven harms to the public.[117] This evolved into the Community Reporting of Algorithmic System Harms project, which aims to mobilize an empowered community to report and advocate for the redress of algorithmic harms.[118] While these kinds of initiatives are still relatively thin on the ground, agencies should capitalize on them where they already exist, and should explore options to provide resources to support and expand their work.

---

[114] Kaminski, Voices In, Voices Out, p. 194.
[115] Ibid.
[116] Margot E. Kaminski & Gianclaudio Malgieri, Algorithmic Impact Assessments Under the GDPR: Producing Multi-Layered Explanations, 11 INT'L DATA PRIV. L. 125, 139 (2021) (encouraging "companies, or regulators, to help fund the involvement" of impacted individuals and to "provide technical expertise or the resources for obtaining technical expertise").
[117] https://medium.com/fast-company/meet-the-computer-scientist-and-activist-who-got-big-tech-to-stand-down-23d95e0347e7
[118] https://www.ajl.org/crash-project

- **Policy Engagement vs. Enforcement Engagement**

External consultation related to the use of AI in regulatory enforcement may be divided into two general categories: policy engagement mechanisms, focused on ensuring that relevant stakeholders have an ability to impact decisions related to the development and deployment of AI systems; and enforcement engagement mechanisms, which generally revolve around allowing the subjects of an enforcement decision, as well as other interested stakeholders, to opine on a particular enforcement process and its outcomes.

Policy engagement mechanisms, which are meant to address broader systemic concerns, can be particularly valuable in helping administrative agencies adopt an approach to the development and deployment of AI and other algorithmic tools which is in line with public values and expectations. Useful functions of this level of engagement may include helping to define key terms, such as "discrimination" and "fairness" or "less favorable treatment".[119] External insights may also feed directly into risk assessment processes, both in shaping how assessments are carried out, as well as responding or providing feedback on draft assessments that are being reviewed. Here, stakeholder input may be invaluable to determine whether, for example, a particular harm is being under-appreciated or mischaracterized.

External participation in policymaking may involve an official rulemaking process or may be carried out on a less formal basis as novel issues arise, for example through a standing committee of community participants. The latter model allows for the development of greater subject matter expertise among participants, which in turn can generate more specific and meaningful guidance for the agencies. However, this model is obviously more labor intensive for both the agency and the community participants, which may push the demands on the latter's time beyond what might be expected on a volunteer basis. It may be worth exploring schemes which compensate community or civil society participants for their engagement, though this needs to be managed carefully in order to avoid creating perverse incentive structures.

By far the most important ingredient in maintaining robust civil society engagement is to ensure that the consultations are meaningful, and that they are perceived as such. Community participants are likely to sour on engagement processes if they feel that their voices are not significantly impacting policy. Where relationships have been forged with trusted and sophisticated community partners, and especially where resources have been invested to provide training to these partners, it is critical that agencies manage these connections carefully.

Enforcement engagement mechanisms target a more easily defined group consisting of, first and foremost, the persons who are subject to decisions where AI is a significant part of the assessment process.[120] While this engagement may overlap with the right to due process and explainability, it extends beyond the specific right to challenge outcomes of decisions to broader procedural dissatisfaction. For example, where regulatory enforcement decisions concern pollution impacting a particular geographic region, residents of that region might also be considered as targets for consultation. While these consultations will naturally be more focused on an individual outcome

[119] Substitute S.5116, 67th Leg., Reg. Sess. § 4 (Wash. 2021).
[120] *See* Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957 (2021).

than any broader structural concern, it is possible for individual review or appeals mechanisms to connect back to broader risk assessment processes or other policy determinations, for example by requiring that a successful appeal (or a number of successful appeals above a certain threshold) should trigger a review of the original risk assessment, or that individual appeal outcomes should be factored into regularly scheduled reassessments.

C. Retirement Must be an Option

Among the most important characteristics of a robust system of oversight for the use of AI in regulatory enforcement is that there must be adequate consideration of when and whether to phase out these tools. AI is a relatively young technology, whose deployment across the public sector is still in its early stages. And yet, already there are plenty of examples of AI failures. These include not only the well documented issues with bias and discrimination documented in Part II, but also examples where AI has simply failed to deliver on the results that its proponents promised.

In their previous publication for the Administrative Conference of the United States (ACUS), Ho et al documented the case of the Sigma system piloted at the USPTO, which was never deployed since it failed to improve efficiency unless its users had a computer science background.[121] A subsequent report on experiences using AI-powered tools to track and clear superfluous regulations, which was authored for ACUS by Catherine Sharkey, noted similar performance challenges in other systems. Staff at the Centers for Medicare and Medicaid Services (CMS) noted numerous false positives, and complained about the labor-intensive process of checking the enormous number of regulations that were flagged.[122] While these failures are not intended to represent the totality of government agencies' experiences in using AI, they demonstrate that, at the very least, the deployment of these technologies has been a mixed bag, though Professor Sharkey's paper suggested that officials were still bullish about the overall potential for AI to improve their agency's operations.[123]

There are a number of factors which may play into a tendency to be unduly optimistic about AI's performance or utility for a given task. Part of this is embedded in human nature, as fascination with new technologies can lead to a natural tendency to over-estimate the capability of AI.[124] In certain contexts, biased assessment standards may be the result of asymmetries in how successes or failures of these systems are tracked and evaluated. For example, a system which under-estimates the threat from a criminal defendant, and recommends their release only to have them reoffend, will receive negative feedback for the mistake.[125] However, if the next defendant's threat level is over-estimated, and they are remanded to custody as a result, there will be no concomitant opportunity to assess whether the system was wrong. A system may therefore learn that it is possible to game its own evaluation, and skew in a biased direction as a result.

---

[121] Ho et al p. 48

[122] Sharkey p. 28.

[123] *See, e.g., Sharkey at 34 quoting a CMS official's opinion that "We absolutely need to add technology into our process." Similarly, Ho et al*

[124] AI RMF p. 4.

[125] Noel L. Hillman, The Use of Artificial Intelligence in Gauging the Risk of Recidivism, Am. Bar Ass'n: Judges' J. (Jan. 1, 2019), https://www.americanbar.org/groups/judicial/publications/judges_journal/2019/ winter/the-use-artificial-intelligence-gauging-risk-recidivism/.

- **Assessing Risk and Assessing Failure**

A core challenge to most risk assessment frameworks, including those proposed under the AI RMF, is that they typically adopt a relatively permissive approach to new technologies, with a heavy emphasis on preserving space for agencies to innovate.[126] Risk regulation implies a tradeoff. The assessment process, and the adoption of certain remedial steps, are necessary costs that entities must bear to access the efficiency and processing gains that AI can provide. This "techno-correctionist" tendency is useful for mitigating problematic aspects of widespread adoption, such as through threats to security, but is less useful for answering broader questions about whether a category of technologies are fit for use among regulatory enforcement functions.[127]

Agencies contemplating the use of AI for regulatory enforcement must be mindful of this gap in the risk-based paradigm, and work to complement their risk assessment framework with a careful and critical assessment of the costs and challenges of pursuing an AI-based solution in the first place. This assessment should include difficult questions about whether the operational context presents too great a challenge, or is otherwise unsuited to automation, or whether the potential institutional harm to legitimacy and reputation are too great. It should also consider whether, to put it bluntly, advocates for a particular system are selling snake oil.[128]

While AI, in general, has enormous potential, it is also buoyed by massive amounts of hype, much of which is generated by stakeholders with a direct financial interest in frothing up enthusiasm for new AI applications. Tech-solutionism, and a desire to appear on the cutting edge of innovation, can be powerful drivers in favor of expanding uses of AI. But there are many instances where AI is unfit for a given application, such as where there is insufficient underlying data to power a system.[129] Likewise, as detailed in previous sections, decisions to delegate increasing agency operations to machines may carry significant institutional costs. In the regulatory enforcement context, this may include hollowing out agency expertise and discretion, undermining popular perceptions of legitimacy. A responsible operational paradigm should include constant reassessment not just of how these technologies may be improved, but of whether they should be retired altogether.

The need to recognize failure is not unique to applications involving regulatory enforcement, though it carries particular salience in a regulatory enforcement context due to the severe and direct impact that enforcement decisions have on their subjects, as well as the long and costly process by which these decisions may be challenged. If an agency waits for a Supreme Court finding that an AI-enabled component of their enforcement process was unreliable or otherwise problematic, the decision may taint years of other enforcement efforts that were based on the same operational paradigm.

---

[126] Kaminski, Risk Regulation Toolkit, p. 2.
[127] Jessica M. Eaglin, "When Critical Race Theory Enters the Law & Technology Frame," Michigan Journal of Race and Law 26 (2021): 151, 155.
[128] https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil#preview..
[129] Sakshi Gupta, When Should You Not Use Machine Learning?, SPRINGBOARD (Sept. 25, 2020), https://www.springboard.com/blog/data-science/when-not-to-use-ml/

The challenge in developing a meaningful standard of review which allows for failure is illustrative of a more foundational problem with using risk regulation as the lodestar of the oversight system. Among the main challenges permeating risk-centric frameworks for assessment is that, at their core, their focus is on developing a structure which will allow a proposed use case to move forward. The obvious gap in this approach is that there will inevitably be cases where incorporating AI is inappropriate or ill-advised regardless of how well it performs. Risks to the democratic legitimacy or public trust in an agency may be impossible to mitigate through more rigorous auditing or ensuring that humans remain at key points in the decision-making process. It is worth noting that the AI RMF presents particular challenges on this front insofar as its focus is on mitigating harms to institutions, as opposed to risks to the public or to democracy more broadly.[130]

This is not to suggest that risk-based approaches should be abandoned, though it does demonstrate the importance of ensuring that the assessments are designed in a way that allows for the conclusion that a system under review should be phased out. It also shows a need to diversify oversight structures through additional mechanisms such as robust engagement with members of the public, including recognizing greater rights to contest decisions and pursue other effective remedies.[131]

D. Structural Oversight Considerations

Although virtually every AI governance framework delegates significant responsibilities to the frontline institutions that develop or deploy these systems, there always is a need for a central organization, or multiple organizations, to play a coordinating and oversight role. One may understand these coordinating functions as lying on a spectrum, from a highly devolved system, where the central agency is little more than a clearinghouse for documentation and reporting, to a more rigorous oversight structure which exercises relatively stringent control over how administrative bodies experiment with AI.

As a baseline, it should be relatively uncontroversial to note the value of a central repository which publicly tracks AI use cases across the administrative state and provides public information about where AI has been deployed, and access to relevant background material such as the results of risk assessments or audits. This appears to be the direction that OMB's AI policy memo is pushing agencies toward, in requiring that enhanced reporting and convening take place.

Beyond a limited publicly-facing function, there is additional value in maintaining a central hub for expertise in AI policy. This may include developing and applying best practices for things like risk assessment, transparency, and external consultation processes. There is a clear need for conceptual work to bolster the baseline standards enumerated in documents like the AI RMF and the Blueprint for an AI Bill of Rights. While aspects of the risk assessment process need to be carried out locally, it clearly does not make sense for every agency to be starting from scratch or working from its own independently developed definitions and benchmarks for things like "bias", "discrimination", and "fairness".[132] Similar performance-based metrics are a core feature of

[130] Kaminski, Risk Regulation Toolkit, p. 16.
[131] *See* Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems, April 2020, https://rm.coe.int/09000016809e1154.
[132] ACUS p. 7.

environmental regulation, among other fields, and as AI governance matures it is essential to develop similar standards which may be implemented as common measures for performance.[133] Relatedly, the federal government will inevitably need to devote additional resources to providing technical support for agencies that are interested in piloting new AI projects, but which lack the resources to build them. These functions do not all need to be consolidated in a single agency, though there are efficiencies in housing them together.

At the further edge of the spectrum, one may envision an agency that performs a more significant oversight function, and even an enforcement role over how AI is being deployed across administrative agencies. There have been numerous scholarly suggestions for more proactive governance structures, though these typically focus on AI as a whole, as opposed to purely regulating public sector or regulatory enforcement functions. These models include Ryan Calo's 2015 proposal for a federal robotics commission, and Andrew Tutt's proposal for an "FDA for algorithms" that would exercise oversight over all such products before they are marketed.[134] Gianclaudio Malgieri and Frank Pasquale suggested an ex-ante licensing regime targeted at certain highly impactful functions, which would presumably include regulatory enforcement.[135] Imposing pre-market licensing requirements across the private sector may require grappling with certain constitutional challenges, though the issue is simpler if licensing is considered purely with regard to administrative regulatory enforcement.[136]

Even if licensing goes too far, the high profile failures detailed through the early chapters of this report suggests the need for a cautious approach which preserves the human-centric nature of government. The expertise and discretion of administrative agencies, and the broader common interest in maintaining perceptions of trust and legitimacy, both speak to the value of an oversight structure which can work towards the safe and responsible use of AI in regulatory enforcement, rather than acting as merely a clearinghouse of public information.

- **Defining Effective Oversight**

It is possible to identify several features that will be essential for an effective AI oversight body for administrative agency functions, including regulatory enforcement. First, and most importantly, it should possess adequate subject matter expertise related to AI, as well as broader issues within its remit, such as laws and standards around discrimination, procedural fairness, etc. To achieve this, the agency will need to be adequately staffed with a diversity of experts from different backgrounds, including law, public policy, computer science and engineering, etc. The latter is particularly important if the oversight body is going to play a technical supporting function for agencies which are seeking to develop new AI systems. Though, it is important that subject matter experts from computer science and engineering be complemented by staff with a background in law, public policy, and the humanities, to ensure that there are robust conversations about ethics and public policy which transcend pure performance-based audits.

---

[133] Lauren E. Willis, *Performance-Based Consumer Law*, 82 U. CHI. L. REV. 1309 (2015).
[134] Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L REV. 83 (2017); Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 556 (2015).
[135] Gianclaudio Malgieri & Frank Pasquale, "From Transparency to Justification: Toward Ex Ante Accountability for AI," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4099657.
[136] See, e.g., https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4431251.

There are a number of existing agencies which possess robust technical expertise, including the Office of Science and Technology Policy (OSTP) and the General Services Administration AI Center of Excellence, as well as NIST. The latter has been an early mover in the AI space, especially through the AI RMF discussed throughout this report, though there have been concerns raised about their efficacy in performing a direct oversight function.[137] There are also concerns that an expanded role for NIST into a politically contentious realm, like AI in regulatory enforcement, may jeopardize its other vital roles related to standard setting and technological advancement.[138]

A second important quality is that the agency should be capable of public engagement, both in terms of a robust network of collaborators across academia and civil society, and as a key vector for transparency reporting. Rather than merely providing a clearinghouse of information, a strong coordinating body should have a mandate to promote engagement with the administrative state, while also cultivating awareness of opportunities for civil society and the public to get involved. The Privacy and Civil Liberties Oversight Board (PCLOB) is one model for an agency with a robust mandate to engage with public feedback, and there have been proposals to create a parallel body for AI applications, although several factors have limited PCLOB's record as an effective oversight body.[139]

Effective oversight can be challenging if the responsible agency is not sufficiently independent to enable it to push back against prevailing political priorities to cut costs, and to effectively advocate in support of longer-term institutional and public interest priorities. Ideally, an oversight agency can be empowered with some sort of mechanism to compel compliance with its standards and recommendations. There can be a tension between independence and effective enforcement powers. An agency which is placed outside of the executive branch has the advantage of greater independence, but will be less likely to be able to mandate that executive branch agencies follow a suggested course of action.[140]

It is difficult to find parallel examples across the executive branch where an agency maintains sufficient binding oversight powers to be effective in the face of concerted resistance. For example, the Office of Information and Regulatory Affairs (OIRA) plays a role in supporting compliance with the Regulatory Flexibility Act and Congressional Review Act, including through mandated impact analyses of significant regulatory actions. However, it is up to the agencies to determine whether a particular use case for AI is high risk enough to warrant review.

One option, which was proposed by Margot Kaminski in a 2023 paper, would be to impose a revocable licensing scheme on new applications of AI within regulatory enforcement structures. Kaminski's proposal, which is not specifically targeted towards government applications, envisions a robust post-market surveillance scheme to ensure that systems which failed to perform

[137] Bryan H. Choi, *NIST's Software Un-Standards*, LAWFARE (Mar. 7, 2024), https://www.lawfaremedia.org/article/nist's-software-un-standards.
[138] https://www.techpolicy.press/regulating-artificial-intelligence-must-not-undermine-nist-integrity/
[139] https://www.justsecurity.org/94999/an-oversight-model-for-ai-in-national-security-the-privacy-and-civil-liberties-oversight-board/
[140] Unzen p. 253.

(or which manifested other problems) could be efficiently removed from the market.[141] In order to avoid chilling innovation, the government could ensure that these licenses were relatively easy to obtain, upon completion of a standardized set of requirements to report on the new use and conduct an initial risk assessment. However, an oversight body could retain the power to suspend or revoke these licenses if the implementing agency failed to maintain an appropriately robust risk assessment process through the life-cycle of the system, or if additional information or complaints were brought to light which suggested a problem.

The question of whether oversight and coordination of such a scheme should be bundled into the responsibilities of an existing agency, or handed over to an entirely new statutory creation, is complex. There are several candidates across the executive branch, including OIRA, OSTP, and OMB, to name a few.[142] Although handing these responsibilities to an existing agency would be simpler, it would require a significant expansion in responsibilities, requiring additional funding and resources, and potentially a change in the agency's mandate.

At the end of the day, the likely scale of impact that AI technologies will have across the administrative state, and the importance of ensuring that transitions to greater automation are carefully managed in line with the public interest, suggest that the creation of a standalone oversight body may be justified. However, the question of how to construct effective oversight of the use of AI for regulatory enforcement includes complicated political calculations which go beyond the scope of this paper. Congress has been struggling to pass an update to federal privacy legislation for decades now.[143] Given the pace with which AI is moving, America cannot afford a similarly painful and drawn-out process for effective regulation of government uses of this technology. While a new standalone oversight body may be the best option in the longer term, it is important to focus on solutions which are politically workable in the short term, to avoid a future where AI becomes deeply embedded across sensitive government functions without any meaningful check on its deployment.

[141] Kaminski, Risk Regulation Toolkit, p. 21.
[142] Aram A. Gavoor & Raffi Teperdjian, A Structural Solution to Mitigating Artificial Intelligence Bias in Administrative Agencies, 89 GEO. WASH. L. REV. ARGUENDO 71, 87 (2021).
[143] https://www.brookings.edu/articles/after-20-years-of-debate-its-time-for-congress-to-finally-pass-a-baseline-privacy-law/

V. Conclusion: Grappling with the Human-Machine Paradigm

Beyond considering the pros and cons of each individual use case, among the most important conceptual questions that agencies need to grapple with over the coming decade concerns the interplay between humans and automated systems, and the appropriate role and limits of automation. The results of this process are likely to shape the administrative state for the next generation.

Addressing this challenge requires a long-term view of agency priorities and values, including balancing the pressure to cut budgets and improve capacity against the need to safeguard perceptions of legitimacy, and the human-centric nature of the administrative state. There are also more subtle impacts that are at play, such as the broader datafication of agency decision-making. As AI becomes more deeply ingrained in regulatory enforcement, it naturally centralizes data flows in order to maximize the efficiency and efficacy of these systems.[144] This, in turn, is likely to prompt a more concentrated agency structure. There are no easy answers, and the questions are likely to become more difficult and complicated as technology continues to advance.

As a starting point, it is useful for agencies to isolate and identify areas of governance where "humanity" is of particular importance, such as policing, or other areas where datafication is viewed as fundamentally or intuitively problematic. Ryan Calo and Danielle Citron propose that a key guide to assessing the wisdom of incorporating an AI system should be whether the use of AI furthers key substantive commitments and values, such as access, quality, and self-assessment.[145] Whether a particular regulatory enforcement use case is designed to further these values, as opposed to simply reducing costs, is the critical question: will the proposed tools enhance the capabilities of implementing agency, or is it mainly a means to outsource human discretion and expertise to a cheaper and inexhaustible automated decision-maker? Ho et al included a similar call for human centered AI in their article, advocating for decision tools which complement, rather than replace, the human element in the process.[146]

It is also important to think strategically about the relationship between humans and machines, and the interplay between these two. This can include the need to manage data being collected now in a way that leaves the door open to future automation, but also to figure out impacts that a reduction of staff in favor of AI will have from a human resources perspective, both in terms of likely reactions from the current workforce and the resulting risk to human capacity and expertise within those agencies.

An additional consideration for agencies concerns the need to grapple with apportioning an appropriate role for technical experts in the development and deployment of these technologies, and assessing their performance, without losing sight of the underlying values that are meant to be guiding their work. Given the scientific and mathematical nature of assessments around qualities

[144] Unzen 234-235.
[145] Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L. J. 798, 840 (2021).
[146] Government by Algorithm at 83.

like fairness and bias, it is easy to forget that both values require a fundamentally human element to ensure that they are meaningful.[147]

From this perspective, one final consideration concerns a persistent skill shortage across the government in terms of AI capacity, and the need for budgets and salaries to reflect the competition for technical expertise in these fields. In their previous publication for ACUS, Ho et al relay a story about early government agencies seeking to experiment with AI who, due to hiring rules, were forced to find lawyers with a coding background to develop software for them.[148] While agencies over the coming years will hopefully enjoy a freer hand to hire technical experts to support their operations, there are likely to be lingering challenges around developing an appropriate pipeline to onboard STEM graduates into public service, and how to integrate these new streams alongside traditional policymaking staff.

Administrative agencies face a rapidly changing and dynamic environment, as political, judicial and technological changes combine to fundamentally reshape their role in the constitutional order. Although there are no easy answers, it is possible for agency leadership to cleave to fundamental values, and their core mission, as they attempt to navigate a course which harnesses the benefits of AI while preserving the essential humanity of democratic governance.

[147] Deborah Hellman, "Measuring Algorithmic Fairness," 106 Virginia Law Review 811, 834 (2020).
[148] Ho et al p. 44.

VI. Recommendations

Recommendations for Implementing Agencies:

- While risk assessments are a valuable tool, implementing agencies should understand their limitations in the realm of regulatory enforcement, including by factoring in the tendency of AI systems to produce unpredictable outcomes, and to expand beyond their initial approved uses. They should also ensure that worst case scenarios are considered, even if their likelihood is relatively remote.
- Risk assessment processes should think beyond immediate harms to stakeholders, and include thorough consideration of the potential for an AI system to undermine public confidence in an agency or perceptions of legitimacy.
- In a risk assessment process over the use of AI in regulatory enforcement, the influence of the system over the final decision is a major risk factor. This level of influence should be understood as a spectrum, rather than a binary question that hinges on the presence of human review.
- Risk assessment processes should be complemented by a careful and critical assessment of the costs and benefits of pursuing an AI-based solution, and systems which are failing to perform to an acceptable standard should be shelved. Agencies should be wary of sunk cost fallacies in considering the worthiness or performance of a system and should be cautious of overly ambitious or optimistic assessments of AI capabilities.
- Implementing agencies should institute robust external consultation processes related to any uses of AI in regulatory enforcement, which span the full lifecycle of the decision-making process, and which include participatory forums and the establishment of co-decision-making bodies with leading civil society and other relevant participants. Key goals for this engagement should include identifying and quantifying potential or manifested harms, and generating buy-in from impacted communities. Strong engagement will depend on agencies' ability to offer meaningful opportunities to impact policy.
- Robust consultation depends on robust transparency, so that external stakeholders can get a complete picture of how AI systems are being implemented and how they are performing. At a minimum, this requires publishing risk assessments and related documentation. At a more advanced level, resources may be required to translate reporting into more accessible formats, or to upskill community partners to engage with more advanced questions.
- In assessing the appropriate scope for disclosure of information, agencies should err on the side of transparency, and only withhold information which would clearly harm the effective administration of justice, and where the harm from this disclosure would outweigh the public interest in its release.

Structural Recommendations for the Federal Administrative Apparatus:

- Responsible use of AI for regulatory enforcement may require that risk assessments be complemented by stronger measures to prevent or mitigate particularly severe harms. These may include prohibitions against particularly problematic use cases, such as where

fundamental rights are impacted or particularly sensitive data is involved. Agencies should consider the role of individual remedies in addressing specific harms, as well as their strengths and weaknesses as a vehicle for structural change.

- Effective oversight of AI across the administrative state, and particularly with regards to regulatory enforcement, suggests that there is value in centralized oversight related to AI development and risk assessment, potentially including a revocable licensing scheme. The oversight structure should be staffed with a diverse range of experts across multiple disciplines, and should be equipped with sufficient independence and enforcement powers to play an effective role.