Federal Procurement of Artificial Intelligence:

Perils and Possibilities

REPORT BY **DAVID S. RUBENSTEIN**

DECEMBER 2020

THE GREAT DEMOCRACY INITIATIVE

THE GREAT DEMOCRACY INITIATIVE

ABOUT THE GREAT DEMOCRACY INITIATIVE

The Great Democracy Initiative develops policy blueprints that offer solutions to the most pressing problems of our time. From taming the concentration of power in our economy to fundamentally reforming our broken government, GDI aims to generate policy ideas that confront the forces that have rigged our society in favor of the powerful and connected.

ABOUT THE AUTHOR

David S. Rubenstein is James R. Ahrens Chair in Constitutional Law, and Director of the Robert J. Dole Center for Law & Government, at Washburn University School of Law. Prior to teaching, Professor Rubenstein clerked for The Honorable Sonia Sotomayor when she was a judge on the United States Court of Appeals for the Second Circuit, and for The Honorable Barbara Jones in the United States District Court for the Southern District of New York. Prior to clerking, he served for three years as Assistant United States Attorney in the Southern District of New York, and was a litigation associate for five years at King & Spalding LLP.

ACKNOWLEDGEMENTS

The author is grateful to Daniel Ho, Raj Nayak, Suzanne Kahn, and Eric Jacobs for their incisive comments and feedback; to Matt Hughes for tight editing; to Anna Smith for logistical support; and to Kaitlyn Bull, Ande Davis, Penny Fell, Barbara Ginsberg, Creighton Miller, and Zach Smith, for invaluable research assistance; and to Leah for everything always. All errors and omissions are the author's alone.

I. Introduction

Artificial intelligence (AI) is transforming how government operates.[1] For example, the Federal Bureau of Investigation uses AI in law enforcement; the Social Security Administration uses AI to adjudicate benefits claims; the Food and Drug Administration uses AI to support its rulemaking processes;[2] the Department of Homeland Security uses AI to regulate immigration;[3] and countless other agencies are experimenting with AI for the delivery of government services, customer support, research, and regulatory analysis.[4] This small sampling presages a new era of "algorithmic governance," in which government tasks assigned to humans will increasingly migrate to machines.[5]

Algorithmic governance brims with promise and peril. Under the right conditions, AI systems can solve complex problems, reduce administrative burdens, and optimize resources. Under the wrong conditions, AI systems can lead to widespread discrimination, invasions of privacy, dangerous concentrations of power, and the erosion of democratic norms.

The possibilities and perils of AI's social disruptions have led the United States and institutions across the globe to propagate principles of trustworthy and ethical AI.[6] Although the particulars vary, ethical AI envisages a cluster of principles relating to transparency, accountability, fairness, privacy, and security.[7] Translating these principles into practice is the next step. Currently, dozens of pending congressional

---

1 AI has no singular definition. Here, I use the term AI to describe a range of computer-enabled abilities and methods to perform tasks that would otherwise require human intelligence, such as learning, adaptation, reasoning, prediction, optimization, and sensory understanding. *See infra* Subpart I.A (explaining the definitional problem and offering some refinements).

2 *See generally* DAVID FREEMAN ENGSTROM, DANIEL E. HO, CATHERINE M. SHARKEY, MARIANO-FLORENTINO CUÉLLAR, GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES (2020) [hereafter GOVERNMENT BY ALGORITHM], https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.

3 *See* Aaron Boyd, *CBP Is Upgrading to a New Facial Recognition Algorithm in March*, NEXTGOV.COM (Feb. 7, 2020), https://www.nextgov.com/emerging-tech/2020/02/cbp-upgrading-new-facial-recognition-algorithm-march/162959/.

4 *See, e.g.*, GOVERNMENT BY ALGORITHM, *supra* note 2; Cheryl Ingstad, *A Message from Leadership*, AI AT DOE NEWSLETTER (Aug. 2020), https://www.energy.gov/sites/prod/files/2020/10/f80/AI%20Newsletter%202020%2002%2011.pdf; PR Newswire, *Geospark Analytics Awarded FEMA Contract for Use of Hyperion and AI-Driven Risk Models*, AP NEWS (May 28, 2020), https://apnews.com/PR%20Newswire/29397df99616a5ea6413cc70caf7ca68.

5 *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 636 (2017); *see also* Will Hurd and Robin Kelly, Rise of the Machines: Artificial Intelligence and its Growing Impact on U.S. Policy (2018), https://www.hsdl.org/?view&did=816362.

6 *See* Anna Jobin et al., *The Global Landscape of AI Ethics Guidelines*, 1 NATURE MACHINE LEARNING 389 (2019), https://www.nature.com/articles/s42256-019-0088-2.pdf (mapping and analyzing the corpus of principles and guidelines on ethical AI).

7 *Id.*; *see also* Jessica Fjeld et al., Berkman Klein Ctr., Res. Publ'n No. 2020-1, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482.

bills would regulate private and governmental uses of AI, the data that fuels this technology, and the infrastructures needed to sustain it.[8]

Meanwhile, the executive branch is revving its engines. In December 2020, President Trump issued an Executive Order to "promote the innovation and use of AI" in government operations "in a manner that fosters public trust, builds confidence in AI, protects our Nation's values, and remains consistent with all applicable laws, including those relating to privacy, civil rights, and civil liberties."[9] This builds on a February 2019 Executive Order, which projects that AI will affect the missions of nearly all executive departments and agencies, and sketches a plan for "maintaining American leadership" in innovative and trustworthy AI.[10]

Wide swaths of law and public administration will need retrofitting to accommodate algorithmic governance.[11] This report focuses critical attention on one regulatory domain that requires immediate attention: *federal procurement law*.

For a variety of reasons, the government's pent-up demand for AI systems far exceeds its in-house capacity to design, develop, field, and monitor this powerful technology.[12] Accordingly, many (if not most) of the tools and services of algorithmic governance will be procured by contract from the technology industry. This is highly concerning, in part, because AI systems are virtually unregulated in the private market.[13] Without intervention, the government will be acquiring unregulated technology for government functions. Moreover, when procured from the private market, AI systems

---

[8] *See* Center for Data Innovation, *AI Legislation Tracker–United States*, https://www.datainnovation.org/ai-policy-leadership/ai-legislation-tracker/ (last visited July 17, 2020); Yoon Chae, *U.S. AI Regulation Guide: Legislative Overview and Practical Considerations*, ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (Jan.–Feb. 2020) (reporting that in 2015–2016, only two introduced bills contained the term "artificial intelligence"; that increased to 51 bills by the end of 2019).

[9] Exec. Order No. 13,960, 85 Fed. Reg. 78,939 (Dec. 8, 2020). By sheer happenstance, the executive order was issued the same day this report was finalized. For that reason, time and space did not permit further discussion here. It is worth noting, however, that the executive order aligns in many ways with the normative and prescriptive thrust of this report.

[10] Exec. Order No. 13,859, 84 Fed. Reg. 3,967 (Feb. 14, 2019).

[11] *See* Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399 (2017) (outlining several policymaking issues that will need to be addressed in the near term).

[12] As recently reported by the National Security Commission on Artificial Intelligence (NSCAI): "[T]here is a severe shortage of AI knowledge in [Department of Defense] and other parts of government . . . Current initiatives are helpful, but only work around the edges, and are insufficient to meet the government's needs." Nat'l Sec. Comm'n on Artificial Intelligence, Second Quarter Recommendations 34 (2020) [hereafter NSCAI, Second Quarter Recommendations], https://drive.google.com/file/d/1hgiA38FcyFcVQOJhsycz0Ami4Q6VLVEU/view. *Cf.* GOVERNMENT BY ALGORITHM, *supra* note 2, at 18, 89 (finding that approximately half of AI applications currently in use were developed in-house by federal agency personnel, but acknowledging the government's in-house capacity challenges); Rudy Mehrbani, Tess Byars, Louis Katz, *A Time to Serve: Proposals for Renewing the Civil Service*, GREAT DEMOCRACY INITIATIVE (Aug. 2020), https://greatdemocracyinitiative.org/wp-content/uploads/2020/08/Personnel-Policy-Final-Copy.pdf (arguing for the "need to change hiring practices to create more and better pathway into government for diverse and talented workers," including by "modernizing the civil service system").

[13] *See* Russell T. Vought, Office of Mgmt. & Budget, Guidance for Regulation of Artificial Intelligence Applications (2020), https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf.

may be shrouded in trade secrecy protection, which can impede public transparency and accountability.[14]

Beyond these concerns lies another. Acquiring AI is not business as usual: It often entails the procurement of policy choices from nongovernmental actors. AI systems are embedded with value-laden tradeoffs between what is technically feasible, socially acceptable, economically viable, and legally permissible. Thus, without intervention, the government will be acquiring technology with embedded policies from private actors whose financial motivations and legal sensitivities may not align with the government or the people it serves.

AI systems are embedded with value-laden tradeoffs between what is technically feasible, socially acceptable, economically viable, and legally permissible.

Of course, the risks of harm are contextually contingent. It is one thing when an AI system misclassifies emails as spam or recommends purchasing more office supplies than needed. It is quite another when an AI system mistakenly deprives individuals of unemployment benefits,[15] automates the illegal seizure of tax refunds,[16] encroaches on personal privacy,[17] leads to wrongful arrest,[18] perpetuates racial and gender biases,[19] deprives access to government food programs,[20] impedes the right to travel,[21] and so on.

---

[14] *See* David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135 (2007); Sonia Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV. 1183, 1186–87 (2019) (explaining how "source code that underlies and governs automated decision making is hidden from public view, comprising an unregulated 'black box' that is privately owned and operated").

[15] *See* Stephanie Wykstra & Undark, *It Was Supposed to Detect Fraud. It Wrongfully Accused Thousands Instead: How Michigan's Attempt to Automate its Unemployment System Went Terribly Wrong*, THE ATLANTIC (June 7, 2020), https://www.theatlantic.com/technology/archive/2020/06/michigan-unemployment-fraud-automation/612721/.

[16] *Id.*

[17] *See, e.g.*, Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716–27 (2010) (showing that an individual's identity may be reverse-engineered from a small number of data points); Tristan Greene, *Horrific AI Surveillance Experiment Uses Convicted Felons as Human Guinea Pigs*, TNW (Aug. 14, 2020, 5:40 PM), https://thenextweb.com/neural/2020/08/14/horrific-ai-surveillance-experiment-uses-convicted-felons-as-human-guinea-pigs/.

[18] Kashmir Hill, *Wrongfully Accused by an Algorithm*, N.Y. TIMES (Jun. 24, 2020), https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

[19] *See, e.g.*, Rashida Richardson, Jason Schultz, Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 NYU L. REV. ONLINE 15 (2019); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); Josh Feast, *4 Ways to Address Gender Bias in AI*, HARV. BUS. REV. (Nov. 20, 2019), https://hbr.org/2019/11/4-ways-to-address-gender-bias-in-ai; *Amazon Ditched AI Recruiting Tool that Favored Men for Technical Jobs*, THE GUARDIAN (Oct. 10, 2018), https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine.

[20] *See* Florangela Davila, *USDA Disqualifies Three Somalian Markets from Accepting Federal Food Stamps*, SEATTLE TIMES (Apr. 10, 2002), http://community.seattletimes.nwsource.com/archive/?date=20020410&slug=somalis10m.

[21] *See generally* JEFFREY KAHN, MRS. SHIPLEY'S GHOST: THE RIGHT TO TRAVEL AND TERRORIST WATCHLISTS (2013); *see also* Latif v. Holder, 28 F. Supp. 3d 1134, 1153 (D. Or. 2014) (ordering the agency to "fashion new procedures that provide plaintiffs with the requisite due process . . . without jeopardizing national security").

More than a marketplace, the acquisition gateway must be reimagined as a policymaking space for promoting trustworthy and ethical AI. Toward that objective, this report offers a set of legal prescriptions that aim to align federal procurement law with the imperatives of ethical algorithmic governance.

More than a marketplace, the acquisition gateway must be reimagined as a policymaking space for promoting trustworthy and ethical AI.

First, federal lawmakers should mandate the creation of a government-wide inventory report that includes clear information on each AI system used by federal agencies. Currently, policymakers and stakeholders are wrangling about algorithmic governance, including whether AI tools such as facial recognition should even be permitted.[22] But an informed policy debate is impossible without knowledge about which AI tools have already been adopted by which agencies, for what purposes, from which vendors, and at what cost.

Second, federal lawmakers should require that agencies prepare "AI risk assessment" reports prior to the government's acquisition of AI tools and services. These risk assessments would foreground several challenges and vulnerabilities that inhere in AI systems—most notably, relating to transparency, accountability, fairness, privacy, and safety.

Third, federal lawmakers should integrate ethical AI considerations into existing regulations for source selection and contractual award. Currently, nothing prevents federal contracting officials from soliciting and evaluating competing bids with an eye toward ethical AI. That is not the general practice, however, and it should be required as matter of law. Doing so will force agency officials and vendors to think more critically—and competitively—about the AI systems passing through the acquisition gateway. Less directly, yet as importantly, the government's purchasing power and virtue signaling can spur market innovation and galvanize public trust in AI technologies.

[22] The use of facial recognition AI technology in law enforcement, for example, is arguably inappropriate because of technological and human limitations. Recent proposals in Congress would create a moratorium on the use of such technology by law enforcement. *See* Facial Recognition and Biometric Technology Moratorium Act of 2020, S. 4084, 116th Cong. (2020) (as referred to S. Comm. on the Judiciary, June 25, 2020).

II. AI Today

AI has no universally accepted definition.[23] That dissensus owes in part to the wide cache of technologies that AI envisages. AI's definitional problem also reveals something about the concept itself: AI sweeps across fields of computer science, mathematics, psychology, sociology, neuroscience, and philosophy, and intersects with countless more. While disagreement persists about what AI means, there is wide consensus that civilization as we know it will never be the same.[24] Whether for better or worse is not the question. Instead, the questions are *whose lives* will be better and worse, in *which ways*, and under *what rules or conditions*.[25]

While disagreement persists about what AI means, there is wide consensus that civilization as we know it will never be the same. Whether for better or worse is not the question. Instead, the questions are *whose lives* will be better and worse, in *which ways*, and under *what rules or conditions*.

AI is disrupting every major market and facet of society. The technology is used in our phones, homes, cars, police stations, schools, social platforms, news feeds, satellites, workplaces, voting booths, and weapons systems. The unprecedented growth and dissemination of AI over the past decade owes to the conflation of several sociotechnical

---

[23] *See* U.S. Gov't Accountability Office, GAO-18-142SP, Artificial Intelligence: Emerging Opportunities, Challenges, and Implications (2018) [hereafter GAO, Artificial Intelligence] (observing "there is no single universally accepted definition of AI, but rather differing definitions and taxonomies"). One provision of U.S. law broadly defines AI to include the following:

> (1) Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.
>
> (2) An artificial system developed in computer software, physical hardware, or another context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.
>
> (3) An artificial system designed to think or act like a human, including cognitive architectures and neural networks.
>
> (4) A set of techniques, including machine learning, designed to approximate a cognitive task.
>
> (5) An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision-making, and acting.

John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115–232, 132 Stat. 1636, 1695 (Aug. 13, 2018) (codified at 10 U.S.C. § 2358, note).

[24] *See, e.g.*, Organization for Economic Cooperation and Development, Artificial Intelligence in Society 122 (2019), https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_eedfee77-en [hereafter OECD, Artificial Intelligence] (providing a comprehensive survey of the many ways that AI is projected to transform social structures and power dynamics across markets and borders); GAO, Artificial Intelligence, *supra* note 23.

[25] *See* VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR 180–88 (2018); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 29–31 (2016).

developments: the availability of exponentially more data; increases in computing power; the democratization of the "internet of things" (mobile phones, tablets, etc.); and breakthroughs in "machine learning" research and development.

A full exposition of machine learning is beyond the scope of this report. But it will be important to unpack some of the key attributes of machine learning for the project ahead because, in many ways, the promises and perils of algorithmic governance are anchored to how machine learning systems are designed and operationalized.[26]

Stripped to its essentials, machine learning is (i) a statistical technique (ii) that learns from data (iii) to make classifications or predictions (iv) for new data inputs. For example, if the objective is to train a machine that distinguishes between pictures of cats and dogs, the machine can be fed thousands of labeled pictures of cats and dogs, learn the difference between them by finding correlations in the data, and generate an algorithmic model that can then be used to identify cats and dogs in real-world settings. Or, to predict home values, datasets of past home sales can be used to train an algorithmic model to predict the sales value of homes that are not in the training data.

Stripped to its essentials, machine learning is (i) a statistical technique (ii) that learns from data (iii) to make classifications or predictions (iv) for new data inputs.

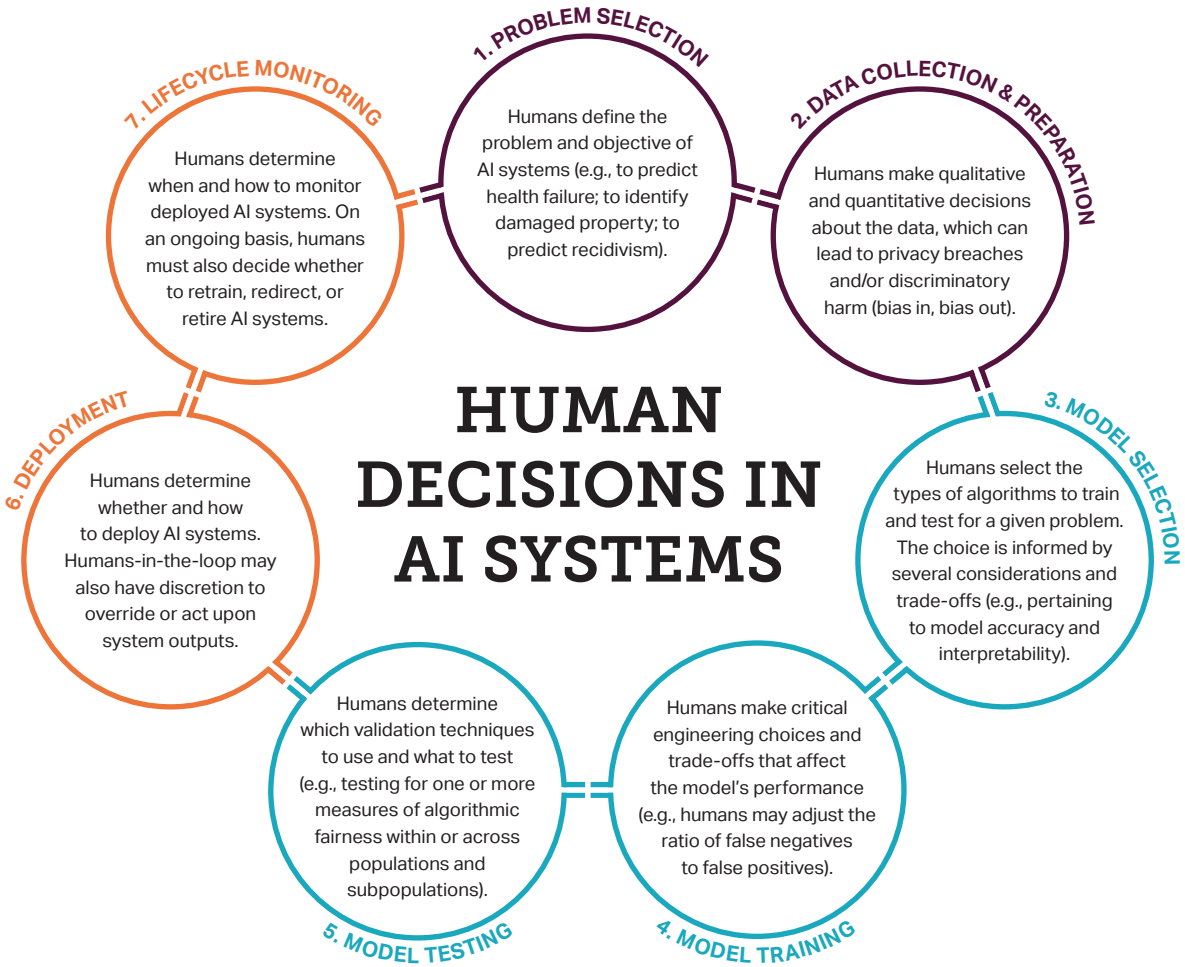
As these examples illustrate, machine learning is not a free-floating enterprise. Rather, it is part of a larger ecosystem comprised of data, humans, and human-computer interactions. For instance, humans generally select and clean the data, train and optimize machine learning algorithms, and deploy the algorithms in real-world or virtual settings. Moreover, humans make a wide variety of choices and trade-offs throughout the process. Just to name a few, humans must make choices about which datasets to include and exclude to train the model, which algorithmic model or models to use for a given task, which validation techniques to use, and which performance metrics to test for.[27]

---

[26] There are several different approaches to machine learning. For a short overview of the approaches, see Jatinder Singh et al., *Responsibility & Machine Learning*: Part of a Process, SSRN 4–9 (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860048.

[27] *Id.*; see also Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 773, 778 (2019) (discussing embedded values choices in AI system design); *see also* NSCAI, Second Quarter Recommendations, *supra* note 12, at 129–31 (discussing "trade-off decisions for AI systems [that] must be made about internal representations, policies of usage and controls, run-time execution monitoring, and thresholds").

Moreover, once a model is deployed, humans may be required to select and input new data. For example, an AI system designed to predict whether a criminal defendant is a high-risk recidivist will require a human to input features about the defendant (e.g., criminal history, age, home address).[28] Humans may also decide what to do, if anything, with the AI model's prediction or classification. In such systems, there is a so-called "human-in-the-loop." The human might be a judge, for example, who takes an algorithm's risk assessment into account when setting bail for a criminal defendant. Other AI systems, such as email spam filters, are called "autonomous" because human input is not required after the tool is deployed. In both types of systems, however, humans are responsible for many critical and consequential decisions throughout the AI lifecycle.



[28] AI systems are currently used throughout the country for this and other functions in the criminal justice system. *See, e.g.*, Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018); Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659 (2018).

III. Challenges of Algorithmic Governance

Under the right conditions, AI can make government more efficient and effective across a range of regulatory domains: from law enforcement to adjudication; from health care to building maintenance; from environmental protection to national defense; from policymaking to ministerial tasks.[29] Moreover, AI systems could potentially enhance government transparency, accountability, and fairness.[30] For example, AI decision-making systems might be more transparent and accountable than government officials who might conceal, or who might not be aware of, their decisional biases.[31] AI systems may also enable faster, more accurate, and more consistent decisions than humans, in contexts like social security, veterans benefits, immigration, and more. Furthermore, algorithmic governance may result in dramatic cost savings. According to one recent estimate, the federal government might save upwards of $500 billion over the next decade by "automating repetitive tasks" and by "augmenting" the capabilities of public-sector workers.[32]

But glaring challenges persist. As discussed in more detail below, AI tools are inherently risky, irrespective of whether public or private actors are utilizing them. Yet the risks are exacerbated when AI tools are wielded by the federal government—not only because of the types of harms that can occur, but because expectations and legal requirements differ between public and private action. Federal action is governed by the Constitution, administrative law, freedom of information laws, and federal procurement laws, in ways that do not apply to private action.[33] The point is not that private actors have free rein;

---

[29] *See* GOVERNMENT BY ALGORITHM, *supra* note 2, at 6 ("Rapid developments in AI have the potential to reduce the cost of core governance functions, improve the quality of decisions, and unleash the power of administrative data, thereby making government performance more efficient and effective."); Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 6 (2019) (describing how machine learning algorithms produce "unparalleled accuracy" compared to other statistical methods and human judgment).

[30] *See* David Freeman Engstrom & Daniel E. Ho, *Artificially Intelligent Government: A Review and Agenda*, in BIG DATA LAW (Roland Vogl ed., forthcoming 2020) (manuscript at 10) ("The perhaps counter-intuitive result is that the displacement of enforcement discretion by algorithm might, on net, yield an enforcement apparatus that is less opaque and more legible to agency heads and reviewing courts alike than the existing system."); Kroll et al., *supra* note 5, at 656–77 (explaining how, through proper design, AI systems can be made more transparent and accountable).

[31] *See* Daniel Castro, *Data Detractors Are Wrong: The Rise of Algorithms is a Cause for Hope and Optimism*, CTR. FOR DATA INNOVATION (Oct. 25, 2016), http://www.datainnovation.org/2016/10/data-detractors-are-wrong-the-rise-of-algorithms-is-a-cause-for-hope-and-optimism/.

[32] *See* Christina Bone et al., *The Coming of AI Productivity Boom: And How Federal Agencies Can Make the Most of It*, ACCENTURE, 1, 4 (2020).

[33] *See* Daniel Guttman, *Public Purpose and Private Service: The Twentieth Century Culture of Contracting Out and the Evolving Law of Diffused Sovereignty*, 52 ADMIN. L. REV. 859, 862, 881–90 (2000) (explaining that "in practice, two different sets of regulations have come to govern those doing the basic work of government": those that apply to federal officials, on the one hand, and those that apply to federal contractors, on the other). For an incisive treatment of the constitutional state action doctrine as applied to private AI vendors, see Kate Crawford & Jason Schultz, *AI Systems as State Actors*, 119 COLUM. L. REV. 1941, 1943–44 (2019) (arguing that courts should adopt a version of the state action doctrine to apply to vendors who supply AI systems for government decision-making).

they do not. Rather, the point is that legal norms around public and private action differ in ways that matter for algorithmic governance, especially as pertains to questions of transparency, accountability, privacy, safety, and fairness (broadly defined).[34]

A. RISK OF HARM

Unlike calculators, algorithmic classifications and predictions can be wrong. Of course, human classifications and predictions can be wrong too. But the efficiencies and scalability of AI systems make them especially risky.[35] One coding error, engineering choice, unfounded assumption, or lapse in human oversight can cause widespread harm. Moreover, when exposed to real-world elements, AI systems make mistakes that most humans never would.[36] Compounding the risk, AI systems may interact with other technologies, humans, or environmental conditions in ways that can negatively bear on those surrounding systems.[37]

B. TRANSPARENCY

AI systems raise a host of transparency challenges for algorithmic governance.[38] Technologically, some algorithms and design configurations are more scrutable than others. Machine learning "neural networks," which are some of the most powerful, sophisticated, and useful, are also the most difficult for humans to comprehend. The inputs and outputs can be known, but the so-called neural network that turns inputs into outputs can entail millions of data correlations, at scales that the smartest minds on earth cannot understand, much less accurately explain to anyone else. Machine learning

---

[34] The term "fairness" is borrowed here from the AI field and has no agreed-upon meaning. *See* Abigail Z. Jacobs & Hannah Wallach, *Measurement and Fairness* 1 (Microsoft, Working Draft No. 1912.05511, 2019), https://arxiv.org/pdf/1912.05511.pdf. A concept like "justice" may work just as well or better. For present purposes, what matters is the breadth of concerns that fairness (or justice) captures, including nondiscrimination, privacy, procedural due process, human rights, and more. *See infra* Subpart II.D.

[35] *See* Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE. J. L. & TECH. 103, 129 (2018) ("The ability of these algorithmic processes to scale, and therefore to influence decisions uniformly and comprehensively, magnifies any error or bias that they embody"); O'NEIL, *supra* note 25, at 29–31 (discussing the "scalability of algorithms"). For example, an efficient AI system that makes 100,000 predictions at a 10 percent error rate may negatively affect 1,000 individuals. An inefficient human who makes 100 recommendations at a 20 percent error rate may negatively affect 20 individuals.

[36] *See* Fabio Kepler, *Why AI Fails in the Wild*, UNBABEL (Nov. 15, 2019), https://unbabel.com/blog/artificial-intelligence-fails/; see also Colin Smith et al., *Hazard Contribution Modes of Machine Learning Components*, 2020 AAAI WORKSHOP ON ARTIFICIAL INTELLIGENCE SAFETY 4 (2020) (discussing unexpected performance, for example, "through unanticipated feature interaction . . . that was also not previously observed during model validation"), https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20200001851.pdf.

[37] Singh et al., *supra* note 26, at 15.

[38] *See* Jenna Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOC'Y 1, 3–5 (2016).

models of this type are metaphorical "black boxes." While they can perform well on the dimensions of efficiency and functionality, the complexity of the algorithmic models can eclipse human-scale reasoning.

Worth emphasizing here, model inscrutability is a feature of complex AI systems, not a bug. After all, the ambition and promise of machine learning is not to think as humans do, but rather to find statistical correlations in big datasets beyond human cognition or intuition. In some settings, the inscrutability of the model's decisional pathway may be of little concern. But for consequential governmental decisions, the inscrutability of AI systems is highly concerning. "[T]he algorithm made me do it" will not satisfy expectations for a human-centric explanation.[39]

For consequential governmental decisions, the inscrutability of AI systems is highly concerning.

Without clarity about how or why an AI system makes a prediction, the government may fail in its responsibility to provide a legally or morally sufficient reason for acting on it. Thus, the government may want or need to deploy "interpretable" and "explainable" AI tools.[40] Its ability to do so will depend on a variety of dynamic—and yet unsettled—legal, technological, and sourcing contingencies.

Trade secrecy can also interfere with algorithmic transparency.[41] In some contexts, the government may not have access to a vendor's trade secrets—for example, when the government purchases the technology as a "commercial item off the shelf."[42] In other contexts, the government may have access to a vendor's trade secrets, but federal law

---

[39] *See, e.g.*, Fairness, Accountability, and Transparency in Machine Learning, www.fatml.org ("[T]here is increasing alarm that the complexity of machine learning may reduce the justification for consequential decisions to 'the algorithm made me do it.'"); Victoria Burton-Harris & Philip Mayor, *Wrongfully Arrested Because Face Recognition Can't Tell Black People Apart*, ACLU (June 24, 2020) ("One officer responded, 'The computer must have gotten it wrong.'"), https://www.aclu.org/news/privacy-technology/wrongfully-arrested-because-face-recognition-cant-tell-black-people-apart/.

[40] *See, e.g.*, P. Jonathan Phillips et al., Nat'l Ins. Sci. & Tech, *Four Principles of Explainable Artificial Intelligence* (Aug. 2020), https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf; GOOGLE LLC, AI Explanations Whitepaper 1–28 (2019), https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf; The Royal Society, Explainable AI: The Basics Policy Brief (2019), https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf.

[41] *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); Levine, *supra* note 14, at 180-81.

[42] *See, e.g.*, 48 C.F.R. § 12.212 (providing, for the acquisition of commercially available computer software, that vendors generally "shall not be required" to "[r]elinquish to, or otherwise provide, the Government rights to use, modify, reproduce, release, perform, display, or disclose commercial computer software or commercial computer software documentation except as mutually agreed to by the parties").

or nondisclosure agreements may prevent the government from revealing the secrets to third parties.[43] As a result, information about an AI system may be withheld from individuals affected by the system, government watchdogs, and perhaps even judges and lawmakers.[44]

The right to privacy is another friction point.[45] Most notably, the data ingested by AI systems may contain sensitive personal information that can be traced to individuals (even if names and other identifying attributes are scrubbed from the data).[46] AI systems may thus run headlong into privacy laws, in a wide variety of contexts where the government is provided personal information for specific purposes that may not be used or disclosed for other purposes.[47]

Moreover, in law enforcement and national security contexts, full transparency about AI systems may be self-defeating or yield bad outcomes[48] because AI systems can be "gamed" or "hacked" by adversarial actors in ways that humans cannot.[49] Of course, humans can be manipulated, bribed, or spied upon in ways that undermine or endanger the public interest. The rub, however, is that AI systems have similar human vulnerabilities (e.g., data scientists, computer programmers, technical engineers,

---

[43] *See id.*; Katherine Fink, *Opening the Government's Black Boxes: Freedom of Information and Algorithmic Accountability*, 21 INFO., COMM. & SOC'Y 1453 (2017) (reviewing current state of law and practice with respect to whether algorithms would be considered "records" under the Freedom of Information Act (FOIA) and reviewing agency bases for withholding algorithms and source code under FOIA requests).

[44] *See* Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1299–1302 (2020) (discussing impediments to algorithmic transparency under FOIA and trade secrecy laws, which hinder public interest groups and watchdogs from obtaining information about AI tools used by government).

[45] *See* Engstrom & Ho, *Artificially Intelligent Government*, *supra* note 30, at 11–12 (canvassing a range of federal privacy laws to explain how "privacy and data security constraints, while designed to safeguard privacy and minimize public burdens, can also impose significant costs on agencies, reduce the efficacy of algorithmic tools, and stymie agency innovation").

[46] *See* Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, 2008 IEEE SYMPOSIUM ON SECURITY AND PRIVACY 111 (2008). This is a major concern, especially in light of malicious threats to information security. *See, e.g.*, Zolan Kanno-Youngs & David E. Sanger, *Border Agency's Images of Travelers Stolen in Hack*, N.Y. TIMES (June 10, 2019), https://www.nytimes.com/2019/06/10/us/politics/customs-data-breach.html; Julie Hirschfield Davis, *Hacking of Government Computers Exposed 21.5 Million People*, N.Y. TIMES (July 9, 2015), https://www.nytimes.com/2015/07/10/us/office-of-personnel-management-hackers-got-data-of-millions.html.

[47] *See, e.g.*, 5 U.S.C. §§ 552a(b) & (e)(3) (prohibiting disclosure of records without the prior written consent of the person whom the records pertain to, excepting for reasons such as routine use for, inter alia, census purposes, matters of the House of Congress or any of its committees or subcommittees, etc.); Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104–191, 110 Stat. 1936 (1996) (codified as amended in scattered sections of 18, 26, 29 and 42 U.S.C. (1996) (setting forth privacy and security standards for protecting personal health information).

[48] *See* Engstrom & Ho, *supra* note 30 (explaining that "in enforcement context[s] . . . transparency facilitates strategic action that can drain [AI] tools of their value"); *see also* 5 U.S.C. § 552(b)(7)(E) (exempting "records or information compiled for law enforcement purposes" whose disclosure "could reasonably be expected to risk circumvention of the law"). *But cf.* Ignacio N. Cafone & Katherine J. Strandburg, *Strategic Games and Algorithmic Secrecy*, 64 MCGILL L.J. (forthcoming 2020) (arguing that the range of situations in which people are able to game decision-making algorithms is narrow, even when there is substantial disclosure), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3440878.

[49] In machine learning literature, the gaming problem is associated with "adversarial learning"—the problem of developing models when it is anticipated from the beginning that adversaries will try to defeat them. Guofu Li et al., *Security Matters: A Survey on Adversarial Machine Learning*, ARXIV (2018), https://arxiv.org/abs/1810.07339.

domain experts) *plus* technical vulnerabilities in the code, cloud, or hardware that can be exploited in ways that elide easy detection. For similar reasons, cybersecurity concerns will generally trump algorithmic transparency around critical infrastructure: in energy, transportation, telecommunication, voting systems, waterways, etc.

> **AI systems have similar human vulnerabilities (e.g., data scientists, computer programmers, technical engineers, domain experts) *plus* technical vulnerabilities in the code, cloud, or hardware that can be exploited in ways that elide easy detection.**

As these examples demonstrate, there is a lot of algorithmic governance that could be off-limits to full transparency. Yet AI systems are already being utilized across all of these domains, with many more in the works.

C. ACCOUNTABILITY

The transparency concerns in algorithmic governance are directly related to a set of accountability concerns. The less stakeholders know, the more difficult it becomes to ascertain whether the human inputs and machine outputs are accurate, fair, and legal. And without those determinants, stakeholders cannot know which actors, if any, should be held accountable for any resulting harms.

One way that our system holds government actors accountable is through judicial review. The use of AI systems in government decision-making, however, can stymie a court's ability to know or understand the reasons for an agency's action if the agency (or its vendors) cannot adequately explain why an AI tool made a particular prediction or classification that led to a government decision.[50] Beyond judicial settings, government watchdogs, journalists, and stakeholders are similarly constrained in their ability to "look under the hood" of AI tools affecting the polity's rights and interests.[51] This is highly

---

[50] *See* Rebecca Wexler, *When a Computer Program Keeps You in Jail*, N.Y. TIMES (June 13, 2017) ("The root of the problem is that automated criminal justice technologies are largely privately owned and sold for profit."), https://www.nytimes.com/2017/06/13/opinion/howcomputers-are-harming-criminal-justice.html; *see also* Sonia K. Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV. 1183, 1186–87 (2019) (expounding on this concern beyond litigation settings).

[51] *See* Bloch-Wehba, *supra* note 44 (emphasizing the importance of open records laws for third-party stakeholders to hold government accountable outside of litigation settings); Brauneis & Goodman, *supra* note 35, at 159 (complaining that "the information allegedly protected by trade secret law may lie at the heart of essential public functions and constitute political judgments long open to scrutiny").

problematic because it shutters out stakeholder input and breeds public distrust in government uses of the technology.

D. FAIRNESS

The transparency and accountability gaps in algorithmic governance are especially concerning because the use of AI may not be procedurally or substantively fair. Procedurally, AI systems raise due process concerns that, to date, have gone unmet in several high-profile cases.[52] In groundbreaking work, Danielle Citron provides an account of the "automated administrative state" using software to determine whether someone should receive "Medicaid, food stamp, and welfare" benefits, be on a no-fly list, or be identified as owing child support.[53] As she cogently explains, "[a]utomation jeopardizes the due process safeguards owed individuals and destroys the twentieth-century assumption that policymaking will be channeled through participatory procedures that significantly reduce the risk that an arbitrary rule will be adopted."[54] These concerns, aired more than a decade ago, have only intensified during the intervening years because machine learning algorithms, which are on the rise, are generally more complex and less explainable.[55]

> **Without proper precautions, and even with precautions, technical bias can refract and reify social biases in the analog world.**

Beyond procedural unfairness, the use of AI tools can be unfair for technological, legal, or moral reasons.[56] Algorithmic models are statistical simplifications that cannot consider all possible relevant facts about subjects. Generalizations and profiling thus typify AI systems.[57] While AI-generated predictions are not inherently unfair,

---

[52] *See* Rashida Richardson, Jason M. Schultz, Vincent M. Southerland, *Litigating Algorithms*, AI NOW INSTITUTE (2019), https://ainowinstitute.org/litigatingalgorithms-2019-us.pdf.

[53] Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1256 (2008).

[54] *Id.* at 1256–57.

[55] Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. (forthcoming 2020), https://papers.ssrn.com/abstract_id=3553590 ("In the decade since the publication of *Technological Due Process*, governments have doubled down on automation despite its widening problems."); *see also* Mulligan & Bamberger, *supra* note 27, at 814–18 (explaining how the machine learning tools of today are even more problematic from a legal standpoint than the AI tools that were the focus of Citron's original study).

[56] *See, e.g.*, Ninaresh Mehrabi et al., *A Survey on Bias and Fairness in Machine Learning*, ARXIV (2019), https://arxiv.org/pdf/1908.09635.pdf.

[57] *See* Mulligan & Bamberger, *supra* note 27, at 787 ("Predictive algorithms are essentially autonomous profiling by a machine-learning system.").

they will always be biased, in a technical sense, because the predictions are based on generalizations mined from data. Without proper precautions, and even with precautions, technical bias can refract and reify social biases in the analog world.

For example, an algorithm trained on historical data to predict criminal recidivism may exhibit higher false positive rates for Black people if the training data is an artifact of past discriminatory policing against that population.[58] Likewise, an AI system designed to predict a "good hire" for a government position may make predictions based on promotion rates and employee evaluations of past government hires. If those input variables in the training data are biased toward men, then algorithmic predictions for future "good hires" will be biased against women.[59]

Separately, but relatedly, choices in model engineering can result in discrimination. Of particular concern are choices about feature selection in the training and testing data.[60] For example, an AI model may be explicitly trained to account for race or sex in decision-making contexts where it is illegal to discriminate on those grounds. Even if an algorithm is trained to ignore race or sex, proxies for those attributes might intentionally or unintentionally be extracted from the data and lead to the same results. For example, zip codes are a well-known proxy for race.[61]

For sensitive government decisions, judges and civil servants may be expected to exercise human judgment as a check on algorithmic predictions. In practice, however, studies reveal the risk of "automation bias," whereby humans-in-the-loop exhibit overconfidence in algorithmic predictions and classifications.[62] Automation bias raises the stakes of any unfairness baked into the algorithm itself because humans may not correct for errors. But risks also run in the opposite direction: When a human-in-the-loop compensates for known algorithmic biases, the human interventions may not

---

[58] *See, e.g.*, Richardson et al., *supra* note 19; Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://perma.cc/JRR9-5D29; cf. Avi Feller et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear.*, WASH. POST: MONKEY CAGE (Oct. 17, 2016), https://perma.cc/7M7V-GPKL

[59] *Cf. Amazon Ditched AI Recruiting Tool that Favored Men for Technical Jobs*, THE GUARDIAN (Oct. 10, 2018), https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine.

[60] *See* Nicholas Diakopoulos, *Algorithmic Accountability: Journalistic Investigation of Computational Power Structures*, 3 DIGITAL JOURNALISM 398, 400–02 (2015) (discussing the value choices embedded in data prioritization, classification, association, and filtering).

[61] Rhema Vaithianathan et al., *Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County Methodology and Implementation*, CTR. FOR SOC. DATA ANALYTICS 12 (2017) (discussing zip codes and other proxies for race), https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf.

[62] *See* Citron, *supra* note 53, at 1271–72 (discussing "automation bias"); Kate Goddard, Abdul Roudsari, Jeremy C. Wyatt, *Automation Bias: Empirical Results Assessing Influencing Factors*, 83 INT'L J. MED. INFORMATICS 368 (2014).

be fair or legal. For instance, if an algorithm used for government hiring exhibits bias toward hiring men, to what extent (if any) can the algorithmic prediction be favorably adjusted toward hiring women (without running afoul of equal-protection principles)?[63] As this example illustrates, fairness depends on definitions of fairness—of which there are many—and legal principles that do not map easily on AI systems.[64]

Another source of algorithmic bias stems from the lack of diversity in the technology industry.[65] A notorious example is facial recognition technology that was trained on the faces familiar to the design engineers, which were mostly white.[66] Consequently, the facial recognition software had a greater propensity to misidentify dark-skinned faces.[67] Needless to say, the disparity is especially concerning in high-stakes contexts such as law enforcement. Recently, police in Detroit wrongfully arrested a Black man at home in front of his family, including his young daughters.[68] The charges were subsequently dismissed after the "officers-in-the-loop" acknowledged the misidentification, but the damage was already done.[69]

DHS and the FBI use similar facial recognition technology, in conjunction with other AI-enabled surveillance tools, in immigration and other law enforcement.[70] Even if the technology can be made more accurate—for example, by correcting for known biases in the training data—that would not address the overarching concerns relating to power, autonomy, and liberty. Having the technical capability for highly accurate surveillance says nothing about whether, or for what purposes, that capability should be wielded.

---

[63] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 694–714 (2016) (noting the ways in which algorithmic data mining techniques can lead to unintentional discrimination against historically prejudiced groups).

[64] *See* Deirdre K. Mulligan et al., *This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology*, PROC. 2019 ACM ON HUMAN-COMPUTER INTERACTION 3, 119 (2019). For example, fairness can be operationalized around group metrics or individual metrics, of which there are many types of each. For a discussion, see *id.*

[65] *See* RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2019) (arguing that human social bias is engineered into automated technology because (overwhelmingly white and male) programmers fail to recognize how their understanding of technology is informed by their identities, and the raw data on which robots of all types are trained are products of racist, sexist, and classist societies).

[66] *See, e.g.*, Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots*, ACLU (July 26, 2018), https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28.

[67] *See id.*

[68] *See* Kashmir Hill, *Wrongfully Accused by an Algorithm*, N.Y. TIMES (Jun. 24, 2020), https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

[69] *Id.* ("[T]he Wayne County prosecutor's office said that . . . Williams could have the case and his fingerprint data expunged. 'We apologize,' . . . 'This does not in any way make up for the hours that Mr. Williams spent in jail.'").

[70] Aaron Boyd, *CBP Is Upgrading to a New Facial Recognition Algorithm in March*, NEXTGOV.COM (Feb. 7, 2020), https://www.nextgov.com/emerging-tech/2020/02/cbp-upgrading-new-facial-recognition-algorithm-march/162959/; Kimberly J. Del Greco, Deputy Assistant Dir., Crim. Justice Info. Serv. Div., FBI, Statement Before the House Oversight and Reform Committee, Facial Recognition Technology: Ensuring Transparency in Government Use (Jun. 4, 2019).

More generally, AI may simply be an inappropriate solution for many government problems. This is especially true for AI systems trained to predict "unobservable theoretical constructs," which by definition are neither observable nor verifiable.[71] Instead, such constructs must be inferred from observable properties that a system designer thinks are closely (enough) related.[72] To illustrate, compare the constructs of "crime" and "criminality." Both are government problems, but the use of AI to address them differs dramatically. Using AI to detect *crime* is not inherently problematic because humans can verify whether, in fact, criminal activity has occurred or is occurring.

The same cannot be said about a person's *criminality*. In a recent academic paper, the authors boasted of an AI tool "capable of predicting whether someone is likely going to be a criminal" with "80 percent accuracy and with no racial bias."[73] In response, thousands of AI experts and practitioners published an open letter that condemned the paper and urged the journal not to publish it.[74] As the open letter explains, criminality based on a person's facial features "reproduces injustice and causes real harm."[75] Criminality itself is a racialized construct, and to infer it based on immutable characteristics of a person's face is an affront to moral justice.[76] One can easily imagine how an AI tool to predict criminality based on a person's appearance can be used inappropriately by government actors to surveil, detain, or deny rights and access to government services.[77]

---

[71] *See* Abigail Z. Jacobs & Hannah Wallach, *Measurement and Fairness* 1 (Microsoft, Working Draft No. 1912.05511, 2019), https://arxiv.org/pdf/1912.05511.pdf.

[72] *Id.*

[73] *HU Facial Recognition Software Predicts Criminality*, HARRISBURG UNIV. OF SCI. AND TECH. (May 5, 2020), https://web.archive.org/web/20200506013352/https:/harrisburgu.edu/hu-facial-recognition-software-identifies-potential-criminals/.

[74] *See* Coal. for Critical Tech., *Abolish the #TechToPrisonPipeline: Crime Prediction Technology Reproduces Injustices and Causes Real Harm*, MEDIUM (June 23, 2020), https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16.

[75] *Id.*

[76] *Id.* ("Data generated by the criminal justice system cannot be used to 'identify criminals' or predict criminal behavior. Ever."). Following the backlash, the authors later deleted the paper, and their publisher, Springer, confirmed that it had been rejected. Sidney Fussell, *An Algorithm That 'Predicts' Criminality Based on a Face Sparks a Furor*, WIRED (June 24, 2020, 7:00 AM), https://www.wired.com/story/algorithm-predicts-criminality-based-face-sparks-furor/.

[77] Conceivably, and assuming accuracy, an AI tool like this could be used for putatively beneficent reasons—for instance, to offer community support for those with facial features that bespeak criminality. Still, that type of profiling would almost certainly be an unwelcome (and unlawful) affront to privacy and dignity.

IV. Recommendations

The proliferation of AI use by federal agencies has raised urgent questions about how algorithmic governance itself should be regulated. As noted earlier, several bills recently introduced in Congress relate, in some way, to the government's use of AI technologies. Moreover, a range of policy initiatives are underway or under consideration within the executive branch. To greater and lesser extents, the prescriptions on the table speak to the core challenges of algorithmic governance: risk of harm, transparency, accountability, privacy, security, and fairness.

By no means is federal procurement law the sole solution to these complex problems. But procurement must be part of the solution. Currently, the government is investing huge sums of taxpayer dollars to acquire AI systems that may be unusable, either because they are not trustworthy, or because they fail to pass legal muster. If the government cannot explain how an AI system works, for example, then it may violate constitutional due process[78] or administrative law.[79] Even if an AI system clears those hurdles, it may still violate federal antidiscrimination laws, privacy laws, and domain-specific laws and regulations. Litigation will no doubt surface these risks and harms. But much of that screening can occur, ex ante, through the acquisition gateway.

Currently, the government is investing huge sums of taxpayer dollars to acquire AI systems that may be unusable, either because they are not trustworthy, or because they fail to pass legal muster.

In a recent report, the National Security Commission on Artificial Intelligence recommended a set of best practices "in support of core American values" for the "responsible development and fielding AI technologies" by the government.[80] Of

[78] *See, e.g.*, Citron, *supra* note 53; Aziz Z. Huq, *Constitutional Rights in the Machine Learning State*, 105 CORNELL L. REV. (forthcoming 2020); Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327 (2015); Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871 (2016).

[79] *See, e.g.*, David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. REG. 800 (2020); Mulligan & Bamberger, supra note 28; Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147 (2017).

[80] NSCAI, Second Quarter Report, *supra* note 12, at 120–55 (Appendix 2). Those American values include the "rule of law," "[v]alues established in the U.S. Constitution, and further operationalized in legislation, include freedoms of speech and assembly, the rights to due process, inclusion, fairness, nondiscrimination (including equal protection), and privacy (including protection from unwarranted government interference in one's private affairs), as well as international human rights and dignity." *Id.* at 125. *See also id.* at 126–27 (discussing additional values in the context of the military and warfighting).

particular note here, the commission highlighted the challenges around ethical AI, and stressed that systems acquired from contractors "should be subjected to the same rigorous standards and practices—whether in the acquisitions or acceptance processes."[81] Clearly, there is a demand for ethical AI in government. Just as surely, federal procurement law can be harnessed in service of that critical mission.

A. TAKING STOCK

Currently, there is no easy way to identify which agencies are using which AI technologies, for what purposes, from which vendors, and under what constraints. Without such information, policymakers, watchdogs, and stakeholders cannot know whether the systems in use are transparent, accountable, fair, secure, testable, and so on.[82]

As a first step, federal lawmakers should mandate the creation of a government-wide inventory report that includes clear information about each AI system that is currently in production or use by the federal government. The Office of Management and Budget, the General Services Administration, or some other centralized agency should be assigned responsibility for the oversight and reporting of the inventory, which should be publicly available and updated annually. Exceptions can be made as needed for national security, intelligence agencies, or otherwise.

As a general rule, however, the disclosure requirements should include information about all AI systems that the government is currently using in adjudicatory, rulemaking, or law enforcement settings that affect a person's rights, duties, or access to public benefits and services, or that may otherwise cause property damage or personal injury.[83] The disclosure requirements should also include AI systems that are used for certain internal operations, such as hiring, resource allocation, and other nontrivial ministerial tasks. Too often, internal agency operations get overlooked in discussions about algorithmic governance because they generally are not judicially reviewable under framework statutes such as the Administrative Procedure Act. But judicial reviewability is not the purpose of

---

[81] *Id.* at 124.

[82] Along these lines, the Artificial Intelligence Reporting Act of 2018, H.R. 6090, calls for annual congressional reporting "on the use and status of unclassified machine learning and artificial intelligence applications across the Federal Government," from the subgroup on Machine Learning and Artificial Intelligence of the Committee on Technology of the National Science and Technology Council. This bill has not progressed beyond the House Committee on Science, Space, and Technology.

[83] For examples of pending legislation that offer definitions of "high risk" automated decision-making systems, see Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong. § 2 (2019) (defining the term in connection with the regulation of commercial uses of AI technologies).

the inventory. And, in any event, the triggers for judicial review do not capture the scope of agency activity that policymakers and the public have an interest to know about. For example, if an AI system is biased in hiring decisions, or in how government resources are allocated, those internal managerial tools can surely have a detrimental effect on government performance and public policy.

For any nonexempt systems, the minimum reporting requirements should include:

1. A list of all vendors or nongovernmental partners that have participated in the design, development, testing, auditing, deployment, or use of the AI system. This should include a description of each vendor or partner role, and the total amount of federal funds paid or contractually committed for the work.
2. A description of the AI system's general capabilities and intended purposes, including reasonably foreseeable capabilities and purposes outside the scope of the agency's intended use.
3. A description of the training and testing data of an AI system; how that data was generated, collected, and processed; the type or types of data that the system is reasonably likely to generate; and whether the government, vendors, or third parties have access to or control over that generated data.
4. A description of whether the algorithmic model used in an AI system can be altered by its users; and if so, by whom, under what circumstances and with what disclosers, and with what safeguards to protect the integrity of the model and traceability of any alterations.
5. A description of whether the AI system has been tested for bias (discriminatory or otherwise); whether the testing was performed by a third party; the date and methodology employed for such testing; whether the AI system has any known biases, and if so, how those biases are accounted for, corrected, tracked, or managed.
6. A description of whether the AI system gives notice to impacted individuals when it is used to collect personal information about those individuals or is used in decision-making that involves them.
7. For any piece of information above that is not disclosed, provide specific reason(s) why the information either cannot or will not be disclosed for purposes of this reporting, and whether the information may be available in other forms or by other means.

This inventory report would not be limited to AI systems acquired from nongovernmental vendors. But it would capture those systems that have already

passed through the acquisition gateway, which is notoriously opaque to all but the contracting community (and murky, too, for many within that community).[84] By making this information publicly available, and on a rolling basis, interested stakeholders can help agencies bridge information gaps and identify sociotechnical blind spots in the government's uses of AI systems. As earlier noted, the government does not have the in-house capacity needed to assess, anticipate, and address the challenges of algorithmic governance. By voluntarily sharing this information, the government can leverage the benefit of public input from scientists, academics, laypersons, interest groups, journalists, and members of government who do not have ready access to the information that an inventory report could provide.

B. FEDERAL ACQUISITION POLICY AND PRACTICE

There are many ways that procurement law may be retrofitted to meet the challenges of algorithmic governance. By way of background, the Federal Acquisition Regulation (FAR) is "the primary regulation for use by all Federal Executive agencies in their acquisition of supplies and services with appropriated funds."[85] FAR addresses various aspects of the procurement process, including acquisition planning, contract formation, source selection, auditing, and contractual management. Moreover, FAR captures a range of objectives, including market competition, integrity, transparency, efficiency, government satisfaction, best value, wealth distribution, and risk avoidance.[86] These pluralistic values do not always align; trade-offs among them are necessary and inevitable. FAR makes those trade-offs through a range of procurement processes, contractual provisions, performance incentives, accountability mechanisms, and a mix of delegated discretion and nonnegotiable directives.[87]

In addition to FAR, the federal acquisition system is governed by congressional statutes, agency regulations, guidance documents, and presidential executive orders. Thus, adjustments to federal procurement law—included those recommended here—can come from Congress, the White House, or designated agencies. Moreover, the

---

[84] *Cf.* Jarrod McAdoo, *How New Initiatives Might Make Federal Sales Easier*, WASH. TECH. (Jun. 2, 2020) ("The procurement process creates a jungle of barriers including crushing complexity, awkward communications and significant expense just to try and compete for business.").

[85] Gen. Servs. Admin. et al., *Foreword* to Federal Acquisition Regulation, at v (2019), https://www.acquisition.gov/sites/default/files/current/far/pdf/FAR.pdf.

[86] *See* Steven L. Schooner, *Desiderata: Objectives for a System of Government Contract Law*, 11 PUB. PROCUREMENT L. REV. 103 (2002).

[87] *See* Steven L. Schooner, *Fear of Oversight: The Fundamental Failure of Businesslike Government*, 59 AM. U. L. REV. 627 (2001).

recommendations below can be taken up separately or in combination. Regardless, they are designed to be cohesive and interoperable with other lawmaking initiatives around algorithmic governance.

a. Pre-Acquisition AI Risk Assessment

Federal law requires that agencies conduct safety risk assessments pertaining to information systems.[88] This requirement, and others like it, are designed to focus agency attention on risk factors in sensitive government contexts so that risks can be addressed and mitigated on an ongoing basis.[89] Currently, no such requirement exists for AI systems in particular. But, given the challenges that inhere in algorithmic governance, federal lawmakers should require agencies to develop and utilize AI-specific risk assessments as part of the acquisition process.[90]

This requirement could fit neatly within current procurement law and practice. FAR already mandates that agencies engage in acquisition planning, which incorporates special considerations beyond mere dollars and cents. For example, agency planners are required to comply with pre-established "Guiding Principles" for green-energy building construction and renovation.[91] Likewise, procurement law can explicitly require that agencies prepare AI risk assessments tailored for the unique challenges of algorithmic governance.

Importantly, the risk assessment should be conducted by a multidisciplinary team that includes agency acquisition and IT personnel, domain experts, legal experts, sociotechnical ethicists, and data specialists. Moreover, as much as possible, the team

---

[88] 44 U.S.C. § 3554(b) ("Each agency shall develop, document, and implement an agency-wide information program to provide periodic assessments of the risk and magnitude of the harm that could result from the unauthorized access, use, disclosure, disruption, modification, or destruction of information and information systems that support the operations and assets of the agency . . ."); *see also* 40 U.S.C. § 11331 (delegating to the Office of Management and Budget (OMB) and National Institute of Science & Technology (NIST) the authority to "promulgate information security standards pertaining to Federal information systems"); Office of Mgmt. & Budget, Exec. Office of the President, OMB Circular A-130, Appendix IV: Analysis of Key Sections (2016), https://obamawhitehouse.archives.gov/omb/circulars_a130_a130appendix_iv ("Each agency program official must understand the risk to [information] systems under their control.").

[89] *See* Nat'l Inst. of Standards and Tech., U.S. Dept. of Comm., NIST Guide for Conducting Risk Assessments, Special Publication 800–30 (2012), https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf (providing comprehensive risk assessment guidance for federal information systems).

[90] It bears noting that risk management is considered best practice when private enterprises acquire AI technologies. For a private sector framework, see for example: Deloitte Ctr. for Regulatory Strategy, AI and Risk Management Innovating with Confidence (2018), https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/innovatie/deloitte-nl-innovate-lu-ai-and-risk-management.pdf.

[91] *See* 48 C.F.R. § 7.103(p)(3); *see also* U.S. Envtl. Prot. Agency, *Guiding Principles for Sustainable Federal Buildings*, EPA (2016), https://www.epa.gov/greeningepa/guiding-principles-federal-leadership-high-performance-and-sustainable-buildings (last visited Aug. 12, 2020).

should be comprised of individuals with diverse backgrounds and perspectives, which can help to address potential blind spots relating to bias and other social concerns.

Whether these multidisciplinary teams should be centralized for government-wide deployment or decentralized within agencies is an important question. There are advantages and disadvantages to both approaches.[92] This report does not directly take up this question, except to note that the government's options may be limited by circumstance. Under existing conditions, the government's in-house capacity challenges may necessitate centralizing expertise to provide consulting, advice, and support for AI acquisitions on a government-wide basis. Currently, the General Services Administration (e.g., "18F" and Centers of Excellence)[93] and the Office of Management and Budget (e.g., US Digital Service) are serving these roles.[94]

Substantively, an AI risk assessment should address the issues outlined below and, to promote accountability, should be signed by the agency official overseeing the acquisition.

1. **Will the AI tool impact individuals, businesses, and communities in high-risk or sensitive contexts?**

   If yes, then thorough consideration must be given to issues of transparency, accountability, fairness, performance, privacy, and safety (which are separately addressed below).

2. **To what extent, if any, will the agency need to rely on vendors and outside consultants to design, develop, implement, audit, and monitor the system?**

   Acquiring AI systems is not a one-time decision; designing the system and testing it over its lifecycle are necessary to ensure that the tool is performing as intended, accurately, and within legal bounds. If the government's reliance on vendors and consultants is more than nominal, then considerations about cost, security, and control of the system should be explicitly accounted for in the risk assessment. Moreover, depending on the degree of the government's anticipated reliance on third parties, this risk factor could itself be a reason to forego acquiring an AI

---

[92] The advantages of centralization include coordination, concentrated expertise, and the development and deployment of best practices on a government-wide basis. *See* Mulligan & Bamberger, *supra* note 27, at 830–32 (favoring a centralized approach). The disadvantages of centralization include cultural and informational gaps between a centralized group of AI experts and the agency personnel who will, post-procurement, need to use, understand, and trust the technology. *Cf.* Aaron Boyd, *SBA Spent $30M on a Digital Service-Built App that Doesn't Work*, NEXTGOV (Aug. 3, 2020) (reporting that "confusion over leadership and culture clashes led to a runaway project with little oversight").

[93] *See* Gen. Servs. Admin., Technology Transformation Services, https://www.gsa.gov/about-us/organization/federal-acquisition-service/technology-transformation-services (last visited Sept. 6, 2020).

[94] *See An inside look at USDS*, U.S. Digital Service, https://www.usds.gov/news-and-blog (last visited Sept. 6, 2020).

solution.[95] If the acquiring agency cannot assemble and maintain an appropriately dedicated multidisciplinary team to oversee and manage the acquisition, that risk will exacerbate all the others, and should probably lead to a no-go decision on acquiring an AI system for use in high-risk or sensitive contexts.

3. **Will the data used or generated by the AI system contain sensitive personal information?**

   If yes, then additional questions relating to data privacy, data integrity, and data security should be considered and addressed.

4. **Will there be a human-in-the-loop, and if so, at what point in the operational workflow?**

   If the AI system is used in high-risk or sensitive contexts, then human validation of the inputs and outputs will likely be necessary. Moreover, depending on context, a human-in-the-loop may be required to exercise discretionary judgment, in which case, risks about computer-human interactions must also be accounted for. Those risks include biased *disregard* of an algorithmic prediction, which may occur if the human is predisposed to contrary outcomes, or does not understand (or trust) how the system operates. Risks at the human-computer interface, however, also include a system operator's *overconfidence* in AI predictions, which may result from a lack of appreciation for how the system operates, resource constraints, or the accretion of domain expertise and human judgment over time.

5. **How transparent is the AI system?**

   As earlier explained, AI systems can be more or less transparent for a variety of reasons. Thus, agencies should not treat transparency as a monolithic concern, but rather as a compendium of system features that should separately be accounted for in the AI risk assessment. Below are some common transparency concerns that should be specifically addressed:

   - **"Interpretability" (and "Explainability"):**[96] Numerous efforts are underway to make complex algorithmic models more interpretable and explainable to humans.[97] As yet, there are no industry or government standards—in part

---

[95] *Cf.* NSCAI, Second Quarter Recommendations, *supra* note 12, at 89 ("Despite pockets of excellence, the government lacks wide expertise to envision the promise and implications of AI, translate vision into action, and develop the operating concepts for using AI.").

[96] *See* Gabriel Nicholas, 4 GEO. L. TECH. REV. 711, 715 (2020) (noting disagreement in the literature over the nomenclature around "interpretability" and "explainability"). For present purposes, I use the terms interchangeably to refer to the ability of humans to understand how an AI system generates outputs from inputs.

[97] *See supra* note 40.

because algorithmic explainability carries its own set of risks. The more the model's decisional pathway is reduced for human comprehension (i.e., dumbed down), the more the explanation will depart from the truth of how a decision was actually made. Moreover, the explaining algorithms are vulnerable to gaming and adversarial attack. Consequently, computer-generated explanations can be unintentionally or intentionally misleading,[98] creating risk in both directions.

- **Data:** Because an AI model learns from data, access to information about training data (or lack thereof) is critical for evaluating an AI system's intended purposes and functionality. Not all data is created equal; there are gradients of data integrity, data bias, and associated risks. Moreover, vendors may claim trade secrecy protection over datasets or the process by which data was assembled and used to train the model. The lack of transparency around data can be highly risky.
- **Model Versioning (e.g., Version 1.0, 1.5, 2.0):** Throughout the lifecycle of an AI system, algorithmic models may be updated to account for new data, new workflows, new technology, and so on. An AI system that is modified may improve performance, but may cause the system to perform *differently* than anticipated or originally conceived. Moreover, without proper precautions, model versioning can make it impossible to know how a model performed when a particular decision was made. If that knowledge is needed in a court setting, but is unavailable, an agency will be hard-pressed to legally justify the specific output that is the subject of litigation.

b. Dividends of AI Risk Assessment

Apart from good practice, pre-acquisition AI risk assessments can serve a variety of useful functions across the lifecycle of an AI system. Most obviously, the deliberation and documentation of known and foreseeable risks will force conversations about whether an AI solution is appropriate at all, and if so, under what conditions. Along similar lines, the risk assessment can inform decisions about data needs, whether data is available on the market, and costs relating to data curation, labeling, scrubbing, enrichment, etc.

Further dividends from risk assessments can accrue during the market solicitation phase. More specifically, contracting officials can use the catalogued risks as focal points for market competition in connection with the agency's requests for information (RFIs),[99]

---

[98] *See* Dylan Slack et al., *Fooling LIME and SHAP: Adversarial Attacks on Post Hoc Explanation Methods*, In Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), https://doi.org/10.1145/3375627.3375830.

[99] *See* 48 C.F.R. § 15.201(e) ("RFIs may be used when the Government does not presently intend to award a contract, but wants to obtain price, delivery, other market information, or capabilities for planning purposes. Responses to these notices are not offers and cannot be accepted by the Government to form a binding contract.").

requests for proposals (RFPs),[100] and requests for quotations (RFQs).[101] By way of illustration, contracting officials can solicit information and proposals from vendors along the following lines:

1. Describe your privacy and cybersecurity approach to the proposed AI system, including but not limited to how the data will be protected.
2. Explain your testing and validation processes for performance, fairness, and discrimination, including any special expertise or innovative approaches that you might use to evaluate the AI system throughout its lifecycle.
3. Describe any anticipated data limitations and challenges, and any strategies or solutions that you might implement to address them.
4. Describe any training programs that your team members have undergone, and any official policies or protocols adopted by your organization, that specifically relate to transparency, accountability, fairness, privacy, or other ethical AI principles.
5. Describe how you enable end-to-end auditability of the system by government personnel, and any technical limitations for such auditing. In this regard, would you permit independent third-party audits of the AI system? If yes, explain the conditions or limitations you would impose on independent auditors. If you would not permit third-party audits, then explain why.
6. Describe whether, and to what extent, the AI system outputs will be explainable and interpretable, and by what means, to (i) third-party computer engineers; (ii) agency personnel trained to use the system; and (iii) other stakeholders, including laypersons, judges, and policymakers.
7. Describe any known or foreseeable (i) biases in the AI system; (ii) performance weaknesses; and (iii) vulnerabilities. Further, describe the known and foreseeable sources or causes of those biases, performance weakness, and vulnerabilities (e.g., in the data, algorithm, design process, human-computer interactions, interoperability with other hardware and software, or otherwise).
8. Explain how you will ensure or test that the AI system does not drift from its intended purposes or outcomes.
9. Explain how you will ensure or facilitate usability for government personnel (e.g., training programs, written materials, access to source code, and otherwise).

---

[100] *See id.* § 15.203(a) ("[RFPs] are used in negotiated acquisitions to communicate Government requirements to prospective contractors and to solicit proposals.").

[101] *See id.* § 8.402 (RFQs are used when agencies order goods and services from federal supply schedules).

Incorporating AI risk-related questions in market solicitations, along the lines above, can yield several direct and indirect benefits. Most directly, the answers by market participants will enable the agency to compare the types and degrees of risk associated with a *particular vendor* relative to the field. Anticipating this, strategic and innovative vendors will compete for an ethical edge. In some instances, the agency might even find opportunities for collaboration—for example, between two or more start-up enterprises—to mitigate the overall risk based on their respective strengths and weaknesses. Further, vendor responses to the questions above may shed light on additional risks not previously identified. In such cases, the AI risk assessment can and should be modified accordingly.

c. Source Selection and Contractual Award

Whether independently or in conjunction with the foregoing recommendations, federal procurement law should require that agencies pay due regard for AI principles in source selection and contractual award. For purposes of this discussion, I will mostly bracket the issue of cybersecurity, which is already receiving sustained attention by policymakers. But it is worth pausing to note some recent cybersecurity regulatory interventions because they point to the need, and feasibility, of acquiring ethical AI tools through the procurement process.

Federal law, for example, now prohibits executive agencies from contracting with entities that use any equipment, system, or service that utilizes covered telecommunications equipment from blacklisted companies (such as Huawei and ZTE Corporation).[102] Moreover, the Department of Defense (DoD) will be phasing in a new Cybersecurity Maturity Model Certification (CMMC) program that will require DoD vendors to have adequate controls in place to protect sensitive information and data.[103] Under this program, vendors will be required to obtain CMMC certification from third-party auditors; vendors that fall short of prescribed security levels cannot even compete for those contracts.[104]

---

[102] John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, § 889(a)(1), 132 Stat. 1636, 1917 (2018); *see also* Federal Acquisition Regulation: Prohibition on Contracting with Entities Using Certain Telecommunications and Video Surveillance Services or Equipment, 85 Fed. Reg. 42,665 (July 14, 2020).

[103] Off. of the Under Sec'y of Def. for Acquisition & Sustainment Cybersecurity Maturity Model Certification, *CMMC FAQ's*, https://www.acq.osd.mil/cmmc/faq.html.

[104] *See* Peter Baldwin & Jason G. Weiss, *DoD's Cybersecurity Maturity Model Certification Is Here: What Your Business Needs to Do to Prepare*, NAT'L L. REV. (Apr. 15, 2020), https://www.natlawreview.com/article/dod-s-cybersecurity-maturity-model-certification-here-what-your-business-needs-to-do.

A certification system for ethical AI is within the realm of future possibility but is premature at this time.[105] Standardizing transparency and fairness metrics, for example, could have the unintended consequence of entrenching market incumbents and stifling innovation. Instead, the recommendation here is to require agencies to incorporate ethical AI considerations into the source selection process, and/or to fold such considerations into the prerequisites for contractual award. The discussion below elaborates on these alternatives.

1. Evaluative Criteria

By way of background, the express and overarching purpose of the Federal Acquisition Regulation (FAR) is to "deliver . . . the best value product or service to the customer, while maintaining the public's trust and fulfilling public policy objectives."[106] To those ends, FAR instructs contracting officials to examine the strengths and weaknesses of all relevant factors (such as cost, performance, quality, and schedule) and to make tradeoffs between cost and non-cost factors.[107] FAR instructs that "cost or price may play a dominant role in source selection" in acquisitions where the requirements for performance are "clearly definable" and the "risk of unsuccessful contract performance is minimal."[108] The same provision, however, also explains that non-price considerations—such as "technical or past performance"—may "play a dominant role in source selection" when the performance requirements are less specified and the risk of unsuccessful performance is greater.[109]

Contracting officials have wide discretion in how to achieve "best value" under these standards. Yet FAR does erect some important guardrails. Pertinent here, contracting officials must pre-disclose how they intend to evaluate cost and non-cost criteria, and then must adhere to those criteria when awarding contracts.[110] Currently, nothing prevents contracting officials from incorporating ethical AI considerations into the source selection

---

[105] *Cf.* Institute of Electrical and Electronics Engineers (IEEE) P7000 (proposing a set of standards "for addressing ethical concerns during [AI] system design"), https://standards.ieee.org/project/7000.html.

[106] 48 C.F.R. § 1.102(a).

[107] *Id.* § 15.101-1.

[108] *Id.* § 15.101.

[109] *Id.*

[110] *See id.* § 15.305(a) ("An agency shall evaluate competitive proposals and then assess their relative qualities solely on the factors and subfactors specified in the solicitation."); *see also* Antarctic Support Assocs. v. United States, 46 Fed. Cl. 145, 155 (2000) (noting that reviewing courts must afford contracting officials great deference; contractual awards must only be reasonable and consistent with stated evaluation criteria).

process.[111] But doing so can and should be required as a matter of law. Issues relating to transparency, accountability, privacy, security, and fairness will *always* be relevant in the procurement of AI systems. How relevant will be context-specific. Thus, considerations of ethical AI principles can be weighed as appropriate depending on the circumstances.

Nothing prevents contracting officials from incorporating ethical AI considerations into the source selection process. But doing so can and should be required as a matter of law.

One virtue of this approach is its flexibility. Little or nothing would be sacrificed by requiring contracting officials to consider ethical AI principles in conjunction with price, technical feasibility, and other criteria. Meanwhile, much could be gained from explicitly making ethical AI part of the calculus. Most importantly, doing so can mitigate the special risks that inhere in AI acquisitions. But it will also signal to agency officials, industry players, and stakeholders that AI procurement is not business as usual: The social and governance implications extend beyond technical criteria and model performance. Challenges around transparency, accountability, privacy, security, and fairness may (or may not) be addressed with technical patches.[112] Still, it is a categorical error to conceive of the challenges of algorithmic governance as mere technical problems that can be fixed with better data, mathematical proofs, or more software to "explain" the computations and configurations of black-box AI systems.[113]

AI procurement is not business as usual: The social and governance implications extend beyond technical criteria and model performance.

---

[111] Nonprofit organizations have made similar recommendations as a matter of best practice, but not as a matter of law. *See, e.g.*, Am. Council for Tech.-Indus. Advisory Council, AI Playbook for the U.S. Federal Government, at 15, 22, 29, 35 (2020); World Econ. Forum, AI Procurement in a Box: AI Government Procurement Guidelines (June 11, 2020), https://www.weforum.org/reports/ai-procurement-in-a-box/ai-government-procurement-guidelines#report-nav.

[112] *See, e.g.*, Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J. L. & TECH. 1 (2017) (providing a computer scientist's perspective on algorithmic accountability and calling for specific, tailored solutions).

[113] *See* Coal. for Critical Tech., *supra* note 74, at 2 ("To date, many efforts to deal with the ethical stakes of algorithmic systems have centered mathematical definitions of fairness that are grounded in narrow notions of bias and accuracy. These efforts give the appearance of rigor, while distracting from more fundamental epistemic problems."); Brent Mittelstadt, Chris Russell, Sandra Wachter, *Explaining Explanations in AI*, in *Proceedings of Fairness, Accountability, and Transparency* (2019) ("[T]he bulk of methods currently occupying AI researchers lies in the . . . building of approximate models that are not intended to capture the full behaviour of physical systems but rather to provide coarse approximations of how the systems behave."), https://arxiv.org/pdf/1811.01439.pdf.

2. Vendor Responsibility

For much the same reasons, an additional (or alternative) requirement should be incorporated into a contracting officer's "responsibility" determination. Currently, FAR directs contracting officials to determine whether a vendor satisfies a set of responsibility standards prior to awarding any procurement contract.[114] To meet the standards, a potential awardee must demonstrate, among other things, that it has "adequate financial resources;" a "satisfactory past performance record;" "the necessary organization, experience, facilities," and "technical skills" to perform the contract; and a satisfactory "record of integrity" and "business ethics."[115] These performance standards are designed, in part, to mitigate the risk of vendor misfeasance and contractual nonperformance.[116]

Beyond these performance standards, a responsibility determination also requires that a potential awardee be "otherwise qualified and eligible."[117] These criteria, in turn, encompass a set of collateral requirements. For example, potential awardees must be disqualified if they do not comply with federal equal employment opportunity requirements or if they fail to agree to an acceptable plan for subcontracting with small businesses. Unlike performance standards, which assess whether prospective vendors can be expected to fulfill the contract in a timely and satisfactory manner, collateral requirements ensure that the government's dealings with contractors promote federal socioeconomic goals.

Ethical AI principles could fold neatly into FAR's responsibility framework, either as a performance standard or collateral requirement. As a performance standard, ethical AI principles can be factored into the criteria for "necessary organization, experience, facilities" and "technical skills" to perform the contract, as well as a prospective vendor's "record of integrity and business ethics."[118] As a collateral requirement, and perhaps to greater effect, the imperative of ethical AI could be leveraged to gain important concessions from vendors.

---

[114] 48 C.F.R. § 9.103(a) ("Purchases shall be made from, and contracts shall be awarded to, responsible prospective contractors only"); *id.* § 9.103(b) ("No purchase or award shall be made unless the contracting officer makes an affirmative determination of responsibility.").

[115] 48 C.F.R. § 9.104–1; *see also* Kate M. Manuel, *Responsibility Determinations Under the Federal Acquisition Regulation: Legal Standards and Procedures*, Congressional Research Service, No. R40633 5–13 (Jan. 4, 2013) (providing explanations of FAR's responsibility standards and processes); Orca Northwest Real Est. Servs. v. U. S., 65 Fed. Cl. 1, 6 (2005), on reconsideration, 65 Fed. Cl. 419 (2005) (discussing criteria for responsibility).

[116] 48 C.F.R. § 9.103; Ryan Co. v. U.S., 43 Fed. Cl. 646, 651 (1999).

[117] 48 C.F.R. § 9.104-1(g).

[118] *Id.* § 9.104–1; *see also* Manuel, *supra* note 115, at 5–13; Orca Northwest Real Est. Servs., 65 Fed. Cl. at 6.

For example:

1. Prime contractors could be required to subcontract with "nontraditional" government vendors (to create opportunities for innovative small businesses and start-ups that might otherwise be reluctant to enter the government market).[119] Along similar lines, prime contractors could be required to subcontract with women-owned and "socially and economically disadvantaged" small-business enterprises.[120] Doing so could promote diversity in AI system design, development, and testing.[121]

2. Vendors could be required to share trade secrets with the government for specified purposes and under certain constraints. Intellectual property (IP) rights can be a major sticking point in government contracting, especially in high-tech fields like AI.[122] Indeed, trade secrets are often the most valuable assets for many small and start-up enterprises.[123] Thus, a responsibility requirement tethered to IP rights should be narrowly tailored. It should call for no more than is *foreseeably necessary* and should not substitute for other contractually negotiated IP terms. To be sure, there is always a risk that requiring IP concessions will scare off market participants.[124] But just as surely, the government should not be permitted to spend tax dollars on AI technologies that the government cannot fully use, audit, or explain. The minimally necessary criteria for purposes of a responsibility determination will be context-specific. For example, potential awardees should not qualify as responsible if they are not willing to waive trade secrets in adjudicatory contexts, where the invocation of trade secrecy would effectively preclude the government from utilizing an AI tool, and thus negate the purpose of acquiring the tool in the first place. If a vendor will not agree, then the government should be required to select a competitor that will. And, if none will agree, then the government should rethink whether a privately developed AI solution is appropriate for the government task.

---

[119] *Cf.* Steve Kelman, *Non-Traditional IT Contractors Unite to Watch Each Other's Backs*, FCW, Lectern Blog (Feb. 12, 2018) (discussing how nontraditional IT vendors in the federal marketplace deliver innovative services), https://fcw.com/blogs/lectern/2018/02/kelman-digital-services-coalition.aspx.

[120] Federal law has an established program to promote contracting with any "small business which is unconditionally owned and controlled by one or more socially and economically disadvantaged individuals who are of good character and citizens of and residing in the United States, and which demonstrates potential for success." 13 C.F.R. § 124.101; 15 U.S.C § 637 (defining "socially and economically disadvantaged" individuals under the Small Business Act as "[s]ocially disadvantaged individuals are those who have been subjected to racial or ethnic prejudice or cultural bias because of their identity as a member of a group without regard to their individual qualities"); *see also* 48 C.F.R. § 19.15 (Woman-Owned Small Business Program).

[121] *Cf.* Bi-Partisan Center, AI and the Workforce, at 2 (Jul. 2020) (reporting on the nation's AI talent gap and stating that "AI talent should ideally have a multi-disciplinary" skill set that includes ethics).

[122] *See* Nancy O. Dix et al., *Fear and Loathing of Federal Contracting: Are Commercial Companies Really Afraid to Do Business with the Federal Government? Should They Be?* 33 PUB. CONT. L.J. 5, 8–9 (2003) (providing a review of the relevant contracting requirements and industry concerns around IP provisions in government contracts that depart from general commercial terms).

[123] *See generally* Rob Kitchin, *Thinking Critically About and Researching Algorithms*, 20 INFO. COMM. & SOC'Y 14, 20 (2016) ("[I]t is often a company's algorithms that provide it with a competitive advantage and they are reluctant to expose their intellectual property even with non-disclosure agreements in place.").

[124] *See* Dix et al, *supra* note 122, at 9.

3. Vendors could be required to allow independent third-party auditing of an AI system—if for no other reason, because the government may not have the expertise or resources to do so itself. Given the many ways that AI systems can cause harm, a vendor that will not permit third-party auditing of its AI system should be disqualified from being awarded a government procurement contract. Of course, vendors will need adequate assurances that their trade secrets will not be compromised by third-party auditing. But there are well-established industry practices (for example, nondisclosure agreements with liability provisions) and federal laws[125] that can be used to safeguard vendors against trade secrecy misappropriation.

[125] *See, e.g.*, 18 U.S.C. § 1836 (allowing the owner of a trade secret to sue in federal court when its trade secrets have been misappropriated); *see also* 18 U.S.C. § 1905 (imposing criminal penalties for any government employee who discloses information that "concerns or relates to trade secrets").

V. Objections

Anticipated objections will come from those who think the forgoing recommendations go too far, and from those who think the recommendations fall short.

For those concerned about overreach, it is likely for one of two reasons. First, it may be objected that federal acquisition procedures are already too cumbersome, especially for technology that can become outdated before they are put to use.[126] The recommendations here are sensitive to this concern. Requiring the government to create an AI inventory report is retrospective-facing and thus should create little or no friction in prospective government acquisitions. Moreover, risk assessments are already undertaken for IT systems in the normal course of procurement planning. The recommendation here would simply formalize and standardize that practice to address the *special risks* associated with AI systems. For example, risks relating to bias, model explainability, and data integrity are not the types of risks normally accounted for in technology acquisitions but can make or break the success of AI projects.

A second objection—and a major one—is that raising the bar on vendor responsibility and contractual award will discourage or disqualify innovative vendors from working with the government.[127] Although the industry is generally wary of more procurement regulations,the prescriptions advanced here seize upon areas of shared interest. For the government and industry alike, AI innovation is a complex ambition that scopes well beyond technological capability. Innovation also entails the responsible development and deployment of AI tools. Currently, every major technology company has teams of high-skilled workers and mounds of investment capital dedicated to ethical AI. And the government, for its part, is pouring huge amounts of tax dollars into related research and development.

Despite motivational differences, public and private interests around trustworthy AI merge in the acquisition gateway. That shared reality is a foundation for principled and pragmatic regulatory compromise, which this report aims to advance. Indeed, the

---

[126] *See, e.g.*, Katherine M. John, *Information Technology Procurement in the United States and Canada: Reflecting on the Past With an Eye Toward the Future*, 48 PROCUREMENT LAWYER 4, 5 (2013) ("If procurement regimes overemphasize transparency and competition—or otherwise take too long—then end users might end up saddled with technology that is outdated by the time it reaches them.").

[127] *Cf.* L. Elaine Halchin, Other Transaction Authority, Cong. Res. Serv., RL34760 (Jul. 15, 2011) ("The Government is finding that not only can it not acquire many of the technologies it needs, but also many corporations will not even accept government dollars to help develop new technologies.").

government's demand for ethically designed AI systems may *attract* innovative talent to the government market, given that many technology companies—big and small—have expressed serious concerns about the government's ethical use of this technology (especially as relates to law enforcement and warfighting).[128]

Public and private interests around trustworthy AI merge in the acquisition gateway.

It remains to be seen whether the government can swing these market dynamics more in its favor. But it is easy to see how these dynamics can become measurably worse. Without intervention, the government will only grow more dependent on private industry to provide the tools of algorithmic government. In a world where industry is more responsible and ethical than the government, perhaps that is a good thing. In the meantime, however, the widely shared hope is that the government and industry will be partners in the journey ahead. Suffusing the procurement process with ethical AI is not the only path. But it's the one America should insist upon.

Without intervention, the government will only grow more dependent on private industry to provide the tools of algorithmic government.

Coming from the opposite direction, policymakers and stakeholders may think the recommendations in this report do not go far enough. Emphatically, I agree. In due course, more can and should be done to meet the significant challenges of algorithmic governance. Crafting concrete regulatory mechanisms that meaningfully fulfill the government's legal and moral obligations is an all-hands-on-deck effort. The

---

[128] *See, e.g.*, Arvind Krishna, *IBM CEO's Letter to Congress on Racial Justice Reform*, IBM (June 8, 2020) (announcing that IBM "will not condone uses of any technology . . . for mass surveillance, racial profiling, violations of basic human rights and freedoms, or any purpose which is not consistent with our values and Principles of Trust and Transparency."), https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/; Brad Smith, *Facial Recognition: It's Time for Action*, Microsoft on the Issues: The Official Microsoft Blog (Dec. 6, 2018) ("[W]e don't believe that the world will be best served by a commercial race to the bottom, with tech companies forced to choose between social responsibility and market success."), https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/; *We Are Implementing a One-Year Moratorium on Police Use of Recognition*, Day One: The Amazon Blog (June 10, 2020), https://blog.aboutamazon.com/policy/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition (taking their product off the market for a year to give lawmakers time to "put in place stronger regulations to govern the ethical use of facial recognition technology"); Rosalie Chan, *Google Drops Out of Contention for a $10 Billion Defense Contract Because It Could Conflict with Its Corporate Values*, BUSINESS INSIDER (Oct. 8, 2018), https://www.businessinsider.com/google-drops-out-of-10-billion-jedi-contract-bid-2018-10.

recommendations offered here are part of that larger project. And, constructively, nothing suggested here would preclude additional regulatory interventions—whether in procurement law, administrative law, privacy law, or in domain-specific contexts.

Regarding procurement law, in particular, federal lawmakers should consider whether special acquisition pathways for AI are necessary and appropriate. Given the complexity of the AI development process, it may be impossible or ill-advised for the agency to specify AI performance metrics at the time of contracting. Moreover, many if not most of the important design choices may happen after an award is made. These characteristics of AI development make traditional "waterfall" acquisition pathways quite unsuitable, and risky, because agencies may get locked into contracts for dead-end solutions with no easy exit ramp.[129]

To meet these challenges, the Government Accountability Office (GAO) recently issued a best-practices guide touting the use of "agile" development pathways for federal software acquisition.[130] The Office of Management and Budget and General Services Administration have likewise championed the use of agile acquisition frameworks.[131] Unlike the front-loaded waterfall approach, agile methods are characterized by incremental and iterative processes in which software is produced in close collaboration with the acquiring agency. When properly planned and managed, agile boasts of improved investment manageability, lowers the risk of project failure, shortens the time to realize value, and allows agencies to better adapt to changing needs. Information obtained during these frequent iterations allow developers to respond quickly to feedback from agency customers, thus potentially reducing sociotechnical, legal, and programmatic risk.

Arguably, FAR contains enough flexibility to accommodate agile acquisition pathways. For example, pursuant to FAR, agencies may use modular contracting vehicles for "delivery, implementation, and testing of workable systems or solutions in discrete increments, each of which comprises a system or solution that is not dependent on any subsequent increment in order to perform its principal functions."[132] Moreover,

---

[129] In a typical waterfall acquisition, the technical and design requirements are fixed at the time of contracting.

[130] GAO-20-590G, Agile Assessment Guide: Best Practices for Agile Adoption and Implementation (Sept. 2020) [hereafter, GAO-20-590G, Agile Best Practices], https://www.gao.gov/products/GAO-20-590G.

[131] *See, e.g.*, Office of Management and Budget, Office of Federal Procurement Policy (OFPP), Contracting Guidance to Support Modular Development (Jun. 2012), https://obamawhitehouse.archives.gov/sites/default/files/omb/procurement/guidance/modular-approaches-for-information-technology.pdf; General Services Administration, De-Risking Government Technology, Federal Agency Field Guide (Sept. 2020), https://derisking-guide.18f.gov/assets/federal-field-guide-a245c3a7dcd0a24f619b458fd51e1e490f2299023fd1bd13fddc87318e67cf03.pdf.

[132] 48 CFR § 39.103.

modular contracts "provide an opportunity for subsequent increments to take advantage of any evolution in technology or needs that occur during implementation and use of the earlier increments," and can "reduce risk of potential adverse consequences on the overall project by isolating and avoiding custom-designed components of the system."[133]

Agile acquisition pathways, however, are no panacea for AI systems. These approaches require skill sets, resources, and institutional buy-in that many agencies are currently without. Indeed, GAO's best-practices guide chronicles an array of challenges that agencies have experienced using agile processes. For example, "teams reported difficulty collaborating closely or transitioning to self-directed work due to constraints in organization commitment and collaboration." Moreover, "some organizations reported that they did not have trust in iterative solutions and that teams had difficulty managing iterative requirements."[134]

AI acquisitions would give rise to all the same challenges—and more—given the complexities, decisions, and ongoing monitoring that these systems require. Still, there is reason to hope the federal government will gain the experience and resources to responsibly manage the acquisition of ethically designed AI systems, through modular contracting or otherwise.

---

[133] *Id.*

[134] *See* GAO-20-590G, Agile Best Practices, *supra* note 130, at 14-16; *see also* GAO-16-467, Immigration Benefits System: US Citizenship and Immigration Services Can Improve Program Management (Jul. 2016) (reporting that the United States Citizenship and Immigration Service Transformation program was not setting outcomes for agile software development), https://www.gao.gov/products/GAO-16-467; GAO-18-46, TSA Modernization: Use of Sound Program Management and Oversight Practices is Needed to Avoid Repeating Past Problems (Oct. 2017) (reporting that the Transportation Security Administration's Technology Infrastructure Modernization (TIM) program did not define key agile roles, prioritize system requirements, or implement automated capabilities), https://www.gao.gov/products/GAO-18-46.

VI. Conclusion

It is encouraging that the United States has committed to "AI principles" and the protection of "civil liberties, privacy, and American values . . . in order to fully realize the potential of AI technologies for the American people."[135] But proselytizing is not actualizing.[136] If federal policymakers are truly committed to ethical algorithmic governance, then the acquisition gateway is a prime place to start. This report has shined critical light on the need for procurement reform and has offered a set of recommendations that future work can build upon.

[135] *See* Office of Sci. and Tech. Policy, Exec. Office of the President, American Artificial Intelligence Initiative: Year One Annual Report (2020).

[136] *Cf.* NIST, Second Quarter Report, supra note 12, at 123 (acknowledging that "[t]here is often a gap between articulating high-level goals around responsible AI and operationalizing them").

THE
GREAT
DEMOCRACY
INITIATIVE

GREATDEMOCRACYINITIATIVE.ORG