ACQUIRING ETHICAL AI

David S. Rubenstein[*]

Abstract

Artificial intelligence (AI) is transforming how the federal government operates. Under the right conditions, AI systems can solve complex problems, reduce administrative burdens, improve human decisions, and optimize resources. Under the wrong conditions, AI systems can lead to widespread discrimination, invasions of privacy, and the erosion of democratic norms. A burgeoning literature has emerged to square algorithmic governance with the precepts of constitutional and administrative law. Federal procurement law, however, remains a dangerous blind spot in the reformist agenda. This Article pivots into that neglected space and emerges with comprehensive framework for acquiring ethical AI. Toward that end, the Article makes three main contributions. First, it provides an original account that yokes the ambitions of algorithmic governance, the imperative of ethical AI, and the levers of procurement law. Second, this Article argues that the procurement system is uniquely situated to check and enable algorithmic governance in ways that other legal frameworks miss. Third, the Article prescribes a set of concrete regulatory reforms to instantiate ethical AI throughout the procurement process: from acquisition planning to market solicitation, bid evaluation, source selection, and contract performance. Procurement law will not solve all the challenges of algorithmic governance. Just as surely, those challenges cannot be solved without procurement law.

[*] James R. Ahrens Chair in Constitutional Law and Director, Robert Dole Center for Law and Government, Washburn University School of Law. The author thanks Daniel Ho, Aziz Huq, Martin Murillo, and Nicole Petroff for their very helpful comments and suggestions on earlier drafts. The author also thanks Kaitlyn Bull, Ande Davis, Penny Fell, Barbara Ginsberg, Creighton Miller, Chris Smith, and Zach Smith for excellent research assistance, as well as the *Florida Law Review* for careful and constructive editing.

INTRODUCTION

Artificial intelligence (AI) is transforming how the federal government operates.[1] For example, the Department of Justice uses AI in law enforcement;[2] the Social Security Administration uses AI for adjudicatory functions;[3] the Department of Homeland Security uses AI to regulate immigration;[4] the Internal Revenue Service uses AI to detect tax fraud;[5] the Department of Veterans Affairs uses AI to deliver health services;[6] the Pentagon uses AI to augment its military and intelligence capabilities;[7] the General Services Administration uses AI to streamline

1. *See generally* DAVID FREEMAN ENGSTROM, DANIEL E. HO, CATHERINE M. SHARKEY, MARIONO-FLORENTINO CUÉLLAR, ADMIN. CONF. OF THE U.S., GOV'T BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES (2020) [hereinafter ACUS REPORT], https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf [https://perma.cc/AU59-KGAG] (providing a rich account of this transformation along with its implications for regulatory practice and administrative law).

2. *See Letter to Attorney General Barr RE: The Use of the PATTERN Risk Assessment in Prioritizing Release in Response to the COVID-19 Pandemic*, LEADERSHIP CONF. ON CIVIL & HUM. RTS. (Apr. 3, 2020), https://civilrights.org/resource/letter-to-attorney-general-barr-re-the-use-of-the-pattern-risk-assessment-in-prioritizing-release-in-response-to-the-covid-19-pandemic/ [https:// perma.cc/M349-RZYM] (discussing and critiquing the Department's use of an AI criminal risk-assessment tool); James Vincent, *FBI Used Facial Recognition to Identify a Capitol Rioter from His Girlfriend's Instagram Posts*, VERGE (Apr. 21, 2021), https://www.theverge.com/2021/4/21/22395323/fbi-facial-recognition-us-capital-riots-tracked-down-suspect [https://perma.cc/RZW9-MJ86].

3. *See* ACUS REPORT, *supra* note 1, at 38–40.

4. *See* Aaron Boyd, *CBP Is Upgrading to a New Facial Recognition Algorithm in March*, NEXTGOV (Feb. 7, 2020), https://www.nextgov.com/emerging-tech/2020/02/cbp-upgrading-new-facial-recognition-algorithm-march/162959/ [https://perma.cc/RH9L-6MUT]; Kate Evans & Robert Koulish, *Manipulating Risk: Immigration Detention Through Automation*, 24 LEWIS & CLARK L. REV. 789, 793 (2020).

5. TREASURY INSPECTOR GEN. FOR TAX ADMIN., U.S. DEP'T OF TREASURY, REFERENCE NO. 2017-20-080, THE RETURN REVIEW PROGRAM INCREASES FRAUD DETECTION; HOWEVER, FULL RETIREMENT OF THE ELECTRONIC FRAUD DETECTION SYSTEM WILL BE DELAYED 4 (2017), https://www.treasury.gov/tigta/auditreports/2017reports/201720080fr.pdf [https://perma.cc/LXW4-PP78] (noting "machine learning algorithms" for generating fraud risk scores).

6. *See* Anagha Srikanth, *How the VA Is Using Artificial Intelligence to Improve Veterans' Mental Health*, THE HILL (Sept. 8, 2020), https://thehill.com/changing-america/well-being/mental-health/515536-how-the-va-is-using-artificial-intelligence-to [https://perma.cc/DH6U-WGUH]; Am. Homefront Project, *VA Embraces Artificial Intelligence To Improve Veterans' Health Care*, CPR NEWS (Feb. 19, 2020), https://www.cpr.org/2020/02/19/va-embraces-artificial-intelligence-to-improve-veterans-health-care/ [https://perma.cc/LYT2-8BW8].

7. *See* DEP'T OF DEF., SUMMARY OF THE 2018 DEPARTMENT OF DEFENSE ARTIFICIAL INTELLIGENCE STRATEGY 15 (2018), https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF [https://perma.cc/WK76-SLZD]; Memorandum from Kathleen H. Hicks, Deputy Sec'y of Def., to Senior Pentagon Leadership et al. 1 (May 26, 2021) (discussing the Department of Defense's "embrace[]" of AI and the need to "adopt responsible behavior, processes, and outcomes in a manner that reflects the Department's commitment to its ethical principles").

business operations;[8] the Department of Health and Human Services uses AI for regulatory analysis and reform.[9] The list goes on[10] and is projected to grow exponentially with the government's digital transformation.[11]

This presages a new era of "algorithmic governance," in which federal responsibilities and functions will increasingly migrate from humans to machines.[12] As emergent technology, AI bears the burden of proof—and it's far from an easy case. Under the right conditions, AI systems can solve complex problems, reduce administrative burdens, improve human decisions, optimize government resources, and drive agency missions.[13] Under the wrong conditions, however, AI systems pose serious threats to civil rights and democratic norms.[14] Already, AI systems have wrongly deprived individuals of unemployment and medical benefits,[15] wrongly identified individuals for criminal arrest,[16]

---

8. *See, e.g.*, Dave Nyczepir, *GSA Leads Rise in Automation Projects Governmentwide*, FEDSCOOP (May 11, 2021), https://www.fedscoop.com/automation-projects-rise-gsa/ [https://perma.cc/PVJ5-UQXG] (reporting that The General Services Administration has four fully operational AI projects, with many more in the works).

9. *See* Press Release, U.S. Dep't of Health & Hum. Servs., HHS Launches First-of-Its-Kind Regulatory Clean-Up Initiative Utilizing AI (Nov. 17, 2020), https://www.pressreleasepoint.com/hhs-launches-first-its-kind-regulatory-clean-initiative-utilizing-ai [https://perma.cc/5EEN-LZGW].

10. *See, e.g.*, ACUS REPORT, *supra* note 1, at 25–29 (discussing a "suite of algorithmic tools" used by the SEC "to identify violators of federal security laws"); *id.* at 46–52 (discussing AI use cases by the U.S. Patent and Trademark Office); Jori Heckman, *USPS Gets Ahead of Missing Packages with AI Edge*, FED. NEWS NETWORK (May 6, 2021), https://federalnewsnetwork.com/artificial-intelligence/2021/05/usps-rolls-out-edge-ai-tools-at-195-sites-to-track-down-missing-packages-faster/ [https:// perma.cc/DQS8-RPZ4] (reporting that the U.S. Postal Service is using AI to examine and categorize packages it receives).

11. *See* KEVIN DROEGEMEIER ET AL., OFF. OF MGMT. & BUDGET, FEDERAL DATA STRATEGY 2020 ACTION PLAN 11 (2020), https://strategy.data.gov/assets/docs/2020-federal-data-strategy-action-plan.pdf [https://perma.cc/8KCC-VXV6] (discussing the government's information technology modernization efforts and data initiatives to support the adoption of AI technologies); NAT'L SEC. COMM'N ON A.I., FINAL REPORT 4 (2021) [hereinafter NSCAI FINAL REPORT] ("We envision hundreds of billions in federal spending [for AI technologies] in the coming years.").

12. *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 636 (2017) ("[I]mportant decisions that were historically made by people are now made by computer systems."); WILL HURD & ROBIN KELLY, RISE OF THE MACHINES: ARTIFICIAL INTELLIGENCE AND ITS GROWING IMPACT ON U.S. POLICY 7–8 (2018), https://www.hsdl.org/?view&did=816362 [https://perma.cc/9KYV-59BX] (reporting concerns about job loss due to AI-driven automation).

13. *See infra* Section II.A (discussing the putative benefits of algorithmic governance).

14. *See infra* Section II.B (discussing the causes and manifestations of AI risk).

15. *See, e.g.*, 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*, ATLANTIC (June 7, 2020), https://www.theatlantic.com/technology/archive/2020/06/michigan-unemployment-fraud-automation/ 612721/ [https://perma.cc/8KT7-DWLS].

16. *See* Kashmir Hill, *Wrongfully Accused by an Algorithm*, N.Y. TIMES (Aug. 3, 2020), https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html [https://perma.

and wrongly denied access to government food programs.[17] More generally, the government's adoption of AI can amplify racial and gender biases, encroach on civil liberties, and create barriers to government transparency and accountability.[18]

From this starting position, the question is not whether algorithmic governance will be for better or worse, but rather *whose lives* will be benefitted and burdened, in *which ways*, under *what rules* of engagement, and *who should decide*.[19] Amidst the swirling uncertainty one thing is clear: the outcomes will by heavily influenced by the technology industry. Despite pockets of excellence, the government's demand for AI systems far exceeds its in-house capacity to design, develop, field, and monitor this technology at scale.[20] Accordingly, many if not most of the tools and operational support for algorithmic governance will be procured by contract from technology firms.[21]

The "outsourcing of algorithmic governance" brings many affordances; chief among them is the government's ability to capitalize

---

cc/PP5D-7E9X]; *see also* Complaint ¶¶ 39–49, Williams v. City of Detroit, No. 2:21-cv-10827 (E.D. Mich. filed Apr. 13, 2021), https://www.aclumich.org/sites/default/files/field_documents/001_complaint_1.pdf [https://perma.cc/7D4L-F77Z] (collecting data regarding racial bias and misidentification in facial-recognition systems used by police).

17. *See, e.g.*, Florangela Davila, *USDA Disqualifies Three Somalian Markets from Accepting Federal Food Stamps*, SEATTLE TIMES (Apr. 10, 2002), https://archive.seattletimes.com/archive/?date=20020410&slug=somalis10m [https://perma.cc/MR7K-9JZX] (describing how the U.S. Department of Agriculture's monitoring system for suspicious transactions denied three Somalian-owned markets from accepting food stamps based on "unusual, irregular, and/or inexplicable" activity at each store).

18. *See infra* Section II.B; *see also* Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE. J.L. & TECH. 103, 129 (2018) (explaining that AI's ability to scale government processes and decision-making magnifies any error or bias); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 29–31 (2016) (providing a trenchant account of how AI systems disproportionately harm marginalized and vulnerable populations); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR 180–88 (2017) (arguing that government decision-making systems create a "digital poorhouse"); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677 (2016) ("[D]ata mining holds the potential to unduly discount members of legally protected classes and to place them at systematic relative disadvantage.").

19. *See* KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE 8 (2021) (providing a similar framing of these sociopolitical issues).

20. NAT'L SEC. COMM'N ON A.I., SECOND QUARTER RECOMMENDATIONS 34 (2020) ("'[T]here is a severe shortage of AI knowledge in [the 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." (quoting NAT'L SEC. COMM'N ON A.I., INTERIM REPORT 35 (2019))); c*f.* ACUS REPORT, *supra* note 1, at 18, 89 (finding that approximately half of AI applications covered in the study were developed in-house by federal agency personnel).

21. *See infra* notes 256–61 and accompanying text (discussing the asymmetry between the government's demand for AI tools and its capacity to develop, implement, and monitor, these tools in-house).

on the industry's innovation, institutional know-how, and high-skilled workforce.[22] But these alliances subsume worrisome reliances. Currently, AI technologies are virtually unregulated in the private market.[23] Unless and until that changes, federal agencies will be acquiring unregulated technology for use in high-stakes government contexts.[24] Moreover, AI systems are embedded with value-laden tradeoffs between what is technically feasible, socially acceptable, economically viable, and legally permissible.[25] Without proper planning and precaution, the government may acquire AI with embedded policies from private actors whose financial motivations and legal sensitivities may not align with the government or the people it serves.[26]

Simply put, acquiring AI is not business as usual. The technology is inherently risky, regardless of who develops and deploys it.[27] But the government's risk profile requires special attention.[28] Most notably,

22. *See* David S. Rubenstein, *The Outsourcing of Algorithmic Governance*, YALE J. ON REGUL.: NOTICE & COMMENT (Jan. 19, 2021), https://www.yalejreg.com/nc/the-outsourcing-of-algorithmic-governance-by-david-s-rubenstein/ [https://perma.cc/9G VS-7PXT]; NSCAI FINAL REPORT, *supra* note 11, at 24 ("The government lags behind the commercial state of the art in most AI categories, including basic business automation. It suffers from technical deficits that range from digital workforce shortages to inadequate acquisition policies, insufficient network architecture, and weak data practices.").

23. To some extent, the regulation of AI development and deployment may be regulated under existing federal law. *See* Memorandum from Russell T. Vought, Dir., Off. of Mgmt. & Budget, to the Heads of Exec. Dep'ts & Agencies 2 (Nov. 17, 2020), https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf [https://perma.cc/HW3J-FHQ5]; Elisa Jillson, *Aiming for Truth, Fairness, and Equity in Your Company's Use of AI*, FED. TRADE COMM'N: BUS. BLOG (Apr. 19, 2021, 9:43 AM), https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai [https://perma.cc/2Q6Z-DQV5] (describing potential FTC enforcement in cases of AI-derived discrimination). Still, it is widely acknowledged that existing federal laws and regulations are inadequate to address the novel ways that AI systems are developed and operationalized. *See infra* notes 192–97 and accompanying text (discussing dramatic uptick in legislative proposals to regulate AI). This regulatory void is becoming increasingly difficult to justify, given the ubiquity and social significance of AI in the areas of finance, education, manufacturing, labor and employment, transportation, recreation, journalism, medicine, insurance, agriculture, energy, and countless more.

24. *See* DAVID S. RUBENSTEIN, GREAT DEMOCRACY INITIATIVE, FEDERAL PROCUREMENT OF ARTIFICIAL INTELLIGENCE: PERILS AND POSSIBILITIES 4 (2020).

25. *Id.*

26. *Id.*

27. Sean McGregor, *When AI Systems Fail: Introducing the AI Incident Database*, P'SHIP ON A.I.: BLOG (Nov. 18, 2020), https://www.partnershiponai.org/aiincident database [https://perma.cc/9ZMH-RH7S] ("Failures of [AI] systems pose serious risks to life and wellbeing, but even well-intentioned intelligent system developers fail to imagine what can go wrong when their systems are deployed in the real world.").

28. *Cf.* 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 (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 to federal contractors, on the other).

government action is subject to constitutional and administrative law requirements, whereas private action is not.[29] Further, the polity generally expects the government to serve the public interest in safe, fair, transparent, and accountable ways. But these norms of good governance pose major challenges for machine learning AI systems, which are agnostic to democratic values, and often opaque, fickle, and brittle.[30]

A rich literature has emerged to address the challenges of algorithmic governance. Generally speaking, the reformist agenda is keyed to how law, technology, or both, can be configured in ways that are normatively desirable and operationally feasible.[31] To date, most of this legal scholarship has concentrated on constitutional due process,[32] equal protection,[33] free speech and assembly,[34] and criminal rights.[35] Scholars

29. *See id.*; Lillian BeVier & John Harrison, *The State Action Principle and Its Critics*, 96 VA. L. REV. 1767, 1786 (2010) ("Constitutional rules are almost all addressed to the government."). For an incisive treatment of the constitutional state action doctrine as applied to government AI vendors, see Kate Crawford & Jason Schultz, *AI Systems as State Actors*, 119 COLUM. L. REV. 1941, 1971–72 (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).

30. *See infra* Section I.A (discussing machine learning AI systems), Section II.B (discussing a suite of sociotechnical challenges associated with machine learning AI systems).

31. *See* Ryan Calo & Danielle K. Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797, 835 (2021) (canvassing the "ongoing project that responds to automation's disruption of rights and values through a combination of legal and technical reforms").

32. *See, e.g.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1281–88 (2008) (discussing how government use of automated systems in governmental administrative proceedings raises due process concerns and prescribing reforms); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 121–28 (2014) (same).

33. *See, e.g.*, Aziz Z. Huq, *Constitutional Rights in the Machine Learning State*, 105 CORNELL L. REV. 1875, 1917–27 (2020); Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 193 (2017) ("[A] simple prohibition on the use of protected characteristics such as race and sex in an automated decision process is easy to implement, but would do little to prevent biased outcomes."); Barocas & Selbst, *supra* note 18, at 677 (discussing the specific technical issues that give rise to models whose use in decision-making may have a disproportionately adverse impact on protected classes).

34. *See, e.g.*, Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1273, 1295–306 (2020) (exploring the "procedural and substantive conflicts between proprietary [algorithmic] decision-making on the one hand and government transparency obligations under the First Amendment and [Freedom of Information Act] on the other"); *see also* Woodrow Hartzog & Evan Selinger, *Facial Recognition Is the Perfect Tool for Oppression*, MEDIUM (Aug. 2, 2018), https://medium.com/s/story/ facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66 [https://perma.cc/ 7FEU-HQJE]; Sigal Samuel, *Activists Want Congress to Ban Facial Recognition. So They Scanned Lawmakers' Faces*., VOX (Nov. 15, 2019, 10:10 AM), https://www.vox.com/future-perfect/2019/11/15/20965325/facial-recognition-ban-congress-activism [https://perma.cc/R8JP-BY38].

35. *See, e.g.*, Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 331–32 (2015); Michael L. Rich, *Machine Learning, Automated*

have also begun the necessary work of squaring algorithmic governance with separation of powers doctrine[36] and precepts of administrative law.[37]

Federal procurement law, however, remains a dangerous blind spot in the reformist agenda. It is no novelty to observe, as others have, that the government's market dependencies and information asymmetries exacerbate the challenges of algorithmic governance.[38] And a handful of scholars have urged contracting officials, at all levels of government, to protect and promote public interests when acquiring AI from private vendors.[39] Yet the regulatory hooks and incentive structures required to meet these challenges remain woefully undertheorized, unspecified, and unutilized. This Article pivots into that neglected space and emerges with a comprehensive framework for "acquiring ethical AI." No less than other areas of law, federal procurement law will need retrofitting to

---

Suspicion Algorithms, and the Fourth Amendment, 164 U. PA. L. REV. 871, 878–79 (2016); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1350–53, 1397 (2018).

36. *See, e.g.*, Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1154, 1176–84 (2017) (discussing how federal agency use of machine learning AI systems could potentially implicate the constitutional nondelegation doctrine); Mariano-Florentino Cuéllar, *Cyberdelegation and the Administrative State*, *in* ADMINISTRATIVE LAW FROM THE INSIDE OUT 134, 134, 156–57 (Nicholas R. Parrillo ed., 2017) (discussing how agency use of AI relates to delegation principles).

37. *See* ACUS REPORT, *supra* note 1; Coglianese & Lehr, *supra* note 36; Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 6 (2019); David Engstrom & Daniel Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. ON REG. 800 (2020); Deirdre K. Mulligan & Kenneth A. Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 773, 782 (2019); Wendy Wagner & Martin Murillo, *Is the Administrative State Ready for Big Data?*, *in* DATA & DEMOCRACY (Apr. 30, 2021), https://s3.amazonaws.com/kfai-documents/documents/684b5fd17e/4.29.2021-Wagner-and-Murillo.pdf [https://perma.cc/U6VN-8B9T]. Danielle Citron's seminal account of the "automated administrative state" argued that "[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." Citron, *supra* note 32, at 1281. These concerns, aired more than a decade ago, have only intensified because machine learning AI systems are generally more complex and less transparent than the technologies that Citron interrogated. *See* Calo & Citron, *supra* note 31, at 818 ("In the decade since the publication of *Technological Due Process*, governments have doubled down on automation despite its widening problems.").

38. *See* Brauneis & Goodman, *supra* note 18, at 152–63 (spotlighting the transparency deficits that accrue when state and local government adopt AI systems developed by third parties); Mulligan & Bamberger, *supra* note 37, at 782 (explaining how a "procurement mindset" can forfeit the government's responsibility to make important design choices with public input); *see also* ACUS REPORT, *supra* note 1, at 88–90 (outlining some pros and cons of the government's insourcing and outsourcing for AI solutions); MONA SLOANE ET AL., AI AND PROCUREMENT PRIMER 3 (Summer 2021) (observing that "existing public procurement processes and standards are in urgent need of revision and innovation").

39. *See, e.g.*, Brauneis & Goodman, *supra* note 18, at 164; Cary Coglianese & Erik Lampmann, *Contracting for Algorithmic Accountability*, 6 ADMIN. L. REV. ACCORD 175, 180 (2021).

regularize and legitimize algorithmic governance. Toward those ends, this Article makes three major contributions.

First, it provides an original account that yokes the ambitions of algorithmic governance, the principles of ethical AI, and the levers of procurement law. Broadly conceived, ethical AI envisages a cluster of principles relating to safety, fairness, transparency, accountability, privacy, and human well-being.[40] Hardly an abstract concern, ethical AI is a global imperative backed by the United States,[41] G20,[42] and all the leading technology firms (Amazon, Google, Facebook, Microsoft, and IBM, just to name a few).[43] Make no mistake, the institutional motivations fueling the ethical AI movement are pluralistic and opportunistic.[44] Yet the convergence of public and private interests around core ethical AI principles is what matters most for this discussion.[45] For the government and industry alike, AI innovation is a complex ambition that mediates technical capability and human values. Awful AI does not sell—politically or commercially.[46] Once these sociopolitical dynamics are accounted for, the normative case for acquiring ethical AI is also pragmatic.

---

40. *See infra* Sections II.B–C; Anna Jobin et al., *The Global Landscape of AI Ethics Guidelines*, 1 NATURE MACH. INTEL. 389, 390 fig.1 (2019), https://www.nature.com/articles/s42256-019-0088-2.pdf [https://perma.cc/68YX-NM8Z] (surveying 84 distinct ethical AI frameworks and finding that they have largely converged around a core set of concepts and principles, including safety, fairness, accountability, transparency, and privacy).

41. *See, e.g.*, Exec. Order No. 13,960, 85 Fed. Reg. 78,939, 78,940–41 (Dec. 8, 2020); OFF. OF SCI. & TECH. POL'Y, EXEC. OFF. OF THE PRESIDENT, AMERICAN ARTIFICIAL INTELLIGENCE INITIATIVE: YEAR ONE ANNUAL REPORT, at i (2020) ("In a time of global power competition, our leadership in AI has never been more of an imperative."); ORG. FOR ECON. COOP. & DEV. ARTIFICIAL INTELLIGENCE IN SOCIETY 16–17 (2019) [hereinafter OECD] www.oecd-ilibrary.org/docserver/eedfee77-en.pdf [https://perma.cc/CJP5-JNAM] (providing a comprehensive survey of the many ways that AI is projected to transform social structures and power dynamics across markets and borders).

42. *See* G20 MINISTERIAL STATEMENT ON TRADE AND DIGITAL ECONOMY app. at 11–14 (2019), https://www.mofa.go.jp/files/000486596.pdf [https://perma.cc/MJ9A-5U82].

43. *See infra* Section II.C (discussing proliferation of ethical AI throughout the public and private sectors); *see also* Alex Hern, *"Partnership on AI" Formed by Google, Facebook, Amazon, IBM, and Microsoft*, GUARDIAN (Sept. 28, 2016, 5:00 PM), https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms [https://perma.cc/74S6-EH9E].

44. *See infra* notes 199–05 and accompanying text.

45. *See infra* Section II.C (discussing political and market demand for ethical AI).

46. *See infra* Section II.C; David Dao et al., *Awful AI*, GITHUB, https://github.com/daviddao/awful-ai [https://perma.cc/PTU9-P92Q] (providing a "curated list to track current scary usages of AI—hoping to raise awareness to its misuses in society" (emphasis omitted)); *Artificial Intelligence Incident Database*, https://incidentdatabase.ai [https://perma.cc/2AXQ-BBA7] (providing a systematized collection of incidents where intelligent systems have caused safety, fairness, or other real-world problems, for the express purpose of "learn[ing] from [AI's] failings").

Second, this Article argues that the procurement system is uniquely suited to both check and enable algorithmic governance. Currently, the government is procuring AI systems that may be inoperable, either because the technology is untrustworthy or unlawful in application. The inscrutability of acquired AI systems, for example, might violate constitutional or statutory requirements for government transparency and accountability.[47] Even if those thresholds are met, the inputs and outputs of AI systems may violate anti-discrimination norms, privacy laws, and domain-specific legal constraints.[48] Litigation will no doubt surface these legal tensions.[49] Indeed, the AI docket is already littered with cautionary cases.[50] Yet many of these governance challenges can be addressed

47. *See, e.g.*, Citron, *supra* note 32, at 1281–88 (discussing how agency use of automated systems raises due process concerns); Mulligan & Bamberger, *supra* note 37, at 782 ("[T]he policy choices embedded in system design fail the prohibition against arbitrary and capricious agency actions absent a reasoned decision-making process that enlists the expertise necessary for reasoned deliberation, provides justifications for such choices, makes visible the political choices being made, and permits iterative human oversight and input."). *But cf.* Cary Coglianese, *Using Machine Learning to Improve the U.S. Government*, REGUL. REV. (Aug. 12, 2019), https://www.theregreview.org/2019/08/12/coglianese-using-machine-learning-to-improve-us-government/ [https://perma.cc/Z9PL-YR4W] (arguing that "with proper planning and implementation, the federal government's use of algorithms, even for highly consequential purposes, should not face insuperable or even significant legal barriers under any prevailing administrative law doctrines").

48. *See* Huq, *supra* note 33, at 1881 (explaining that machine learning AI technology "places pressure on the formulation of due process, equality, and privacy interests in subtly different ways").

49. For comprehensive surveys of AI-related litigation and trends, see *Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems*, AI NOW INST. 5 (2018), https://ainowinstitute.org/litigatingalgorithms.pdf [https://perma.cc/L584-83ZQ]; RASHIDA RICHARDSON ET AL., LITIGATING ALGORITHMS 2019 US REPORT: NEW CHALLENGES TO GOVERNMENT USE OF ALGORITHMIC DECISION SYSTEMS, AI NOW INST. 3 (2019), https://ainowinstitute.org/litigatingalgorithms-2019-us.pdf [https://perma.cc/X9H9-B3W5]; *see also* Coglianese & Lampmann, *supra* note 39, at 177 ("Without question, agencies that choose to use AI tools need to be mindful of the possibility that their choices could later come under not just the spotlight of media attention but also the scrutiny of judicial review.").

50. *See, e.g.*, Barry v. Lyon, 834 F.3d 706, 710–11 (6th Cir. 2016) (holding that Michigan's public benefits system erroneously terminated food assistance benefits of more than 20,000 individuals based on crude data matching algorithm in violation of due process guarantees); Cahoo v. SAS Analytics Inc., 912 F.3d 887, 892 (6th Cir. 2019) (noting that defendants "designed, created, and implemented" allegedly flawed software that erroneously terminated unemployment benefits of thousands of Michigan residents without adequate notice); Ark. Dep't of Hum. Servs. v. Ledgerwood, 530 S.W.3d 336, 338–40 (Ark. 2017) (affirming a temporary restraining order against an unlawful switch to computer algorithm that reduced the attendant-care services for multiple patients with severe illnesses by an average of 43%); Hous. Fed'n of Tchrs., Local 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1171–74 (S.D. Tex. 2017) (finding due process violation because teachers had no way to replicate and challenge their algorithmic scores); Latif v. Holder, 28 F. Supp. 3d 1134, 1161–62 (D. Or. 2014) (ordering the Federal Bureau of Investigation to "fashion new procedures" for its no-fly list policy and to "provide plaintiffs with

through the procurement system in ways that are more efficient, effective, and prior to harm.[51]

Third, this Article offers a set of pragmatic recommendations to operationalize ethical AI throughout the procurement process: from acquisition planning to market solicitation, bid evaluation, source selection, and contract performance.[52] By centering ethical AI across the procurement lifecycle, agency officials and vendors will be incented to think more holistically—and competitively—about the AI tools passing through the acquisition gateway for government use. Less directly, though equally important, the government's purchasing power and virtue signaling can spur technological innovation and galvanize public trust in AI technologies inside and outside of government.[53] Thus conceived and constructed, the acquisition gateway is more than a marketplace: it is a policymaking space for mediating the possibilities and perils of modern AI systems.

The Article proceeds as follows. Part I offers a primer on machine learning AI systems and spotlights a range of human value judgments embedded in the technology. Part II expounds upon AI's sociotechnical challenges and the political economies of ethical AI. Part III homes in on the residual gaps between ethical AI principles and practice, the implications of those gaps for algorithmic governance, and the limitations of existing laws and technologies to bridge the gulf. Part IV explicates how the federal procurement system can help—indeed, why it must. Procurement law will not solve all the challenges of algorithmic governance. Just as surely, those challenges cannot be solved without procurement law.

---

the requisite due process . . . without jeopardizing national security"). *But cf.* State v. Loomis, 881 N.W.2d 749, 753 (Wis. 2016) (holding that using an AI risk-assessment tool to aid judges with sentencing decisions "does not violate a defendant's right to due process").

51. *See generally* Part IV (offering a set of recommendations to promote AI safety, fairness, transparency, and accountability).

52. *See infra* Sections IV.A–B.

53. For a similar claim about the potential for positive externalities, see Coglianese & Lampmann, *supra* note 39, at 181 ("[T]he expectations that governments insist upon in their procurement contracts can help set the bar for algorithmic accountability throughout the economy, promoting the diffusion of norms about responsible AI across both the public and private sectors.").

I. AI TODAY

Despite the hype—or perhaps because of it—"artificial intelligence" has no "universally accepted definition."[54] The lexical rifts are largely attributable to the field's evolution and multi-disciplinarity, which spans computer science, mathematics, psychology, sociology, neuroscience, philosophy, linguistics, and intersects with countless more.[55] AI's lexical dissensus also reflects clashing ideologies. As Kate Crawford explains, "[e]ach way of defining artificial intelligence is doing work, setting a frame for how it will be measured, valued, and governed."[56] Sensitive to these concerns, and without normative pretense, this Article employs the term AI to capture a range of computer-based technologies that make predictions, classifications, recommendations, and automated decisions.[57]

Only a decade ago, AI was a fringe subject of academic study with sparse real-world applications. Rather abruptly, however, AI has emerged from research labs to disrupt every major market and facet of society.[58]

---

54. U.S. GOV'T ACCOUNTABILITY OFF., GAO-18-142SP 15 (2018) (observing that "[t]here is no single universally accepted definition of AI, but rather differing definitions and taxonomies"); *see also* Forrest E. Morgan et al., *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World*, RAND CORP. 8–9 & 9 n.4 (2020), www.rand.org/pubs/research_reports/RR3139-1.html [https://perma.cc/4XUX-7PQZ] (explaining the definitional challenges, and noting that "[i]t was striking how averse the experts we interviewed were to providing definitions of artificial intelligence").

55. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-18-142SP, *supra* note 54, at 15.

56. CRAWFORD, *supra* note 19, at 7.

57. 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 other 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, that is 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.

10 U.S.C. § 2358.

58. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-519SP, ARTIFICIAL INTELLIGENCE: AN ACCOUNTABILITY FRAMEWORK FOR FEDERAL AGENCIES AND OTHER ENTITIES 5 (2021) (noting that AI applications "rang[e] from medical diagnostics and precision agriculture, to advanced manufacturing and autonomous transportation, to national security and defense"); OECD, *supra* note 41, at 16–17 (providing a comprehensive survey of the many ways that AI is projected to transform social structures and power dynamics across markets and borders).

AI's ubiquity is masked by its obscurity. Yet, increasingly and impalpably, the technology is embedded or connected to smartphones and drones, cars and cattle, workstations and police stations, classrooms and war rooms, energy grids and traffic grids, social platforms and financial platforms, medical systems and surveillance systems.[59] As such, the technology is radically changing how nations, institutions, and individuals interact, experience, and perceive the world.[60]

A. *Machine Learning Systems*

AI's ascendance and dissemination over the past decade owes to the conflation of several developments: the availability of exponentially more data and computing power; the democratization of the internet of things;[61] and breakthroughs in "machine learning" technologies.[62] Unlike traditional computer algorithms that require manual coding, machine learning algorithms learn and improve from exposure to large amounts of data.[63] A full exposition of machine learning is beyond this Article's remit. But a basic understanding of the technology will be important for the discussion ahead. The challenges of algorithmic governance, and this Article's procurement prescriptions, are anchored to how machine learning systems are designed, developed, and deployed.

Stripped to its essentials, machine learning is a statistical technique that learns from data to make classifications or predictions for new data inputs.[64] For example, if the objective of an AI system is to detect and

---

59. *See* NSCAI FINAL REPORT, *supra* note 11, at 33.

60. *Id.* ("Americans have not yet grappled with just how profoundly the [AI] revolution will impact our economy, national security, and welfare."); Eleonore Pauwels, *The New Geopolitics of Artificial Intelligence*, WORLD ECON. F. (Oct. 15, 2018), https://www.weforum.org/agenda/2018/10/artificial-intelligence-ai-new-geopolitics-un [https://perma.cc/6JBY-X4NB] ("The multilateral system urgently needs to help build a new social contract to ensure that . . . [AI] is deployed safely and aligned with the ethical needs of a globalizing world."); *see also supra* notes 58–59 and accompanying text.

61. "The Internet of Things, or IoT, refers to the billions of physical devices around the world that are now connected to the internet, all collecting and sharing data." Steve Ranger, *What Is the IoT? Everything You Need to Know About the Internet of Things Right Now*, ZDNET (Feb. 3, 2020, 6:45 AM), https://www.zdnet.com/article/what-is-the-internet-of-things-everything-you-need-to-know-about-the-iot-right-now [https://perma.cc/ J2FD-B29T].

62. *See* NSCAI FINAL REPORT, *supra* note 11, at 20–21; *see also* Mireille Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, 20 THEORETICAL INQUIRIES L. 83, 84–85 (2019).

63. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-519SP, *supra* note 58, at 14 (distinguishing "first-wave" and "second-wave" AI technologies on this basis); *see also* David Lehr & Paul Ohm, *Playing with Data: What Legal Scholars Should Learn About Machine Learning*, 51 UC DAVIS L. REV. 653, 678 (2017).

64. *See* IBM Cloud Educ., *Machine Learning*, IBM (July 15, 2020), https://www.ibm.com/cloud/learn/machine-learning [https://perma.cc/X7J9-GWW7]; Sendhil

diagnose cancer, a computer can be fed many thousands of labeled images of malign and benign tumors, learn to distinguish the images based on patterns and correlations in the pixel data, and then generate an algorithmic model that can make diagnostic predictions.[65] Or, if the objective of an AI system is to predict violence within prison populations, a machine learning algorithm can be trained with data of prior incidents, inmate characteristics (e.g., age, education, weight, criminal history, gang affiliations), and other proxy variables deemed to correlate with prison violence. In turn, a warden or prison guard can deploy the trained AI model to predict incidents of inmate violence and take prophylactic action.[66]

The machine learning AI systems in circulation today are powerful but "narrow," insofar as they can handle discrete tasks in bounded domains (like in the examples above).[67] Deep neural networks are a subset of sophisticated machine learning algorithms that have been trained to classify images, recognize faces, translate languages, predict human emotions, personalize online experiences, and much (much) more.[68] Some of the most technologically advanced AI systems coordinate or aggregate multiple algorithmic models. For example, autonomous vehicles utilize a range of algorithms to perform driving and navigation functions.[69] Still, at present, the most advanced AI systems do

---

Mullainathan & Jann Spiess, *Machine Learning: An Applied Econometric Approach*, 31 J. ECON. PERSP. 87, 88 (2017) (defining machine learning in terms of its capacity for "out of sample" prediction). There are several different approaches to machine learning. For a short and accessible overview of the main approaches, see Nicholas Diakopoulos, *Algorithmic Accountability: Journalistic Investigation of Computational Power Structures*, 3 DIGIT. JOURNALISM 398, 399 (2015). For extended treatments, see generally IAN H. WITTEN ET AL., DATA MINING: PRACTICAL MACHINE LEARNING TOOLS AND TECHNIQUES (4th ed. 2017); KEVIN P. MURPHY, MACHINE LEARNING (2012); STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH (4th ed. 2021).

65. *Cf.* Daoud Meerzaman, *Machine Learning and Computer Vision Offer a New Way of Looking at Cancer*, NAT'L CANCER INST. (Jan. 27, 2019), https://datascience.cancer.gov/news-events/blog/machine-learning-and-computer-vision-offer-new-way-looking-cancer [https://perma.cc/7NS2-7JYY].

66. *See* Stefanie Kanowitz, *How Predictive Analytics Keeps Corrections Staff, Inmates Safe*, GCN (Aug. 18, 2021), https://gcn.com/articles/2021/08/18/predictive-analytics-corrections.aspx?s=gcnda_190821&oly_enc_id=&m=1 [https://perma.cc/RW46-WHRP].

67. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-18-142SP, *supra* note 54, at 15–16 (distinguishing between narrow and general AI).

68. *See* Bertrand Leong, *Rise of the Machines: Deep Learning, Machine Learning, AI, and Big Data*, SING. INST. MGMT. (2018), https://m360.sim.edu.sg/article/Pages/Rise-of-the-Machines.aspx [https://perma.cc/AFR4-5FFG]; Bernard Marr, *What Is Deep Learning AI? A Simple Guide With 8 Practical Examples*, BERNARD MARR & CO., https://bernardmarr.com/what-is-deep-learning-ai-a-simple-guide-with-8-practical-examples/ [https://perma.cc/PXV4-EGZZ].

69. *See* Rilind Elezaj, *How AI Is Paving the Way for Autonomous Cars*, MACH. DESIGN (Oct. 17, 2019), https://www.machinedesign.com/mechanical-motion-systems/article/21838234/

not have common sense, causal reasoning,[70] or situational awareness "to determine the relevance of new 'unknowns.'"[71] Moreover, unlike humans, AI systems cannot generalize or reliably transfer knowledge across experiential domains.[72] That type of "artificial general intelligence" does not (yet) exist and is beyond this Article's scope.[73]

B. *Humans in AI Systems*

The math and science behind AI systems can make them seem objective and neutral. However, a critical literature debunks that myth.[74] Beyond bits and bytes, AI systems are social artifacts that embed and project human choices, biases, and values.[75] These human inputs and outputs are hardly obvious. Precisely for that reason, it is crucial for policymakers and stakeholders to appreciate that value-laden choices of great social and legal consequence may be encased in AI systems prior to their adoption and deployment.

The discussion below provides a stylized account of just some of the human choices and tradeoffs that occur behind the AI curtain. This prelude supplies contextual mooring for the social, technical, and institutional challenges of algorithmic governance, which Parts II and III expound upon.

---

how-ai-is-paving-the-way-for-autonomous-cars [https://perma.cc/YH5L-JHH2]; U.S. GOV'T ACCOUNTABILITY OFF., GAO-18-142SP, *supra* note 54, at 65–68 (discussing the technology of automated vehicles).

70. *See* Brian Bergstein, *What AI Still Can't Do*, TECH. REV. (Feb. 19, 2020), www.technologyreview.com/2020/02/19/868178/what-ai-still-cant-do/ [https://perma.cc/3QL9-JYL6] (describing the inability of AI to engage in causal reasoning).

71. David Leslie, *Understanding Artificial Intelligence Ethics and Safety*, ALAN TURING INST. 32 (2019), https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf [https://perma.cc/96KG-VB77].

72. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO 21-519SP, *supra* note 58, at 16.

73. *See id.* (noting the speculative nature of general AI).

74. *See, e.g.*, CRAWFORD, *supra* note 19, at 8 (arguing that AI is neither artificial nor intelligent); MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM: THE SCIENCE OF SOCIALLY AWARE ALGORITHM DESIGN 5–7 (2019) (discussing the emerging science of ethical algorithm design to address the social implications of AI technologies); O'NEIL, *supra* note 18 (arguing that AI systems are "weapons of math destruction" that shape individual action and social dynamics); SAFIYA UMOJA NOBEL, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM 1–2 (2018) ("While we often think of . . . 'big data' and 'algorithms' as being benign, neutral, or objective, they are anything but.").

75. For critical treatments of AI's power dynamics, see *supra* note 74 (collecting sources), *see also* Sarah M. West et al., *Discriminating Systems: Gender, Race and Power in AI.* AI NOW INST. (Apr. 2019), https://ainowinstitute.org/discriminatingsystems.pdf [https://perma.cc/KCR9-CT6B] (describing technology industry's diversity challenges and how those challenges manifest in AI applications and ideations).

1. Problem Formulation and System Objectives

The design of a machine learning system generally begins with the identification of a specific problem that AI may help to solve. To illustrate, suppose a federal agency has a huge backlog of applications for government benefits.[76] There are many ways to solve that problem. But if the plan involves AI, then agency officials must select a relatively specific objective for the AI system.[77] Here, assume that the agency wants an AI system to predict which applications in the backlog are likely to be granted, for the well-intended purpose of expediting the delivery of government benefits to qualifying applicants.

Despite good intentions, the agency's problem formulation and AI objective may have negative collateral effects on various stakeholders. For instance, the unflagged cases might be further delayed because resources are channeled to the flagged cases. Worse still, agency personnel may unfairly (and unwittingly) stigmatize the unflagged cases as unmeritorious because the AI system did not predict a win.[78] Meanwhile, agency personnel may be shuffled or shifted to accommodate the dual-track system. Even if institutionally justified, the disruptions may seed confusion or discontent in the ranks.

Already, this sketch illustration exposes the human underbelly of AI systems. The agency's formulation of the problem (backlog) and chosen objective for the AI system (to flag meritorious cases) were not preordained, much less conscribed to math or science. Furthermore, as discussed below, each of the foregoing choices will lead to countless more choices and path dependencies throughout the AI development lifecycle.

---

76. *See* Aaron Boyd, *VA Wants to Automate Digitization of Its 5-Mile-High Electronic Health Record Backlog*, NEXTGOV (July 9, 2020), https://www.nextgov.com/it-modernization/2020/07/va-wants-automate-digitization-its-5-mile-high-electronic-health-record-backlog/166769/ [https://perma.cc/A33G-36U2]; Abigail Hauslohner, *The Employment Green Card Backlog Tops 800,000, Most of Them Indian. A Solution Is Elusive.*, WASH. POST (Dec. 17, 2019, 5:26 PM), https://www.washingtonpost.com/immigration/the-employment-green-card-backlog-tops-800000-most-of-them-indian-a-solution-is-elusive/2019/12/17/55def1da-072f-11ea-8292-c46ee8cb3dce_story.html [https://perma.cc/THS8-DLR2]; *see also* ACUS REPORT, *supra* note 1, at 37–45 (discussing use of AI system for social security case processing).

77. *See* MURPHY, *supra* note 64, at 2.

78. *Cf.* Citron, *supra* note 32, at 1271–72 (discussing "automation bias"); *infra* notes 153–57 and accompanying text (same).

2. Data Selection and Preparation

Once the system objectives have been established, developers must assemble a corpus of data to train a machine learning algorithm. Because AI models are "only as good as the data" that trains them,[79] data selection and preparation are arguably the most important parts of the development process.

In our hypothetical, the agency's previously decided cases will be a primary source of training data. All else equal, the accuracy of machine learning algorithms improve with exposure to more data.[80] Thus, five years of training data may be better than three years of comparable data. But if the data quality depreciates over time, then a tradeoff between data quality and quantity is unavoidable.[81] These sorts of discretionary judgments will generally be made by data scientists with input from subject matter experts.[82]

In some contexts, data selection may also entail legal judgment, political choice, or some combination thereof. For instance: should the training data exclude cases decided prior to a relevant statutory amendment, and should cases from certain regions or subpopulations be included?[83] Whatever the answers, they will be informed by normative and analytic judgments that will directly impact the AI system's performance in potentially consequential ways. As Michael Kearns and Aaron Roth explain, "[w]hen maximizing accuracy across multiple different populations, an algorithm will naturally optimize better for the majority population, at the expense of the minority population[.]" [84]

3. Model Training

Once the data is ready, the development team can use it to train a machine learning algorithm. In general, the goal of this phase is to optimize an algorithm's objective function, which is a "mathematical expression of the algorithm's goal."[85] In lay terms, as applied to our

79. Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 106 (2014).

80. *See* Lehr & Ohm, *supra* note 63, at 678 ("To reap the predictive benefits of machine learning, a sufficiently large number of observations is required.").

81. *See* WITTEN ET AL., *supra* note 64, at 60 ("Domain experts need to be consulted to explain anomalies, missing values, the significance of integers that represent categories rather than numeric quantities, and so on.").

82. *Id.*

83. *Cf.* John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1794 (2018) (discussing the mismatch that can occur, in the criminal bail context, if training data is collected before the implementation of bail reforms).

84. KEARNS & ROTH, *supra* note 74, at 78. This occurs because, by definition, "there are more people from the majority group, and hence they contribute more to the overall accuracy of the model." *Id.*

85. Lehr & Ohm, *supra* note 63, at 671.

hypothetical, the goal is to train an AI model that can accurately predict which cases in the backlog are meritorious (and, as much as possible, to flag *only* those cases).[86] No attempt will be made here to capture all the ingenuity, craft, analysis, and discretionary judgment that model training entails.[87] But, as one pivotal example, AI developers must make tradeoffs between two types of prediction errors: false positives and false negatives.[88]

In our hypothetical, the *false positives* are the cases that the AI system erroneously flags as likely winners. The *false negatives* are the meritorious cases that the AI system misses.[89] For certain types of algorithms, the ratio of false positives to false negatives can be predetermined and forced by code.[90] For other algorithms, the error rates can be manipulated by adjusting sensitivity thresholds.[91] Whether to err on the side of false negatives or false positives, and how much so, are policy choices of great significance. As such, different stakeholder could choose differently, based on different considerations and calculations. Indeed, as David Lehr and Paul Ohm explain, "it is very rare for a stakeholder to view being wrong in one way as equally harmful as being wrong in the opposite way."[92] Moreover, context matters. The optimal error ratio for our hypothetical case-flagging system, whatever it may be, will probably not be the optimal error ratios for predicting cancer or prison violence. And, certainly, the considerations that inform those judgments are not the same. The key point here is that *humans* decide not only what to shoot for during model training, but also the metrics for "success."

4. Model Testing and Evaluation

After an AI model is trained, it must be tested and evaluated to explore its "fit" between the data and its target objective. The goal of this phase is to determine if, and how well, the trained model can generalize to make accurate predictions for *new* data inputs (i.e., outside of the original

86. *Cf. id*. at 671–72 (discussing objective functions).

87. Entire courses and books are filled with the math, science, and ingenuity of model training. *See, e.g.*, WITTEN ET AL., *supra* note 64; Andrew Ng et al., *Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization*, COURSERA, https://www.coursera.org/learn/deep-neural-network?trk_location=query-summary-list-link [https://perma.cc/RC69-EZ38].

88. *See* Lehr & Ohm, *supra* note 63, at 691–92 (discussing false positives and false negatives).

89. *See id.*

90. *See id.* (discussing opportunities to code an "asymmetric cost ratio").

91. *Id.*

92. *Id.*

training data).[93] There are many different testing and validation methods, each with their own tradeoffs and limitations.[94] Moreover, AI developers must make normative choices about what to test for and why. Thus, for our hypothetical, the developers may (or may not) test the model to determine if it performs equally well on benefit applications filed by men and women, equally well on applications filed by Black men and White women, and so on.[95] In addition, the clustering of testing variables entails its own set of choices and empirical voids. Testing for demographic parity on the dimensions of race and gender, for example, may fail to properly account for intersectional differences within those groups along the dimensions of age and nationality.[96]

5. Model Selection and System Configuration

Throughout the development process, several different AI models may be trained, tested, and evaluated. Selecting which model(s) to deploy in an AI system is generally informed by a range of objectives, performance metrics, risks, and constraints.[97] Moreover, each of those considerations may be influenced by a mix of social, legal, financial, technical, political, and logistical considerations.

One important set of design choices pertain to model "interpretability." In the AI field, interpretability refers to the ability of humans to understand an AI model's logic or decisional pathway from inputs to outputs.[98] Some machine learning models are more scrutable

93. *See* Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, COMMC'NS OF THE ACM, Oct. 2012, at 78, 81–82 (discussing challenges relating to overfitting and underfitting).

94. *See id.*; Lehr & Ohm, *supra* note 63, at 699–701.

95. *Cf.* Barocas & Selbst, *supra* note 18, at 677 (explaining why an AI system might not perform equally across groups and subgroups); *see also infra* Section II.B.2 (discussing technical and non-technical causes of algorithmic bias and other forms of unfairness).

96. *See* Inioluwa Deborah Raji et al., *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing*, *in* FACCT '20: PROCEEDINGS OF THE 2020 CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 39 (2020), https://arxiv.org/pdf/2001.00973.pdf [https://perma.cc/GH6A-HXKC] ("Algorithm development implicitly encodes developer assumptions that they may not be aware of, including ethical and political values.").

97. *See* Lehr & Ohm, *supra* note 63, at 690 (discussing six considerations for model selection: "the kind of output variable, the ability to implement an 'asymmetric cost ratio,' the ability to explain or offer reasons for the predictions, the potential for overfitting, the opportunities for tuning, and practical resource limitations").

98. *See* Brent Mittelstadt et al., *Explaining Explanations in AI*, *in* FACCT '19: PROCEEDINGS OF THE 2019 CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 279, 280 (2019), https://arxiv.org/pdf/1811.01439.pdf [https://perma.cc/NEM3-KHLU] (highlighting how "poorly interpretable models" are unable to reveal how classifications result from the inputs).

than others.[99] Linear regression algorithms, for example, are relatively easy for humans to comprehend but have limited functionality.[100] By contrast, deep learning neural networks drive some of the most powerful, sophisticated, and functional AI systems, but their complexity renders them inscrutable to humans.[101] The inputs and outputs of these black-box systems may be known, but the computational formulas that churn inputs into outputs may entail thousands, millions, or billions of parameters.[102] At that level of complexity, an AI's inner logic defies human comprehension.[103] Thus, in contexts where different machine learning algorithms might be suited for a particular task, AI developers may need to make a tradeoff between model accuracy and interpretability.[104] Put otherwise, models with better overall accuracy may be pitted against simpler models that, if selected, would allow humans to know when and why the AI's predictions are wrong.

Other important design considerations anticipate post-deployment human interactions with the AI model. For example, to run the model, humans may need to collect and input feature variables (e.g., criminal

99. *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 3–4, 9 (2015) (shining critical light on the "black box" nature of algorithmic systems); Jatinder Singh et al., *Responsibility & Machine Learning: Part of a Process* 4–5 (Oct. 27, 2016) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860048 [https://perma.cc/DW2M-5PG7].

100. *See* Singh et al., *supra* note 99, at 4.

101. *See id*.; *see also* Scott Wisdom et al., *Interpretable Recurrent Neural Networks Using Sequential Sparse Recovery* 1 (Nov. 22, 2016) (unpublished manuscript), https://arxiv.org/pdf/1611.07252.pdf [https://perma.cc/TL24-M94R] ("Interpreting the learned features and outputs of machine learning models is problematic. This difficulty is especially significant for deep learning approaches [like neural networks], which are able to learn effective and useful function maps due to their high complexity.").

102. A model parameter is a configuration variable that is internal to the model. OpenAI's GPT-3, a language model capable of natural language processing tasks, "has a whopping 175 billion parameters." Khari Johnson, *OpenAI Debuts Gigantic GPT-3 Language Model with 175 Billion Parameters*, VENTUREBEAT (May 29, 2020, 8:34 AM), https://venturebeat.com/2020/05/29/openai-debuts-gigantic-gpt-3-language-model-with-175-billion-parameters/ [https://perma.cc/4GZQ-SJCL]; *see also* William Fedus et al., *Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity*, ARXIV (Jan. 11, 2021), https://arxiv.org/abs/2101.03961 [https://perma.cc/6S54-BV3U]; Divye Singh, *New Contender in Trillion Parameter Model Race*, MEDIUM (June 8, 2021), https://medium.com/geekculture/new-contender-in-trillion-parameter-model-race-6ef0675ddd46 [https://perma.cc/9UTN-425H].

103. *See* Singh et al., *supra* note 99, at 5–6; *see also* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For*, 16 DUKE L. & TECH. REV. 18, 56–57 (2017) (discussing the "curse of dimensionality" that results in complex machine learning models when large amounts of variables are combined in complex ways so as to defy comprehension).

104. *Cf.* Cynthia Rudin, *Please Stop Explaining Black Box Models for High States Decisions*, 1 NATURE MACH. INTEL. 206, 207–08 (2019) (urging the use of interpretable models in high-stakes contexts and challenging the myth that there is necessarily a tradeoff between accuracy and interpretability).

history, nationality, age, etc.) to generate an AI output. Humans may also need to decide what to do, if anything, with that output. In such systems, there is a so-called human *in-the-loop*[105]—which in our example might be agency adjudicators assigned to the flagged cases. But AI systems can also be designed to work autonomously post-deployment,[106] with or without a human monitor *on-the-loop*.[107] For all configurations, however, AI developers must exercise human judgment and foresight about who will use or control the system, in what contexts, for what purposes, and under what constraints.[108]

The resulting configurations can have major downstream implications.[109] For instance, the design of the human–computer interface may influence a developer's *own* sense of responsibility, insofar as accountability is offloaded—rightly or wrongly—to operators and end-users during deployment.[110] Irrespective of a developer's intentions, a human in-the-loop may be held to account for reasons beyond their control. And, in these or other scenarios, a human may become "a rubber stamp for the machine, providing nothing more than a cosmetic reason to lull [stakeholders] into feeling better about the results."[111]

* * *

As the foregoing discussion hopes to impress, machine learning AI systems are infused with countless value-laden choices and tradeoffs. Far too often, these human decisions are latent, unexpressed, ad hoc, post hoc, or myopically informed. Drawing them to the surface is a critical first step toward understanding why AI systems are less like calculators and

105. *See* Singh et al., *supra* note 99, at 13.

106. *See id.* at 14. An email spam filter is an example of autonomous AI.

107. *See* Joel E. Fischer et al., *In-the-Loop or On-the-Loop? Interactional Arrangements to Support Team Coordination with a Planning Agent*, CONCURRENCY & COMPUTATION PRAC. & EXPERIENCE, Mar. 6, 2017, at 2, https://onlinelibrary.wiley.com/doi/full/10.1002/cpe.4082 [https://perma.cc/5USE-7BW6] (distinguishing between humans in-the-loop and on-the-loop, and studying contexts in which one structuring might be preferable to others).

108. Some or all of this information may not be known during the development stage, or might be known but change over time in unforeseeable ways.

109. *See* Will Orr & Jenny L. Davis, *Attributions of Ethical Responsibility by Artificial Intelligence Practitioners*, INFO. COMMC'N & SOC'Y 719, 725 (2020) (describing a "pattern of ethical dispersion" in machine-learning AI development, whereby "powerful bodies set the parameters, practitioners translate these parameters into tangible hardware and software, and then relinquish control to users and machines, which together foster myriad and unknowable outcomes"); *see also* Meg Leta Jones, *The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles*, 18 VAND. J. ENT. & TECH. L. 77, 90–91 (2015) (revealing how legal approaches that ignore the complex relations between humans and machines fail to protect the values legal approaches sought to protect).

110. *See* Orr & Davis, *supra* note 109, at 7 (discussing how perceptions of ethical responsibility are dispersed in AI development and deployment).

111. Lehr & Ohm, *supra* note 63, at 716.

more like calculated policy. This orientation, in turn, foists pressure on the organizing question of *who decides* what an AI's embedded policies should be.[112] In our case-flagging hypothetical, the government presumably made those important choices. Still, who decides *within* government can impact the functionality and features of deployed AI systems. By the same token, the government's decision to *outsource* AI development or deployment can be highly consequential. Especially if private vendors become the de facto deciders.[113]

II. TOWARD ETHICAL ALGORITHMIC GOVERNANCE

This Part maps the promises and pitfalls of algorithmic governance. The discussion begins with AI's positive potential because that is what drives the demand for algorithmic governance in the first place. From that starting position, the discussion pivots to a range of recursive sociotechnical challenges inhering in machine learning systems: specifically as pertains to safety, fairness, transparency, and accountability. The Part concludes with a contextual rendering of the rise of ethical AI frameworks to address these challenges.[114]

A. *Good (Algorithmic) Governance*

Ideally, AI can make government more efficient and effective across a wide range of functions: law enforcement, adjudication, rulemaking, national security, resource allocation, in-house management, delivery of public services, and beyond.[115] Moreover, with proper design, AI systems can provide greater accuracy than human deciders alone.[116] Further, AI

112. *See infra* notes 250–54 and accompanying text (generally discussing the government's build-or-buy choice).

113. *Cf.* GRY HASSELBACH ET AL., WHITE PAPER ON DATA ETHICS IN PUBLIC PROCUREMENT OF AI-BASED SERVICES AND SOLUTIONS 11 (2020) ("The government's choice among competing market solutions will generally entail "a prioritization of interests and values embedded in [product] design.").

114. *See infra* Section II.C; *see also infra* Part III (canvassing the many challenges of translating ethical AI principles in practice, for both industry and government actors).

115. *See* ACUS REPORT, *supra* note 1, 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."); *see also supra* notes 1–11 and accompanying text (providing examples of federal agency uses of AI).

116. *See* Coglianese & Lehr, *supra* note 37, at 6 (describing how machine learning algorithms produce "unparalleled accuracy" compared to other statistical methods and human judgment).

systems may be more transparent and accountable than government agents, who might conceal or be unaware of their own cognitive biases.[117]

The centralization of AI decision-making may also promote greater consistency across cases, both in public-facing operations (such as adjudication and law enforcement) and inward-facing operations (such as personnel retention).[118] What's more, centralized AI decision-making can facilitate audits of external and internal government programs.[119]

The government's adoption of AI technologies may also be fiscally responsible. By "automating repetitive tasks" and "augmenting" the capabilities of federal workers, taxpayer dollars can be saved or rerouted to better use.[120] According to one rosy estimate, the government's widespread adoption of AI could yield $500 billion in cost reductions over the next decade.[121]

Suffice to say, AI has the potential to augment, enable, and vastly improve government operations. Beyond better, however, the government's rapid uptake of AI is arguably imperative "to protect [the

---

117. *See id.* ("We find reason to be optimistic that, notwithstanding machine learning's black-box qualities, responsible governments can provide sufficient transparency about their use of algorithms to supplement, and possibly even replace, human judgments."); David Freeman Engstrom & Daniel E. Ho, *Artificially Intelligent Government: A Review and Agenda*, *in* RESEARCH HANDBOOK ON BIG DATA LAW 64 (Roland Vogl ed., 2021) ("The perhaps counterintuitive 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 12, at 656–77 (explaining how, through proper design, AI systems can be made more transparent and accountable).

118. The gains in consistency depend, in part, on whether humans-in-the-loop can adjust or override algorithmic scores and under what conditions or constraints. *Cf.* Evans & Koulish, *supra* note 4, at 794 (exposing how immigration enforcement agents manipulated an automated risk classification system used for immigration detention and release).

119. Alice Xiang, *Reconciling Legal and Technical Approaches to Algorithmic Bias*, 88 TENN. L. REV. (forthcoming 2021) (manuscript at 12), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3650635 [https://perma.cc/SUW2-2L5Q] (noting the "potential for algorithms to centralize decision-making, which can make auditing decisions easier," relative to "hundreds or thousands of human decision-makers"). That said, the audibility and oversight of AI decision-making will depend on whether those decisions are documented, traceable, and transparent—none of which can be assumed under current conditions. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-519SP, *supra* note 58 (providing an auditing framework for federal AI systems); *cf.* U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-518, FACIAL RECOGNITION TECHNOLOGY: FEDERAL LAW ENFORCEMENT AGENCIES SHOULD BETTER ASSESS PRIVACY AND OTHER RISKS 20 (2021) (reporting that more than a dozen "federal agencies do not have awareness of what non-federal systems with facial recognition technology are used by [federal] employees," and "have therefore not fully assessed the potential risks of using these systems, such as risks related to privacy and accuracy").

120. CHRISTINA BONE ET AL., THE COMING AI PRODUCTIVITY BOOM AND HOW FEDERAL AGENCIES CAN MAKE THE MOST OF IT 2, 4 (2020).

121. *Id.* That estimate, however, was based on projections of government adoptions of AI systems at a much faster rate than current capabilities.

nation's] security, promote its prosperity, and safeguard the future of democracy."[122] That was a top-line message delivered by the National Security Commission on Artificial Intelligence (NSCAI) to the President and Congress in 2021.[123]

While this Article is principally focused on civilian and domestic contexts, the global "AI arms race" is quite relevant here. Foremost, the global competition exerts gravitational pull on the government's *entire* AI trajectory.[124] More concretely, the race anchors the government's ambition to "retain [America's] innovation leadership"[125]—which depends mightily on the industry's capacities and cooperation. Further, the U.S./China juxtaposition—and narratives around it—crystallize the need for AI innovation and ideation that reflects American values.[126] As put by the NSCAI: If AI systems violate civil rights, or "have significant negative consequences, then leaders will not adopt them, operators will not use them, Congress will not fund them, and the American people will not support them."[127]

---

122. NSCAI FINAL REPORT, *supra* note 11, at 8.

123. *Id*. at 8–9 ("[N]ational security professionals must have access to the world's best [AI] technology to protect themselves, perform their missions, and defend us."). The NSCAI was established by Section 1051 of the John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, 132 Stat. 1636 (2018). The NSCAI's task is to make recommendations to the President and Congress to "advance the development of artificial intelligence [AI], machine learning, and associated technologies to comprehensively address the national security and defense needs of the United States." *Id.*

124. *See, e.g*., Michael Kratsios, Opinion, *Why the US Needs a Strategy for AI*, WIRED (Feb. 11, 2019, 9:00 AM),) https://www.wired.com/story/a-national-strategy-for-ai/ [https://perma.cc/4NEB-S6SH] ("An AI future that enriches the lives of our citizens, promotes innovation, and ensures our national and economic security requires continued American leadership."); Gary Grossman, *The AI Arms Race Has Us on the Road to Armageddon*, VENTURE BEAT (Apr. 19, 2021, 2:10 PM), https://venturebeat.com/2021/04/19/the-ai-arms-race-has-us-on-the-road-to-armageddon/ [https://perma.cc/J3YJ-8YU6] ("For now, the AI arms race is a cold war, mostly between the U.S., China, and Russia, but worries are it will become more than that.").

125. *See* NSCAI FINAL REPORT, *supra* note 11, at 11 ("The United States . . . must do what it takes to retain its innovation leadership and position in the world . . . . and organize to win it by orchestrating and aligning U.S. strengths.").

126. *See* Darren Byler, *China's Hi-tech War on Its Muslim Minority*, GUARDIAN (Apr. 11, 2019, 1:00 PM), https://www.theguardian.com/news/2019/apr/11/china-hi-tech-war-on-muslim-minority-xinjiang-uighurs-surveillance-face-recognition [https://perma.cc/63LM-ZM93]; Paul Mozur, *Inside China's Dystopian Dreams: A.I., Shame, and Lots of Cameras*, N.Y. TIMES (July 8, 2018), https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html [https://perma.cc/8J6E-R4J5]; Drew Donnelly, *An Introduction to the China Social Credit System*, NEW HORIZONS (Aug. 2, 2021), https://nhglobalpartners.com/china-social-credit-system-explained [https://perma.cc/VRZ8-KKSD].

127. NSCAI FINAL REPORT, *supra* note 11, at 133.

B. *Ethical Challenges*

AI's social prospects are simultaneously alluring and alarming: we want efficient and effective AI systems, but not if they efficiently or effectively cause harm. Satisfying *all* of these conditions may not be possible; AI's promises and perils are hard to decouple.

Facial recognition systems, for example, can be deployed to find lost children and terrorists, but can also be used for Orwellian surveillance and social repression.[128] AI can help to mitigate climate change, but the computing resources and raw materials powering large-scale AI models are environmentally unsustainable.[129] AI can uncover and rectify social biases, but can also learn those biases from training data and project them into the future.[130]

If there is any sense in which AI is neutral, it is the technology's dual capacity for good and evil when deployed by humans in real-world settings. That's the rub and root of AI's sociotechnical challenges in general, and for algorithmic governance especially. Although necessarily partial, the discussion below is centered around four pillars of ethical AI: safety, fairness, transparency, and accountability.[131] Broadly conceived, these principles link to the government's legal obligations and norms of good governance.

---

128. *See* Clare Garvie & Laura M. Moy, *America Under Watch: Face Surveillance in the United States*, GEO. L. CTR. ON PRIV. & TECH. (May 16, 2019), https://www.americaunderwatch.com/ [https://perma.cc/4BXQ-8X98] ("When used on public gatherings, face surveillance may have a chilling effect on our First Amendment rights to unabridged free speech and peaceful assembly."); *see also* GEORGE ORWELL, NINETEEN EIGHTY FOUR (1949) (depicting a canonical surveillance state).

129. *See* Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, *in* FACCT '21: PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 610, 612–13 (2020), https://dl.acm.org/doi/10.1145/3442188.3445922 [https://perma.cc/4FNP-CXBF] (providing a critical account of the environmental costs of training large-language AI models); Amy Stein, *Artificial Intelligence and Climate Change*, 37 YALE J. ON REG. 890, 892–93, 919 (2020) (arguing for the enhanced use of AI to address climate change, and discussing the "need to ensure that AI's negative environmental impacts are outweighed by its positive ones"); *see also* Thomas Griffin, *Why We Should Care About the Environmental Impact of AI*, FORBES (Aug. 17, 2020, 9:10 AM), https://www.forbes.com/sites/forbestechcouncil/2020/08/17/why-we-should-care-about-the-environmental-impact-of-ai [https://perma.cc/HV3A-B28Y].

130. *See* Steven Mills, *Foreword* to MONTREAL AI ETHICS INST., *The State of AI Ethics: January 2021*, at 11, 12 (2021), https://montrealethics.ai/wp-content/uploads/2021/01/The-State-of-AI-Ethics-Report-January-2021.pdf [https://perma.cc/AQ8W-H4WB]; *see also* Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2226 (2019) ("Counterintuitively, algorithmic assessment could play a valuable role in a system that targets the risky for support rather than for restraint.").

131. *See* Jobin et al., *supra* note 40, at 391 (mapping and analyzing a corpus of ethical AI principles and guidelines from around the globe, and finding a convergence around the principles of "transparency, justice and fairness, non-maleficence, responsibility and privacy").

1. Safety

In the AI field, "safety" connotes a range of system attributes, including model accuracy, reliability, robustness, and security.[132] When exposed to real-world elements, AI systems make mistakes that most humans never would. In part, that is because the variance and complexity of the real world may not be captured in the training data, which is all the algorithm knows.[133] Of course, humans make mistakes too. But the efficiency and scalability of AI systems make them uniquely concerning.[134] As Robert Brauneis and Ellen Goodman explain, "[t]he ability of these algorithmic processes to scale, and therefore to influence decisions uniformly and comprehensively, magnifies any error or bias that they embody[.]"[135] Here, math is relevant: an efficient AI decision-making system that makes 100,000 predictions at a 10% error rate may harm 10,000 individuals; by contrast, an inefficient human that makes 100 predictions at a 25% error rate may harm 25 individuals.

The safety of AI systems can also be compromised by adversarial attacks. For example, a malicious actor can manipulate data upon which an AI model will be trained and tested.[136] Such "data poisoning" can result in "curated misclassification, systemic malfunction, and poor performance."[137] A malicious actor can also manipulate deployed AI models to induce gross miscalculations.[138] This brittleness, at scale, can have profound consequences in high-stakes and safety-critical contexts. Certainly, humans can be manipulated, bribed, or spied upon by malicious actors in ways that undermine or endanger public interests. Yet AI systems have similar human vulnerabilities throughout the developmental pipeline *plus* attack surfaces in the data, code, cloud, and hardware, across sprawling supply chains.[139]

---

132. *See* Leslie, *supra* note 71, at 30.

133. Fábio Kepler, *Why AI Fails in the Wild*, UNBABEL (Nov. 15, 2019), https://unbabel.com/blog/artificial-intelligence-fails/ [https://perma.cc/VHN9-RNKT] ("When unrepresentative data is used for training, sometimes with no considerations about how the training data was collected or where it came from, it can be very problematic to apply a model to different situations from the ones it knows."); *see also* Colin Smith et al., *Hazard Contribution Modes of Machine Learning Components*, *in* THE AAAI-20 WORKSHOP ON ARTIFICIAL INTELLIGENCE SAFETY 4 (2020), http://ceur-ws.org/Vol-2560/paper41.pdf [https://perma.cc/2D27-GECB] (discussing unexpected performance, for example, "through unanticipated feature interaction . . . that was also not previously observed during model validation").

134. *See* Brauneis & Goodman, *supra* note 18, at 129; *see also* O'NEIL, *supra* note 18, at 29–31 (discussing the scalability of algorithms and consequent risk of widespread harm).

135. Brauneis & Goodman, *supra* note 18, at 129.

136. Leslie, *supra* note 71, at 32–33.

137. *Id.* at 33.

138. *Id.* at 32–33.

139. *Cf.* Exec. Order No. 14,028, 86 Fed. Reg. 26,633, 26,637 (May 12, 2021) (noting, in a

The foregoing safety risks compound when AI systems interact with other technologies, analog systems, or new environmental conditions.[140] For example, the output from one AI system may become another system's input, and so on. The resulting domino effects and feedback loops can cause "systems of systems" to drift from their anticipated performance, often imperceptibly and dangerously.[141] This arguably occurred in Michigan, for example, when the state used an AI system created by a commercial vendor to detect fraudulent claims for unemployment benefits.[142] For a variety of reasons, the system had a high error rate that caused tens of thousands of individuals to suffer life-changing financial harm, not only from the denial of benefits, but also in the form of collateral penalties, interest, and lost wages.[143] The resulting damage, and class-action litigation, spans years and is still ongoing.[144]

2. Fairness

"Fairness" has no agreed-upon meaning.[145] A concept like "justice" may work just as well or better, but it too has no agreed-upon meaning.[146] For present purposes, what matters is the breadth of concerns that fairness (or justice) captures, including nondiscrimination, due process, autonomy, inclusivity, equal opportunity, and fair dealing. To greater and lesser extents, AI may cohere with these human values—but that alignment will not obtain by default.

---

lengthy executive order to improve the nation's cybersecurity posture, that "the development of commercial software often lacks transparency, sufficient focus on the ability of the software to resist attack, and adequate controls to prevent tampering by malicious actors").

140. *See* Singh et al., *supra* note 99, at 15.

141. *Id.* (noting that the interactions can be direct or more indirect, through "butterfly effects," where subtle actions of a system can affect others in potentially dramatic ways).

142. *See* Cahoo v. SAS Inst. Inc., 377 F. Supp. 3d 769, 771 (E.D. Mich. 2019).

143. *See* Kate Crawford et al., *AI Now 2019 Report*, AI NOW INST. 36 (2019), https://ainowinstitute.org/AI_Now_2019_Report.html [https://perma.cc/E4ZL-DXEL].

144. *See* Cahoo v. Fast Enters. LLC, 528 F. Supp. 3d 719, No. 17-10657, 2021 WL 1146119 (E.D. Mich. Mar. 25, 2021); Cahoo v. SAS Analytics Inc., 912 F.3d 887, 892 (6th Cir. 2019); *see also* Alejandro De La Garza, *States' Automated Systems Are Trapping Citizens in Bureaucratic Nightmares with Their Lives on the Line*, TIME (May 28, 2020, 2:24 PM), https://time.com/5840609/algorithm-unemployment/ [https://perma.cc/CA99-9CT3].

145. *See* Abigail Z. Jacobs & Hanna Wallach, *Measurement and Fairness*, *in* FACCT '21: PROCEEDINGS OF THE ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 375, 375 (2021), https://arxiv.org/pdf/1912.05511.pdf [https://perma.cc/3HEG-6YBN].

146. *See* Reuben Binns, *Fairness in Machine Learning: Lessons from Political Philosophy*, *in* FACCT '18: PROCEEDINGS OF THE 2019 CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1, 1 (2018), http://proceedings.mlr.press/v81/binns18a/binns18a.pdf [https://perma.cc/CA4Z-VHRL] (drawing parallels between justice and fairness, and noting that "attempts to formalise 'fairness' in machine learning contain echoes of these old philosophical debates").

Intentionally or not, AI models absorb social biases contained in training data.[147] For example, an AI system deployed to predict criminal activity will exhibit higher false positive rates for Black defendants if the training data is an artifact of discriminatory policing.[148] Likewise, AI language models that learn from a corpus of text scraped from the internet will exhibit the anti-Semitic, anti-Muslim, and gender biases captured in online content.[149] Especially in complex AI systems, it can be difficult to identify algorithmic biases until they manifest, and quite difficult to fix post hoc.[150]

Of course, social bias is not specific to AI; humans are biased too. This cynical observation, however, only sharpens the point: human biases throughout society get captured in data, which gets imbued in AI models that make biased classifications and predictions. Through these dynamics—and with objective veneer—historical data is effectively laundered into the future and dispersed through networks of AI and analog systems.[151]

For sensitive government decisions, humans in-the-loop may be expected to exercise human judgment as a check on algorithmic classifications and predictions.[152] Studies show, however, that humans

147. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO 21-519SP, *supra* note 58, at 9 ("Biases arise from the fact that AI systems are created using data that may reflect preexisting biases or social inequities.").

148. *See, e.g.*, Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 15, 20 (2019) (finding that nine of thirteen police departments that were studied likely used "dirty data" to train and used predictive policing algorithms, many of which were acquired from the private sector with federal funding); Letter from Members of Congress to Merrick Garland, Att'y Gen. (Apr. 15, 2021) (on file with author) (requesting information about federal funding and oversight of predictive policing algorithms and asserting that these algorithms "likely . . . amplify biases against historically marginalized groups").

149. Tom Brown et al., *Language Models Are Few-Shot Learners* 36–39 (July 22, 2020) (unpublished manuscript), https://arxiv.org/pdf/2005.14165.pdf [https://perma.cc/X6K2-L6ES] ("[I]nternet-trained models have internet-scale biases; models tend to reflect stereotypes present in their training data."); *see also* Abubakar Abid et al., Persistent Anti-Muslim Bias in Large Language Models 9–10 (Jan. 18, 2021) (unpublished manuscript), https://arxiv.org/abs/2101.05783 [https://perma.cc/D7H5-5KGU].

150. *See* Brown et al., *supra* note 149, at 32–39 (discussing preliminary findings of bias in GPT-3 along the dimensions of gender, race, and religion); *id.* at 39 (noting that efforts to "remove" bias have "been shown to have blind spots"); Karen Ho, *This Is How AI Bias Really Happens—and Why It's So Hard to Fix*, MIT TECH. REV. (Feb. 4, 2019), https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix [https://perma.cc/C39Y-GELM].

151. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO 21-519SP, *supra* note 58, at 9 (observing that AI systems have the "potential to amplify existing biases and concerns related to civil liberties, ethics, and social disparities").

152. *See supra* notes 106–07 and accompanying text (discussing humans in-the-loop and on-the-loop).

are prone to over rely on computer recommendations—a phenomenon known as "automation bias."[153] Such complacency can result in human failure to identify or rectify AI errors.[154] Risks at the human–computer interface also run in the opposite direction—a phenomenon known as "algorithmic aversion."[155] More specifically, humans that do not trust or understand the technology might seek to compensate for actual or perceived AI failures and shortfalls.

Even if well intended, these human compensations may be unwarranted, unfair, or illegal in some settings.[156] For instance, if an AI system used for government hiring exhibits bias toward men, to what extent can or should the hiring official upwardly adjust the algorithmic score for women?[157] Likewise, if an AI risk-assessment system used for bail determinations exhibits bias against Black defendants, to what extent can or should judges ignore or discount AI recommendations for Black (or White) defendants?[158] As these examples lay bare, correcting for

153. *See* Kate Goddard et al., *Automation Bias: Empirical Results Assessing Influencing Factors*, 83 INT'L J. MED. INFORMATICS 368, 368–69 (2014); Citron, *supra* note 32, at 1271–72.

154. Moreover, in the long run, systemic automation bias can have atrophying effects on domain expertise and human judgment. *See* ACUS REPORT, *supra* note 1, at 8 ("Managed poorly, government deployment of AI tools can hollow out the human expertise inside agencies with few compensating gains, widen the public-private technology gap, increase undesirable opacity in public decision-making, and heighten concerns about arbitrary government action and power.").

155. *Cf.* Berkeley J. Dietvorst et al., *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCH. 114, 114 (2015) (finding that "people are especially averse to algorithmic forecasters after seeing them perform, even when they see them outperform a human forecaster").

156. Alice Xiang & Daniel E. Ho, *From Affirmative Action to Affirmative Algorithms: The Legal Challenges Threatening Progress on Algorithmic Fairness*, P'SHIP ON A.I.: BLOG (Nov. 9, 2020), https://www.partnershiponai.org/affirmativealgorithms [https://perma.cc/B7GA-X83X] ("The ways in which many of us in the AI community have moved to mitigate bias in the algorithms we develop may pose serious legal risks of violating equal protection.").

157. *Compare* Kim, *supra* note 33, at 191 (arguing that, "despite its limitations, auditing for discrimination should remain an important part of the strategy for detecting and responding to biased algorithms," and moreover, that "the law permits the use of auditing to detect and correct for discriminatory bias."), *with* Kroll et al., *supra* note 12, at 694–95 (expressing skepticism about auditing as a strategy for detecting and correcting algorithmic bias, on both technological and legal grounds).

158. *Cf.* Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/JRR9-5D29] (finding that one system erroneously categorized Black defendants as future criminals at nearly twice the rate as it did for White defendants); Sam Corbett-Davies et al., *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear.*, WASH. POST (Oct. 17, 2016, 5:00 AM), https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/ [https://perma.cc/7M7V-GPKL] (countering that the problem ProPublica identified is "mathematically guaranteed" given historical data showing disparate recidivism rates for Black and White offenders combined with a particular definition of fairness).

perceived or actual unfairness in algorithmic governance requires a mix of legal, technical, social, and pragmatic considerations—none of which are settled.[159]

Another source of algorithmic bias stems from cultural and experiential blind spots in the technology industry. According to recent studies, only 26% of computing related jobs are held by women at the leading technology firms,[160] and the share of technical workers who are Black sits below 4%.[161] While the causes of these diversity challenges are complex and contestable, the consequences are widely acknowledged: demographic hegemony "affects how AI companies work, what products get built, who they are designed to serve, and who benefits from their development."[162]

Facial recognition technology is perhaps the most notorious manifestation of these sociotechnical challenges.[163] In 2018, pioneering work by Joy Buolamwini and Timnet Gerbu demonstrated that three prominent facial recognition systems performed significantly worse on people of color, especially women of color.[164] The disparities were not intentional; the AI models were optimized to fit the training data of predominantly White faces. This revelation prompted additional studies,

---

159. *See, e.g.*, Engstrom & Ho, *supra* note 37, at 806 (finding it "far from certain" that current doctrine "will resolve the most pressing cases" in algorithmic governance); Huq, *supra* note 33, at 1917–27 (discussing the difficulties that arise in transposing the equal protection doctrine to the machine learning context). For additional sources and viewpoints on these issues, see *supra* note 33.

160. Sam Daley, *Women in Tech Statistics Show the Industry Has a Long Way to Go*, BUILT IN (May 5, 2021), https://builtin.com/women-tech/women-in-tech-workplace-statistics [https://perma.cc/N4VF-SZ4Q].

161. Michael Ellison, *This Is How Big Tech Is Failing Its Black Employees*, FAST CO. (Oct. 21, 2020), https://www.fastcompany.com/90565387/why-big-techs-lofty-diversity-reports-fell-so-far-from-expectations [https://perma.cc/VP4S-8JP4] ("The share of technical workers who are Black at Facebook, Google, and Microsoft has inched up less than one percentage point since 2014 and still sits below 4% at each company.").

162. West et al., *supra* note 75, at 5; *see also* RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE 4 (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").

163. There are many more examples of this problem in hiring, lending, medicine, and beyond. *See, e.g.*, Xiang, *supra* note 119, at 17–18 (discussing biased algorithms and applications in hiring and healthcare); Ted Knutson, *AI Lending Discrimination Needs To Be Tackled with Legislation Says House Financial Services Chair*, FORBES (May 7, 2021, 2:29 PM), https://www.forbes.com/sites/tedknutson/2021/05/07/ai-lending-discrimination-needs-to-be-tackled-with-legislation-says-house-financial-services-chair [https://perma.cc/XAR9-Z4VZ] (discussing discrimination in AI systems used for lending decisions).

164. Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. OF MACH. LEARNING RSCH. 1, 1 (2018); *see also Overview of Gender Shades Project*, MIT MEDIA LAB, https://www.media.mit.edu/projects/gender-shades/overview/ [http://perma.cc/AFB3-GHVF].

including by the National Institute of Standards and Technology (NIST), which found that several facial recognition systems—including those used for law enforcement—were exponentially more likely to misidentify people of color.[165]

Algorithmic bias can be mitigated with better data practices, technical patches, and workarounds.[166] But even if the technology can be made perfectly accurate, that would not address structural concerns relating to power, autonomy, and liberty.[167] For example, the government's ability to engage in sprawling and accurate digital surveillance says nothing about whether that capability should be wielded, for what purposes, in what contexts, and over what communities or subpopulations.[168]

Moreover, even when working as intended, AI models are statistical simplifications that necessarily treat people as members of groups, not as individuals.[169] That might be of less consequence or concern when an algorithm recommends songs based on group characteristics. But when deployed in high-stakes or sensitive government contexts, algorithmic stereotyping can undermine a person's sense of autonomy, unfairly

165. Drew Harwell, *Federal Study Confirms Racial Bias of Many Facial-Recognition Systems, Casts Doubt on Their Expanding Use*, WASH. POST (Dec. 19, 2019), https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/ [https://perma.cc/AK25-NYNT] ("Asian and African American people were up to 100 times more likely to be misidentified than white men, depending on the particular algorithm and type of search."); *see also* Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots*, ACLU (July 26, 2018, 8:00 AM), https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28 [https://perma.cc/8WCR-42J6] (reporting that Amazon's facial recognition system wrongly matched 28 members of Congress to criminal mug shots).

166. In 2015, for example, Google's image recognition system was found to classify African-Americans as "gorillas." Tom Simonite, *When It Comes to Gorillas, Google Photos Remains Blind*, WIRED (Jan. 11, 2010, 7:00 AM), https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/ [https://perma.cc/S2HE-RRNZ]. Google "fixed" the problem by removing gorillas from the service's lexicon. *Id.* This workaround is highly revealing of the *technical* challenges of retrospective solutions to algorithmic bias, as well as the *social* challenges of grappling with structural bias.

167. *See* West et al., *supra* note 75, at 18 ("[T]hough improving the performance of AI systems might be a necessary step toward making them more inclusive, there are some contexts in which 'fixing' such inaccuracies may not fix the overall problems presented by such systems—and some problems that cannot be fixed by a technical solution at all.").

168. *See id.*; Julia Powles & Helen Nissenbaum, *The Seductive Diversion of "Solving" Bias in Artificial Intelligence*, ONEZERO (Dec. 7, 2018), https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53 [https://perma.cc/TT4E-4JZC].

169. *See* Mulligan & Bamberger, *supra* note 37, at 787.

deprive them of benefits and opportunities,[170] or affirmatively cause harm.[171]

3. Transparency

AI transparency can be frustrated by a host of technical and non-technical reasons.[172] Regardless of the cause, the opacity of AI systems can be highly problematic. Most basically, if stakeholders do not understand how an AI system works, they may not understand when or why it fails. Without transparency, moreover, the government may lack moral or legal justification to act upon an AI model's outputs.[173] "[T]he algorithm made me do it" will simply not do in contexts where an individual's rights or well-being are compromised.[174]

---

170. This sense of unfairness was on full display after the British government cancelled the annual "A-Level" qualification exams for university placements due to the COVID-19 pandemic. Bryan Walsh, *How an AI Grading System Ignited a National Controversy in the U.K.*, AXIOS (Aug. 19, 2020), https://www.axios.com/england-exams-algorithm-grading-4f728465-a3bf-476b-9127-9df036525c22.html [https://perma.cc/298X-F8KA]; *see* Kelsey Piper, *The UK Used a Formula to Predict Students' Scores for Canceled Exams. Guess Who Did Well.*, VOX (Aug. 22, 2020, 7:30 AM), https://www.vox.com/future-perfect/2020/8/22/21374872/uk-united-kingdom-formula-predict-student-test-scores-exams [https://perma.cc/6H7D-RYFU]. As a "stand-in for actual scores," an AI system predicted how students would have performed on the exam. *Id.* Because the algorithm placed significant weight on the past performance of the students' schools, students "lost the chance to be treated as individuals." Walsh, *supra*. Moreover, because many of the lower performing schools were also less affluent, students lost the opportunity to outscore their predicted performance and earn a place in more affluent learning institutions. *Id.*; *see also Exam Algorithms: Some Lessons*, FTI CONSULTING (Aug. 21, 2020), https://www.fticonsulting.com/emea/insights/articles/exam-algorithms-some-lessons [https://perma.cc/H92M-8KRU].

171. *See* Reva Schwartz et al., *Draft NIST Special Publication 1270: A Proposal for Identifying and Managing Bias in Artificial Intelligence*, NAT'L INST. STANDARDS & TECH. 9 (June 2021), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270-draft.pdf [https://perma.cc/H3ZF-L6GJ] ("[AI] tools that are designed to use aggregated data about groups to make predictions about individual behavior—a practice initially meant to be a remedy for non-representative datasets—can lead to biased outcomes.").

172. *See* Jenna Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y, Jan.–June 2016, at 1, 3–5.

173. *See* Mulligan & Bamberger, *supra* note 37, at 782 (arguing that "the policy choices embedded in system design fail the prohibition against arbitrary and capricious agency actions absent a reasoned decision-making process"); *see also supra* note 47; *infra* notes 267–268 and accompanying text (discussing possible legal implications of model inscrutability for constitutional and administrative law).

174. *See* FAIRNESS, ACCOUNTABILITY & TRANSPARENCY MACH. LEARNING, www.fatml.org [https://perma.cc/EE2L-SNSY] ("[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), https://www.aclu.org/news/privacy-technology/wrongfully-arrested-because-face-recognition-cant-tell-black-people-apart/ [https://perma.cc/6AL9-ENJ8] (following the arrest of a Black man that was misidentified by a facial recognition system, "[o]ne officer responded, 'The computer must have gotten it wrong.'").

Some AI models *cannot* be explained because of their complexity.[175] Other AI models *will not* be explained because they don't have to be.[176] Most pertinent here, federal agencies that procure AI solutions may not be privy to the vendor's trade secrets—for example, when the technology is acquired as a "commercial item off the shelf."[177] Even if the government has access to vendor trade secrets, federal law or nondisclosure agreements may prevent disclosure of those secrets to litigants, lawmakers, or other stakeholders.[178]

Moreover, AI transparency may be self-defeating or dangerous in many government contexts. For example, full transparency of the input variables and source code of a fraud-detection system might facilitate "gaming" or "hacking" by adversarial actors.[179] For similar reasons, cybersecurity concerns may trump AI transparency in safety-critical domains, such as energy, transportation, telecommunication, voting systems, waterways, and more.[180] Finally, AI transparency may run

175. *See supra* notes 101–04 and accompanying text.

176. *See* PASQUALE, *supra* note 99, at 180–81; Sonia K. 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"); Brauneis & Goodman, *supra* note 18, 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").

177. *See infra* Section IV.D.1 (discussing commercial off-the-shelf acquisitions); *see also* 48 C.F.R. § 12.212 (2021) (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").

178. Katherine Fink, *Opening the Government's Black Boxes: Freedom of Information and Algorithmic Accountability*, 21 INFO. COMMC'N & SOC'Y 1453, 1456–59 (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); Wexler, *supra* note 35, at 1396–417 (describing and critiquing how trade secrecy has been used to prevent criminal defendants to gain access to information about AI risk-assessment tools used in the criminal justice system).

179. *See, e.g.*, Engstrom & Ho, *supra* note 117, at 68 (discussing the risk of gaming the suite of tools under development at the U.S. Patent and Trademark Office to help examiners adjudicate patent applications); Leslie, *supra* note 71, at 32–34 (discussing a range of adversarial attacks and failure modes); NSCAI FINAL REPORT, *supra* note 11, at 47 (same). *But cf.* Ignacio N. Cofone & Katherine J. Strandberg, *Strategic Games and Algorithmic Secrecy*, 64 MCGILL L.J. 623, 623 (2019) (detailing the relationship between gaming and transparency, and arguing that the threat of gaming is overblown in many contexts and often addressable in ways that do not require secrecy).

180. *See, e.g.*, David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135, 135 (2007) (describing and critiquing how government outsourcing creates transparency and accountability gaps around critical public infrastructures); Brauneis & Goodman, *supra* note 18, at 175–76; *see also* Exec. Order No. 14,028, 86 Fed. Reg.

headlong into privacy laws in a wide variety of contexts where the government is provided personal data.[181] This tension arises, for example, when the inputs or outputs of AI systems contain protected or sensitive information that can be traced to individuals (even when names and other identifying information are scrubbed from the data).[182]

In short, AI transparency is undercut by a mix of technical, commercial, and legal issues that coalesce to keep much of algorithmic governance in the dark. Whether justifiably so is a matter of debate—normatively, doctrinally, and contextually.[183]

26,633, 26,633 (May 17, 2021) ("[T]he trust we place in our digital infrastructure should be proportional to how trustworthy and transparent that infrastructure is, and to the consequences we will incur if that trust is misplaced.").

181. *See, e.g.*, 5 U.S.C. § 552a(b) (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 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C.) (setting forth privacy and security standards for protecting personal health information); *see also* Engstrom & Ho, *supra* note 117, at 65 (explaining that "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").

182. *See* KEARNS & ROTH, *supra* note 74, at 30–33 (discussing how data can be extracted from AI models and de-anonymized); Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, 2008 IEEE SYMPOSIUM ON SECURITY AND PRIVACY 111. This is a major concern, especially because 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 [https://perma.cc/GW6H-9TNQ]; 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 [https://perma.cc/FF2P-NBQB].

183. *See, e.g.*, Coglianese & Lehr, *supra* note 37, at 40–49 (arguing that government use of AI can generally comport with constitutional due process, as well as administrative law's reason-giving and transparency norms); Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1273, 1295–306 (2020) (exploring the "procedural and substantive conflicts between proprietary [algorithmic] decision-making on the one hand and government transparency obligations under the First Amendment and [Freedom of Information Act] on the other"); Citron, *supra* note 32, at 1281–88 (discussing how agency use of automated systems raises due process concerns); Mulligan & Bamberger, *supra* note 37, at 782 (arguing that "policy choices embedded in system design fail the prohibition against arbitrary and capricious agency actions absent a reasoned decision-making process that enlists the expertise necessary for reasoned deliberation, provides justifications for such choices, makes visible the political choices being made, and permits iterative human oversight and input"). *See also* Brauneis & Goodman, *supra* note 18, at 152–63 (spotlighting the transparency deficits that accrue when state and local government adopt AI systems developed by third parties).

4. Accountability

The foregoing transparency challenges have major implications for government accountability. The less stakeholders know, the more difficult it becomes to ascertain whether an AI system is being used, and if so, whether that use is properly authorized, justified, and legal. As of this writing, there is no publicly available register of AI systems currently used by which federal actors, for what purposes, from what sources, and under what authority.[184] This is highly problematic for two related reasons: first, the opacity shutters out stakeholder input; second, the opacity breeds public distrust around the government's use of AI systems (including benign and potentially beneficial uses).[185]

Judicial review is another way that our legal system might hold government actors accountable for their use of AI systems. Conceivably, courts could also hold government actors accountable for the technical and non-technical value judgments embedded in or emanating from AI systems. The opacity of AI systems, however, can stymie a court's ability to perform these functions.

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.[186] As the Government Accountability Office (GAO) acknowledged in a 2021 report: "The U.S. government, industry leaders, professional associations, and others have begun to develop principles and frameworks to address [transparency and fairness] concerns, but there is limited information on how these will be implemented to allow for third-party assessments and audits of AI systems."[187]

Lines of accountability, moreover, are frequently tangled because AI systems are assemblages of datasets, technology stacks, and complex human networks.[188] When things go wrong, it can be far from clear which

184. *See* Rubenstein, *supra* note 24, at 20–21 (urging the creation of a federal registry of federal AI use cases). A 2020 executive order calls for such a catalog. *See* Exec. Order No. 13,960, 85 Fed. Reg. 78,939, 78,941 (Dec. 8, 2020). A few European cities have launched AI registries, with more jurisdictions likely to follow this type of proactive disclosure. *See* Khari Johnson, *Amsterdam and Helsinki Launch Algorithm Registries to Bring Transparency to Public Deployments of AI*, VENTURE BEAT (Sept. 28, 2020, 11:41 AM), https://venturebeat.com/2020/09/28/amsterdam-and-helsinki-launch-algorithm-registries-to-bring-transparency-to-public-deployments-of-ai/ [https://perma.cc/E5Y8-YJ2F].

185. *See* Rubenstein, *supra* note 24, at 14–15; Schwartz et al., *supra* note 171, at 5 ("A consistent finding in the literature is the notion that trust can improve if the public is able to interrogate systems and engage with them in a more transparent manner.").

186. *See* Brauneis & Goodman, *supra* note 18, at 159 (expounding on this concern); Katyal, *supra* note 176, at 1259 (same).

187. U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-519SP, *supra* note 58, at 9–10.

188. The use of "open" data and source code in AI system is common, and double-edged.

actors and institutions (if any) should be held accountable or to what extent.

C. *The Rise of Ethical AI*

The foregoing challenges around AI safety, fairness, transparency, and accountability are beginning to register in social and political discourse. This reckoning may be credited to a series of high-profile AI episodes that do not require technical savvy to appreciate. In particular, the 2018 media coverage of the Cambridge Analytica–Facebook scandal was a watershed moment that exposed how AI ecosystems secretly exploit consumer data for commercial and political ends.[189]

The news was hardly surprising to a small cadre of academics, journalists, and industry insiders who—years prior—had foretold the dangers of digital surveillance and the power of AI to shape human behaviors.[190] When the Cambridge Analytica–Facebook scandal broke, however, "techlash" went mainstream.[191] As just one measure, only a

---

A.I. Now Inst., *A New AI Lexicon: OPEN*, MEDIUM (July 12, 2021), https://medium.com/a-new-ai-lexicon/a-new-ai-lexicon-open-3ec7daa300a [https://perma.cc/H33H-99EH?type=image] ("Despite the potential benefits of open data, there has been little research or discussion on the assumptions and applications of open data in the context of AI technologies, specifically how data is collected and made available.").

189. *See* Alvin Chang, *The Facebook and Cambridge Analytica Scandal, Explained with a Simple Diagram*, VOX (May 2, 2018, 3:25 PM), https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram [https://perma.cc/2X9K-VGKQ] (discussing the shutdown of a political consulting firm that harvested user data from Facebook); Alex Hern, *Cambridge Analytica: How Did It Turn Clicks into Votes?*, GUARDIAN (May 6, 2018, 3:00 PM), https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie [https://perma.cc/CB57-RPRE].

190. *See, e.g.*, WOLFIE CHRISTL & SARAH SPIEKERMANN, NETWORKS OF CONTROL: A REPORT ON CORPORATE SURVEILLANCE, DIGITAL TRACKING, BIG DATA & PRIVACY 7 (2016) ("While the media and special interest groups are aware of these developments for a while now, we believe that the full degree and scale of personal data collection, use and—in particular—abuse has not been scrutinized closely enough."); PASQUALE, *supra* note 99, at 8–10; O'NEIL, *supra* note 18, at 13; WENDY HUI KYONG CHUN, CONTROL AND FREEDOM 1 (2008); Citron, *supra* note 32, at 1262; *see also* Shoshana Zuboff, *Surveillance Capitalism and the Challenge of Collective Action*, NEW LAB. F. (Jan. 24, 2019), https://journals.sagepub.com/doi/full/10.1177/1095796018819461 [https://perma.cc/V5P6-PEQE].

191. Matthew Le Bui & Safiya Umoja Noble, *We're Missing a Moral Framework of Justice in Artificial Intelligence: On the Limits, Failings, and Ethics of Fairness*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 163–67 (Markus D. Dubber et al. eds., 2020) (connecting the rise of techlash to the Cambridge Analytica—Facebook scandal); *see also* Rana Foroohar, *Year in a Word: Techlash*, FIN. TIMES (Dec. 16, 2018), https://www.ft.com/content/76578fba-fca1-11e8-ac00-57a2a826423e [https://perma.cc/HW8D-HG93] (defining "[t]echlash" as the "growing public animosity towards large Silicon Valley platform technology companies and their Chinese equivalents").

handful of AI-related bills were pending in Congress in 2017.[192] Since then, more than 100 distinct pieces of AI-related bills have been introduced in Congress.[193] This trend is paralleled in U.S. state and local jurisdictions (and across the globe).[194]

1. Ethical AI in Industry

Ethical AI was not unheard of in 2017.[195] But its embrace as an industry *movement* occurred in 2018.[196] By 2019, a spate of ethical AI frameworks were promulgated or adopted by technology firms, trade groups, and non-government organizations.[197] While the particulars vary, ethical AI principles generally coalesce around a set of values relating to safety, fairness, transparency, accountability, privacy, and human well-being.[198]

---

192. *See* STAN. INST. FOR HUMAN-CENTERED A.I., ARTIFICIAL INTELLIGENCE INDEX REPORT 172 (2021); *see also* Yoon Chae, *U.S. AI Regulation Guide: Legislative Overview and Practical Considerations*, 3 J. ROBOTICS, A.I. & L. 17, 17 (2020) (reporting that from 2015–2016, only two bills were introduced that contained the term "artificial intelligence," which increased to fifty-one bills by the end of 2019).

193. *See* STAN. INST. FOR HUMAN-CENTERED A.I., *supra* note 192, at 172*; see also AI Legislation Tracker–United States*, CTR. FOR DATA INNOVATION (June 19, 2020), https://www.datainnovation.org/ai-policy-leadership/ai-legislation-tracker/ [https://perma.cc/E2U7-LEBQ].

194. *See Legislation Related to Artificial Intelligence*, NAT'L CONF. OF STATE LEGISLATURES (Apr. 16, 2021), https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx [https://perma.cc/T9US-2FWC] (tracking state AI-related legislation); *State Facial Recognition Policy*, ELEC. PRIV. INFO. CTR., https://epic.org/state-policy/facialrecognition/ [https://perma.cc/UZL5-XLFL] (tracking state and local laws pertaining to facial recognition); *National AI Policies & Strategies*, OECD.AI (2021), https://oecd.ai/en/dashboards [https://perma.cc/AVX7-N75Z] (tracking global AI policies and strategies).

195. In 2016 and 2017, Apple, Amazon, Google, Facebook, Microsoft, IBM, joined to form the Partnership for Artificial Intelligence to Benefit People and Society. *See* Hern, *supra* note 43; James Vincent, *Apple Joins Research Group for Ethical AI with Fellow Tech Giants*, VERGE (Jan. 27, 2017, 7:02 AM), https://www.theverge.com/2017/1/27/14411810/apple-joins-partnership-for-ai [https://perma.cc/6R2V-HFA5]. Then, as now, the consortium's express purpose is to develop industry best practices for promoting "ethics, fairness and inclusivity; transparency, privacy, and interoperability; collaboration between people and AI systems; and the trustworthiness, reliability and robustness of the technology." *See* Hern, *supra* note 43.

196. *Cf.* STAN. INST. FOR HUMAN-CENTERED A.I., *supra* note 192, at 129 ("In terms of rolling out ethics principles, 2018 was the clear high-water mark for tech companies—including IBM, Google, and Facebook . . . .").

197. *Id.* at 129–30; JESSICA FJELD ET AL., PRINCIPLED ARTIFICIAL INTELLIGENCE: MAPPING CONSENSUS IN ETHICAL AND RIGHTS-BASED APPROACHES TO PRINCIPLES FOR AI, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y (2020), http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420 [https://perma.cc/2A6N-N2HX].

198. *See generally* Jobin et al., *supra* note 40 (mapping and analyzing the corpus of principles and guidelines on ethical AI); FJELD ET AL., *supra* note 197.

The motivations driving the ethical AI movement are ideological and instrumental. No doubt, altruism and corporate social responsibility are playing a part.[199] Just as surely, ethical AI is a political pitch to forestall government regulation,[200] and a market pitch to placate consumers and investors.[201] But it must also be appreciated that ethical AI is a grassroots movement, curated and cultivated in significant part by the high-skilled, and highly in-demand, technology workforce.[202]

In 2018, thousands of technologists signed open letters and staged headline-generating protests, urging corporate leaders to end law enforcement and military contracts with the government.[203] Heeding the

199. *See* Andrew Charlesworth, *Regulating Algorithmic Assemblages: Looking Beyond Corporatist AI Ethics*, *in* DATA-DRIVEN PERSONALISATION IN MARKETS, POLITICS AND LAW 243, 245–46 (Uta Kohl & Jacob Eisler eds., 2021) (linking the proliferation of ethical AI frameworks in the technology industry to the corporate social responsibility movement). For a discussion of some of these initiatives, see JESSICA CUSSINS NEWMAN, DECISION POINTS IN AI GOVERNANCE (2020); Kathy Baxter, *Ethical Frameworks, Tool Kits, Principles, and Oaths—Oh My!*, SALESFORCE (Oct. 19, 2020), https://blog.einstein.ai/frameworks-tool-kits-principles-and-oaths-oh-my/ [https://perma.cc/898E-KLTQ]. For examples of ethical AI toolkits, see *AI Fairness 360*, IBM RSCH. TRUSTED AI, https://aif360.mybluemix.net/ [https://perma.cc/EU5N-3J7S]; SARAH BIRD ET AL., FAIRLEARN: A TOOLKIT FOR ASSESSING AND IMPROVING FAIRNESS IN AI 1 (2020), https://www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_WhitePaper-2020-09-22.pdf [https://perma.cc/77T6-JJAJ]; Rachel K. E. Bellamy et al., AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. (Oct. 3, 2018) (unpublished manuscript), https://arxiv.org/abs/1810.01943 [https://perma.cc/Q4ZH-X4Z8]; Google Research, *What if Tool* (2019), https://pair-code.github.io/what-if-tool/ [https://perma.cc/8UMW-J8R8].

200. *See* Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 407–08 (2017) (noting that "ethically motivated 'self-commitments' can discourage policymakers from pursuing legally binding principles and constraints"); Orr & Davis, *supra* note 109, at 8 ("[O]rganizations and industry professionals have been careful to set their own standards to avoid control at the hands of non-expert forces."); *cf.* Charlesworth, *supra* note 199, at 245 ("The establishment of ethics boards, ethics oversight committees and codes of practice for AI by corporate entities follows a familiar regulatory pattern, well established in the technology sphere, whereby industries seek to head off formal governmental regulatory intervention by providing putatively self-regulatory mechanisms to address the problematic impacts of their services or corporate activities.").

201. *See* NEWMAN, *supra* note 199, at 14 (discussing how ethics committees are "viewed with some suspicion, and in some cases have been called out as 'AI ethics-washing'" (quoting Karen Hao, *In 2020, Let's Stop AI Ethics-Washing and Actually Do Something*, MIT TECH. REV. (Dec. 27, 2019))).

202. Nataliya Nedzhvetskava & JS Tan, *What We Learned From over a Decade of Tech Activism*, GUARDIAN (Dec. 23, 2019), https://www.theguardian.com/commentisfree/2019/dec/22/tech-worker-activism-2019-what-we-learned [https://perma.cc/M6H3-HQTX] (discussing the rise of activism within the technology industry).

203. *See* NEWMAN, *supra* note 199, at 18 ("A group called Microsoft Workers 4 Good, whose mission is 'to empower every worker to hold Microsoft accountable to their stated values,' has called on Microsoft leadership to end certain contracts."); Daisuke Wakabayashi & Scott Shane, *Google Will Not Renew Pentagon Contract That Upset Employees*, N.Y. TIMES (June 1, 2018),

call, Google declined to renew its contract with the Department of Defense (DoD) on Project Maven (which uses AI for drone strikes), and declined to compete for a major DoD cloud-computing contract (which was worth up to $10 billion).[204] More recently, in the Summer of 2020, three leading technology firms stopped selling facial recognition technology to law enforcement agencies.[205] In a telling letter to Congress, IBM explained that it "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 [IBM's] values and Principles of Trust and Transparency."[206]

More than ironic, these corporate displays of social responsibility are instructive here for three related reasons. First, the government is susceptible to techlash, including from technologists. Second, ethical AI speaks loudly in the market, and the government market is no exception. Third, public anxieties around AI systems will not be neatly cabined into government and commercial spheres. Nor should the polity draw sharp distinctions, given that the AI technologies used in the private sector are, by and large, the same technologies deployed for government functions.

2. Ethical AI in Government

In 2019, the White House issued an Executive Order that sketched an agenda for "[m]aintaining American leadership" in innovative and trustworthy AI.[207] Soon after, the United States joined with other global leaders to adopt a set of "[p]rinciples for responsible stewardship of

---

https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html [https://perma.cc/3PA7-XPVE] ("About 4,000 Google employees signed a petition demanding 'a clear policy stating that neither Google nor its contractors will ever build warfare technology.' A handful of employees also resigned in protest, while some were openly advocating the company to cancel the Maven contract.").

204. *See* Wakabayashi & Shane, *supra* note 203. In the wake of protracted litigation, the DoD cancelled the contract in 2021. *See* Kate Conger & David E. Sanger, *Pentagon Cancels a $10 Billion Technology Contract*, N.Y. TIMES (July 6, 2021, 12:52 PM), https://www.nytimes.com/2021/07/06/technology/JEDI-contract-cancelled.html [https://perma.cc/7W3S-PKBC].

205. *See* Jay Greene, *Microsoft Won't Sell Police Its Facial-Recognition Technology, Following Similar Moves by Amazon and IBM*, WASH. POST (June 11, 2020, 2:30 PM), https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/ [https://perma.cc/6CJ2-VDKJ]; *We Are Implementing a One-Year Moratorium on Police Use of Rekognition*, DAY ONE: AMAZON BLOG (June 10, 2020), https://blog.aboutamazon.com/policy/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition [https://perma.cc/Q2UF-5PZH].

206. Arvind Krishna, *IBM CEO's Letter to Congress on Racial Justice Reform*, IBM (June 8, 2020), https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/ [https://perma.cc/BD5G-5R8R].

207. Exec. Order No. 13,859, 84 Fed. Reg. 3967, 3967 (Feb. 14, 2019).

trustworthy AI."[208] These principles, promulgated by the Organization for Economic Cooperation and Development (OECD), were the first intergovernmental standards on AI.[209] Although the OECD framework is not binding on member states, the core ethical AI principles are beginning to germinate in U.S. policy.

In 2020, the DoD and U.S. Intelligence Community formally adopted ethical AI principals.[210] Later that year, the White House issued another Executive Order, with the eponymous aim of "Promoting the Use of Trustworthy [AI] in the Federal Government."[211] Like the foregoing ethical AI initiatives, this White House directive espouses principles relating to safety, fairness, transparency, and accountability.[212] Moreover, it instructs agencies to "design, develop, acquire, and use AI in a manner that exhibits due respect for our Nation's values and is consistent with the Constitution and all other applicable laws and policies, including those addressing privacy, civil rights, and civil liberties."[213]

Thus far, Congress has been slow to act on a multitude of pending AI-related bills.[214] However, the National Defense Authorization Act of 2021 provides an early glimpse of Congress's wide bipartisan support for responsible AI uses by government and industry alike.[215] Most pertinent here, the Act directs NIST to support the development of technical

208. *Recommendation of the Council on Artificial Intelligence*, OECD LEGAL INSTRUMENTS (May 21, 2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [https://perma.cc/6MAK-PGG2] (espousing (1) inclusive growth, sustainable development and well-being; (2) human-centered values and fairness; (3) transparency and explainability; (4) robustness, security and safety; and (5) accountability); *see also* Michael Kratsios, *White House OSTP's Michael Kratsios Keynote on AI Next Steps*, U.S. MISSION TO ORG. FOR ECON. COOP. & DEV. (May 21, 2019), https://usoecd.usmission.gov/white-house-ostps-michael-kratsios-keynote-on-ai-next-steps/ [https://perma.cc/2QZ8-L2MT] (discussing the principles adopted by the OECD).

209. *Recommendation of the Council on Artificial Intelligence*, *supra* note 208.

210. Press Release, U.S. Dep't of Def., DOD Adopts Ethical Principles for Artificial Intelligence (Feb. 24, 2020), https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence [https://perma.cc/X82V-SLHD]; *see also* Hicks, *supra* note 7, at 1 ("As the DoD embraces [AI], it is imperative that we adopt responsible behavior, processes, and outcomes in a manner that reflects the Department's commitment to its ethical principles, including the protection of privacy and civil liberties.").

211. *See* Exec. Order No. 13,960, 85 Fed. Reg. 78,939, 78,939 (Dec. 8, 2020).

212. *Id.* at 78,940–41.

213. *Id.* at 78,940. The executive order also establishes a common policy for implementing the principles, instructs agencies to catalog their AI use cases, and directs the General Services Administration and the Office of Personnel Management to enhance AI implementation expertise within the executive branch. *Id.* at 78,941–43.

214. *See supra* notes 191–92 and accompanying text.

215. Pub. L. No. 116–283; *see also Summary of AI Provisions from the National Defense Authorization Act 2021*, STAN. INST. FOR HUMAN-CENTERED A.I., https://hai.stanford.edu/policy/policy-resources/summary-ai-provisions-national-defense-authorization-act-2021 [https://perma.cc/LB64-QVWF] (providing a summary of AI-related provisions from the Act).

standards, guidelines, and risk-management frameworks to promote "trustworthy" AI systems.[216] The Act also creates a new National AI Initiative Office (tasked with coordinating federal AI activities and supporting AI research),[217] and a National AI Advisory Committee (which will advise the President on a range of matters pertaining to ethical AI, including the use of facial recognition by law enforcement authorities).[218] Undoubtedly, this is just the opening salvo of much more federal policymaking and oversight to come.

III. FROM PRINCIPLES TO PRACTICE

The proliferation of ethical AI principles, along with the government's high-level support, are generally viewed as steps in the right direction.[219] As this Part explains, however, ethical AI is easier said than done. Most assuredly, the extant frameworks and declarations of ethical AI do not address, much less resolve, an open set of challenges and tradeoffs at the fulcrum of law, society, and technology.[220] This Part spotlights those tensions and their implications for algorithmic governance.

A. *The Gap Between Ethical AI Principles and Practice*

This first Section homes in on the impediments to ethical AI in workaday practice. The challenges manifest somewhat differently within industry and across the public/private divide. To tease out some of those differences, the discussion begins with industry before turning to the government. This ordering tracks reality: the government's AI journey is effectively bootstrapped to industry, and the government's ethical AI challenges are mostly derivative.

---

216. National Defense Authorization Act of 2021 § 5301.

217. *Id.* §§ 5101–03.

218. *Id.* § 5104.

219. To say that these are steps in the right direction is not to say they are sufficient. *See* Lee Rainie et al., *Experts Doubt Ethical AI Design Will Be Broadly Adopted as the Norm Within the Next Decade*, PEW RSCH. CTR. (June 16, 2021), https://www.pewresearch.org/internet/2021/06/16/experts-doubt-ethical-ai-design-will-be-broadly-adopted-as-the-norm-within-the-next-decade/ [https://perma.cc/3ST9-789D]; JONATHAN ROTNER, HOW CAN ETHICS MAKE BETTER AI PRODUCTS? 6 (2020) ("Skeptics might see declarations, frameworks, and toolkits as virtue signaling, resulting in words without action." (footnote omitted)).

220. *See infra* Section III.A.1 (discussing a range of challenges and gaps between ethical AI principles and practice); Raji et al., *supra* note 96, at 2 ("The AI industry lacks proven methods to translate principles into practice."); *see also infra* Section III.A.2 (discussing the government's reliance on industry for translating ethical AI principles into practice).

1. Industry Challenges

To start, the voluntary nature of ethical AI allows competing market incentives to dominate.[221] For instance: if the choice is between algorithmic auditing and rushing a product to market, many if not most firms will choose the latter. To be clear, some firms may choose ethical AI principles over profits and growth. But most firms don't because they don't have to.

The principles-to-practice challenge is exacerbated by the AI ecosystem's distributed network of responsibilities and domain expertise.[222] For example, data scientists and software engineers may not anticipate or feel responsible for the social impacts of their digital creations.[223] Meanwhile, social scientists and lawyers may not appreciate the technical challenge of translating nebulous concepts like fairness and nondiscrimination into code.[224] More generally, "the ethical development and deployment of AI systems typically involves decisions that no individual practitioner can make on their own."[225] Consequently, the amount of influence or responsibility that any individual has might be preempted or superseded by others in the AI pipeline. For example,

---

221. *See* Charlesworth, *supra* note 199, at 2 ("'AI ethics' is far removed from 'AI law' and broadly captures the idea of self-policing by private corporate actors in their use of AI systems, as sanctioned by government."); Michael A. Madaio et al., *Co-Designing Checklists to Understand Organizational Challenges and Opportunities Around Fairness in AI*, *in* CHI '20: PROCEEDINGS OF THE 2020 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 1, 10 (2020) ("[O]rganizational culture typically prioritizes 'moving fast' and shipping products over pausing to consider fairness."); *see also* Schwartz et al., *supra* note 171, at 4 ("Often a technology is not tested—or not tested extensively—before deployment, and instead deployment may be used as testing for the technology."); Emanuel Moss & Jacob Metcalf, *The Ethical Dilemma at the Heart of Big Tech Companies*, HARV. BUS. REV., Nov. 14, 2019, https://hbr.org/2019/11/the-ethical-dilemma-at-the-heart-of-big-tech-companies [https://perma.cc/ B95N-EDGC] (highlighting the tension between the race to market and the race to ethical AI).

222. *See, e.g.*, Kenneth Holstein et al., *Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?*, *in* CHI 2019: PROCEEDINGS OF THE 2019 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, at 5–12 (2019), https://arxiv.org/pdf/1812.05239.pdf [https://perma.cc/HV4D-Q8AC] (assessing the practical needs of private sector AI practitioners in relation to ethical AI); Orr & Davis, *supra* note 109, at 10 ("[P]ractitioners play the part of (highly skilled) technicians, rather than morally autonomous agents."); *see also* Michael Veale et al., *Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making*, *in* CHI 2018: PROCEEDINGS OF THE 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, at 3–7 (2018), https://arxiv.org/pdf/1802.01029.pdf [https://perma.cc/8FPA-A8GW] (surveying public sector AI practitioners and cataloguing challenges they face achieving fairness standards).

223. *See* Orr & Davis, *supra* note 109, at 9 (describing the disconnect between practitioners and those that commission their work).

224. *See* KEARNS & ROTH, *supra* note 74, at 18 ("[T]he first challenge in asking an algorithm to be fair or private is agreeing on what those words should mean . . . in so precise a manner that they can be 'explained' to a machine.").

225. *See* Madaio et al., *supra* note 221, at 2.

system engineers may have no control over the decisions of subject matter experts, and vice versa. Sales representatives may be informed about system deficiencies but bury those problems in market pitches. Last, but not least, corporate leadership can ignore, marginalize, or terminate ethical AI champions that do not tow the company bottom line.[226]

Even under the right conditions, ethical AI frameworks are too generalized to resolve more specific issues that commonly arise in practice.[227] When ethical AI principles collide, the problem can be particularly acute.[228] For example, ethical AI frameworks generally do not resolve conflicts between AI safety and transparency, transparency and accuracy, accuracy and fairness, and so on.[229] Nor do the frameworks resolve incoherencies within specific principles.[230] Fairness, for example, has upwards of a dozen formulations in the AI field.[231] For virtually all

226. Unfortunately, the marginalization of ethical AI voices within industry is reportedly common. *See* Madaio et al., *supra* note 221, at 5 (reporting that "[i]ndividual advocates [for AI fairness] face both sociocultural barriers to speaking up and structural barriers to having their teams address AI fairness issues"); *see also id.* at 6 ("[T]he disconnect arising from rhetorical support for AI fairness efforts coupled with a lack of organizational incentives that support such efforts is a central challenge for practitioners."). Recently, Google made national headlines when it ousted the co-leads of its ethical AI team, Timnet Gebru and Melanie Mitchell. *See Google to Change Research Process After Uproar Over Scientists' Firing*, GUARDIAN (Feb. 26, 2021, 2:32 PM), https://www.theguardian.com/technology/2021/feb/26/google-timnit-gebru-margaret-mitchell-ai-research [https://perma.cc/A7KL-Z737]. This was an especially shocking display of capitalism cancelling ethical AI. *See* Alex Hanna & Meredith Whitaker, *Timnit Gebru's Exit from Google Exposes a Crisis in AI*, WIRED (Dec. 31, 2020, 7:00 AM), https://www.wired.com/story/timnit-gebru-exit-google-exposes-crisis-in-ai/?redirectURL=https%3A%2F%2Fwww.wired.com%2Fstory%2Ftimnit-gebru-exit-google-exposes-crisis-in-ai%2F [https://perma.cc/2DFG-5RZD]; Tom Simonite, *What Really Happened When Google Ousted Timnit Gebru*, WIRED (June 8, 2021, 6:00 AM), https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/ [https://perma.cc/F2BG-YAN7] (providing an in-depth, behind-the-scenes account).

227. Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 NATURE MACH. INTEL. 501, 504 (2019) ("Norms and requirements cannot be deduced directly from mid-level principles without accounting for specific elements of the technology, application, context of use, or relevant local norms."); Raji et al., *supra* note 96, at 2 (noting that "AI principles have been criticized for being vague and providing little to no means of accountability").

228. *Cf.* Mittelstadt, *supra* note 227, at 504; Madaio et al., *supra* note 221, at 2 ("AI ethics principles can fail to achieve their intended goal if they are not accompanied by other mechanisms for ensuring that practitioners make ethical decisions.").

229. Jess Whittlestone et al., *The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions*, *in* AIES '19: PROCEEDINGS OF THE 2019 AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY 196–97 (2019), https://dl.acm.org/doi/pdf/10.1145/3306618.3314289 [https://perma.cc/NL4A-9BKA].

230. *See id.*

231. *See* Arvind Narayanan, *Translation Tutorial: 21 Fairness Definitions and Their Politics*, YOUTUBE (Mar. 1, 2018), https://www.youtube.com/watch?v=jIXIuYdnyyk [https://perma.cc/ZX3Q-GJV6]; *see also* Jacobs & Wallach, *supra* note 145, at 382–83 (discussing and distinguishing conceptions of "individual fairness" and "group fairness").

data distributions, however, it is mathematically impossible to simultaneously satisfy the three most commonly used fairness metrics.[232] Likewise, transparency and accountability are flexible ideals; there are different conceptions, expressions, and degrees of each.[233] Ethical AI frameworks could be more prescriptive and precise (and arguably should be). The claim here, however, is purely descriptive: ethical AI practice is unavoidably noisy and uneven because of the generality and incoherence of the frameworks themselves.

To some extent, AI systems can be ethically designed at inception with disciplined procedures and protocols.[234] The curation of "datasheets for datasets,"[235] "model cards for model reporting,"[236] and "fairness checklists"[237] are examples of responsible design practices. Moreover,

232. *See* Jon Kleinberg et al., *Inherent Trade-offs in the Fair Determination of Risk Scores*, *in* PROCEEDINGS OF 8TH INNOVATIONS IN THEORETICAL COMPUTER SCIENCE CONFERENCE (2016), https://arxiv.org/pdf/1609.05807.pdf [https://perma.cc/YHZ5-XSD4]; *see also* Kailash Karthik Saravanakumar, *The Impossibility Theorem of Machine Fairness: A Casual Perspective* (Jan. 29, 2021) (preprint), https://arxiv.org/pdf/2007.06024.pdf [https://perma .cc/PK9E-3R2Y]; KEARNS & ROTH, *supra* note 74, at 85 ("There are certain combinations of fairness criteria that—although they are each individually reasonable—simply cannot be achieved simultaneously, even if we ignore accuracy considerations.").

233. *See, e.g.*, Engstrom & Ho, *supra* note 117, at 61 (discussing different types and conceptions of transparency in the AI literature); Deven R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 9–11 (2017) (comparing "technical accountability" to its legal and political forms).

234. *See* INST. OF ELEC. & ELECS. ENG'RS, IEEE P7000: DRAFT STANDARD FOR MODEL PROCESS FOR ADDRESSING ETHICAL CONCERNS DURING SYSTEM DESIGN (2020); IBM, EVERYDAY ETHICS FOR ARTIFICIAL INTELLIGENCE 6 (2019) ("Ethics must be embedded in the design and development process from the very beginning of AI creation."); WORLD ECON. F., ETHICS BY DESIGN: AN ORGANIZATIONAL APPROACH TO RESPONSIBLE USE OF TECHNOLOGY 6–8 (Dec. 2020) (discussing ethical design principles); *see also* BATYA FRIEDMAN & DAVID G. HENDRY, VALUE SENSITIVE DESIGN: SHAPING TECHNOLOGY WITH MORAL IMAGINATION 1 (2019) ("[A]ctively engaging with values in the design process offers creative opportunities for technical innovation as well as for improving the human condition.").

235. Timnit Gebru et al., *Datasheets for Datasets*, ARXIV 6 (2020), https://arxiv.org/pdf/1803.09010.pdf [https://perma.cc/YW29-LGTG] (proposing the use of these instruments to provide information about the providence of data used to train and develop an AI system).

236. Margaret Mitchell et al., *Model Cards for Model Reporting*, *in* FACCT '19: PROCEEDINGS OF THE 2019 CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 220 (2019), https://dl.acm.org/doi/pdf/10.1145/3287560.3287596 [https://perma.cc/KN6J-9CL5] (proposing the use of model cards that provide information about the intended use of the model, along with its known limitations and risks); *see also* Galen Harrison et al., *Towards Supporting and Documenting Algorithmic Fairness in the Data Science Workflow*, WORKSHOP ON TECH. & CONSUMER PROTECTION (May 23, 2019), https://www.ieee-security.org/TC/SPW2019/ConPro/papers/harrison-conpro19.pdf [https://perma .cc/VPZ9-B4UN] (proposing documentation and visualization of algorithms in data science processes).

237. *See generally* Madaio et al., *supra* note 221 (proposing the use of structured

some ethical AI concerns may be ameliorated with technical patches and toolkits.[238] But it is a categorical error to treat ethical AI as a suite of problems that can be resolved with procedural protocols and technical solutions.[239] To be sure, computers can help. The point here, however, is that many AI challenges cannot, or should not, be resolved by computers.

For example, software tools that detect AI bias and optimize for fairness are readily available.[240] Still, humans must determine which fairness metrics to utilize in which contexts.[241] Explainable AI software is another example of faux techno-solutionism. This software may provide useful insights to data scientists for purposes of model training and evaluation; however, those insights may be meaningless or unsatisfactory for end users, auditors, adjudicators, and policymakers.[242] Even more concerning, studies show that explainable AI tools can be manipulated or misleading.[243] Needless to say, if these tools or their uses are untrustworthy, then the AI explanations will be untrustworthy too.

The burgeoning field of algorithmic auditing offers additional promise for actualizing and incenting ethical AI.[244] Such audits can be internal or

---

considerations pertaining to AI fairness to instigate dialogue and deliberation during AI development).

238. *See, e.g.*, *supra* note 199 and accompanying text (referencing AI debiasing tools); *see also* Kroll et al., *supra* note 12, at 636–41 (providing a computer scientist's perspective on algorithmic accountability and calling for specific tailored solutions); ASS'N FOR COMPUTING MACH. U.S. PUB. POL'Y COUNCIL, STATEMENT ON ALGORITHMIC TRANSPARENCY AND ACCOUNTABILITY 2 (2017), http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf [https://perma.cc/7U6F-ETB6].

239. *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 [https://perma.cc/93EQ-3AKD] ("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.").

240. *See supra* note 199 and accompanying text.

241. *See* BIRD ET AL., *supra* note 199, at 2 ("Because fairness in AI is a sociotechnical challenge, there is no software tool that will 'solve' fairness in all AI systems."); KEARNS & ROTH, *supra* note 74, at 63 ("Good algorithms can specify a menu of solutions, but people still have to pick one of them."); *see also see also* Coal. for Critical Tech., *supra* note 239.

242. *See Explainable AI*, THE ROYAL SOC'Y 12 (2019), https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf [https://perma.cc/R6XE-3N36].

243. *See* Himabindu Lakkaraju & Osbert Bastani, *"How Do I Fool You?": Manipulating User Trust Via Misleading Black Box Explanations*, *in* AIES '20: PROCEEDINGS OF THE AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY 79, 85 (2020), https://dl.acm.org/doi/pdf/10.1145/3375627.3375833 [https://perma.cc/WJ6H-HBJ7] (empirically establishing how user trust in black-box AI models can be manipulated by misleading explanations).

244. *See, e.g.*, Raji et al., *supra* note 96, at 1 (introducing "a framework for algorithmic auditing that supports artificial intelligence system development end-to-end, to be applied

external. In both settings, algorithmic auditing generally entails the inspection of technical and non-technical aspects of AI systems.[245] Under the right conditions and constraints, algorithmic audits can be highly beneficial. But under current conditions, the constraints are neither standardized nor regularized. As such, the reliability and social value of algorithmic audits are highly contingent. Certainly, the opacity and partiality of some audits have prompted justified concern that the process may be exploited to legitimize dubious AI systems.[246] Because the nascent AI auditing industry is unregulated, the audits themselves lack the patina of legitimacy enjoyed in more mature markets.[247]

All told, the disciplined practice of ethical AI requires time, resources, and institutional buy-in that many firms may not have—or feel the need to have—unless compelled by market forces or binding norms.[248] Moreover, to greater and lesser extents, firms will externalize the costs of

---

throughout the internal organization development lifecycle"); Jennifer Cobbe et al., *Reviewable Automated Decision-Making*, COMPUT. L. & SEC. REV., Nov. 2020, at 1 (calling for a "reviewability framework" to promote accountability); MILES BRUNDAGE ET AL., TOWARD TRUSTWORTHY AI DEVELOPMENT: MECHANISMS FOR SUPPORTING VERIFIABLE CLAIMS 8–10 (2020); James Guszcza et al., *Why We Need to Audit Algorithms*, HARV. BUS. REV., Nov. 28, 2018, https://hbr.org/2018/11/why-we-need-to-audit-algorithms [https://perma.cc/8ZPJ-9EQW]; Rumman Chowdhury & Narendra Mulani, *Auditing Algorithms for Bias*, HARV. BUS. REV., Oct. 24, 2018, https://hbr.org/2018/10/auditing-algorithms-for-bias [https://perma.cc/ULJ5-FSDR] (discussing a fairness tool to audit outcomes developed by Accenture); Bruneis & Goodman, *supra* note 18, at 339 (identifying eight criteria that developers would need to identify for external review, including: the predictive goals of the algorithm and the problem it is meant to solve; the training data considered relevant to reach the predictive goal; the training data excluded and the reasons for excluding it; the actual predictions of the algorithm as opposed to its predictive goals; the analytical techniques used to discover patterns in the data; other policy choices encoded in the algorithm besides data exclusion; validation studies or audits of the algorithm after implementation; and a plain language explanation of how the algorithm makes predictions); *see also* INST. OF INTERNAL AUDITORS, GLOBAL PERSPECTIVES AND INSIGHTS: THE IIA'S ARTIFICIAL INTELLIGENCE AUDITING FRAMEWORK 2–3 (2017), https://na.theiia.org/periodicals/Public%20Documents/GPI-Artificial-Intelligence-Part-II.pdf [https://perma.cc/6U59-SMAR].

245. *See, e.g.*, Shea Brown et al., *The Algorithm Audit: Scoring the Algorithms that Score Us*, BIG DATA & SOC'Y, Jan.–June 2021, at 2.

246. *See* Mona Sloane, *The Algorithmic Auditing Trap*, MEDIUM (Mar. 17, 2021), https://onezero.medium.com/the-algorithmic-auditing-trap-9a6f2d4d461d [https://perma.cc/DY9T-SVJD].

247. *Cf.* Raji et al., *supra* note 96, at 4, 10 (comparing the unsystematized nature of AI audits to the maturity of other auditing systems); BRUNDAGE ET AL., *supra* note 244, at 25 (noting that auditing "standards are not yet established for AI systems").

248. *Cf.* Orr & Davis, *supra* note 109, at 8 ("Giving primacy to legal mandates renders ethical considerations a relative luxury—something 'nice to think about,' but ultimately subservient to the formal codes and regulations in place."); Rainie et al., *supra* note 219, at 4 (reporting on widespread concern among AI experts that "main developers and deployers of AI are focused on profit-seeking and social control, and there is no consensus about what ethical AI would look like").

AI risks to clients, consumers, or communities of (un)interest. If this is a failure, then it is both a market failure and a regulatory failure to fix.

2. Government Challenges

The government, for its part, faces many of the same challenges as industry and arguably more. True, the government is relieved of the industry's duty to shareholders and market drivers. Yet the government has sovereign duties that offset the difference along all the relevant dimensions: safety, fairness, transparency, and accountability. That is not to say that technology firms have free rein. But it is to acknowledge the asymmetrical laws and expectations that attach to public and private action.[249] Those distinctions are important, insofar as they entail special government responsibilities.

But the immediate focus here is another asymmetry: namely, the government's market dependency on industry to supply the tools of algorithmic governance.

In theory, when the government makes sourcing decisions, "[it] can either hire and train personnel and assemble the raw materials needed to perform government tasks, or it can contract through the procurement process to buy them."[250] Far from a typical "build-or-buy" decision, however, the government has nowhere near the in-house capacity to build and field AI systems at scale.[251] While AI prototypes and pilot programs are plentiful in some agencies, the government is in short supply of the technical resources and know-how required for enterprise-level AI ideation, development, integration, deployment, monitoring, and sustainment.[252]

The government's technological debt profoundly affects how agency demand for ethical AI solutions will be fulfilled, or perhaps unfulfilled. As a threshold matter, the industry's commercially oriented research agenda only partially aligns with the government's needs and

249. *See* Crawford & Schultz, *supra* note 29, at 1944; *see also infra* Section II.A.4.

250. *See* ACUS REPORT, *supra* note 1, at 88; *see also* Oliver E. Williamson, *Public and Private Bureaucracies: A Transaction Cost Economics Perspective*, 15 J.L. ECON. & ORG. 306, 319 (1999) (discussing make-or-buy sourcing decisions).

251. *See* NAT'L SEC. COMM'N ON A.I., INTERIM REPORT 22 (2019) [hereinafter NSCAI INTERIM REPORT] ("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.").

252. NSCAI FINAL REPORT, *supra* note 11, at 32 ("Successful development and fielding of AI technologies depends on a number of interrelated elements that can be envisioned as a stack," the integration of which "can be daunting and historically has been underestimated."); *see also* MARK TREVEIL ET AL., INTRODUCING MLOPS: HOW TO SCALE MACHINE LEARNING IN THE ENTERPRISE 4 (2020) (explaining that the "machine learning life cycle in an enterprise setting is much more complex, in terms of needs and tooling").

responsibilities.[253] Thus, commercially available AI tools and services may not exist, or may be unsuitable for government use. When AI solutions are available, customer agencies may be unaware of the functional limits, biases, and value judgments embedded in acquired AI systems. Indeed, as earlier discussed, information about a vendor's design choices may be legally or contractually insulated from disclosure.[254] Meanwhile, on the supply side, vendors may not fully appreciate the government's legal constraints, institutional protocols, or use contexts in which acquired AI tools will be deployed.

To a considerable extent, the informational asymmetries may be overcome (Part IV makes recommendations for how). Yet, more broadly, the government's market dependencies for AI solutions are structurally entrenched and hard to rectify; certainly not on a timescale that matches the government's projected demand for ethical AI solutions. This predicament is the culmination of decades of structural reforms that are coming home to roost.

Widespread bipartisan support to "shrink" and "reinvent" the federal government in the 1990s was propelled by aspirations to make government more business-like and efficient.[255] These reforms yielded certain successes, but they drained the federal workforce.[256] Decades of federal hiring caps, cuts, and freezes have left the federal government

253. *See* REBECCA GELLES ET AL., CTR. FOR SEC. & EMERGING TECH., MAPPING RESEARCH AGENDAS IN U.S. CORPORATE AI LABORATORIES 3 (2021), https://cset.georgetown.edu/wp-content/uploads/CSET-Mapping-Research-Agendas-in-U.S.-Corporate-AI-Laboratories.pdf [https://perma.cc/T6XL-ANHU] (finding a "potential mismatch" between "private research investments and national priorities"); *id.* at 8 ("The major private labs that have invested aggressively in [machine learning] in recent years may not be investing in the specific areas that are most beneficial to the overall U.S. position in the technology.").

254. *See supra* notes 172–78 and accompanying text (discussing transparency challenges relating to commercial trade secrecy).

255. *See* Guttman, *supra* note 28, at 881–90 (discussing the ideological and political shift toward federal outsourcing in the mid-to-late twentieth century); Steven L. Schooner, *Fear of Oversight: The Fundamental Failure of Businesslike Government*, 50 AM. U. L. REV. 627, 636 (2001) ("The mid-1990s witnessed a tsunami of procurement reforms heralded as the most successful aspect of [Vice President] Gore's reinventing government initiative, which were intended to make the procurement system less bureaucratic and more businesslike."); Steven Kelman, *Strategic Contracting Management*, *in* MARKET-BASED GOVERNANCE 88, 89–91 (John D. Donahue & Joseph S. Nye Jr. eds., 2002). *See generally* AL GORE, CREATING A GOVERNMENT THAT WORKS BETTER & COSTS LESS (1993) (describing the Clinton administration's plans to reduce regulatory barriers and governmental waste).

256. Dan Guttman, *Governance by Contract: Constitutional Visions; Time for Reflection and Choice*, 33 PUB. CONT. L.J. 321, 324–25 (2004) (discussing the challenges associated with the privatization of the federal workforce); *see also* Shelly Roberts Econom, *Confronting the Looming Crisis in the Federal Acquisition Workforce*, 35 PUB. CONT. L.J. 171, 189 (2006); PAUL R. VERKUIL, OUTSOURCING SOVEREIGNTY 162 (2007).

with little choice but to use contract and grant employees to achieve its goals.[257]

Over the same stretch, federal spending on research and development for new technologies declined precipitously.[258] The technological waves that ushered in home computers, pocket computers, and the internet-of-things, were mostly sourced with private capital, free and clear of government rights.[259] Even with the recent uptick in federally spending on AI research and development, private capital investments still "dwarf" federal funding.[260] Consequently, the government is in "perpetual catch-up mode," with "limited control over how AI technologies are developed, shared, and used."[261]

None of this is lost on the government, which remains clear-eyed about its in-house capacity challenges. Recently, the government has established several programs to recruit and build an AI workforce,[262] and

257. *See* PAUL C. LIGHT, THE GOVERNMENT INDUSTRIAL COMPLEX 65–69 (2019) (explaining how increased government services, combined with personnel ceilings, led to a rise in private contracting by the government).

258. Federal R&D dropped from a height of near 1.9% of GDP in 1964 to just 0.62% in 2018. Anne Q. Hoy, *Increases in U.S. Federal R&D Needed in a Global Crisis*, AM. ASS'N FOR ADVANCEMENT SCI. (Aug. 31, 2020), https://www.aaas.org/news/increases-us-federal-rd-needed-global-crisis [https://perma.cc/522L-LBU3].

259. *See* NAT'L RCH. COUNCIL, FUNDING A REVOLUTION: GOVERNMENT SUPPORT FOR COMPUTER RESEARCH 179–81 (1999).

260. NSCAI, INTERIM REPORT, *supra* note 20, at 15–16.

261. *Id.* at 15.

262. For example, the General Services Administration (GSA) offers a variety of programs including: "18F" ("[a] digital consulting office that partners with agencies to help them build or buy digital services") and "IT Modernization Centers of Excellence" ("[a] centralized team of technical experts that accelerate agency-wide IT modernization"). *See Technology Transformation Services*, GEN. SERVS. ADMIN., https://www.gsa.gov/about-us/organization/federal-acquisition-service/technology-transformation-services [https://perma.cc/KB6H-JJH7]. GSA also hosts a Presidential Innovation Fellowship, which is "[a] program that pairs top technologists with civil-servants to spend 12 months tackling some of our nation's biggest challenges." *Id.* Tellingly, these capacity building programs lean on private industry. For example, the GSA's AI Center for Excellence is designed to provide acquisition consulting and assistance to agencies on a government-wide basis. *See* Kathleen Walch, *How The Federal Government's AI Center of Excellence Is Impacting Government-Wide Adoption of AI*, FORBES (Aug. 8, 2020, 1:00 PM), https://www.forbes.com/sites/cognitiveworld/2020/08/08/how-the-federal-governments-ai-center-of-excellence-is-impacting-government-wide-adoption-of-ai/#7da611206660 [https://perma.cc/D5TW-CYYK]. Additionally, "18F" offers a centralized team of private-sector technology experts to consult and work with agencies on specific projects. *See About 18F*, GEN. SERVS. ADMIN, https://18f.gsa.gov/about/ [https://perma.cc/QD9D-9AKV]. Likewise, DoD's Joint Artificial Intelligence Center (JAIC) has transitioned from a product building unit into an AI training, acquisition, and platform hub. Jackson Barnett, *"JAIC 2.0" Moves Away From Building Products to Focus on DOD-wide AI Transformation*, FEDSCOOP (Nov. 6, 2020), https://www.fedscoop.com/jaic-2-0-moving-away-from-products-artificial-intelligence/ [https://perma.cc/U5SS-VM7R]; Joint A.I. Ctr., *JAIC Completes Responsible AI Champions Pilot*, AI IN

is piloting programs for AI acquisition reform.[263] As matters currently stand, however, "the vast majority of IT leaders say their agency is struggling to incorporate AI into overall IT operations."[264]

B. *The Gap Between Ethical AI and Algorithmic Governance*

A rich scholarship has emerged to square AI systems with U.S. legal structures, institutions, and democratic norms that have not yet been coded for algorithmic governance.[265] As Aziz Huq explains, "present doctrinal formulations" do not necessarily mesh with, or address, a range of constitutional values that are implicated "when the focus shifts from human to machine action."[266] Moreover, as David Engstrom and Daniel Ho explain, there is a dearth of ready-made legal or technological tools to cope with the breadth of governance challenges in the digital era: "[J]udicial review of agency action using AI is unlikely to yield systematic scrutiny," "a thicket of reviewability and related doctrines largely insulate algorithmic decision making," and "the current [administrative law] mechanisms remain ill-suited for providing meaningful accountability over rapid advances in AI."[267]

Legal scholars have proposed a variety of doctrinal and institutional reforms to address the foregoing mismatches and maladaptation of AI for government use. For good reason, much of that prescriptive work is focused on constitutional law, administrative law, and norms of good

---

DEF. BLOG (July 8, 2020), http://www.ai.mil/blog_07_08_20-jaic_completes_responsible_ai_champions_pilot.html [https://perma.cc/EQ73-GCRR]; Jackson Barnett, *With $106M Contract, JAIC Takes Major Step Building Central AI Platform for DOD*, FEDSCOOP (Aug. 13, 2020), https://www.fedscoop.com/jaic-ai-development-platform-dod-joint-common-foundation-deloitte/ [https://perma.cc/8VXT-J67B].

263. *See* Press Release, JAIC Pub. Affs., Joint Artificial Intelligence Center to Pilot a Responsible AI Procurement Process (July 27, 2021), https://www.ai.mil/news_07_27_21-jaic_to_pilot_a_responsible_ai_procurement_process.html [https://perma.cc/333Y-AV45]; *see also infra* notes 369–72 and accompanying text.

264. *See From Pilots to Proficiency: Operationalizing Federal AI*, MERITALK 17 (2021), https://www.meritalk.com/wp-content/uploads/2021/05/pilots-to-proficiency.pdf [https://perma.cc/P5A4-LTZN].

265. *See supra* notes 31–37 and accompanying text.

266. Huq, *supra* note 33, at 1881.

267. Engstrom & Ho, *supra* note 37, at 828, 844–45. In a similar vein, Wendy Wagner and Martin Murillo argue that the incentives and doctrines around agency rulemaking not only cut against the grain of current AI best practices but may also "tacitly reward[] agencies for developing and using algorithmic tools that are opaque and potentially biased." Wagner & Murillo, *supra* note 37, at 3. Moreover, Deirdre Mulligan and Kenneth Bamberger argue that inexplainable AI systems may run afoul of the Administrative Procedure Act's prohibition against "arbitrary and capricious" agency action. *See* Mulligan & Bamberger, *supra* note 37, at 773–74 (explaining that AI may violate the arbitrary and capricious agency standard); 5 U.S.C. § 706(2)(A) (providing that courts shall "hold unlawful and set aside [an] agency action" they deem to be "arbitrary [or] capricious"); Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 42–43 (1983).

governance—most notably pertaining to substantive and procedural regularity, transparency, and accountability.[268]

This Article aligns with that larger project: namely, to reconcile the ideals of our constitutional democracy with the sociotechnical challenges inhering in AI systems. But the injunctions and sanctions of constitutional and administrative law can only obliquely address a cache of governance challenges that originate and disseminate through the acquisition gateway. Without procurement law, the reformist agenda is dangerously incomplete.

This Article is not the first to sound the alarm. Deidre Mulligan and Kenneth Bamberger, for example, offer a trenchant account of how a "procurement mindset" can forfeit the government's responsibility to make important design choices with public input.[269] Yet their proposed solutions are adjacent to procurement law itself. Specifically, they prescribe: (1) the use of "algorithmic impact assessments," which would allow for public deliberation and input for AI systems that embed certain types of policy decisions; and (2) specialized in-house technical teams to provide consulting and support services to agencies adopting AI tools.[270]

Closer to home, some scholars and advocates have proposed more contract-based solutions. Robert Brauneis and Ellen Goodman, for example, urge procurement officials to use their "contracting powers to insist on appropriate record creation, provision, and disclosure."[271] Along similar lines, Cary Coglianese and Erik Lampmann argue that "careful drafting of contracts for AI services paired with suitably robust public input over [contract] provisions . . . can allow procurement officers to assure the public that agencies are using AI tools responsibly."[272] These prescriptions, too, angle in the right direction. Still, they only scratch the surface.

As yet, federal procurement law offers a reservoir of untapped possibilities. Indeed, as elucidated below, the acquisition gateway is primely situated to check and enable ethical algorithmic governance. As importantly, procurement law is uniquely suited for these purposes in ways that other legal frameworks miss.

IV. OPERATIONALIZING ETHICAL AI THROUGH PROCUREMENT LAW

This final Part drills into procurement's positive potential. The recommendations here are keyed to four phases of the procurement pipeline: (1) acquisition planning; (2) market solicitation; (3) bid

268. *See supra* notes 36–37, 266–67, and accompanying text (observing the mismatch between machine learning AI systems and legal doctrine).

269. Mulligan & Bamberger, *supra* note 37, at 782.

270. *Id.* at 774.

271. *See, e.g.*, Brauneis & Goodman, *supra* note 18, at 164.

272. Coglianese & Lampmann, *supra* note 39, at 180.

evaluation and source selection; and (4) contract performance. While each recommendation can be adopted in isolation, their full value will accrue in combination. Thematically, the approach here aims to capitalize on the merger of public and private interests around ethical AI. Toward that end, the recommendations exploit the procurement system's monetary and regulatory levers to incent market competition and responsible innovation. By centering ethical AI across the procurement lifecycle, the hope is that federal buyers and their AI suppliers will think more critically and holistically about the AI tools passing through the acquisition gateway for government use.

By way of background, federal procurement is subject to an elaborate body of regulations and practices designed to advance myriad objectives.[273] Most pertinent here, those objectives include market competition, transparency, efficiency, socioeconomic policy, risk avoidance, and best value to the government customer.[274] 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."[275] Federal procurement is also governed by agency-specific supplements to the FAR,[276] congressional statutes,[277] presidential Executive Orders,[278] and agency guidance documents.[279]

---

273. Schooner, *supra* note 255, at 634–37 ("The laws, regulations, and policies controlling the award and performance of government contracts present a dense thicket reflective of a large, complex bureaucracy."); *see also infra* note 274 and accompanying text.

274. *See generally* KATE M. MANUEL ET AL., CONG. RSCH. SERV., RS2826, THE FEDERAL ACQUISITION REGULATION (FAR): ANSWERS TO FREQUENTLY ASKED QUESTIONS (2015) (providing an overview of various procurement regulations and the values they serve); Steven L. Schooner, *Desiderata: Objectives for a System of Government Contract Law*, 11 PUB. PROCUREMENT L. REV. 103 (2002) (discussing several goals commonly associated with procurement systems).

275. *Foreword* to FAR (2021), https://www.acquisition.gov/sites/default/files/current/far/pdf/FAR.pdf [https://perma.cc/WQ7W-M93R].

276. *See* FAR 1.301(a)(1) (2021) (authorizing agencies to issue "agency acquisition regulations that implement or supplement the FAR, and incorporate . . . agency policies, procedures, and contract clauses, solicitation provisions, and forms" that govern the contract); *id.* 1.301(a)(2) (allowing for "internal agency guidance").

277. *See, e.g.*, 41 U.S.C. §§ 3101–06 (governing the procurement of supplies and services for most civilian agencies); 10 U.S.C. §§ 2302–39c (governing the procurement procedures for the DoD, National Aeronautics and Space Administration, and Coast Guard).

278. *See, e.g.*, Exec. Order No. 11,246, 30 Fed. Reg. 12,319, 12,319 (Sept. 28, 1965) (requiring government contractors not to discriminate and to develop affirmative action plans); Exec. Order No. 14,026, 86 Fed. Reg. 22,835, 22,835 (Apr. 27, 2021) (calling for increase in hourly minimum wage paid by the parties that contract with the federal government).

279. *See, e.g.*, FAR 1.301(a)(2) (2021) (allowing for "internal agency guidance"). *See generally* OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, OMB CIRCULAR A-76, PERFORMANCE OF COMMERCIAL ACTIVITIES (1999) (setting forth the procedures for determining

Two caveats before proceeding. First, this Article's recommendations for acquiring ethical AI are mostly agnostic to policymaking form—statutory, regulatory, or otherwise. To be sure, the choice of policymaking form can be consequential. For example, statutory mandates are generally more stable than Executives Orders, and amendments to the FAR would apply more broadly than agency-specific policies and practice. As a first pass, however, the discussion below focuses on substance and leaves questions of policymaking form to future work.[280] This approach allows for variation and experimentation, which can be a virtue, given AI's sociotechnical and contextual sensitivities.

Second, the recommendations below are most directly applicable to FAR-based procurement contracts—which account for the great bulk of federal acquisitions. Other Transaction Agreements (OTAs) are beyond this Article's immediate purview.[281] By regulatory design, a main purpose of OTAs is to provide alternative acquisition pathways for (mostly) nontraditional vendors to conduct research and prototype development for the government, unencumbered by the FAR's regulatory strictures, procedures, and contractual clauses.[282] While many of the recommendations advanced below can be adopted or adapted for OTAs, this Article leaves those questions for future work.

A. *Acquisition Planning: AI Risk Assessments*

Federal procurement begins with acquisition planning.[283] During this phase, the agency's product or service requirements are established, the personnel responsible for the acquisition are coordinated, costs and risks relating to the acquisition are assessed, and an overall acquisition strategy is developed.[284] When acquiring information technology (IT), agency officials must conduct specialized risk assessments pertaining to schedule

---

whether commercial activities should be outsourced or performed in-house using government facilities and personnel); OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, OMB CIRCULAR A-130, MANAGING INFORMATION AS A STRATEGIC RESOURCE (establishing general policy for the planning, budgeting, governance, acquisition, and management of federal IT systems).

280. Thus, insofar as the recommendations below would require agency officials to take certain actions, those mandates could come from Congress in the form of a statute, the White House in the form of an Executive Order, amendment to the FAR, or other forms of binding federal policy. Other recommendations are not intended to be binding, but call for new or modified procurement practice. Those recommendations can be implemented and supported by informal agency guidance.

281. For a useful overview of OTAs, and their use by the DoD in particular, see MOSHE SCHWARTZ & JEODO M. PETERS, CONG. RSCH. SERV., R45521, DEPARTMENT OF DEFENSE USE OF OTHER TRANSACTION AUTHORITY: BACKGROUND, ANALYSIS, AND ISSUES FOR CONGRESS (2019).

282. *Id.* at 2 ("[OTAs] are legally binding contracts that are generally exempt from federal procurement laws and regulations such as the Competition in Contracting Act and the [FAR]").

283. *See generally* FAR Part 7 (2021) (governing acquisition planning).

284. *See id.* at 7.105 (2021).

and cost overruns, security and privacy, and interoperability with existing government systems.[285] IT risk assessments, however, are not styled or suited for the unique challenges of acquiring AI.[286] The recommendation here aims to fill that void with mandatory "AI risk assessments." [287]

Because AI risks are contextually contingent, a one-size-fits-all approach is neither necessary nor advisable. But acquiring AI will always entail certain types of risks that can and should be accounted for during the planning phase. If nothing else, forecasting and logging AI risks will force conversations about whether an AI solution is necessary or appropriate to meet the agency's needs, and if so, under what conditions and constraints.

By way of illustration, below is a non-exhaustive set of considerations that an AI risk assessment could capture:

- ❖ To what extent, if any, will the agency need to rely on third parties to design, develop, deploy, audit, or monitor the AI system? All else equal, the more the government must rely on third parties for these lifecycle needs, the less control the government will have over a system's operations—both when the technology is working as intended and not.

- ❖ What are the transparency gaps in the AI system? As earlier explained, AI systems can be more or less transparent for a variety of technical and non-technical reasons.[288] The government should

---

285. *See id.* at 39.102(b) (2021); *see also* 44 U.S.C. § 3554(b); *id.* § 11331 (delegating authority 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"); *Appendix IV to OMB Circular No. A-130*, OFF. OF MGMT. & BUDGET, https://obamawhitehouse.archives.gov/omb/circulars_a130_a130appendix_iv [https://perma.cc/HF67-332Z] ("Each agency program official must understand the risk to [information] systems under their control.").

286. *See* Raji et al., *supra* note 96, at 5 ("[T]the design, prototyping and maintenance of AI systems raises many unique challenges not commonly faced with other kinds of intelligent systems or computing systems more broadly."); Exec. Order No. 13,960, 85 Fed. Reg. 78,939, 78,941 (Dec. 8, 2020) (observing that "[e]xisting OMB policies currently address many aspects of information and information technology design, development, acquisition, and use that apply, but are not unique, to AI."). *See generally infra* Sections II.B, Section III.A (mapping the sociotechnical challenges of AI systems).

287. It bears noting that risk-management is considered best practice when private enterprises acquire AI technologies. For frameworks utilized in the private sector, see for example: *AI and Risk Management Innovating with Confidence*, DELOITTE CTR. FOR REGUL. STRATEGY (2018), https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/deloitte-gx-ai-and-risk-management.pdf [https://perma.cc/3Y47-4N38]; *AI Risk and Controls Matrix*, KPMG LLP (2018), https://assets.kpmg/content/dam/kpmg/uk/pdf/2018/09/ai-risk-and-controls-matrix.pdf [https://perma.cc/XNB7-F9VR].

288. *See supra* Section II.B.3.

tease out those differences, and assess each risk separately. For example, transparency risks may relate to model interpretability, data provenance, trade secrecy, model versioning, or some combination thereof.[289] Moreover, in many government settings, the interpretability and explainability of AI systems may be operationally or legally required. The costs of mitigating or overcoming these transparency risks will necessarily be context specific. But, at least in some contexts, transparency gaps might render an AI system unusable for its intended purpose.

- ❖ Will there be a human in-the-loop (or on-the-loop)? If so, what roles and responsibilities will be assigned to the human? In high-stakes and sensitive contexts, human validation of system inputs and outputs will generally be necessary before further government action is taken.[290] Moreover, in contexts where human judgment or discretion is required, risks relating to automation bias, automation aversion, model interpretability, etc., must also be forecasted and assessed.[291]

- ❖ Will the AI system be used in contexts that may have a discriminatory effect, or that may inflict special burdens or hardships on marginalized groups? Most public-facing AI use cases will carry these risks. But internal and back-office AI uses can also be risky. For example, AI systems used for government

289. *See supra* notes 167–71 and accompanying text; *see also* JONATHON PHILLIPS ET AL., NAT'L INST. STANDARDS & TECH., FOUR PRINCIPLES OF EXPLAINABLE ARTIFICIAL INTELLIGENCE 2–6 (2021), https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf [https://perma.cc/YPE7-3X43]; P. Jonathon Phillips et al., Nat'l Inst. Standards & Tech., Four Principles of Explainable Artificial Intelligence 10–11 (Aug. 2020), https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf [https://perma.cc/X26Y-FMGW] (discussing different dimensions and considerations relating to AI explainability).

290. *See* Singh et al., *supra* note 99, at 13–14 ("[H]aving a human in the loop represents a clear point for exercising judgement, intervention and control."). In time-critical contexts, a human in-the-loop might obstruct optimal system performance. For example, in the realms of cybersecurity and military tactical engagement, human oversight of the system (i.e., a human *on*-the-loop) may lead to better outcomes than human validation of AI outputs in real-time. *Cf.* Joel E. Fischer et al., *In-the-Loop or On-the-Loop? Interactional Arrangements to Support Team Coordination with a Planning Agent*, CONCURRENCY & COMPUTATION PRAC. & EXPERIENCE, Apr. 25, 2021, at 1, https://onlinelibrary.wiley.com/doi/10.1002/cpe.4082 [https://perma.cc/4DEW-EMUH] (distinguishing between humans in-the-loop and on-the-loop, and studying contexts in which one structure might be preferrable to others); NSCAI FINAL REPORT, *supra* note 11, at 9 ("Human operators will not be able to keep up with or defend against AI-enabled cyber or disinformation attacks, drone swarms, or missile attacks without the assistance of AI-enabled machines.").

291. *See supra* notes 109–11 and accompanying text (discussing human-computer interactions as a feature of system design).

hiring and promotion, resource allocation, language translation, text generation, and building security,[292] may not work equally or sufficiently for certain subpopulations.

- ❖ Will the data used or generated by the AI system contain sensitive personal information? If yes, then a slew of considerations relating to data privacy, data integrity, and data security must be assessed and accounted for in the risk portfolio.[293]

- ❖ How might the AI system fail, or drift from its intended uses or performance standards? Relatedly, what protocols exist (or need to exist) to identify and rectify failure modes? As earlier explained, AI systems can fail or drift for myriad reasons—malign and benign, technical and non-technical.[294] Due to network effects, AI failure modes may also infect surrounding systems. To mitigate harm, agencies must be prepared for these contingencies in advance.

- ❖ Will the AI system require frequent updating, and if so, what protocols exist (or need to exist) to ensure the traceability and reliability of model versioning over time? A modified AI system may improve performance along one or more metrics but impair performance in other ways. Moreover, without proper precautions, model versioning can make it impossible to know how a model performed at the point in time that a particular government decision was made.[295] Without that information, an agency may be hard pressed to justify any actions based on the algorithmic output.

---

292. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-526, FACIAL RECOGNITION TECHNOLOGY: CURRENT AND PLANNED USES BY FEDERAL AGENCIES 12–13 (2021) (discussing the use of facial recognition by several federal agencies for the purpose of digital access, cybersecurity purposes, and building security).

293. Existing risk-management frameworks for securing data and sensitive personal information could be used and tailored as necessary to capture AI-specific risks. *Cf. NIST Risk Management Framework*, NAT'L INST. STANDARDS & TECH., https://csrc.nist.gov/Projects/Risk-Management [https://perma.cc/7FKF-T7Q7] (linking to a suite of NIST standards and guidelines to support implementation of risk management programs to meet the requirements of the Federal Information Security Modernization Act); NIST JOINT TASK FORCE, NAT'L INST. STANDARDS & TECH., SPEC. PUBL'N 800-53 REV. 5, SECURITY AND PRIVACY CONTROLS FOR INFORMATION SYSTEMS AND ORGANIZATIONS (2020), https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf [https://perma.cc/XA8D-7HU5].

294. *See supra* Section II.B.1.

295. *See* Cuéllar, *supra* note 36, at 135–36 (cautioning that "heavy reliance on computer programs—particularly adaptive ones that modify themselves over time—may complicate public deliberation about administrative decisions, because few observers may be entirely capable of understanding how a given decision was reached").

- ❖ In addition to the foregoing, risks relating to system access, change management, compute resources, system interoperability, sustainability, and lifecycle costs should also be assessed (to the extent not already accounted for in other acquisition planning documents).[296]

The value of AI risk assessments will depend, in large measure, on the people responsible for their curation. In general, cross-disciplinary teams will be necessary for *all* procurement phases. At minimum, the AI risk assessment team should include subject matter experts, IT personnel, data scientists, lawyers, and ethical AI champions (who could be specially trained or certified for that function).

Skeptics may question whether this investment in human capital is necessary, but the answer is unequivocally yes. A diversity of experience and expertise mitigates contextual blind spots and cultivates systematic thinking about sociotechnical risks.[297] Skeptics may also worry that AI risk assessments will create bureaucratic drag on AI acquisitions.[298] To some extent, however, that is the point: to carve time and space for critical deliberations that otherwise may not occur or come too late.

Of course, time is a valuable resource that should not be squandered. But the benefits of AI risk assessments are likely to outweigh the costs of conducting them. More importantly, the benefits of curating AI risk assessments are likely to outweigh the costs of forgoing them.[299] Especially under current market conditions,[300] it would be irresponsible for the government to acquire AI solutions without rigorously screening for risks relating to safety, discrimination, privacy, transparency, and accountability. Future AI regulation may mitigate these concerns;

---

296. Even if these risks are captured elsewhere, collecting them in the AI risk assessment may be useful so that they can be managed and mitigated systematically.

297. *See* Schwartz et al., *supra* note 171, at 8 ("A consistent theme from the literature is the benefit of engaging a variety of stakeholders and maintaining diversity along social lines where bias is a concern (racial diversity, gender diversity, age diversity, diversity of physical ability)."); *id.* ("Technology or datasets that seem non-problematic to one group may be deemed disastrous by others.").

298. This is a long-running concern in government contracting, especially for rapidly evolving technologies. *Cf.* Katherine M. John, *Information Technology Procurement in the United States and Canada: Reflecting on the Past with an Eye Toward the Future*, PROCUREMENT L., Summer 2014, at 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.").

299. Empirically, this may prove not to be the case. For the reasons indicated in the text above, however, it seems fair to assume that the costs of not doing risk assessments will be greater than the costs required to conduct them.

300. *See supra* note 23 and accompanying text.

nevertheless, AI risks will still endure to some significant extent, both in general and in government contexts more specifically.[301]

Last but not least: humans interacting with an AI system may reject it or engage in (risky) compensating behaviors if they do not trust the technology. Done right, AI risk assessments can set the foundations for that trust, inside and outside of government.[302]

B. *Market Solicitations: Calling for Ethical AI*

The dividends of AI risk assessments extend beyond the planning phase. Most pertinent here, the government can recast the identified risks as focal points in the government's market solicitations. These solicitations may come in the form of requests for proposals (RFPs),[303] quotations (RFQs),[304] or information (RFIs).[305] Despite their legal and

301. *See supra* Section II.B (discussing a cache of latent risks and challenges in machine learning AI systems); KEARNS & ROTH, *supra* note 74, at 64 ("[A]nywhere machine learning is applied, the potential for discrimination and bias is very real—not in spite of the underlying scientific methodology but often because of it.").

302. It is worth noting that this Article's prescriptions for AI risk assessments may cohere with, but are different than, "algorithmic impact assessments" (AIAs). *See, e.g.*, DILLON REISMAN ET AL., ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY 5–6 (2018), https://ainowinstitute.org/aiareport2018.pdf [https://perma.cc/U26X-EDMQ] (arguing for the use of AIAs to promote government accountability and public deliberation); Mulligan & Bamberger, *supra* note 37, at 842–45 (same); *see also* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 110, 168 (2017) (same in the context of predictive policing in particular). In general, proposals for AIAs aim "to engage the public and proactively identify concerns, establish expectations, and draw on expertise and understanding from relevant stakeholders." *See* REISMAN ET AL., *supra*, at 7. Some proposals for AIAs include opportunities for third-party auditing, as well as mechanism to challenge the agency's impact assessment, including through judicial review. *See id.* at 10. This Article takes no position on AIAs or their ideal design features. For present purposes, the more important point is that the two instruments can harmonize toward the same general objectives: namely, safe, fair, transparent, and accountable algorithmic governance. Because of this alignment, information curated in AI risk assessments can be incorporated into an AIA covering the same system. And working in reverse, the public-facing requirements of AIAs can incent agencies to undertake robust AI risk assessments. (To the extent that AIAs are intended for agency self-assessments only, and not for public participation and external review, then pre-acquisition AI risk assessments and AIA might be functional equivalents).

303. *See* FAR 15.203(a) (2021) ("[RFPs] are used in negotiated acquisitions to communicate Government requirements to prospective contractors and to solicit proposals.").

304. *Id.* at 8.402(d)(1) (explaining how RFQs are used when agencies order goods and services from federal supply schedules).

305. *See* FAR 15.201(e) (2021) ("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."). Market participants are not required to respond to RFIs. But, for strategic reasons, they often do. For example, a vendor can hope to draw attention to its products and capabilities, which may influence the requirements on a future government contract.

contextual distinctions, each of these instruments serve important dialogic functions. First, as discussed further below, the government can strategically utilize market solicitations to smooth out information asymmetries. Second, and as importantly, the government can craft these solicitations to drive innovation and commercial competition around ethical AI.

By way of illustration, and with the foregoing objectives in view, the government's solicitations can prompt vendors along the following lines:[306]

❖ Describe any training programs that your team members have undergone, and any official policies or protocols adopted by your company that specifically relate to AI safety, fairness, transparency, accountability, or other ethical AI principles.

❖ Describe how your developmental protocols or practices enable end-to-end auditability of the proposed AI solution, and any technical or proprietary limitations that may inhibit auditability. In this regard, would you permit auditing by independent third parties? If yes, explain the conditions or limitations you would impose. If such audits would not be allowed, then explain why.

❖ Describe any known or foreseeable performance weaknesses and vulnerabilities in your proposed AI solution, and explain the source or causes of those vulnerabilities (e.g., in the data, algorithm, design process, human–computer interface, interoperability with other hardware and software, or otherwise).

❖ Describe whether and how your proposed AI solution will be explainable and interpretable to end users, operators, auditors, and other stakeholders, including lay persons, judges, and policymakers.

❖ Describe any anticipated data-related limitations and challenges for your proposed AI solution. What strategies or protocols, if any, might you implement or recommend to address those challenges?

306. Additional questions and prompts, tailored to specific use cases, can and should be included in the government's solicitations. The World Economic Forum provides useful templates and suggestions that can be tailored for federal acquisitions. *See generally* SABINE GERDON ET AL., WORLD ECON. F., AI PROCUREMENT IN A BOX: WORKBOOK (2020), http://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_Workbook_2020.pdf [https://perma.cc/LE73-NGR8].

❖ Describe your privacy and cybersecurity approach to the proposed AI solution, including but not limited to how the data and model will be protected from adversarial attack and human error.

Agencies have little (or nothing) to lose and much to gain from this information exchange. To start, vendor responses may shed light on previously unidentified AI risks. In such cases, the government can and should modify the AI risk assessment to capture those additional concerns. As importantly, vendor responses will enable the agency to make side-by-side comparisons of the risks and capabilities associated with a particular vendor (or AI solution) relative to the field.

Beyond obtaining information, the government can use these instruments to provide information about the government's requirements, constraints, and AI use contexts. More generally, however, centering ethical AI in market solicitations will signal to prospective vendors that they will need to compete on the field of ethical AI to win federal contracts. That signaling is important for three related reasons. First, strategic and innovative vendors may embrace ethical AI as a competitive differentiator. Second, the government's ethical voicing may draw responsible and innovative firms *into* the government market.[307] Third, as discussed below, the inclusion of ethical AI criteria in contract solicitations carries legal significance.

C. *Evaluation and Source Selection: Requiring Ethical AI*

Under existing regulations, agencies must evaluate vendor proposals solely on the criteria pre-specified in the relevant contract solicitation.[308] Thus, to capitalize on this opportunity, agency officials will *need* to include ethical AI requirements in contract solicitations.[309] Separately, or

307. *See supra* notes 202–06 and accompanying text (discussing the technology industry's activism and the reticence of some firms to partner with the government in the areas of national security and law enforcement).

308. *See* FAR 15.305(a) (2021) ("An agency shall evaluate competitive proposals and then assess their relative qualities solely on the factors and subfactors specified in the solicitation."); *id.* at 15.304(d) (2021) ("All factors and significant subfactors that will affect contract award and their relative importance shall be stated clearly in the solicitation."); *id.* at 13.106-1(a)(2)(i) ("When soliciting quotations or offers [for simplified acquisitions,] the contracting officer shall notify potential quoters or offerors of the basis on which award will be made (price alone or price and other factors, e.g., past performance and quality)."); *id.* at 13.106-2(a)(2) ("Quotations or offers shall be evaluated on the basis established in the solicitation."); *see also* Antarctic Support Assocs. v. United States, 46 Fed. Cl. 145, 155 (2000) (noting that contractual awards must be consistent with stated evaluation criteria).

309. Advocacy groups and 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 15, 22, 29, 35 (2020); World Econ. Forum, *AI Procurement in a Box: AI Government Procurement Guidelines*, WORLD ECONOMIC

additionally, ethical AI considerations could be factored into pre-award "responsibility" determinations of prospective vendors.[310] The discussion below elaborates on these recommendations and situates them within existing procurement policy.

1. Evaluation Criteria

As prefaced above, agencies must "evaluate competitive proposals and then assess their relative qualities solely on the factors and subfactors specified in the solicitation."[311] This regulatory constraint promotes competition by steadying the target for prospective vendors. Moreover, this constraint promotes the integrity and transparency of the acquisition process by committing agency officials to the specified evaluative criteria. When crafting solicitations, agencies have discretion over which evaluation criteria to include and prioritize.[312] But certain evaluative criteria, such as price and vendor past performance, generally must be included as a matter of law in competitive procurements.[313] The recommendation here is to create a similar requirement for ethical AI when the government acquires AI solutions.

Specifically, under this proposal, agency officials would be legally required to evaluate vendor proposals on ethical AI grounds. Discretionary waivers of this general rule could be allowed in exceptional circumstances or in specific contexts where AI risks are negligible. In such cases, however, contracting officials should be required to justify the waiver in writing.[314]

Like price and past performance, ethical AI principles will almost always be relevant in AI acquisitions. And, like price and past performance, the relative weight afforded to ethical AI can be determined on a contract-by-contract or contextual basis.[315] To be clear, ethical AI need not be paramount. But including ethical AI among the evaluative criteria will be necessary if the government intends to award contracts even partly on that basis.[316]

FORUM (June 11, 2020), https://www.weforum.org/reports/ai-procurement-in-a-box/ai-government-procurement-guidelines#report-nav [https://perma.cc/7L9L-DGFM].

310. *See* FAR 9.103(a) ("Purchases shall be made from, and contracts shall be awarded to, responsible prospective contractors only."); *see also infra* notes 326–334 and accompanying text (discussing the regulatory framework for responsibility determinations).

311. FAR 15.305(a) (2021).

312. *Id.* at 15.304(c) (2021).

313. *See id.* at 15.304(c)(1), (2) (2021); *see also id.* at 13.106-1(a)(2).

314. Requiring a written justification for norm deviations has a pedigree in procurement law. *See, e.g.*, FAR 6.303 (2021) (requiring written justifications under certain circumstances); *id.* at 13.501 (2021) (same). Without such a requirement, there is a real concern that contracting officials will not include ethical AI criteria in a systematic way and in contexts when they should.

315. *See* FAR 15.101–1 (2021).

316. *See supra* note 308 and accompanying text.

Currently, this is not the government's general practice—far from it. However, there are some encouraging signs of positive change. In 2021, for example, the DoD's Joint Artificial Intelligence Center (JAIC)[317] issued an RFP "to form multiple Blanket Purchase Agreements" with vendors who can provide AI testing and evaluation services to support the DoD and entire U.S. government.[318] As one of the first publicly available RFPs that even *mentions* ethical AI, it provides a useful baseline and template to build upon.

The Performance of Work Statement for this RFP plainly indicates that vendor solutions must account for the "DoD's AI Ethical Principles."[319] Moreover, the RFP explains that blanket purchase agreements will be awarded to a pool of the most highly qualified offerors based on their responses to the accompanying questionnaire.[320] In turn, that questionnaire asks vendors to describe their AI capabilities and developmental processes, along with examples of past performance to support their claims.[321] So far so good.

As pertains to ethical AI, however, the questionnaire contains only one question. To wit: "Is your company willing to incorporate responsible AI methodologies, such as the Department of Defense's AI Ethical Principles . . . into your company's testing and evaluation approach. (Yes or No)."[322] For this question, prospective vendors are instructed that "the Government will consider 'Yes' answers Acceptable and 'No' answers Unacceptable. The purpose of the [yes/no] question is to build awareness

317. The JAIC "serves as the DoD's acquisition hub and coordinator for the development and implementation of the Department's 'Responsible AI' strategy, guidance, and policy." Memorandum from Kathleen H. Hicks, *supra* note 7, at 2; *see also* Barnett, *supra* note 262 (discussing the JAIC's transition from development to acquisition and coordination).

318. JOINT AI CTR., U.S. DEP'T OF DEF., NOTICE I.D. W52P1J21R0029, JOINT ARTIFICIAL INTELLIGENCE CENTER TEST AND EVALUATION BLANKET PURCHASE AGREEMENT REQUEST FOR PROPOSAL (Feb. 11, 2021), https://sam.gov/opp/93bc03aa061e43c0b5567ae8e33e9c2b/view?keywords= [https://perma.cc/4LEU-AJAC].

319. *See* JOINT AI CTR., U.S. DEP'T OF DEF., NOTICE I.D. W52P1J21R0029, PERFORMANCE WORK STATEMENT §§ 1.2.2, 3.4.2, 4 (Feb. 11, 2021), https://sam.gov/opp/93bc03aa061e43c0b5567ae8e33e9c2b/view?keywords= [https://perma.cc/4LEU-AJAC] (downloadable as "Attachment 0002-JAIC and E BPA PWS 11Feb21.pdf") (describing compliance in testing and evaluation, quality control, and tasks).

320. JOINT AI CTR., U.S. DEP'T OF DEF., NOTICE I.D. W52P1J21R0029, INSTRUCTIONS AND EVALUATION CRITERIA 5–7 (Feb. 11, 2021) [hereinafter "JAIC INSTRUCTIONS"], https://sam.gov/opp/93bc03aa061e43c0b5567ae8e33e9c2b/view?keywords= [https://perma.cc/4LEU-AJAC] (downloadable as "Attachment 0003-JAIC TE BPA Instructions and Eval Criteria 11Feb21.pdf").

321. JOINT AI CTR., U.S. DEP'T OF DEF., NOTICE I.D. W52P1J21R0029, TEST & EVALUATION OF AI BLANKET PURCHASE AGREEMENT QUESTIONNAIRE 1–3 (Feb. 11, 2021) [hereinafter "JAIC QUESTIONNAIRE"], https://sam.gov/opp/93bc03aa061e43c0b5567ae8e33e9c2b/view?keywords= [https://perma.cc/4LEU-AJAC] (downloadable as "Attachment 0001-JAIC TE BPA Questionnaire 11Feb21.docx").

322. *Id.* at 3.

for DoD's AI Ethical Principles and to begin incorporating aspects of the principles in future call orders."[323]

The JAIC, of course, is acutely aware that ethical AI cannot be captured in "Yes or No" terms. In fairness, perhaps the industrial base is not yet prepared to compete for government contracts on ethical AI grounds. Or perhaps the government is not ready to evaluate vendor proposals on those grounds. Either way, this early snapshot exposes the current gap between ethical AI in principle and practice, which this Article's recommendations aim to bridge.[324]

As is, the government's needs may be underserved. Worse still, the reduction of ethical AI to yes/no box ticking might be self-defeating if prospective vendors perceive, rightly or wrongly, that the government is not treating ethical AI as a *differentiator* in sourcing decisions. Per the RFP instructions, answering "No" to the ethical AI prompt will automatically disqualify prospective vendors from consideration.[325] Presumably, therefore, all responsive vendors checked "Yes." But without more particulars, it is far from clear how the JAIC can or will differentiate among competing proposals on ethical AI grounds when awarding blanket purchase agreements under the RFP. Nor is it clear how customer agencies can or will do so at the call-order level. At minimum, however, the government will need to ensure that vendor proposals that reflect the costs of ethical AI will not be competitively *disadvantaged* (which potentially could occur, for example, if price is treated as an evaluative criteria, but ethical AI is not). Surely, this is not what the government intends.

Emphatically, the JAIC's leadership in acquiring ethical AI is commendable, and the constructive critique here is not meant to suggest otherwise. Rather, the point is that even the government leaders in this space have a long haul ahead. Requiring contracting officials to include ethical AI among the evaluation criteria, and differentiating vendors on that basis, will be pivotal to progress. Acquiring ethical AI is something

323. *See* JAIC INSTRUCTIONS, *supra* note 320, at 6–7.

324. In May 2021—after the RFP for blanket purchase agreements was issued—the Deputy Secretary of Defense issued a memorandum titled "Implementing Responsible [AI] in the Department of Defense." *See generally* Memorandum from Kathleen H. Hicks, *supra* note 7. Among other things, the memorandum calls for the incorporation of ethical AI principles into the DoD's AI requirements and acquisition processes. To that end, the memorandum directs the JAIC to take specific steps, including the establishment of a Responsible AI Working Council that will provide recommendations to integrate ethical AI into the acquisition lifecycle. *See id.* at 3 (instructing the RAI Working Council to "provide recommendations on the integration of RAI into the AI acquisition requirements, on process, and on any policy modifications to enable RAI considerations within existing supply chain risk management practices."). As of this writing, those recommendations are pending.

325. *See* JAIC INSTRUCTIONS, *supra* note 320, at 6–7.

that the government should insist upon. Otherwise, vendors cannot be expected to compete upon that basis.

2. Responsibility Determination

Additionally, or alternatively, ethical AI criteria could be integrated into contracting officials' pre-award responsibility determinations of prospective vendors. Like all the forgoing recommendations, this one builds upon pre-existing regulatory structure.

By way of background, longstanding procurement law requires vendors to satisfy a set of "responsibility" requirements,[326] which fall into three general categories. *First*, contracting officials must assess whether prospective vendors can fulfill the contract in a timely and satisfactory manner.[327] Toward those ends, a prospective vendor must demonstrate that it: has adequate financial resources; can meet the delivery schedule; has a satisfactory record of past performance; has a satisfactory record of business integrity and ethics; and has the necessary organization, technical skills, and production capabilities to perform the contract.[328] *Second*, prospective vendors must be "otherwise qualified and eligible to receive an award under applicable laws and regulations."[329] This general requirement, in turn, incorporates a range of socioeconomic polices effectuated through procurement law.[330] For example, a potential awardee must be deemed ineligible if it has not complied with federal equal employment opportunity requirements,[331] or fails to agree to an acceptable subcontracting plan with small businesses under the contract.[332] *Third*, the government may establish "special standards of responsibility" in contract solicitations when "necessary for a particular acquisition or class of acquisitions."[333] Per regulation, special standards "may be particularly desirable when experience has demonstrated that unusual expertise" is needed "for adequate contract performance."[334]

---

326. FAR 9.103(a) ("Purchases shall be made from, and contracts shall be awarded to, responsible prospective contractors only."). A vendor's failure to meet the responsibility threshold is disqualifying as a matter of law. *Id.* at 9.103(b) ("No purchase or award shall be made unless the contracting officer makes an affirmative determination of responsibility."); *id.* at 9.103(c) ("A prospective contractor must affirmatively demonstrate its responsibility . . . .").

327. *See* KATE M. MANUEL, CONG. RSCH. SERV., R40633, RESPONSIBILITY DETERMINATIONS UNDER THE FEDERAL ACQUISITION REGULATION: LEGAL STANDARDS AND PROCEDURES 6–13 (2013) (providing explanations of the FAR's responsibility standards and processes); Ryan Co. v. United States, 43 Fed. Cl. 646, 651 (1999).

328. FAR 9.104-1 (2021).

329. *Id.* at 9.104-1(g) (2021).

330. *See* MANUEL, *supra* note 327, at 5.

331. *See* FAR 22.802 (2021); *id.* at 52.222–26 (2021); 41 C.F.R. § 60-1.1 (2021).

332. *See* 15 U.S.C. § 637(d)(4)(C); *see also* MANUEL, *supra* note 327, at 9 (listing these and other collateral responsibility requirements).

333. FAR 9.104-2 (2021).

334. *Id.*

The foregoing responsibility framework offers several points of ingress for ethical AI. Here are some possibilities, keyed to the typology above. *First*, as a general performance standard, ethical AI could be factored into a prospective vendor's "necessary organization, experience . . . [and] technical skills" to perform the contract, or the vendor's "record of integrity and business ethics."[335] *Second*, or alternatively, ethical AI standards could be established (e.g., by NIST), and then be required by law for vendors doing business with the federal government.[336] *Third*, ethical AI can be the basis for special standards of responsibility in connection with a particular contract or class of acquisitions.

For example, prospective vendors could be required to allow independent third-party auditing of their proposed AI solutions. Vendors could also be required to waive trade-secrecy claims under certain conditions or in certain contexts (e.g., in adjudicatory settings where the government must provide an explanation for an AI output). Furthermore, to support a diverse and robust AI ecosystem, prime contractors could be required to agree to subcontracting plans that include small businesses, or socioeconomically disadvantaged businesses, with ethical AI expertise.[337]

The foregoing suggestions come with important qualifiers and caveats. To start, intellectual property (IP) rights can be a major sticking point for AI vendors.[338] Indeed, for many small businesses and startups, trade secrets are their most valuable assets.[339] Thus, responsibility requirements related to IP should be limited to what is foreseeably

335. *See id.* at 9.104-1 (2021); *see also supra* notes 327–28 and accompanying text (discussing general performance standards).

336. *See supra* notes 329–44 and accompanying text.

337. 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 (2020); *see also* 15 U.S.C. § 637(a)(5) (defining "[s]ocially disadvantaged individuals" under the Small Business Act as "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"); FAR 19.15 (2019) (discussing Woman-Owned Small Business Program); Exec. Order No. 13,985 § 7, 86 Fed. Reg. 7009, 7011 (Jan. 25, 2021) ("Government contracting and procurement opportunities should be available on an equal basis to all eligible providers of goods and services.").

338. *See generally* Rob Kitchin, *Thinking Critically About and Researching Algorithms*, 20 INFO. COMMC'N. & 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."); 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 (2003) (providing a review of the relevant contracting requirements and industry concerns around IP provisions in government contracts that depart from general commercial terms).

339. *See* Kitchin, *supra* note 338, at 20.

necessary.[340] If the government has no better options, it may need to pay vendors for their trade secrets, whether upfront, on a contingency basis, or otherwise. But, so long as a critical mass of innovative and responsible market participants are available to compete for the work, the disinclination of *some* vendors to meet the government's legal and operational needs is a poor reason to lower the bar for *all*.

Another potential friction is the government's use of procurement requirements to advance collateral socioeconomic policies. This is a longstanding and generalized concern, with varying degrees of intensity depending on context.[341] However, in this context, objections of this sort miss the mark. Procurement requirements pertaining to ethical AI are not, in the main, collateral socioeconomic policies detached from a vendor's ability to execute the contract. Rather, ethical AI requirements are foremost directed at meeting the government's operational and legal needs. The fact that ethical AI requirements may also promote socioeconomic objectives is a testament to procurement law's breadth of purpose and regulatory value. Lest there be any doubt, the procurement system's express and overarching objective is to "deliver . . . the *best value* product or service to the [government] customer, while maintaining the *public's trust* and fulfilling *public policy* objectives."[342] Ethical AI strikes all of those notes.[343] Conversely, spending many millions (or billions) of taxpayer dollars for ethically agnostic AI solutions would be antithetical to best value, could undermine public trust, and leave public policy objectives unfilled.

If a prospective vendor is unable or unwilling to satisfy responsibility thresholds relating to ethical AI, then the government should be required to select a competitor that will. And if no vendor will, then the

340. There are well-established industry practices (e.g., nondisclosure agreements with liability provisions) and federal trade secrecy laws that can be utilized to safeguard vendors against trade secrecy misappropriation. *See, e.g.*, Defend Trade Secrets Act of 2016, Pub. L. No. 114-153, 130 Stat. 376 (codified at 34 U.S.C. § 41310 (2017)).

341. The use of procurement law to influence socioeconomic policy has been the subject of controversy, at various times at to various degrees. *See, e.g.*, Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 211 (1995) (analyzing the validity of the Federal Government's practice of providing general contractors on federal projects with incentives to hire subcontractors operated by socially and economically disadvantaged individuals); OFF. OF SEN. ELIZABETH WARREN, BREACH OF CONTRACT: HOW FEDERAL CONTRACTORS FAIL AMERICAN WORKERS ON THE TAXPAYER'S DIME 2 (2017). For generations, the government has leveraged the procurement system to advance national policy objectives, and has generally been able to do so because of its special relationship with federal contractors and raw spending power. *See* Schooner, *supra* note 274, at 108–09 (explaining that "government spending can influence behav[ior] and infuse growth in communities and economic sectors").

342. FAR 1.102(a) (2021) (emphasis added).

343. *See* Exec. Order No. 13,960, 85 Fed. Reg. 78,939, 78,940 (Dec. 8, 2020) (instructing agencies to abide to ethical AI principles when acquiring AI); *see also supra* Section II.C.2 (discussing ethical AI initiatives within the federal government).

government should rethink whether a market solution is appropriate, endeavor to fill the need in-house, or seek other (non-AI) solutions.[344]

D. *Contract Performance: Pathways and Pitfalls*

Thus far, the discussion has focused on acquisition planning, market solicitation, proposal evaluation, and contract award. This final section turns to contract performance.[345] Before proceeding, it must be emphasized that the challenges and opportunities for acquiring ethical AI will depend on what transpires during the preceding phases. But, even if the recommendations above are duly implemented, contract performance will be key to mission success.

The ethical AI challenges during this phase depend on countless variables. One important organizing distinction is between (1) commercial off-the-shelf (COTS) and (2) customized AI solutions. Although necessarily partial, this dichotomy provides useful framing to address various challenges that may arise or manifest after a contract is awarded.

1. COTS AI Solutions

COTS acquisitions aspire to transactions in the commercial market.[346] For that reason, COTS items are offered to the government "without modification."[347] Moreover, COTS vendors are relieved of certain regulatory terms and conditions unique to government contracting.[348] This lowers the market barriers for commercial vendors that otherwise might not do business with the government. And by reducing red tape, COTS acquisitions are generally more efficient for the government as well.

The regulatory pretenses around COTS acquisitions are fundamentally pragmatic: if the product is good enough for commercial

344. The use of RFIs, discussed above, is one way the government can gauge whether sufficient competition exists in the market to meet the government's legal, operational, and sustainment needs. *See generally supra* Section IV.B (discussing market solicitations and offering illustrations).

345. *See generally* FAR Part 42 (2021) (governing contract administration and audit services).

346. *See* Christopher F. Corr & Kristina Zissis, *Convergence and Opportunity: The WTO Government Procurement Agreement and U.S. Procurement Reform*, 18 N.Y.L. SCH. J. INT'L & COMP. L. 303, 314–15 (1999) (discussing the statutory genesis and motivations behind COTS acquisitions).

347. *See* FAR 2.201 (2021) (defining "[c]ommercially available off-the-shelf item").

348. *See id.* at 12.503–04 (2021) (listing laws that are not applicable to federal contracts and subcontracts for commercial items); *id.* at 12.505 (2021) (listing additional laws that are not applicable to COTS items).

consumption, then it should be good enough for the government too.[349] Through an ethical AI lens, however, these pretenses are precarious. Unlike virtually all other COTS solutions, the risks of AI product failure and related harms have *not* been meditated by regulatory or market forces. Quite the contrary, the design and development of commercially available plug-and-play AI solutions are virtually unregulated.

In these routinized transactions, moreover, the government will "acquire only the technical data and the rights in that data customarily provided to the public with a commercial item or process."[350] Because those data rights are generally quite limited, the government may forfeit crucial opportunities to address its ethical AI needs, both at the time of purchase and thereafter. Although COTS products can be configured to agency needs, doing so can lead to long term support and maintenance challenges, insofar as customized functionality is not supported by the COTS vendor.[351]

2. Customized AI Solutions

Customized AI systems may be better suited for the government's missions and lifecycle needs but give rise to different challenges. At the outset, traditional "waterfall" acquisition pathways are not viable for custom AI solutions. Under a typical waterfall approach, the government's technical and design requirements are fixed at the time of contracting.[352] Given the complexity of the AI development process, however, it may be impossible for the agency to specify conditions of performance at the time of contracting. Even if possible, front-loading decisions about system features and configurations is antithetical to the trial-and-error nature of AI development.

For these reasons, "agile" acquisition methodologies are better suited for custom AI solutions. Agile methodologies are characterized by incremental, modular, and iterative processes in which software is produced in close collaboration with the end user.[353] Information obtained during these frequent iterations allow developers to respond quickly to feedback from agency customers, thus potentially reducing sociotechnical, legal, and programmatic risk. Moreover, modular

349. *Cf.* Laura Gerhardt et al., *When to Use Commercial Off-the-Shelf (COTS) Technology*, 18F BLOG (Mar. 26, 2019), https://18f.gsa.gov/2019/03/26/when-to-use-COTS/ [https://perma.cc/SN2N-DN2Q] (outlining some general considerations for agencies to consider when choosing between COTS and customized software solutions).

350. FAR 12.211 (2021) (technical data rights); *see also id.* at 12.212 (2021) (computer software documentation).

351. *See* Gerhardt et al., *supra* note 349.

352. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-20-590G, AGILE ASSESSMENT GUIDE: BEST PRACTICES FOR AGILE ADOPTION AND IMPLEMENTATION 7 (2020), https://www.gao.gov/products/GAO-20-590G [https://perma.cc/KZ7S-F3V6].

353. *See id.*

contracting vehicles "provide an opportunity for subsequent increments to take advantage of any evolution in technology or needs that occur during implementation," and can "reduce risk of potential adverse consequences on the overall project by isolating and avoiding custom-designed components of the system."[354]

The built-in flexibilities of agile processes may be well suited for AI development—certainly more so than a waterfall approach. But better is not sufficient. Like so much else in AI's domain, agile methodologies will need to be retrofitted and retooled for the unique challenges of acquiring ethically designed AI systems.

This is no small matter. Even for conventional (non-AI) software acquisitions, agile approaches require skillsets, resources, and institutional buy-in that many agencies currently lack. In 2020, the Government Accountability Office (GAO) chronicled an array of challenges that agencies have experienced using agile acquisition processes in the past.[355] For example, "teams reported difficulty collaborating closely or transitioning to self-directed work due to constraints in organization commitment and collaboration."[356] Moreover, GAO reported that some agency organizations "did not have trust in iterative solutions and that teams had difficulty managing iterative requirements."[357]

This GAO report does not directly address the unique challenges of AI systems, much less AI ethics. Nor do other recently issued government best-practices guides.[358] But it must be assumed that the government's

---

354. 48 C.F.R. § 39.103 (2021).

355. *See* GAO-20-590G, *supra* note 352, at 14–16.

356. *Id.* at 14.

357. *Id.*; *see also* U.S. GOV'T ACCOUNTABILITY OFF., GAO-16-467, IMMIGRATION BENEFITS SYSTEM: US CITIZENSHIP AND IMMIGRATION SERVICES CAN IMPROVE PROGRAM MANAGEMENT 24 (2016), https://www.gao.gov/products/GAO-16-467 [https://perma.cc/R279-REQW] (reporting that the United States Citizenship and Immigration Service Transformation program was not setting outcomes for Agile software development); U.S. GOV'T ACCOUNTABILITY OFF., GAO-18-46, TSA MODERNIZATION: USE OF SOUND PROGRAM MANAGEMENT AND OVERSIGHT PRACTICES IS NEEDED TO AVOID REPEATING PAST PROBLEMS 57 (2017), https://www.gao.gov/products/GAO-18-46 [https://perma.cc/CS84-J42U] (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).

358. The U.S. Digital Service and the General Services Administration (GSA) have likewise championed agile methodologies for acquisitions of customized software. *See, e.g.*, OFF. OF MGMT. & BUDGET, CONTRACTING GUIDANCE TO SUPPORT MODULAR DEVELOPMENT 7, 12 (2012), https://obamawhitehouse.archives.gov/sites/default/files/omb/procurement/guidance/modular-approaches-for-information-technology.pdf [https://perma.cc/4S9G-JWBC]; HANDBOOK FOR PROCURING DIGITAL SERVICES USING AGILE PROCESSES, at i, 1 (2014), https://playbook.cio.gov/assets/TechFAR%20Handbook_2014-08-07.pdf [https://perma.cc/R3YE-FBZJ]; GEN. SERVS. ADMIN., DE-RISKING GOVERNMENT TECHNOLOGY: FEDERAL AGENCY FIELD GUIDE 7, 10–12

agility challenges will be amplified in the AI context, given the data-centric dynamics, value-laden judgments, cross-disciplinarity, procedural discipline, and institutional buy-in required to acquire ethical AI solutions.[359] Indeed, under the status quo, agile methodologies may cut against the grain of ethical AI. As Joshua Kroll explains, agile workflows and development sprints can lead to path-dependent and shallow thinking about the social implications of technical designs.[360] Moreover, Deb Raji et al. note that agile AI development presents unique auditing challenges, especially if the developers are not scrupulous about managing the data and documenting key decisions throughout the iterative process.[361]

To be sure, synching ethical AI with agile methodologies is an ongoing challenge. While some promising approaches to this challenge exist,[362] none can lay claim to standard industry practice. Regardless, what may work in the private sector may not translate in the government sector. The government's general struggles with agile methodologies, and its AI talent shortages, are again of relevant concern. But, in addition, the government's agility is limited by regulatory constraints. For example, the government cannot lawfully outsource "inherently governmental functions."[363] Although this limitation is notoriously fuzzy and forgiving, it could—and arguably should—prevent the government from devolving policy choices to vendors in the AI development process.[364] Moreover, the government is generally prohibited from entering into "personal services contracts."[365] This is a fuzzy and forgiving standard as well, but it generally forbids the government from

(2020), https://derisking-guide.18f.gov/assets/federal-field-guide-a245c3a7dcd0a24f619b458fd51e1e490f2299023fd1bd13fddc87318e67cf03.pdf [https://perma.cc/DE6S-QJUA]. These best-practices guides are generalized for software. Thus, they do not speak to the special challenges of AI development.

359. Raji et al., *supra* note 96, at 5 ("The design, prototyping and maintenance of AI systems raises many unique challenges not commonly faced with other kinds of intelligent systems or computing systems more broadly.").

360. Joshua A. Kroll, *The Fallacy of Inscrutability*, 376 PHIL. TRANSACTIONS ROYAL SOC'Y 1, 5 (2018), http://dx.doi.org/10.1098/rsta.2018.0084 [https://perma.cc/7EC4-UT6Z].

361. Raji et al., *supra* note 96, at 4.

362. *See, e.g.*, Dorian Peters, Karina Vold, Diana Robinson & Rafael A. Calvo, *Responsible AI—Two Frameworks for Ethical Design Practice*, 1 IEEE TRANSACTIONS ON TECH. & SOC'Y 34 (2020).

363. *See* FAR 7.503 (2021); *see also id.* at 37.104(c)(2) ("Each contract arrangement must be judged in the light of its own facts and circumstances, the key question always being: Will the Government exercise relatively continuous supervision and control over the contractor personnel performing the contract.").

364. *See generally* KATE M. MANUEL, DEFINITIONS OF "INHERENTLY GOVERNMENTAL FUNCTION" IN FEDERAL PROCUREMENT LAW AND GUIDANCE (2014), https://sgp.fas.org/crs/misc/R42325.pdf [https://perma.cc/QD7J-E7YW] (analyzing the bounds of these constraints).

365. *See id.* at 37.104 (2021).

micromanaging vendors during contract performance.[366] These regulatory bounds are seldom a problem for government acquisitions. Yet they may prove to be in this context, whether because of the value judgments embedded in AI technologies, or the process by which those decisions are made.

To put it mildly, it is too soon to know—and dangerous to assume—that industry approaches for integrating AI ethics into agile workflows will be effective, scalable, and suitable for government contexts. To set the right conditions, agencies will first need to be much more attentive to the risks inherent in ethics-agnostic agile methodologies, and then foster market competition around that particular challenge.

With those related aims, the government's market solicitations should contain prompts and requirements that explicitly tether ethical AI to agile methodologies. Just for example, the government could ask potential vendors to:

- ❖ Identify agile methodologies that your company has used that incorporate ethical AI principles, and provide two (or three) examples that demonstrate capability in those methodologies.

- ❖ Describe any challenges or lessons learned from those past experiences, and any anticipated challenges, strategies or solutions that your company might implement in future work.

These prompts are no panacea; they are merely preludes to contract performance. But if the government expects ethical AI *and* agile methodologies, the government must recognize that the sum of the two is greater than its parts. Moreover, the government should make sourcing decisions that account for the difference, and secure funding to pay for that difference. Otherwise, the government will have plenty of industry partners that are fluent in agile AI development, and "willing to incorporate" ethical AI principles,[367] but that have no plans or protocols to synchronize these ambitions.

---

366. *See id.* at 37.104(c)(2) ("Each contract arrangement must be judged in the light of its own facts and circumstances, the key question always being: Will the Government exercise relatively continuous supervision and control over the contractor personnel performing the contract."); *see also* ASI Gov't, *A COR's Guide to Personal Services Contracts* (2011), https://www.navsup.navy.mil/site/public/flcph/documents/contracting/cor_guides/A_CORs_Guide_to_Personal_Services_Contracts.pdf [https://perma.cc/62BX-DH3U] (providing guidance in navigating these murky constraints).

367. *See* JAIC QUESTIONNAIRE, *supra* note 321, at 3 (requiring such willingness as a condition of a procurement proposal).

***

The foregoing discussion has provided a blueprint for acquiring ethical AI which spans the procurement process. Still, and again, it must be emphasized that parchment policies are not enough: acquiring ethical AI requires intentionality, additional resources, industry collaboration, and government coordination. Moreover, robust implementation of this Article's recommendations will depend on a cadre of talented and dedicated personnel that can execute the mission. It is one thing if the government does not have the in-house capacity to meet its demand for ethical AI solutions. It is quite another matter, however, if the government does not have a federal acquisition workforce with the skills, resources, and cultural commitment to *acquire* ethical AI from the private market. In this regard, the NSCAI has admonished that agencies which "rely solely on contractors for digital expertise will become incapable of understanding the underlying technology well enough to make successful acquisition decisions independent of contractors."[368]

As this Article nears completion, there are some promising signs of meaningful government progress toward acquiring ethical AI. Of special note, the JAIC has recently launched a program to align AI acquisitions with ethical AI as "part of a holistic approach that focuses not only on the technology, but on organizational operating structures and culture to advance Responsible AI within the DoD."[369] The GSA's AI Center of Excellence, which provides governmentwide acquisition and development support, is also championing ethical AI in its offerings.[370] More generally, a recently introduced Senate bill, titled the "Artificial Intelligence Training for the Acquisition Workforce Act,"[371] would establish an AI training program to ensure that acquisition personnel,

368. NSCAI FINAL REPORT, *supra* note 11, at 123.

369. Press Release, JAIC Pub. Affs., Joint Artificial Intelligence Center to Pilot a Responsible AI Procurement Process (July 27, 2021), https://www.ai.mil/news_07_27_21-jaic_to_pilot_a_responsible_ai_procurement_process.html [https://perma.cc/333Y-AV45]. Moreover, the program aspires to "establish clear guidance and expectations for those who are interested in working with DoD to ensure that they are providing AI systems designed, developed, deployed, and used responsibly." *Id*.

370. *See* A.I. CTR. OF EXCELLENCE, GEN. SERVS. ADMIN., COE GUIDE TO AI ETHICS 1, https://coe.gsa.gov/docs/CoE Guide to AI Ethics.pdf [https://perma.cc/3TJ6-ZFVX]; *Accelerate Adoption of Artificial Intelligence to Discover Insights at Machine Speed*, A.I. CTR. OF EXCELLENCE, https://coe.gsa.gov/coe/artificial-intelligence.html - coe-updates [https://perma.cc/5MLP-EVEL]; A.I. CTR. OF EXCELLENCE, GEN. SERVS. ADMIN., PROMOTE ADOPTION OF MODERN PRACTICES TO INCREASE SUCCESS OF IMMEDIATE & FUTURE MODERNIZATION EFFORTS 1 (2021), https://coe.gsa.gov/docs/2020/CoE Innovation Adoption Service Catalog 2021.pdf [https://perma.cc/P76Z-8MDE]; *see also* Dave Nyczepir, *IT Centers of Excellence Program Is Signed into Law*, FEDSCOOP (Dec. 3, 2020), https://www.fedscoop.com/gsa-centers-of-excellence-codified/ [https://perma.cc/X26N-828P].

371. S. 2551, 117th Cong. (2021).

program managers, system evaluators, and others agency officials, have "knowledge of the capabilities and risks associated with AI."[372] Time will tell, but initiatives of this sort are precisely what is needed across government.

CONCLUSION

It is encouraging that the United States has committed to ethical 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."[373] But proselytizing is not actualizing. If federal officials are truly committed to ethical algorithmic governance, then the federal procurement system must be recalibrated for that purpose. This Article has provided a principled and pragmatic agenda for acquiring ethical AI that future work can utilize and build upon.

372. *Id.* § 2(b)(2)–(3).
373. Exec. Order No. 13,859, 84 Fed. Reg. 3967, 3967 (Feb. 14, 2019).