ADJUDICATING ALGORITHMS: ACCOUNTABILITY IN REGULATION OF SURVEILLANCE, PRIVACY, AND DISCRIMINATION

Peter Margulies†

TABLE OF CONTENTS

INTRODUCTION 2
I. PROBLEMS IN THE DIGITAL DOMAIN 6
A. *Cybersecurity: Data Breaches and Disinformation* 6
B. *Mass Surveillance* 8
C. *Algorithmic Bias* 9
D. *Functional Tradeoffs* 12
E. *Acknowledging Baselines* 13
F. *Regional Tensions* 14
II. MODELS OF REGULATION 15
A. *Prohibition* 15
B. *Regulatory Requirements* 16
1. Setting Standards 16
2. Assessments 17
3. Transparency and Disclosure 17
4. Use Restrictions 18
III. ADJUDICATION'S RISKS AND BENEFITS 20
A. *The Posture of Adjudication About Algorithmic Activity* 20
B. *Adjudication's Risks* 21
C. *A Programmatic Approach to Adjudication* 23
D. *The Meta OB's Programmatic Pivot* 24
1. Programmatic Perspective and the *Gender Identity* Decision 24

† Professor of Law, Roger Williams University. B.A., Colgate University; J.D., Columbia Law School.

2. *Cross-Check* and Favoritism in Content Moderation....26
3. Comparing Adjudication and Administration....29
IV. STEWARDSHIP AS A NORM....30
A. *Iterative Review*....30
B. *Layered Accountability*....32
C. *Institutionalized Opposition*....35
V. APPLYING THE STEWARDSHIP APPROACH....36
A. *The U.S. Data Protection Review Court*....36
B. *Data Breaches and Disinformation*....42
1. Stewardship, Data Breaches, and Privacy....43
2. Stewardship and the Dilemmas of Combating Online Disinformation....46
C. *Discrimination in Credit, Marketing, Employment, and Housing*....48
CONCLUSION....50

INTRODUCTION

The movement for accountable algorithms has attained critical mass.[1] That momentum includes a range of areas where the collection of data plays a key role, including privacy, online disinformation, surveillance, and screening for credit, housing, employment, and government benefits. For example, the White House has released an Artificial Intelligence (AI) Bill of Rights that outlines standards and recourse for a host of AI applications that touch human needs and endeavors.[2] Assessments, disclosure, and procedures for filing

---

1 *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); *see also* Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1552–63 (2019) (discussing importance of combining different modes of accountability, including government oversight and individual remedies, based in part on exploration of European Union (EU) experience with General Data Protection Regulation (GDPR)).

2 *See* THE WHITE HOUSE, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE (2022), [hereinafter WHITE HOUSE, AI BILL OF RIGHTS] https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf [https://perma.cc/5JPJ-GX4G]. The Biden administration has issued rules establishing a Data Protection Review Court (DPRC) that will hear complaints from European Union (EU) residents that U.S. surveillance has wrongly targeted them or incidentally collected their communications. *See* Data Protection Review Court, 87 Fed. Reg. 62303 (Oct. 14, 2022) (to be codified at 28 C.F.R. pt. 201). The Meta (formerly Facebook) Oversight Board (OB) has issued decisions that find bias in both human and algorithmic procedures for combating disinformation. *Case Decision 2022-009-IG-UA and 2022-010-IG-UA*, OVERSIGHT BD. (Jan. 17, 2023) [hereinafter

complaints about abuse are frequent features in this turn toward accountability.[3] However, at least in the United States, accountability is still a set of tropes, rather than a consistent approach across domains.[4]

Even among scholars proposing reforms, disagreements and omissions are frequent calling cards. Scholars are divided about the utility of procedures for review of individual complaints, as ongoing controversy over the value of the Meta Oversight Board (OB) reveals.[5] In addition, few, if any, accountability frameworks respond to the ubiquity of functional trade-offs between different values or even within values. For example, attempts to make algorithms more transparent can reduce their accuracy.[6] This Article outlines a consistent approach, which the Article calls stewardship,[7] for regulation of data abuse.

Gender Identity and Nudity], https://www.oversightboard.com/decision/BUN-IH313ZHJ [https://perma.cc/6QE3-NJGZ]; *Policy Advisory Opinion on Meta's Cross-Check Program*, OVERSIGHT BD. (Dec. 6, 2022) [hereinafter *Cross-Check*], https://oversightboard.com/attachment/512630074120983 [https://perma.cc/34L9-BP5W].

3 *See* Kaminski, *supra* note 1, at 1548–51.

4 The EU has provided a more comprehensive approach, although there are reasons to think that a wholesale importation of the EU model to the United States would not meet U.S. needs. *See* Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. REV. 1687, 1719–20 (2020) (discussing patchwork quilt of U.S. privacy regulation); Ira Rubinstein & Peter Margulies, *Risk and Rights in Transatlantic Data Transfers: EU Privacy Law, U.S. Surveillance, and the Search for Common Ground*, 54 CONN. L. REV. 391, 395–96 (2022) (proposing path to resolving the friction between the U.S. and EU on surveillance and privacy).

5 Meta is the new corporate name for Facebook. On the debate regarding the Meta OB, compare Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 535–37 (2022) (criticizing individual complaint mechanisms as distracting from a more productive systemic approach), with Lawrence R. Helfer & Molly K. Land, *The Meta Oversight Board's Human Rights Future*, 44 CARDOZO L. REV. 2233 (2023) (praising Meta OB as addressing both systemic and individual issues). *Cf.* Kate Klonick, *Of Systems Thinking and Straw Men*, 136 HARV. L. REV. F. 339, 343–53 (2023) (criticizing certain premises of Douek's critique).

6 *See* Cary Coglianese & Alicia Lai, *Algorithm v. Algorithm*, 71 DUKE L.J. 1281, 1312–13 (2022) (asserting that critics of algorithms focus unduly on the lack of explainability of certain AI models without considering the trade-off between explainability and accuracy; critics also do not acknowledge that human judgment can also be opaque in certain settings, such as reviewing large volumes of employment or credit applications, and that lack of explainability of algorithms may reflect temporary limits that data scientists will overcome over time).

7 The term "stewardship" has a more generic meaning in governance of algorithms, describing sound governance. *See* General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO), *Recommendation on the Ethics of Artificial Intelligence*, at 27–28, U.N. Doc. SHS/BIO/PI/2021/1 (Nov. 23, 2021) [hereinafter UNESCO, *Recommendation on the Ethics of Artificial Intelligence*]. The approach taken in this Article is more concrete and specific, outlining three crucial components of stewardship: procedural safeguards, public engagement, and an oppositional voice within data governance institutions. This more concrete model builds on the broader applications of stewardship that appear in the literature. Scholars have used this term in other contexts involving U.S. executive power. *See* Peter Margulies, *Taking Care of Immigration Law: Presidential Stewardship, Prosecutorial Discretion, and the Separation of Powers*, 94 B.U. L. REV. 105 (2014).

It is not surprising that unified approaches have not emerged regarding the whole spectrum of algorithmic activity, including foreign surveillance, data breaches, platform content moderation, and applicant screening. Each of these domains has a vast literature and disparate technical challenges.[8] Moreover, each has a different set of regulators and stakeholders. On the regulatory front, the U.S. Federal Trade Commission (FTC) has long used its authority to litigate against unfair and deceptive practices to combat data breaches that expose the personal information of millions of customers.[9] However, the FTC has no jurisdiction to monitor or regulate government surveillance, although the FTC's new initiative on privacy suggests that it is open to issuing privacy rules that will go beyond its traditional limited enforcement role. Similarly, U.S. agencies that enforce antidiscrimination laws have thus far had only a modest role in addressing algorithms that have discriminatory effects in access to credit, housing, or employment.[10] This balkanization also reflects the United States' traditional sectoral approach, in which regulation of individual sectors such as the power grid predominates, while no single agency has overarching jurisdiction.[11]

Time is overdue for an approach that will work across sectors and subject areas whose common feature is algorithmic activity. Others have called for a unified approach in the United States that resembles the approach of the European Union (EU).[12] However, these proposals often suffer from a hostile view of adjudication and skepticism toward the most prominent example of adjudication, the Meta OB.[13] This Article takes a fresh look at very recent decisions of the OB in the *Cross-Check* and *Gender Identity and Nudity* opinions recommending limits on bias in Meta's content moderation policies and practices.[14] It also articulates a broader view of stewardship that fits different sectors.

---

8 *See* Emily Berman, *When Database Queries Are Fourth Amendment Searches*, 102 MINN. L. REV. 577 (2017) (discussing surveillance); Douek, *supra* note 5 (discussing regulation of disinformation and hate speech on social media); Coglianese & Lai, *supra* note 6 (considering regulation of machine learning, including models used in screening credit, housing, and employment applications).

9 *See* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583 (2014).

10 *See* Talia B. Gillis, *The Input Fallacy*, 106 MINN. L. REV. 1175 (2022); Michael Selmi, *Algorithms, Discrimination and the Law*, 82 OHIO ST. L.J. 611 (2021).

11 *See* Eldar Haber & Tal Zarsky, *Cybersecurity for Infrastructure: A Critical Analysis*, 44 FLA. ST. U. L. REV. 516, 535–36 (2017) (discussing role of Federal Energy Regulatory Commission (FERC) in formulating and implementing cybersecurity policies for the energy sector).

12 *See* Kaminski, *supra* note 1.

13 *See* Douek, *supra* note 5, at 535–37.

14 *Cross-Check*, *supra* note 2; *Gender Identity and Nudity*, *supra* note 2, § 8.3.

Stewardship flows from the view that regulators, including entities regulating themselves, are fiduciaries for the data and communications they oversee. That fiduciary conception does not necessarily include the vast array of technical obligations that common law prescribes.[15] However, it does entail three robust norms: iterative review, layered accountability, and institutionalized opposition. Iterative review refers to adjudication that considers the validity of systems and programs. A programmatic approach installs practices as a benchmark, such as the Meta OB's recommendation of content-moderation policies that do not discriminate against marginalized groups.[16] Iterative review also monitors compliance with those benchmarks.[17] Layered accountability supplements adjudication with public disclosure and broad participation by interested parties.[18] Institutionalized opposition establishes a voice within each subject domain that counters the narratives of technology companies or the government.[19] To apply the stewardship model, this Article considers three challenging areas: (1) transatlantic data privacy and government surveillance; (2) data breaches and privacy in the private sector; and (3) regulation of applicant-screening algorithms regarding credit, housing, employment, and government benefits.

This Article refines earlier approaches in several ways. First, it provides a broad conception of programmatic adjudication encompassing both individual and aggregate cases. Some prominent scholars have derided individual adjudication of algorithmic activity as narrow and inconsequential.[20] This Article argues, in contrast, that adjudication counts as programmatic because of its effects on practices and discourse. Those effects matter more than the procedural vehicle that the adjudication takes. In making this argument, the Article relies on recent Meta OB decisions that previous scholarship has not had a chance to address. Second, this Article seeks to disrupt the silos that have hitherto contained analysis of algorithmic activity. Most previous works in the literature have addressed one topic such as privacy and corporate data

---

15 *See* Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1160–61 (2018); *cf.* Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497, 502–05 (2019) (suggesting that the theory of information fiduciaries, with its analog to common law evidentiary privileges linked to professions such as law and medicine, is too narrow to address issues posed by social media and other technology companies in the Information Age).

16 *Gender Identity and Nudity*, *supra* note 2, § 1.

17 *Cross-Check*, *supra* note 2, ¶¶ 123–186.

18 *Gender Identity and Nudity*, *supra* note 2, § 8.3 (discussing public comments received on importance of free online expression by marginalized groups).

19 *See* Peter Margulies, *Searching for Accountability Under FISA: Internal Separation of Powers and Surveillance Law*, 104 MARQ. L. REV. 1155, 1206–07 (2021).

20 *See* Douek, *supra* note 5, at 535–39.

practices, instead of the entire landscape, which also includes government surveillance and applicant-screening algorithms.[21] That focused approach can yield depth of insight but can also miss common threads linking topics together. Third, because this Article provides a comprehensive overview of the algorithmic landscape, it also addresses functional trade-offs between fields that much other work misses, such as the trade-off between preserving privacy from hackers' exploits and curbing government surveillance of hackers' activity. Through these three refinements, the Article helps inform current debates.

This Article has five Parts. Part I outlines the problems of government surveillance, cyber breaches and disinformation, and algorithmic screening. Part II outlines proposed measures for resolving those problems, including standards, disclosure, and assessments. Part III considers procedural safeguards, centering on the Meta OB's recent decisions. Part IV sets out the stewardship approach, centering on iterative review, layered accountability, and institutionalized opposition. Part V applies the stewardship model to the proposed Data Protection Review Court on EU complaints regarding U.S. government surveillance; private-sector data breaches and online disinformation; and algorithmic applicant screening.

I. PROBLEMS IN THE DIGITAL DOMAIN

Digital media have produced great benefits but have also spawned major risks. This Part concisely lays out current problems in cybersecurity and social media; government surveillance; and the use of algorithms in commerce and adjudication. After outlining these pervasive issues, the next Part discusses current approaches, including prohibition of these risky technologies and various modes of regulation.

A. *Cybersecurity: Data Breaches and Disinformation*

The internet is the world's hub for information, communication, and essential goods and services. Yet the internet also is vulnerable to manipulation by humans and the machines they control. Managing the internet's evolving risks is crucial for reaping its benefits.[22]

Internet intrusions form a very long menu. Distributed denial of service (DDoS) intrusions wield a vast array of computers (botnets) to

---

21 *See, e.g.*, Margulies, *supra* note 19 (discussing U.S. foreign intelligence collection).

22 *See* U.S. CYBERSPACE SOLARIUM COMM'N, REPORT 8–16 (2020), http://www.fdd.org/wp-content/uploads/2020/03/CSC-Final-Report.pdf [https://perma.cc/MT7J-3ULD].

flood websites with email or other communications, temporarily shutting down those sites.[23] Moreover, states and nonstate actors also can launch malicious software (malware) that can exfiltrate data for purposes of identity theft, pilfering of intellectual property, or espionage.[24] Hackers working with states or nonstate actors can target critical infrastructure such as the water supply, health care, or energy, holding those crucial services for ransom.[25]

Other problems in the same vein are just as daunting. States and others can use malware to covertly alter code or destroy data.[26] In addition, social media is vulnerable to large-scale information operations that use botnets to impersonate persons and groups and spread both hate speech and misinformation. Those information operations can distort democratic elections, as Russian entities sought to do in the 2016 U.S. presidential campaign.[27] Moreover, politicians, celebrities, and groups with political agendas can use social media for the same purpose.

The hate speech that viral social media posts generate can prompt discrimination, repression, and even genocide.[28] While some defend such posts as an exercise of free speech, a cascade of hostile and noxious social media speech can also inhibit the free speech of marginalized groups, driving them further into the shadows. Whistleblowers have detailed the scope of such problems and their relationship to the business models of social media companies.[29] But whistleblowing is a perilous exercise with scant legal protection. Furthermore, even whistleblowers provide an

---

23 *See* Andrea M. Matwyshyn & Miranda Mowbray, *Fake*, 43 CARDOZO L. REV. 643, 725–26 (2021) (discussing mechanics and operation of DDoS incursions).

24 *See* Daniel M. Filler, David M. Haendler & Jordan L. Fischer, *Negligence at the Breach: Information Fiduciaries and the Duty to Care for Data*, 54 CONN. L. REV. 105, 122–24 (2022) (discussing identity theft through hacking of computer networks and tort theories that could provide remedies to victims of these acts as well as less concrete harms to privacy).

25 *See* H. Justin Pace & Lawrence J. Trautman, *Mission Critical:* Caremark*, Blue Bell, and Director Responsibility for Cybersecurity Governance*, 2022 WIS. L. REV. 887, 892–93 (2022) (discussing ransomware attack against Colonial Pipeline, which supplies gasoline to many service stations on the East Coast).

26 *See* Thomas Eaton, *Self-Defense to Cyber Force: Combatting the Notion of 'Scale and Effect,'* 36 AM. U. INT'L L. REV. 697, 711–13 (2021) (discussing Stuxnet and other state efforts to destroy another state's data).

27 *See* Elena Chachko, *National Security by Platform*, 25 STAN. TECH. L. REV. 55, 68 (2021) (discussing 2016 election misinformation operations by Russia); Michael N. Schmitt, *"Virtual" Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law*, 19 CHI. J. INT'L L. 30 (2018); Sean Watts & Theodore Richard, *Baseline Territorial Sovereignty and Cyberspace*, 22 LEWIS & CLARK L. REV. 771, 790 (2018) (discussing Russian efforts to use human trolls to influence Ukrainian elections).

28 Richard Ashby Wilson & Molly K. Land, *Hate Speech on Social Media: Content Moderation in Context*, 52 CONN. L. REV. 1029, 1032–33 (2021).

29 *See Cross-Check*, *supra* note 2, ¶ 2 (discussing Facebook whistleblower Frances Haugen).

incomplete picture of threats to privacy and related harms, limited by their own access within the larger firm or entity.

B. *Mass Surveillance*

Privacy and free speech are also at risk due to the increased scope of government surveillance. Some uses of surveillance are necessary to counter threats to national security, public safety, public health, and privacy—understanding the tradecraft of hackers and holding them accountable requires information about hackers' practices.[30] Similarly, while the threat of terrorism has receded to a degree, states need the ability to track persons who, for ideological reasons, target civilians for harm. However, that imperative has led to substantial intrusions on privacy from government and from certain private companies that supply governments with surveillance tools.[31]

In practicing surveillance, the United States and other technologically sophisticated states scan huge numbers of communications by computers to ferret out evidence of terrorism, espionage, and proliferation of weapons of mass destruction.[32] Under section 702 of the Foreign Intelligence Surveillance Act (FISA), the United States has designated over 100,000 foreign individuals, entities, and electronic user accounts for surveillance.[33] Surveillance of those targets can entail acquisition of information on U.S. persons' communications.[34] While an independent federal court approves on an

---

30 *See* David E. Pozen, *Privacy-Privacy Tradeoffs*, 83 U. CHI. L. REV. 221, 229–32 (2016).

31 Asaf Lubin, *Selling Surveillance* (Maurer Sch. of L., Legal Stud. Rsch. Paper Series, Rsch. Paper No. 495, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4323985 [https://perma.cc/GA9G-84R8]. Tribunals such as the Court of Justice of the European Union have addressed the use of algorithms in mass surveillance to detect patterns of suspicious activity. *See* Joined Cases C-511/18, C-512/18 & C-520/18, La Quadrature du Net v. Premier Ministre, ECLI:EU:C:2020:791, ¶ 178 (Oct. 6, 2020); *see also* Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1282–83 (2018) (discussing use of algorithms in mass surveillance); Margaret Hu, *Crimmigration-Counterterrorism*, 2017 WIS. L. REV. 955, 993 (same); Peter Margulies, *Surveillance by Algorithm: The NSA, Computerized Intelligence Collection, and Human Rights*, 68 FLA. L. REV. 1045, 1055–56 (2016) (same).

32 United States v. Hasbajrami, 945 F.3d 641, 661–62 (2d Cir. 2019); United States v. Muhtorov, 20 F.4th 558, 585–89 (10th Cir. 2021).

33 *See* Robert Chesney, *The Supreme Court, 2021 Term—Comment: No Appetite for Change: The Supreme Court Buttresses the State Secrets Privilege, Twice*, 136 HARV. L. REV. 170, 203–05 (2022); Foreign Intelligence Surveillance Act of 1978 Amendments Act of 2008, Pub. L. No. 110-261, 122 Stat. 2436 (2008) (codified as amended at 50 U.S.C. § 1881a (2023)).

34 This occurs in so-called "one-end foreign communications" between U.S. persons and foreign section 702 targets. United States officials can pose queries related to U.S. persons to the vast database of section 702 information, although those queries must be reasonably likely to produce foreign intelligence information. *See Muhtorov*, 20 F.4th at 589; Berman, *supra* note 8, at

annual basis the procedures that the United States uses under section 702, the court does not approve in advance all foreign targets. Moreover, abuses have also occurred under a different FISA surveillance provision that requires a specific court order for surveillance of U.S. persons and certain other persons physically present in the United States. In this "traditional FISA" setting, officials failed to provide complete information to the court in a 2016 request to conduct surveillance on Carter Page, a onetime foreign policy advisor to then-candidate Donald Trump.[35]

C. *Algorithmic Bias*

In the domain of artificial intelligence and machine learning, problems of accuracy and bias have been salient.[36] Algorithms, including those premised on machine learning in which software draws new inferences from data based on data that developers have used to train the software, have major benefits. They can analyze material at speeds exponentially beyond human capabilities and discern patterns in a blizzard of variables that would be impossible for humans to parse.[37] That said, the bias, brittleness, and unintelligibility of certain algorithms trigger substantial concerns.

593–94. The Foreign Intelligence Surveillance Court (FISC), consisting of independent, life-tenured federal judges, monitors officials' compliance with rules governing section 702. Every year, the FISC must approve the Justice Department's certification that it is compliant with section 702. On these and other matters raising important and sometimes complex legal issues, the FISC has appointed an experienced lawyer to act as an amicus curiae, often countering the government's assertions. *See* Faiza Patel & Raya Koreh, *Amici Curiae in the FISA Courts: A Civil Liberties Impact Assessment*, 76 N.Y.U. ANN. SURV. AM. L. 499, 539–42 (2021) (suggesting that amici are too narrow in their arguments and limited by their own experience in past legal work for the intelligence community); *see also* U.S. DEP'T OF JUST. & OFF. OF THE DIR. OF NAT'L INTEL., SEMIANNUAL ASSESSMENT OF COMPLIANCE WITH PROCEDURES AND GUIDELINES ISSUED PURSUANT TO SECTION 702 OF THE FOREIGN INTELLIGENCE SURVEILLANCE ACT, REPORTING PERIOD DEC. 1, 2019–MAY 31, 2020, at 42–44 (2021), https://www.intelligence.gov/assets/documents/702%20Documents/declassified/23rd_Joint_Assessment_of_FISA_for_Public_Release.pdf [https://perma.cc/BT5K-B5SS] (reporting reduction in compliance incidents by FBI, including inappropriate querying of U.S. person information and safeguards imposed by FISC to spur compliance). On prospects for congressional reauthorization of section 702 and reforms that would address the concerns of civil liberties advocates on the Left and libertarians who often have conservative views of other legal issues, see Adam Klein, *FISA Section 702 (2008–2023?)*, LAWFARE (Dec. 27, 2022, 8:30 AM), https://www.lawfareblog.com/fisa-section-702-2008-2023 [https://perma.cc/NQ8X-NPBZ].

35 Margulies, *supra* note 19, at 1188–98; Bernard Horowitz, *FISA, the "Wall," and Crossfire Hurricane: A Contextualized Legal History*, 7 NAT'L SEC. L.J. 1, 80–98 (2019).

36 *See* Andrew Keane Woods, *Robophobia*, 93 U. COLO. L. REV. 51 (2022).

37 *See id.*

Bias occurs because machine learning relies on large datasets to train AI agents. Those sets can be unrepresentative of different populations, raising issues of bias. For example, if developers fail to include women's health conditions in a set of training data on medical diagnoses or fail to include faces of persons of color in a dataset that is training facial recognition technology, those results will not be accurate. The technology will compound inequity instead of easing it.[38] In addition, certain AI models are opaque: they do not provide a conventional verbal explanation for their outputs.[39]

Machine learning can also be brittle because the data that developers use to train algorithms often fails to include context. AI agents lack the vast store of contextual understandings that human beings—even young children—possess.[40] AI agents can become easily confused or attach disproportionate weight to factors that are unimportant. In the employment or housing context, an AI agent assessing an applicant's

---

38 *See* Kristin N. Johnson, *Automating the Risk of Bias*, 87 GEO. WASH. L. REV. 1214, 1239–42 (2019) (discussing AI bias in private sector); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 77 (2018) (unpacking flaws of facial recognition technology (FRT) in identifying women of color); David S. Rubenstein, *Acquiring Ethical AI*, 73 FLA. L. REV. 747, 775–77 (2021) (outlining manner in which government procurement protocols can combat risk of bias); Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117, 128–30 (2021) (discussing bias and approaches to assessing and ameliorating risk in private sector).

39 *See* Coglianese & Lai, *supra* note 6, at 1312–13 (2022); Cary Coglianese & Kat Hefter, *From Negative to Positive Algorithm Rights*, 30 WM. & MARY BILL RTS. J. 883, 892–93 (2022) (discussing discourse around transparency in AI); Jane R. Bambauer, Tal Zarsky & Jonathan Mayer, *When a Small Change Makes a Big Difference: Algorithmic Fairness Among Similar Individuals*, 55 U.C. DAVIS L. REV. 2337, 2388 (2022) (discussing virtues and risks of requiring explainable results for algorithms, including concern that more explainable models yield less accurate outputs); Katherine J. Strandburg, *Rulemaking and Inscrutable Automated Decision Tools*, 119 COLUM. L. REV. 1851, 1877–78 (2019) (discussing dynamics of explainability and how policymakers should assess trade-offs between explainability and other values). Developers and scholars are working toward making AI models more explainable, including methods such as allowing developers to submit counterfactual inputs to models to ascertain which inputs alter outputs. *See* Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1834–37 (2019) (discussing different conceptions of explainability and ways to implement each conception); Tobias Clement, Nils Kemmerzell, Mohamed Abdelaal & Michael Amberg, *XAIR: A Systematic Metareview of Explainable AI (XAI) Aligned to the Software Development Process*, 5 MACH. LEARNING & KNOWLEDGE EXTRACTION 78, 86–94 (2023) (providing typology of different approaches to explainable AI); Prashant Gohel, Priyanka Singh & Manoranjan Mohanty, Explainable AI: Current Status and Future Directions (July 12, 2021) (unpublished manuscript), https://arxiv.org/pdf/2107.07045.pdf [https://perma.cc/HJ75-WYJX] (discussing trends). The challenges of providing explanations for certain AI outputs also complicate the issue of accountability for AI decisions. *See* Margot Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 2012–40 (2021) (discussing ways of structuring grievance mechanisms regarding adverse impacts of AI decisions).

40 *See* Coglianese & Lai, *supra* note 6, at 1327–28; Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 CORNELL L. REV. 1875, 1889–90 (2020).

resume may attach disproportionate weight to a gap in the applicant's employment history, which could lead to rejecting the application at that stage. A more nuanced approach would consider other cues, including whether the applicant had experienced an adverse medical condition or a family emergency. A human application reviewer could more readily turn to the nuanced approach.[41]

AI brittleness springs from developers' blind spots. Often developers themselves take crucial elements of context for granted and fail to input data that will train the model to draw similar inferences. Consider the familiar red stop sign, which usually takes a hexagonal or octagonal shape and features the near-universal one-word traffic command against a red background. Next, consider whether specks, dirt, or small stickers on the sign change its meaning. Most human beings, including children in grade school, would interpret specks and dirt on the sign as having no impact on the sign's role. Even a dirty stop sign still directs the driver to halt and look both ways for crossing pedestrians or vehicular traffic. However, unless a developer has trained the AI model on this elementary, contextual understanding by inputting examples of clean and dirty stop signs with drivers stopping for both, the model may draw an inference that even a child would find to be bizarre.[42] An AI agent might interpret the presence of specks on a stop sign as connoting a substantive change in the sign's underlying function and perceive the stop sign as a yield sign. Carnage could result. In this sense, an algorithm is brittle, with the accuracy of its inferences hinging on a developer's ability to identify and compensate for gaps in the model's contextual understandings. Sometimes, as in our dirty stop sign example, the elementary nature of those contextual understandings contributes to the developer's complacency and failure to adjust the model's training set.

An insular clique of software developers may not even be aware of these issues. Sharpening awareness requires input from more stakeholders. In addition, if AI processes are significant for life chances

---

41 Humans also need training and review for compliance, since such shortcuts for high-volume applications may be tempting for human reviewers too. *See* Woods, *supra* note 36, at 69–70; Coglianese & Lai, *supra* note 6, at 1286–87.

42 *See* Yonathan A. Arbel & Shmuel I. Becher, *Contracts in the Age of Smart Readers*, 90 GEO. WASH. L. REV. 83, 121–24 (2022) (discussing so-called "adversarial attacks" in which developers fool models by making minor changes in images; those minor changes spur radical changes in the model's interpretation of the images, although even a child would still interpret the image correctly); David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability and the Administrative State*, 37 YALE J. ON REG. 800, 842–44 (2020) (discussing the use of adversarial examples to deceive enforcement algorithms); JP Vähäkainu, MJ Lehto & AJE Kariluoto, *Adversarial Attack's Impact on Machine Learning Model in Cyber-Physical Systems*, 19 J. INFO. WARFARE 57, 60–63 (2020) (explaining the technical basis for adversarial attacks and defenses against them).

and human flourishing, a range of stakeholders should participate as a matter of human dignity and representation, independent of the accuracy of the results that these processes reach.

D. *Functional Trade-offs*

One of the more difficult quandaries in the AI space is the existence of tension between solutions to the individual problems described above. For example, automation is useful in discerning imminent and ongoing cyber threats.[43] However, the use of automation can increase the risk of errors, due to the brittleness, bias, and lack of explainability of automated methods. An entity needs to have structural and operational institutions to address this tension. Finding the appropriate balance between the scale that automation provides and the need to enhance equity and fairness requires difficult trade-offs.[44]

Social media content moderation illustrates the depth of this dilemma. To curb hate speech and other harms on their platforms, social media companies have resorted to large-scale content moderation.[45] These content-moderation efforts can involve both automated and human review. Whatever the modality of review, content moderation can generate huge numbers of false positives, such as speech by marginalized groups that provides valuable information but that either a human or automated reviewer wrongly classifies as spreading injurious images and messages.[46] By the same token, social media companies can cause false-negative errors, in which their content-moderation policies fail to remove harmful posts. Most disturbingly, a recent decision by the Meta OB—an institution that Meta created and funds to address these issues—recently found that Meta systematically underenforced its content-moderation policies for politicians, celebrities, and others who generate traffic and revenue for Meta or could create regulatory headaches for the company.[47] The Meta OB has devoted substantial resources and attention to this question. However, in other contexts, rules barely acknowledge the tension, let alone seek to resolve it.

---

43 *See* Pozen, *supra* note 30.

44 *See* Selbst, *supra* note 38.

45 *See Cross-Check*, *supra* note 2, at 3; Douek, *supra* note 5, at 550–55.

46 *Gender Identity and Nudity*, *supra* note 2, § 8.3(III)(b).

47 *Cross-Check*, *supra* note 2, at 4–5; *see also* Douek, *supra* note 5, at 535–37 (discussing rationale for content moderation on social media and checks on effects of content moderation); Hannah Bloch-Wehba, *Automation in Moderation*, 53 CORNELL INT'L L.J. 41, 45 (2020); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018). *See generally* Klonick, *supra* note 5, at 351–54 (discussing trade-offs in content moderation); Helfer & Land, *supra* note 5 (same).

Contradictions between privacy, law enforcement, and the accuracy of algorithmic tools are pervasive. Consider algorithmic tools used to assess applications for credit, housing, employment, immigration status, and government benefits.[48] Certain design parameters of these tools will result in intrusions on privacy. For example, suppose algorithms do not merely consider an applicant's own materials but also search online data more broadly.[49] AI models trained to conduct credit checks can search a wide variety of sources with an attenuated relationship to the applicant's finances. Similarly broad ranging searches may occur in applications for housing, employment, and various government programs. A search that is not appropriately tailored to relevant information can lead to an accurate output—a correct decision about a loan application or other item sought by the applicant—but may do so at the price of substantial intrusions on the applicant's privacy and the privacy of others with ties to the applicant.

In addition, internal assessments of the accuracy of algorithmic tools for credit, housing, or other items can result in privacy intrusions. Those assessments require consideration of the details of individual cases. That individualized consideration increases privacy risks. Some privacy impacts of this consideration can be mitigated through informed consent of users, anonymization techniques that measure aggregate data without disclosing individual information, and use restrictions on persons and processes that collect the data. However, consent is a problematic concept, because of the importance of housing and other goods sought by individual applicants. When needs are great, individuals understandably feel pressure to agree to conditions that would allow them to fulfill those needs.[50] Moreover, the possibility of a breach or unauthorized use remains. In addition, even the mitigation measures discussed above require responsible parties to be mindful of the problem and formulate and execute solutions. Follow-up, including monitoring of implementation and input from stakeholders, is also imperative.

E. *Acknowledging Baselines*

A related problem occurs when evaluation of algorithms fails to address the performance of traditional alternatives, such as human decision-making. Human judgment is deeply flawed due to pervasive cognitive biases. For example, confirmation bias pushes individual

48 *See* Hu, *supra* note 31.

49 *See* Gillis, *supra* note 10.

50 *See* Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 630 (2019).

decision-makers to view all new information as bolstering a position that the decision-maker has already reached.[51] A sound analysis would compare the performance of algorithms to this or other relevant baselines.[52]

F. *Regional Tensions*

There are also significant regional tensions about the threats outlined above. Europe, in legislation such as the GDPR and the Digital Services Act, is both protective of privacy and skeptical about the use of algorithms to make decisions about credit, housing, and other goods.[53] Moreover, while the United States has long preferred a sectoral approach to privacy that focuses on individual elements of the economy, such as energy or health care, Europe takes a national and transnational approach with data protection authorities that operate across economic fields.[54] Although states that belong to the European Union can engage in bulk surveillance, the primary adjudicative body for the EU believes that U.S. surveillance ranges more broadly and lacks recourse that may be available in the EU.[55] As a result, the Court of Justice of the European Union (CJEU) has ruled on two occasions that data-sharing agreements between the United States and the EU do not meet EU requirements that non-EU parties have privacy protections that are substantially equivalent to EU safeguards.[56] These regional differences in perspectives about surveillance, algorithms, and safeguards against excesses in each space

---

51 DANIEL KAHNEMAN, OLIVIER SIBONY & CASS R. SUNSTEIN, NOISE: A FLAW IN HUMAN JUDGMENT 172 (2021) (explaining that confirmation bias "leads us, when we have a prejudgment . . . to disregard conflicting evidence"); Simone Galperti, *Persuasion: The Art of Changing Worldviews*, 109 AM. ECON. REV. 996, 1016 (2019) (observing that in empirical studies, "[e]xperimental subjects show 'a clear tendency to resist [falsifying] evidence' inconsistent with their hypotheses" (quoting JONATHAN ST. B. T. EVANS, BIAS IN HUMAN REASONING: CAUSES AND CONSEQUENCES 50 (1989))); *see also* Michael A. Bruno, Eric A. Walker & Hani H. Abujudeh, *Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction*, 35 RADIOGRAPHICS 1668, 1671–72 (2015) (noting that studies show that radiologists regularly make an X-ray reading error called "satisfaction of search," in which the doctor misses a key abnormality requiring medical attention "because of a failure to continue to search" after spotting an initial anomaly that is less serious; the doctor wrongly infers that any further abnormalities will merely bolster the preliminary diagnosis).

52 *See* Selmi, *supra* note 10, at 616–17; Coglianese & Lai, *supra* note 6, at 1288–304 (discussing flaws in individual and group decision-making).

53 *See* Kaminski & Urban, *supra* note 39.

54 *See* A. Michael Froomkin, Phillip J. Arencibia & P. Zak Colangelo-Trenner, *Safety As Privacy*, 64 ARIZ. L. REV. 921, 926–29 (2022).

55 Case C-311/18, Data Prot. Comm'r v. Facebook Ireland Ltd. (*Schrems II*), ECLI:EU:C:2020:559, ¶¶ 198–201 (July 16, 2020).

56 *Id.*; Rubinstein & Margulies, *supra* note 4.

have increased uncertainty about the future of transnational commercial and legal regimes.

II. MODELS OF REGULATION

Several different, albeit overlapping, models have been proposed for addressing the problems described above. This Part discusses the virtues and disadvantages of those models. It notes that few, if any, models deal adequately with the problem of functional contradictions. Moreover, most models have gaps, which this Part explores.

A. *Prohibition*

The prohibition approach has gained momentum in a number of areas. On this view, certain technologies and uses of technology, such as facial recognition technology (FRT), autonomous weapons that make targeting decisions without human preapproval, and AI in employment screening are so potentially dangerous, intrusive, or inaccurate that only prohibition will address the risks involved.[57] Debates about prohibition have the benefit of bringing to the surface important arguments about the risks of new technologies involving algorithms. However, prohibition has significant drawbacks.

First, advocates of prohibition usually fail to adequately acknowledge that the status quo prior to a new technology is often deeply flawed. That status quo involves greater reliability on a problematic decision mechanism: human judgment. Critiques of FRT stress the false positives and racial bias in some current applications of that technology.[58] However, the status quo prior to FRT—eyewitness identification—is notoriously inaccurate, especially in identification across racial groups.[59] Moreover, human judgment is often inconsistent within, between, and among individuals and collectives.[60] Second, prohibition stifles technological advances that can improve on current defaults.[61] Third,

57 *See* Coglianese & Hefter, *supra* note 39, at 893–95.

58 *Id.* at 893. On local efforts to ban FRT, see Ira S. Rubinstein, *Privacy Localism*, 93 WASH. L. REV. 1961, 2037–42 (2018).

59 *See* Coglianese & Hefter, *supra* note 39, at 899.

60 *See* KAHNEMAN, SIBONY & SUNSTEIN, *supra* note 51, at 24–27, 248–53 (observing that organizations either acquiesce in or fail to spot wide variations caused by irrelevant factors or "noise," among decision-makers); *id.* at 17 (discussing variations in decisions by juvenile court judges in cases with similar facts that parallel performance of a local football team: if the team loses, the judge issues more severe decisions).

61 *See* Woods, *supra* note 36.

prohibition is prone to demagoguery, including demagoguery with a biased edge: consider current measures designed to combat China's influence, from the clamor to shoot down a Chinese spy balloon over the United States to recent measures to ban the app TikTok from U.S. government phones.[62] In particular contexts, prohibition or a pause such as a moratorium on deployment of a technology may be appropriate. However, in most cases prohibition yields more costs than benefits.

B. *Regulatory Requirements*

While prohibition constitutes an outright ban on a technology, another approach subjects the technology's use to certain regulatory requirements.[63] Those requirements can entail standards, periodic audits and assessments, and mandated disclosure. The following paragraphs discuss these methods.

1. Setting Standards

To manage the risks of new technologies, setting standards is crucial.[64] Any new technology is susceptible to abuse, particularly because of the superior knowledge of its purveyors and purveyors' agendas, which can include acquiring money, fame, and power. Setting standards can curb those risks.

Recently, the Biden administration has taken a public stance favoring standards for AI.[65] For example, the administration's AI Bill of Rights cautions against algorithms that discriminate in their design or execution. UNESCO's recommendations also focus on diversity, urging

---

62 *See* Alex W. Palmer, *How TikTok Became a Diplomatic Crisis*, N.Y. TIMES (Dec. 20, 2022), https://www.nytimes.com/2022/12/20/magazine/tiktok-us-china-diplomacy.html [https://perma.cc/2QLP-UZUB].

63 *See* Andrew Guthrie Ferguson, *Surveillance and the Tyrant Test*, 110 GEO. L.J. 205, 212 (2021) (discussing the "technocratic" approach to regulation of surveillance technologies).

64 Coglianese & Hefter, *supra* note 39.

65 WHITE HOUSE, AI BILL OF RIGHTS, *supra* note 2; *see also* NAT'L INST. OF STANDARDS & TECH., U.S. DEP'T OF COM., NIST AI 100-1, ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) 13 (2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [https://perma.cc/W49X-BRXU] (providing technical detail on assessing AI risks along several axes, including validity and reliability); UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, *supra* note 7, at 18–20, 28 (stating general principles including respect for human rights, fairness, and nondiscrimination, and recommending the establishment of general standards).

states to forge policies that will promote inclusive datasets and equal access to the benefits of AI.[66]

Standards can include restrictions on the use of a technology. For example, proposals for the accreditation of companies making powerful hacking tools require that hacking be limited to serious national security threats, instead of allowing use for eavesdropping on journalists.[67] In U.S. national security surveillance, use restrictions bar targeting U.S. persons without a specific court order.[68] As another measure, licensing can also impose geographic limits on the dissemination of a technology, including its export to other states.[69]

2. Assessments

Regulation may also require that the creator or user of technology perform an assessment of the human rights and privacy impacts of a particular application.[70] An assessment can yield benchmarks to measure future progress or decline. As a follow-up, a regulator may also require that a user conduct periodic audits that measure harmful impacts, gauge compliance, and suggest remedies.[71]

3. Transparency and Disclosure

Regulation can also entail greater transparency, including required disclosures to government officials, investors, or the public.[72] Disclosure can mitigate information asymmetries that give an edge to companies and the developers of technology and impede public understanding. For example, a recent executive order issued by President Biden requires that government contractors that use software provide a software bill of materials (SBOM).[73] The SBOM discloses the origins of each component of software in the contractor's products and operations. The process of compiling and disclosing an SBOM generates greater awareness of the vulnerabilities in software. It also allows other groups to assess the risk of purchasing or using products that contain vulnerabilities.

---

66 UNESCO, *Recommendation on the Ethics of Artificial Intelligence, supra* note 7, at 28.

67 *See* Lubin, *supra* note 31, at 43.

68 Rubinstein & Margulies, *supra* note 4, at 414.

69 *Id.* at 418–47.

70 *See* Lubin, *supra* note 31, at 46; Selbst, *supra* note 38.

71 *See* Douek, *supra* note 5, at 601.

72 *See* Lubin, *supra* note 31; Rubinstein & Margulies, *supra* note 4, at 403.

73 Exec. Order No. 14028, 86 Fed. Reg. 26633 (May 12, 2021).

The U.S. Securities and Exchange Commission (SEC) has recently proposed heightened disclosure duties on cybersecurity for publicly traded companies.[74] Under the SEC's proposed rule, corporate disclosure regarding cybersecurity risks should include both disclosure of processes for assessing such risks and disclosure to investors of material cyber breaches.[75] Companies should make their disclosures accessible to a broad audience of stakeholders. To facilitate that goal, companies should use machine-readable XBRL for disclosures under the proposed rule.[76]

4. Use Restrictions

Regulators can also limit the use of particular technologies.[77] For example, proposals for the accreditation of companies making powerful hacking tools require that hacking be limited to serious national security threats, instead of allowing use for eavesdropping on journalists.[78] While national security surveillance by government agencies is not subject to express licensing, use restrictions that bar targeting U.S. persons without a specific court order are similar to use restrictions in the licensing domain. In crafting such restrictions, regulators should be aware that restrictions can be unduly prescriptive, locking in techniques that may become outdated in the near future.[79]

A closer examination of use restrictions highlights the pitfalls of some regulatory strategies. Use restrictions can vary from industry to industry or even firm to firm, leading to confusing and contradictory regulatory regimes. Consider the vaunted U.S. effort to "decouple" U.S. technology from China. Recognition of how the United States' interests diverge from China's is useful. But efforts to limit the operation of Chinese technology companies have been volatile, lurching from curbs on military uses of technology to broad-based attacks on products such

---

74 Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure, 87 Fed. Reg. 16590, 16603 (proposed Mar. 23, 2022) (to be codified at 17 C.F.R. pts. 229, 232, 239, 240, 249).

75 *Id.* at 16595–600.

76 *Id.* at 16603.

77 *See* Lubin, *supra* note 31; Rubinstein & Margulies, *supra* note 4, at 403.

78 *See* Lubin, *supra* note 31, at 43.

79 *See* Michael C. Dorf & Charles F. Sabel, *A Constitution of Democratic Experimentalism*, 98 COLUM. L. REV. 267, 267 (1998); Charles F. Sabel & William H. Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 HARV. L. REV. 1015, 1021–22 (2004); Douek, *supra* note 5, at 605 (warning against approaching regulation of social media as adoption of "one-size-fits-all checklist" (quoting Woodrow Hartzog & Daniel J. Solove, *The Scope and Potential of FTC Data Protection*, 83 GEO. WASH. L. REV. 2230, 2259 (2015))); *cf.* Dave Owen, *The Negotiable Implementation of Environmental Law*, 75 STAN. L. REV. 137, 144–49 (2023) (noting this critique of environmental regulation as unduly prescriptive, while arguing that the critique overestimates the nature and scope of the problem).

as supercomputers with a broad array of uses including health care.[80] As in the United States, some experts employed by Chinese firms also publish important theoretical work in academic journals, where U.S. government restrictions would be ill-advised and possibly unconstitutional. Yet, U.S. officials and politicians rarely acknowledge these disparate contexts and the imperatives unique to each. These controls often fail to recognize that technological interdependence is both a fact of life and a vital U.S. interest in its own right. For example, Apple's iPhones are largely made in China in factories owned and run by FoxConn, a Taiwanese company. Disrupting the web of interdependence that produces this crucial product would injure the world economy.[81]

Moreover, the focus on China can also lead to invidious targeting of Chinese nationals in the United States, including university professors and students. This targeting, which often lacks a factual basis, discriminates against Chinese nationals.[82] It may also seep into policies and attitudes about Chinese-Americans. Such results promote discrimination and distract from effective policies. They may also lead to reciprocal targeting of U.S. nationals abroad, producing a vicious cycle of harm.

As of July 2023, the results of this dynamic on the operation of TikTok—a Chinese company—in the United States remain to be seen. Concerns have multiplied that TikTok is sharing U.S. persons' information with the Chinese government. The company has responded with a proposal that would keep U.S. persons' information on servers in the United States run by Oracle, a U.S. company.[83] The Biden administration has been weighing how to impose limits on the app's use of U.S. persons' personal data without outright prohibition of the service. The latter approach would echo the Chinese government's own

---

80 *See* JON BATEMAN, CARNEGIE ENDOWMENT FOR INT'L PEACE, U.S.-CHINA TECHNOLOGICAL "DECOUPLING": A STRATEGY AND POLICY FRAMEWORK 59–63 (2022).

81 *Id.* at 63 (noting that unduly punitive U.S. moves could drive Chinese efforts to limit U.S. access to needed Chinese technology).

82 *See* Xiaoxing Xi v. Haugen, No. 17-2132, 2021 WL 1224164, at *2–3 (E.D. Pa. Apr. 1, 2021) (detailing arrest and indictment on espionage charges of U.S. citizen who had emigrated from China; law enforcement officers held scientist's family at gunpoint while they searched his home, but acknowledged months later, after tarnishing scientist's reputation and causing employee discipline by his U.S. university, that charges were unfounded). The storm of anger and apprehension caused by a single Chinese surveillance balloon over U.S. territory sums up the volatile politics that can undermine a stable licensing regime and turn productive competition into a ruinous zero-sum game. *See* David E. Sanger, *Balloon Incident Reveals More than Spying as Competition with China Intensifies*, N.Y. TIMES (Feb. 13, 2023) https:// www.nytimes.com/2023/02/05/us/politics/balloon-china-spying-united-states.html [https://perma.cc/6W6G-8G5T].

83 *See* Cecilia Kang, Sapna Maheshwari & David McCabe, *TikTok's New Defense in Washington: Going on the Offense*, N.Y. TIMES (Jan. 26, 2023), https://www.nytimes.com/2023/01/26/technology/tiktok-bytedance-data-security.html [https://perma.cc/LM42-48Z7].

restrictive approach to the internet. Indeed, China actually bars access to TikTok within Chinese territory. Whatever the Biden administration decides, it will not resolve all of the current anomalies in U.S. policy on Chinese technology.

Use restrictions and other regulatory requirements can also devolve into perfunctory check-the-box gestures with little or no substantive impact. Companies or agencies can recycle impact assessments, cutting and pasting from earlier submissions. Audits can be dutiful, rather than probing. The signs of activity that regulatory models demand can be just that—signs and empty gestures, rather than indicia of meaningful candor, reflection, and change.[84]

III. ADJUDICATION'S RISKS AND BENEFITS

Regulation also includes avenues for adjudicating issues about the legality and validity of algorithmic action. Adjudication, which merits a separate discussion because of its importance, can address problems concretely and specifically. It can address purely individual complaints that lack any broader impact. In addition, through a range of procedural models, it can address programmatic features of algorithmic action, including accuracy, bias, and explainability. This Part first addresses the posture of adjudication. It then turns to adjudication's risks and scope, centering on whether adjudication is solely about individual claims processing or instead has broad, programmatic effects. The Part concludes with a closer look at the programmatic turn in the jurisprudence of the Meta OB.

A. *The Posture of Adjudication About Algorithmic Activity*

Some measure of recourse for individuals wronged by algorithmic action is a mainstay of many models of regulation.[85] For example, Asaf Lubin's proposal for regulating companies that produce spyware includes a provision for an individual grievance mechanism.[86] The soon-to-be-operational U.S. Data Protection Review Court will field individual

84 *See* Selbst, *supra* note 38, at 145 (observing that the "compliance industry" has emerged to monitor and assess AI models, but that this industry has not necessarily produced uniform, positive change in the design and operation of AI models; the author nevertheless argues that assessments channel conversation and shape habits in useful ways).

85 *See* WHITE HOUSE, AI BILL OF RIGHTS, *supra* note 2.

86 Lubin, *supra* note 31, at 44.

complaints from EU residents regarding U.S. surveillance.[87] The Meta OB handles individual complaints that Meta's algorithms and human reviewers have wrongly taken down individual posts.[88]

Some individual adjudication follows a request by the party conducting algorithmic or surveillance activities. The Foreign Intelligence Surveillence Court (FISC) adjudicates requests by the U.S. Department of Justice to conduct surveillance of individuals whom the government believes are "agent[s] of a foreign power."[89] Because of the concern that notice to the target of surveillance will undermine the purpose of the investigation, such proceedings in the United States and the EU usually only involve submissions by the government.[90] Fair and accurate determinations by the adjudicator in ex parte proceedings require careful vetting and adequate disclosure by the party requesting the activity. In at least one prominent case—a matter involving Trump 2016 campaign advisor Carter Page—the Justice Department and FBI did not practice due diligence and supplied incomplete information to the FISC, leading to a mistaken result.[91]

B. *Adjudication's Risks*

The Carter Page debacle illustrates three risks in individual adjudication. First, individual adjudication can proceed without attention to systemic problems, missing the forest for the trees.[92] That

---

87 *See* Data Protection Review Court, 87 Fed. Reg. 62303 (Oct. 14, 2022) (to be codified at 28 C.F.R. pt. 201).

88 *Case Decision 2020-004-IG-UA*, OVERSIGHT BD. § 3 (Jan. 28, 2021) [hereinafter *Breast Cancer Symptoms and Nudity*], https://www.oversightboard.com/decision/IG-7THR3SI1 [https://perma.cc/4VG3-5KK8].

89 50 U.S.C. §§ 1801–1885c; *see also* United States v. Duggan, 743 F.2d 59 (2d Cir. 1984) (upholding constitutionality of FISA).

90 *See, e.g.*, Kennedy v. United Kingdom, App. No. 26839/05, ¶¶ 139–140 (May 18, 2010), https://hudoc.echr.coe.int/eng?i=001-98473 [https://perma.cc/8S8W-JZRP].

91 In the Carter Page case, the FBI agents who prepared the factual basis for the request to the FISC to authorize surveillance of Page as an alleged Russian agent failed to disclose to Justice Department lawyers overseeing the request that Page had previously served as a contact for a U.S. intelligence agency. Information about Page's history of cooperation with U.S. intelligence would have provided a counterweight to the nebulous assertions in the request that Page was working for the Russians in 2016. *See* Horowitz, *supra* note 35 (discussing errors in FISA request regarding Carter Page); Margulies, *supra* note 19 (same); OFF. OF THE INSPECTOR GEN., U.S. DEP'T OF JUST., REVIEW OF FOUR FISA APPLICATIONS AND OTHER ASPECTS OF THE FBI'S CROSSFIRE HURRICANE INVESTIGATION 157–58 (2019), https://www.justice.gov/storage/120919-examination.pdf [https://perma.cc/ULS6-4SC3].

92 *See* Douek, *supra* note 5, at 532; Klonick, *supra* note 5, at 347–52. It is less clear that the individual process of requesting surveillance of suspected agents of a foreign power under FISA has systemic problems, apart from the failures in the Page case. *See* Margulies, *supra* note 19. However,

concern is particularly compelling when secrecy shrouds the adjudication and related activity, as is true with surveillance and with much algorithmic activity that is protected by trade secrets law. While revelations about the lack of due diligence in preparation of the Carter Page FISA request eventually emerged, that emergence reflected political dynamics that do not apply to the vast run of cases. In particular, discrimination against Muslim Americans and South Asians in U.S. surveillance decisions is a substantial concern that often fails to receive sustained attention.[93]

Moreover, a focus on individual complaints that arise during the operation of a program is vulnerable to information asymmetries.[94] Because of the opacity of surveillance and algorithmic activities, individuals and entities lack awareness that they have been affected. As a result, those individuals will lack the knowledge they need to pursue a complaint. Adjudicators often depend on submissions from entities conducting algorithmic activities.[95] If those entities are not forthcoming, adjudicators also lack adequate information.

Finally, individual adjudication—like programmatic adjudication or any kind of regulation—is prone to capture. Capture entails successful efforts of regulated parties to influence the perspectives of regulators.[96] That influence involves the information asymmetries described above, as well as other considerations such as a shared elite pedigree and values, the regulated entity's access to regulators, and the regulated entity's superior resources. Capture is a pervasive problem with any kind of regulation. It plays a role in adjudication, even when decision-makers are structurally

---

legislators, advocates, and scholars should view the Page episode as raising legitimate questions about government surveillance programs. *See* Klein, *supra* note 34.

93 *Cf.* Sahar F. Aziz, *Policing Terrorists in the Community*, 5 HARV. NAT'L SEC. J. 147, 195–96 (2014) (arguing that U.S. counterterrorism tactics target Muslim American communities but fail to combat most ideologically motivated violence); Maryam Jamshidi, *Bringing Abolition to National Security*, JUST SEC. (Aug. 27, 2020), https://www.justsecurity.org/72160/bringing-abolition-to-national-security/ [https://perma.cc/UA59-PA4Y] (asserting that both substantive counterterrorism laws and surveillance frameworks discriminate against Muslim Americans, Arabs, and South Asians).

94 Douek, *supra* note 5, at 547–48; *see also* Klonick, *supra* note 5, at 352–53 (discussing lack of public awareness of the nature and operation of content moderation on social media platforms).

95 *See* Peter Margulies, *FISA and the FBI: Fixing Material Omissions, Overbroad Queries, and Antiquated Technology*, U.S. PRIV. & C.L. OVERSIGHT BD. (Sept. 2021), https://documents.pclob.gov/prod/Documents/Projects/d726421e-deac-4a0e-be51-783ff0999b3f/Fixing%20FISA-Margulies.09.21a.pdf [https://perma.cc/LXY3-BCFD] (discussing reliance of FISC on Justice Department submissions in Carter Page FISA request).

96 *See* Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621, 669–70 (2021) (discussing effects of regulatory capture); Michael A. Livermore & Richard L. Revesz, *Regulatory Review, Capture, and Agency Inaction*, 101 GEO. L.J. 1337, 1367 (2013); Margot E. Kaminski, *The Capture of International Intellectual Property Law Through the U.S. Trade Regime*, 87 S. CAL. L. REV. 977, 994 (2014).

independent. In certain high-stakes cases, adjudicators may appoint amici curiae to question the algorithmic entity's assumptions.[97] However, those amici may also share the experience and values of the algorithmic entity.[98] In addition, amici may lack the institutional support to provide a meaningful ongoing check. As a result, capture is still a primary concern.

C. *A Programmatic Approach to Adjudication*

Programmatic decisions address broad-based issues such as the bias and accuracy of algorithmic or related activity. They can also impose various attributes of regulation, including requirements for standards, use restrictions, and assessments or audits. For example, the FISC addressed revelations that FBI agents had queried the vast section 702 database for U.S. person information without a sufficient predicate. In response, the FISC attached new conditions to FBI queries. FBI personnel had to contemporaneously document the justification for all queries and obtain the approval of the FBI's General Counsel.[99] That requirement amounts to a mini-assessment that FBI personnel must complete before executing a U.S.-person query.

The programmatic effect of a given decision turns on the consequences of the decision, not on its procedural vehicle. As in cases under the Constitution, federal statutes, or common law, a case involving a single individual can be important if enough parties view the decision as providing useful guidance. That point is particularly compelling when actors are repeat players in adjudication. Repeat players who are active in a given arena over time pay close attention to prior cases.[100]

Those prior cases are part of a system that includes the repeat players and their legal advisors and advocates, as well as insurers, bankers, and others with a stake in future controversies. For example, over time, the FTC has reached settlements with many private companies involving lax cybersecurity practices. In these cases, companies have failed to follow basic cybersecurity hygiene, such as the use of encryption and robust passwords, even when companies have held themselves out to the public on their websites as taking effective steps to safeguard data. When data

---

97 *See* Patel & Koreh, *supra* note 34, at 539–42; Klein, *supra* note 34.

98 Patel & Koreh, *supra* note 34, at 540–42.

99 Memorandum Opinion and Order at 62, *In re* Section 702 2018 Certification (FISA Ct. Oct. 18, 2018), https://www.intelligence.gov/assets/documents/702%20Documents/declassified/2018_Cert_FISC_Opin_18Oct18.pdf [https://perma.cc/UF7S-Q7CY].

100 *See* Andrew D. Bradt & D. Theodore Rave, *It's Good to Have the "Haves" on Your Side: A Defense of Repeat Players in Multidistrict Litigation*, 108 GEO. L.J. 73 (2019).

breaches expose the personal information of a company's customers, the FTC has successfully asserted that the company has engaged in unfair and deceptive trade practices by not living up to its public assurances.[101] Typically, FTC enforcement proceedings against companies culminate in a settlement, in which the company agrees to implement cybersecurity best practices.[102] Since the settlements are accessible to lawyers in the field, those lawyers have studied these settlements and used them to provide concrete guidance to their own corporate clients.

D. *The Meta OB's Programmatic Pivot*

Two recent decisions of the Meta OB are crucial examples of the programmatic approach. In one decision that is nominally about two individual complaints—*Gender Identity and Nudity*[103]—and another that the OB expressly styled as a policy decision—*Cross-Check*[104]—the OB charted interrelated programmatic norms for Meta's vast platforms. Because these two recent decisions set a high standard for programmatic review of algorithmic activity, they are worth reviewing in depth.

1. Programmatic Perspective and the *Gender Identity* Decision

In the *Gender Identity* case, the Meta OB reiterated its concern in *Breast Cancer Symptoms and Nudity* that the combination of algorithms and human moderators that Meta has used to identify, monitor, and combat noxious speech has curbed freedom of expression.[105] The two joined complaints in *Gender Identity* entailed videos by transgender individuals providing information on medical procedures appropriate for this group. Both videos focused on the role of breast-reduction surgery for trans men assigned female at birth. To illustrate the nature of this gender-affirming surgery, each video featured a subject who appeared bare-chested, covering their nipples.[106] Meta's algorithms flagged each post as violating Meta's standards on sexual solicitation and adult nudity. Because of continued complaints about the posts by Meta users—some of whom may have objected to any content that addressed trans issues—

---

101 *See* Solove & Hartzog, *supra* note 9.

102 *Cf.* FTC v. Wyndham Worldwide Corp., 799 F.3d 236, 241–42 (3d Cir. 2015) (documenting a corporation's fundamental failures in cybersecurity and upholding FTC's authority to initiate proceedings).

103 *Gender Identity and Nudity*, *supra* note 2.

104 *Cross-Check*, *supra* note 2.

105 *Breast Cancer Symptoms and Nudity*, *supra* note 88, § 8.3.

106 *Gender Identity and Nudity*, *supra* note 2, § 2.

human reviewers ultimately agreed that the posts violated Meta's standards.

In a wide-ranging decision that cited Article 19, the free-expression guarantee in the International Covenant on Civil and Political Rights (ICCPR),[107] the OB found that Meta's actions in the joined cases violated human rights law. The OB found that, as implemented by the company, Meta's standards on sexual solicitation were vague and led to overenforcement.[108] The Board made similar findings about Meta's implementation of its standard on "Adult Nudity and Sexual Activity."[109] In addition, the Board cited the gender bias built into Meta's "default to female" approach.

Under this approach, when either algorithms or human reviewers had doubts about the gender identity of a person portrayed in a posted image, the review coded that person as female. That default then triggered Meta's own gendered rules on nudity. Those rules included restrictions on images of females that did not apply to images of males.[110] The net result of Meta's policies and its application of those policies was overenforcement that stifled content related to trans issues, women's health, and HIV education.[111] This overenforcement failed the ICCPR's tests, including proportionality and necessity, legality (entailing transparency and notice), and legitimate aim (requiring a link to national security or public safety/health).[112]

The OB's remedies were far-reaching. The Board recommended that Meta conduct a "comprehensive human rights impact assessment" to shape its formulation and implementation of standards regarding images related to gender identity.[113] That assessment should encompass broad

---

107 International Covenant on Civil and Political Rights art. 19(2)–(3), *opened for signature* Dec. 16, 1966, 999 U.N.T.S. 171, 178 (entered into force Mar. 23, 1976); *see also* Evelyn Mary Aswad, *Taking Exception to Assessments of American Exceptionalism: Why the United States Isn't Such an Outlier on Free Speech*, 126 DICK. L. REV. 69, 92–95 (2021) (discussing the ICCPR); Evelyn Mary Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26, 42–57 (2018) (arguing for applying human rights standards to moderation of online speech).

108 *Gender Identity and Nudity*, *supra* note 2, § 8.3(I)(a).

109 *Id.* § 8.3(I)(b).

110 *Id.* § 4. The core difference in rule application involved the portrayal of subjects that appeared topless. The Board suggested that this approach was biased, although it stopped short of saying that Meta had to treat male and female nudity under the same standard. *Id.* § 8.3(I)(b). In declining to require strict gender equality in this area, the Board cited to thousands of community comments that it had received, including comments that cited the internet's use for victimization of female children and adults. *Id.* § 8.3(II)(b). However, the Board viewed Meta's formulation and implementation of standards as unclear and biased. *Id.* § 8.3.

111 *Id.* § 8.3(III).

112 *Id.* § 8.3.

113 *Id.* § 8.3(III)(b).

participation of interested parties.[114] In addition, the Board recommended that Meta provide reviewers with clearer instructions. To ensure compliance with this recommendation, the Board requested that Meta supply it with those new guidelines when Meta completed the drafting process.[115] That recommendation indicated that the *Gender Identity* decision was not a one-off; rather, it was part of a sustained engagement with Meta's practices.

2. *Cross-Check* and Favoritism in Content Moderation

In the landmark *Cross-Check* Policy Advisory Opinion,[116] the OB took on the bookend of the overenforcement for marginalized groups cited in *Gender Identity*: underenforcement for the rich and famous. The OB's *Cross-Check* opinion considered a prominent Meta policy that imposed obstacles to the takedown of offensive posts—such as posts that degraded women or targeted trans individuals or minority ethnic or religious groups—by certain previously designated persons and entities. These "entitled" entities were often Meta advertisers, governments, or celebrities such as the soccer star Neymar that drove traffic to Meta's sites.[117]

The cross-check program's underenforcement for advertisers and celebrities worked in the following way: Meta's ordinary processes, including extensive automated content-moderation tools that the OB also alluded to in its *Gender Identity* case, flagged content as harmful.[118] As part of the cross-check program, human reviewers considered market-based factors in assessing the harmful content of posts by the entitled entity. During the duration of the review by human content moderators, Meta kept the material online, where it could "go viral" with millions of visitors to the site.[119]

In a candid opinion of sweeping scope, the OB described the cross-check program as placating advertisers and celebrities who drove increased revenue for the company.[120] That favoritism constituted underenforcement of Meta's content policy, without any basis in Meta's

---

114 *Id.*

115 *Id.* § 10.

116 *Cross-Check*, *supra* note 2.

117 *Id.* at 3–4.

118 *Id.* ¶ 17.

119 *Id.* ¶ 181.

120 *Id.* at 3 (noting that the cross-check program "appears . . . structured to satisfy business concerns" and provides "extra protection to users selected largely according to business interests").

stated values, which included voice, dignity, and privacy.[121] Indeed, the OB observed, the effect of the harmful posts as they reached millions of viewers was further suppression of marginalized groups.[122]

Surveying the effects of the cross-check program, the OB contrasted its underenforcement of Meta's norms with the pervasive overenforcement that *Gender Identity* critiqued.[123] These two problems both stemmed from Meta's disparate responses to algorithmic flagging of content. For the favored entities that received human review under cross-check, algorithmic decisions were often correct: much of this material was harmful. However, due to the elaborate and time-consuming human review that cross-check mandated for this material and the stay of a takedown pending the completion of human review, the net result was a flood of false negatives. In other words, harmful material that Meta should have removed remained on the site for protracted periods.[124] Calculation of the duration of exposure to that harmful material should also include its extended half-life, once users embed the content on other sites and use texting and email to reach still more people.

At the same time, the lack of cross-checking for content from marginalized groups, who were not on Meta's "entitled" list, left algorithmic decisions in place. The resulting takedowns triggered a cascade of false positives. Marginalized groups' material was relegated to imprecise algorithmic decisions and the biased processes critiqued in *Gender Identity*. As a result, Meta took down material it should have retained. The OB described the situation forthrightly: the conjunction of false negatives for the rich and famous and false positives for marginalized groups meant that Meta was "cross-checking the wrong content."[125]

As in the *Gender Identity* decision, the OB based its critique of Meta's practices on international human rights law, including Article 19 of the ICCPR, which guarantees freedom of expression. Article 19, which is less absolute in its free speech provisions than the U.S. Constitution's First Amendment, includes carveouts for national security, public safety, and public health. Challenged practices and their fit with these carveouts

---

121 *Id.* ¶ 70–78. By voice, Meta meant participation, particularly by groups that otherwise feel shut out of mainstream discourse and decision-making. *Id.* ¶ 70.

122 *Id.* ¶ 66 (observing that cross-check may "contribute to an environment that inhibits expression from those who may be targeted" by harmful content). The OB contrasted that with cross-check's operation as a "system . . . designed primarily to protect or prioritize the expression of people who are already powerful." *Id.* ¶ 65.

123 *Cf. Cross-Check* ¶ 109 (juxtaposing problems of "over and under-enforcement").

124 *Id.* ¶¶ 35, 103 (noting that potentially harmful material subject to human review under cross-check remained up in the United States for an average of twelve days, and remained up for longer periods in other areas where human reviewers were limited due to language deficits).

125 *Id.* ¶ 107.

are reviewed under the principles of legality, legitimate aim, and necessity and proportionality. The principle of legality refers to transparency and notice. To have a legitimate aim, the program would have to be tailored to public health, safety, and security. To be necessary and proportionate, Meta would have to tailor the program's formulation and execution to Meta's avowed goals. The OB found that the cross-check program was too opaque to meet the requirements of legality and too underinclusive to fit Meta's own stated goals of voice and dignity.[126] Moreover, the OB found that, because the cross-check program included material that could incite violence against marginalized groups and was not required for national security or public safety and health, the program failed to serve a legitimate aim.[127]

On a methodological level, although the OB did not discuss the technical details of automated processes, the Board did discuss the cross-check program's overall contours precisely and comprehensively. Moreover, the Board included a rigorous discussion of the cross-check program's flaws, including the abject lack of transparency in both the composition of the entitled-entity list and the process that Meta used in designating such entities.[128] That discussion included Meta's efforts to keep the cross-check program secret. The OB noted the role of the whistleblower Frances Haugen in disclosing the program.[129] It also outlined Meta's misleading initial answers to the Board's queries and its continuing failure to provide complete responses to basic questions such as the identity of entitled entities and the basis for their inclusion on this favored list.[130]

Finally, the Board made robust recommendations about reforms. These included adjusting the contours of the cross-check program; allowing a greater range of entities, including those from marginalized groups, to apply for inclusion; publicizing criteria for selection; and clearly separating reviewers from the revenue concerns that had driven the program in the past.[131] A crucial goal that the OB cited and Meta acknowledged was the reduction of both false negatives for the rich and famous and false positives for groups that lacked those privileges.[132] Stressing that these reforms were not a mere wish list, the Board noted

---

126 *Id.* ¶ 118.

127 *Id.* ¶ 120.

128 *Cf. id.* ¶¶ 43–44.(discussing details of cross-check program).

129 *Id.* ¶ 2.

130 *Id.* ¶¶ 6, 79–80.

131 *Id.* ¶¶ 123–160. Separation of content moderation from market concerns is a key concern of scholars critiquing current social media approaches. *See* Douek, *supra* note 5, at 586–87.

132 *Cross-Check*, *supra* note 2, ¶ 179 (observing critical role of reducing false positives for "historically over-enforced entities").

that Meta had committed to report biannually to senior officials on their progress toward complying with the *Cross-Check* decision's recommendations.[133]

3. Comparing Adjudication and Administration

The Meta OB's recent decisions have shown the value of programmatic adjudication. That value transcends the procedural posture of a case. Both *Gender Identity*—a case arising from two related complaints—and *Cross-Check*—an opinion expressly about policy—had programmatic scope and consequences. The Board's performance was imperfect in this strand of decisions, as the next Part will show. However, the Board's recent decisions nonetheless serve as a reminder of adjudication's value.

Moreover, assuming that it is independent, a court or other adjudicative tribunal has advantages over an agency decision-maker. A tribunal, because it is independent, can give a straightforward account of the facts.[134] The Meta OB's candid account of the favoritism that Meta showed in the cross-check program is a case in point.[135] The iterative process that courts follow, including painstaking adherence to precedent, provides some assurance that a tribunal's decisions will not be arbitrary.[136] In contrast, administrative agencies, as political creatures by design and implementation, lack these guardrails.[137]

---

133 *Id.* at 49; *see also* Klonick, *supra* note 5 at 359 (arguing that the Meta OB is at least potentially a "dynamic solution for . . . reform, [albeit] not a panacea").

134 *See* THE FEDERALIST NO. 78, at 465 (Alexander Hamilton) (Clinton Rossiter ed., 1961) (describing courts as having "neither [force] nor [will] but merely judgment").

135 *Cross-Check*, *supra* note 2, ¶ 45. Admittedly, courts may be disingenuous and results-driven in discussing facts and law. *See* Brett M. Kavanaugh, *Fixing Statutory Interpretation*, 129 HARV. L. REV. 2118, 2139 (2016) (reviewing ROBERT A. KATZMANN, JUDGING STATUTES (2014)) (discussing temptation faced by courts to manipulate facts or legal tests). When courts are at their best, however, independence provides a hedge against such proclivities.

136 *See* THE FEDERALIST NO. 78, *supra* note 134, at 471. Here, too, one must guard against overstatement of the impact of judicial habits. As the Supreme Court has recently shown, adherence to precedent is not absolute. *See* Dobbs v. Jackson Women's Health Org., 142 S. Ct.. 2228, 2272 (2022) (asserting that stare decisis—the prudential doctrine counseling adherence to precedent—did not require preserving the right to terminate pregnancy as previous Supreme Court decisions had defined that right).

137 As an example, consider the U.S. intelligence community's distinction between "bulk collection" of intelligence, which can entail acquisition and storage of all information for future analysis, and wide-ranging targeted collection, which can entail similar treatment of the communications corpus of tens of thousands of electronic accounts. David S. Kris, *On the Bulk Collection of Tangible Things*, 7 J. NAT'L SEC. L. & POL'Y 209, 218–21 (2014); Steven G. Bradbury, *Understanding the NSA Programs: Bulk Acquisition of Telephone Metadata Under Section 215 and Foreign-Targeted Collection Under Section 702*, at 2–3 (Lawfare Rsch. Paper Series, Vol. 1, No.

IV. STEWARDSHIP AS A NORM

The previous Sections have shown that the problems posed by algorithmic activity are substantial. Models of regulation have promise but also show flaws. In response, this Part suggests a stewardship approach.

Stewardship on this view connotes responsibility, recourse, and participation. All algorithmic activity reposes trust in the actor to safeguard data that the actor uses and obtains. Discharging that trust is an algorithmic entity's key responsibility. When an algorithmic entity fails to fulfill that responsibility, victims require recourse from an independent reviewer. Otherwise, the algorithmic entity can violate its charge with impunity. To protect against capture of the reviewer by the subjects of regulation, mechanisms in place should impel the reviewer to question assumptions and challenge dominant narratives.

Stewardship in this conception has three prongs: iterative review, layered accountability, and institutionalized opposition. This Section explains these elements. The next Section applies the stewardship model to three current issues: (1) U.S. surveillance in transatlantic data sharing; (2) cybersecurity, data breaches, and online disinformation; and (3) algorithms in credit, housing, and employment screening.

A. *Iterative Review*

Scholars of innovation in business have identified an intriguing model of iterative product and contract design.[138] Under this model, stakeholders start with benchmarking, which is where participants accurately describe the status quo and describe possible improvements.[139]

3, 2013), https://s3.documentcloud.org/documents/7276726/UNDERSTANDING-THE-NSA-PROGRAMS-BULK-ACQUISITION.pdf [https://perma.cc/98V9-7NZZ]. These differences are significant—the collection of tens of thousands of communications, while it seems like a substantial endeavor, is less intrusive than the collection of millions of communications. However, an independent tribunal may be more inclined to see similarities between such substantial programs and thus track the reasonable response of ordinary people without intelligence agencies' stake in fine distinctions. *See* United States v. Hasbajrami, 945 F.3d 641, 671 (2d Cir. 2019) (discussing vast scale of section 702 program in course of holding that querying of U.S. person information in section 702 database constitutes a Fourth Amendment search).

138 Ronald J. Gilson, Charles F. Sabel & Robert E. Scott, *Contracting for Innovation: Vertical Disintegration and Interfirm Collaboration*, 109 COLUM. L. REV. 431, 447 (2009).

139 *Id.*; Peter Margulies, *Benchmarks for Using Technology to Protect Civilians in Armed Conflicts: Learning Feasible Lessons About Systemic Change*, MINN. J. INT'L L. (forthcoming 2023), https://www.ssrn.com/papers/abstract=4186087. Special questions emerge regarding the role of adjudication in the use of AI for military targeting. Controversy surrounds the use of AI as a substitute for human judgment in targeting, as in autonomous weapons systems in which a

Participants assemble an initial plan and propose adjustments. The plan includes provisions for monitoring progress. As implementation of the plan proceeds, participants assess the successes and failures of both the plan and implementation. Benchmarking requires that participants continuously identify assumptions and question them.[140] Adjustments are expected and necessary, ranging from wholesale rethinking of the initial plan to more modest revisions.[141] While I address the participatory component of many theories of iterative processes below, this subsection centers on the temporal dimension and on the importance of benchmarking as a guide to measurable, reliable progress.

In coupling iteration and review, this Section also envisions a court, which we could call the Algorithmic Rights Court (ARC).[142] A court can embody independence from the agendas of other participants and interested parties. It can then sift through inputs without direct pressure to provide a particular outcome.[143] Provisions for independence can entail a range of methods, some more formal than others. The gold

computer makes a decision to use lethal force without advance human approval. *See id.*; PAUL SCHARRE, ARMY OF NONE: AUTONOMOUS WEAPONS AND THE FUTURE OF WAR (2018); Ashley Deeks, Noam Lubell & Daragh Murray, *Machine Learning, Artificial Intelligence, and the Use of Force by States*, 10 J. NAT'L SEC. L. & POL'Y 1 (2019). The planner of an attack that uses AI should have a legal responsibility for taking reasonable steps to ensure that the AI's algorithms comply with legal requirements. *See* ALFONSO SEIXAS-NUNES, THE LEGALITY AND ACCOUNTABILITY OF AUTONOMOUS WEAPONS SYSTEMS: A HUMANITARIAN LAW PERSPECTIVE 204–07 (2022); Peter Margulies, *Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts*, *in* RESEARCH HANDBOOK ON REMOTE WARFARE 405 (Jens David Ohlin ed., 2017); *see also* Rebecca Crootof, *War Torts*, 97 N.Y.U. L. REV. 1063 (2022) (discussing prospects for tort liability for civilian harm in wartime). Compliance entails, *inter alia*, taking feasible precautions against civilian harm, including facilitating an attack planner's receipt of accurate information and sound interpretation. *See* Geoffrey S. Corn & Michael W. Meier, *Enhancing Civilian Risk Mitigation by Expanding the Commander's Information Aperture*, *in* THE GLOBAL COMMUNITY: YEARBOOK OF INTERNATIONAL LAW AND JURISPRUDENCE 2019 159, 183–89 (Giuliana Ziccardi Capaldo ed., 2020). Investigation of suspected war crimes is an essential component of compliance with the law of armed conflict. *See id.*; Michael N. Schmitt, *Investigating Violations of International Law in Armed Conflict*, 2 HARV. NAT'L SEC. J. 31, 80 (2011). Domestic, transnational, and international avenues of legal recourse, including proceedings under the U.S. Uniform Code of Military Justice, may provide a mechanism that is consistent with the iterative review described in this Article. Additional avenues of recourse, such as those suggested in this Article, might be appropriate for military uses of AI. However, those avenues for adjudication might also unduly chill attack planners' initiative and discretion. Balancing the need for recourse in a military context against the cost of such recourse on attack planners' initiative and discretion is a subject beyond the scope of this Article.

140 *See* KAHNEMAN, SIBONY & SUNSTEIN, *supra* note 51, at 306–24.

141 *See* Gilson, Sabel & Scott, *supra* note 138, at 447; *see also* Douek, *supra* note 5, at 533 (noting importance of iterative process).

142 *See* Rubinstein & Margulies, *supra* note 4, at 447–50.

143 *See* Douek, *supra* note 5, at 586–87 (discussing importance of separating content moderation reviewers from the business side of social media).

standard is the independent, life-tenured status of U.S. federal judges.[144] In reviewing both individual and programmatic features of FISA, judges of the FISC possess independence in this axiomatic way.

There may be other ways of establishing independence that are not quite as ironclad but provide a significant measure of protection through the needs of actors in the process to preserve a reputation for fairness and sound governance. For example, in the federal government, inspectors general are often independent, even though they have no special protections against dismissal.[145] Independence in this sense entails a pragmatic inquiry into what works for institutions, not a textual inquiry into the job description of an adjudicator.

A related model for iterative review is the traditional model of equitable discretion over remedies. In a lawsuit seeking relief against an entity or official's illegal practices, the court may enter an injunction.[146] The party seeking the injunction may then seek to compel compliance by alleging that the party subject to the injunction has not been implementing its terms. The court that has ordered this equitable relief can also seek to ascertain or compel compliance on its motion (sua sponte) through its own inquiries of the party or appointment of a special master.

B. *Layered Accountability*

Layered accountability includes a range of means and audiences. In addition to adjudication, layered accountability requires several of the approaches listed above, including assessments, disclosure, and audits. It also includes a prime ingredient in iterative conceptions: participation by interested parties. The layering of these approaches has a conceptual advantage: it obliges both regulators and regulated entities to constantly brainstorm about new forms of accountability. In a more instrumental fashion, layered accountability also entails redundancy, so that one mode of accountability can compensate if another flounders.

In some situations, layered accountability will include both a reviewing body, such as a court, and a regulating body, such as an administrative agency. For example, in the domain of privacy, the EU has

144 U.S. CONST. art. III, § 1.

145 *See* 5 U.S.C. § 403(b); *see also* Amy C. Gaudion, *Recognizing the Role of Inspectors General in the U.S. Government's Cybersecurity Restructuring Task*, 9 BELMONT L. REV. 180, 203–09 (2021) (discussing the importance of inspectors general).

146 *See* Horne v. Flores, 557 U.S. 433, 447–48 (2009) (discussing injunctions and the need to modify relief in appropriate circumstances); F. Andrew Hessick & Michael T. Morley, *Interpreting Injunctions*, 107 VA. L. REV. 1059, 1067–72 (2021) (discussing standards for granting and modifying equitable relief).

data protection authorities whose rulings are subject to review both by domestic courts and by the CJEU.[147] In the United States, while no one federal agency has full power over algorithmic activities, the FTC and SEC have exerted authority. The FTC is currently in the early stages of an effort to promulgate privacy rules that would complement its well-established enforcement efforts targeting firms' flawed privacy policy implementation as an unfair and deceptive trade practice.[148] The extensive corpus of blunders by firms on the privacy front may provide sufficient predicate for the FTC to exercise its limited rulemaking authority in this domain.

Broad participation is both a means and an end to layered accountability. Parties that participate can provide a check on algorithmic entities' excesses outside the context of formal adjudicative review. For example, communities that view facial recognition technology as unduly intrusive can mobilize.[149] That mobilization can lead to prohibition, which this Article rejects as an unduly blunt form of regulation. However, mobilization can also make algorithmic entities sit up and take notice, as Meta did in the wake of concerns that its business practices facilitated disinformation during the 2016 U.S. election campaign and in the targeting of marginalized groups, such as the Rohingya in Myanmar. That mobilization, sometimes filtered through sympathetic elected officials such as U.S. legislators who support a national data protection law, can reap broad dividends.

Disclosure by algorithmic entities enables participation. By mandating disclosure, courts, agencies, and legislative bodies overcome information asymmetries that stifle inputs from consumers. Armed with information, interested parties can make more effective arguments. The Meta OB has continually relied on comments by interested parties.[150] Perhaps the OB should do even more outreach to "street-level" groups that often lack elite representation.[151] That would open up the process to an ever greater degree.

---

147 *See* Kaminski, *supra* note 1.

148 *See* Press Release, Fed. Trade Comm'n, FTC Explores Rules Cracking Down on Commercial Surveillance and Lax Data Security Practices (Aug. 11, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/08/ftc-explores-rules-cracking-down-commercial-surveillance-lax-data-security-practices [https://perma.cc/VCU8-TAE4]. Full discussion of the FTC's rulemaking authority is beyond the scope of this Article.

149 *See* Michele Estrin Gilman, *Beyond Window Dressing: Public Participation for Marginalized Communities in the Datafied Society*, 91 FORDHAM L. REV. 503, 521–22 (2022).

150 *See Gender Identity and Nudity*, *supra* note 2, § 8.3.

151 *See* Gilman, *supra* note 149; Scott L. Cummings & Doug Smith, *Policy by the People, for the People: Designing Responsive Regulation and Building Democratic Power*, 90 FORDHAM L. REV. 2025, 2040–42 (2022).

Auditing is another element of layered accountability.[152] Auditing of and by algorithmic entities can inform regulation and review. Identification of successes and flaws in compliance efforts are necessary for each process. To facilitate review, entities should put audits in XBRL or other formats that AI models can read.[153] The push for documentation that is conducive to automated audits has been a longtime work in progress for the FBI in the area of U.S.-person queries of section 702 information.[154] Further remedial efforts by the courts, the Justice Department, and Congress may be necessary to resolve these compliance issues.

Layered accountability can be particularly useful in addressing the functional trade-offs in algorithmic activity. For example, while interested parties can provide input on the privacy impacts of surveillance, other groups can discuss how hackers' efforts frustrate privacy.[155] Those combined inputs on the costs of certain kinds of algorithmic activity, such as surveillance versus the costs of doing nothing, can produce a more balanced adjudication. As in the Meta OB, layered accountability can prompt inputs on the benefits of algorithms in content moderation at scale versus the cost of overenforcement. Pluralistic inputs can compare the virtues and costs of algorithmic activity with the virtues and costs of human judgment, given the distinctive errors characteristic of each process. Similarly, pluralistic inputs can allow adjudicators to weigh the trade-offs between explainability and accuracy in algorithms. With a range of inputs, adjudicators can discern whether decreases in accuracy are steep as models become more explainable, or whether advances in design can produce a more gradual roll-off or no negative effect.[156] Access to inputs on both axes can nudge courts and policymakers out of a fruitless debate weighted with false dichotomies to an acknowledgement of the dynamic potential of inclusive algorithms.[157] That effect can also push public discourse away from the polar extremes of prohibition and laissez-faire toward a focus on substantial reforms.

---

152 *See* Douek, *supra* note 5, at 601.

153 *See* Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure, 87 Fed. Reg. 16590, 16603 (proposed Mar. 23, 2022) (to be codified at 17 C.F.R. pts. 229, 232, 239, 240, 249).

154 *See* Margulies, *supra* note 19, at 1203.

155 *See* Pozen, *supra* note 30, at 229–32.

156 *See* Clement, Kemmerzell, Abdelaal & Amberg, *supra* note 39; Gohel, Singh & Mohanty, *supra* note 39.

157 *See* Woods, *supra* note 36.

C. *Institutionalized Opposition*

To promote iterative review and layered accountability, stewardship also requires that any algorithmic activity install institutionalized opposition. By opposition, I mean a voice that will counter the algorithmic entity's contentions. That voice should be distinct from the neutral posture of a reviewing body such as the Meta OB. Rather than being neutral, the role of an institutionalized opposition body will be adversarial.[158] That voice should be institutionalized by a charter and resources that will allow it to stay the course and by access to information that will allow it to make effective counterarguments.

Institutionalized opposition can help stave off capture of the entire process by the algorithmic entity. As noted above, capture of a regulator by a regulated entity can undermine robust review. Critics have charged that the current system for review of U.S. foreign surveillance suffers from this problem.[159] Meta's constraints on its OB make capture a constant threat in that context as well.[160] An opposing voice can keep regulators honest and counter the resources, information asymmetries, and converging perspectives that enable regulatory capture.

Wide-ranging participation, such as the inputs received by Meta's OB, is important to layered accountability but is not a substitute for the institutionalized opposition outlined here. Similarly, the FISC's appointment of an amicus in certain matters involving U.S. surveillance is valuable but does not constitute the institutional presence that stewardship requires. Amici, although they often oppose the government's arguments, lack the authority or access to assess matters such as the National Security Agency's criteria for selecting targets under section 702.[161] The Privacy and Civil Liberties Oversight Board (PCLOB), which has produced an important and revealing study of section 702 and is poised to issue another substantial report, has a limited charter to study intelligence collection regarding counterterrorism and is not suited for the institutionalized-opposition role outlined here.[162] Congress should

---

158 An adversarial voice need not be vexatious or oppositional for its own sake. Rather, an institutional voice should rigorously question an algorithmic entity's assumptions on a variety of planes, involving conception, values, and execution.

159 *See* Elizabeth Goitein, *The Year of Section 702 Reform, Part I: Backdoor Searches*, JUST SEC. (Feb. 13, 2023), https://www.justsecurity.org/85068/the-year-of-section-702-reform-part-i-backdoor-searches/ [https://perma.cc/EU3H-3ZMU].

160 *See* Chinmayi Arun, *Facebook's Faces*, 135 HARV. L. REV. F. 236, 245 (2022).

161 *See* Patel & Koreh, *supra* note 34.

162 *See* PRIV. & C.L. OVERSIGHT BD., REPORT ON THE SURVEILLANCE PROGRAM OPERATED PURSUANT TO SECTION 702 OF THE FOREIGN INTELLIGENCE SURVEILLANCE ACT 93 (2014) (discussing section 702), https://irp.fas.org/offdocs/pclob-702.pdf [https://perma.cc/85VP-JJ4T]; Goitein, *supra* note 159 (noting pending 2023 PCLOB section 702 report).

create a Public Advocate that will serve that function. Institutionalized opposition will also be useful in the other contexts addressed in this Article.

V. APPLYING THE STEWARDSHIP APPROACH

Having outlined the principles of the stewardship approach, this section applies those principles to current issues involving algorithmic activity. Those issues include the newly established U.S. Data Protection Review Court on EU complaints regarding U.S. government surveillance; private-sector data breaches and online disinformation; and algorithmic applicant screening.

A. *The U.S. Data Protection Review Court*

The Biden administration's establishment of a Data Protection Review Court (DPRC) is an important step forward in the dialectic between EU data protection law, as interpreted robustly by the CJEU, and the legal framework of U.S. surveillance. The U.S. framework includes checks that many individual EU states lack, but nonetheless historically has provided few if any formal remedies to aggrieved EU residents.[163] The independence of the DPRC and the appointment of a Special Advocate who will advocate for "the complainant's interest"[164] are key building blocks that harmonize with the stewardship approach. Nevertheless, questions remain about the DPRC's structure, including the contained role of the Special Advocate, which provides a measure of institutionalized opposition but not the constant presence that stewardship envisions.[165]

The DPRC works in the following way. Aggrieved individuals abroad will first file a complaint with the data protection authorities in their country of origin or "regional economic integration organizations"

---

163 Prior to creation of the DPRC, many scholars had urged the creation of a dedicated and independent tribunal to hear EU residents' complaints. *See, e.g.*, Rubinstein & Margulies, *supra* note 4, at 447–52.

164 Data Protection Review Court, 87 Fed. Reg. 62303, 62307 (Oct. 14, 2022) (to be codified at 28 C.F.R. pt. 201).

165 *Id.* (describing substantial evidence standard); *id.* at 62306–07 (noting that Special Advocate is appointed on a case-by-case basis, instead of serving as an ongoing watchdog); *id.* at 62307 (limiting communication of Special Advocate with complainant to written questions and responses subject to vetting by U.S. Department of Justice Office of Privacy and Civil Liberties and representatives of U.S. intelligence community).

such as the EU, which the regulations define as a "qualifying state."[166] The qualifying state will, after whatever screening it determines to be necessary, forward the complaint to the Civil Liberties Protection Officer (CLPO) of the Office of the Director of National Intelligence (ODNI).[167] The CLPO is an official within the executive branch whose statutory role entails a degree of independence.[168] Nevertheless, the CLPO may be dismissed for any reason by the Director of National Intelligence who in turn serves at the pleasure of the President. The CLPO may determine that the complaint is well-founded based on consideration of U.S. law.[169] Under the Executive Order that announced the new transatlantic data privacy framework and paved the way for the DPRC, the surveillance at issue must be necessary to achieve a substantial government objective and proportionate in light of that purpose.[170] Those tests are robust and also

166 *Id.* at 62303–04.

167 *Id.* at 62306.

168 *See, e.g.*, 50 U.S.C. § 3029(b)(2) (noting that the CLPO shall "oversee compliance" by ODNI with "requirements under the Constitution and all laws, regulations, Executive orders, and implementing guidelines relating to civil liberties and privacy"). The nature and logic of this oversight role encompass scenarios in which other officials who focus on national security may seek to formulate and implement policies that endanger civil liberties and privacy, requiring pushback from the CLPO. Ensuring robust pushback requires a privacy official who can do their job without constant worry about dismissal. However, this logic does not entail special job security for the CLPO, such as a requirement that dismissal be only for good cause. Rather, the statutory language in the CLPO's job description is a signal by Congress to the executive branch about the latitude that the CLPO needs to fulfill the scheme that Congress has enacted into law. Unduly hasty or retaliatory dismissal of a CLPO for doing their job might well trigger congressional hearings, inquiries about ODNI's budget and priorities, and other interactions with Congress that senior officials would generally strive to avoid. *See generally* Margo Schlanger, *Intelligence Legalism and the National Security Agency's Civil Liberties Gap*, 6 HARV. NAT'L SEC. J. 112, 141–42 (2015) (discussing role of National Security Agency CLPO, who has input into policy and also communicates with independent boards, such as the PCLOB, which has produced reports about U.S. foreign surveillance). However, a privacy officer's role does not extend to veto power over decisions by senior officials, who may attach a lower priority to civil liberties concerns. *See id.* at 116–18; Shirin Sinnar, *Institutionalizing Rights in the National Security Executive*, 50 HARV. C.R.-C.L. L. REV. 289, 304–09 (2015) (discussing benefits and drawbacks of civil liberties officer's agency role); *see also* Engstrom & Ho, *supra* note 42, at 847–49 (discussing benefits of creating boards within executive agencies to review use of algorithms). In a world where politics, relationships, and past practice matter, that signal is significant, even though it does not accompany provisions for special job security.

169 Data Protection Review Court, 87 Fed. Reg. at 62307.

170 Exec. Order No. 14086, § 2(a)(ii)(A)–(B), 87 Fed. Reg. 62283 (Oct. 7, 2022). The CJEU has outlined a test based on necessity, proportionality, and independent review. *See* Case C-311/18, *Schrems II*, ECLI:EU:C:2020:559, ¶¶ 76, 168, 180–81, 185 (July 16, 2020); Joined Cases C-511/18, C-512/18 & C-520/18, La Quadrature du Net v. Premier Ministre, ECLI:EU:C:2020:791, ¶ 137 (Oct. 6, 2020); Big Brother Watch v. United Kingdom, App. Nos. 58170/13, 62322/14 & 24960/15, ¶¶ 292, 350 (Sept. 13, 2018), https://hudoc.echr.coe.int/?i=001-210077 [https://perma.cc/NA7B-JVSU] (in decision by European Court of Human Rights (ECHR), engaging in deferential review of United Kingdom surveillance, but finding that system needed more internal contemporaneous review); Rubinstein & Margulies, *supra* note 4, at 406–18 (discussing CJEU and ECHR surveillance

fit European law requirements. Having made this finding, the CLPO may recommend "appropriate remediation" of the problem identified in the complaint.[171] If the complainant is dissatisfied with the CLPO's finding, the complainant may file a request through a qualifying state with the DPRC.

Under the U.S. Department of Justice regulation establishing the DPRC, this tribunal will consist of six or more judges outside the U.S. government. To bolster the independence of the tribunal, DPRC judges will not serve under the regular supervision of the Attorney General and may be removed only for cause.[172] Judges of the DPRC will have access to classified information. Each panel will select a "Special Advocate" who will advocate for the complainant's position and inform the judges about the relevant issues.[173] The DPRC, like the CLPO, will apply U.S. law.[174]

---

jurisprudence); Kenneth Propp & Peter Swire, *Geopolitical Implications of the European Court's* Schrems II *Decision*, LAWFARE (July 17, 2020, 11:31 AM), https://www.lawfaremedia.org/article/geopolitical-implications-european-courts-schrems-ii-decision [https://perma.cc/Q96Y-4QV6]. *See generally* TIMOTHY H. EDGAR, BEYOND SNOWDEN: PRIVACY, MASS SURVEILLANCE, AND THE STRUGGLE TO REFORM THE NSA 160 (2017) (discussing effect in the EU of revelations of Edward Snowden on vast, hitherto secret U.S. surveillance operations). President Obama laid the groundwork for this approach with an earlier policy. *See* Press Release, Off. of the Press Sec'y, The White House, Presidential Policy Directive—Signals Intelligence Activities (Jan. 17, 2014), https://obamawhitehouse.archives.gov/the-press-office/2014/01/17/presidential-policy-directive-signals-intelligence-activities [https://perma.cc/U5RU-NJMG]. The European Commission has found that the new transatlantic privacy pact, including the rules establishing the U.S. DPRC, provides adequate privacy protections for EU residents. *See* Commission Implementing Decision Pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the Adequate Level of Protection of Personal Data Under the EU-US Data Privacy Framework, COM (2023) 4745 final (July 10, 2023). The CJEU will eventually determine whether the pact is consistent with EU law.

171 Data Protection Review Court, 87 Fed. Reg. at 62304.

172 *Id.* Legislation, rather than regulation, would have been optimal for the DPRC judges' job security. Congress can protect members of multimember adjudicative bodies within the executive branch with "for cause" limits on dismissal. *See* Seila Law LLC v. Consumer Fin. Prot. Bureau, 140 S. Ct. 2183, 2194, 2197–200 (2020) (citing Article II prerogatives of the President in invalidating a legislative provision creating the Consumer Financial Protection Bureau with a single director whom the President could only dismiss "for cause," while recognizing Congress's power to protect multi-member quasi-adjudicative bodies such as the FTC or SEC).

173 Data Protection Review Court , 87 Fed. Reg. at 62304.

174 *Id.* (providing that DPRC will rely on "United States law and legal traditions . . . [including] decisions of the United States Supreme Court"). A court applying U.S. law could readily apply the proportionality and necessity tests identified as key in European decisions because of U.S. precedent requiring narrow tailoring of means to ends in certain executive and legislative measures. *See* City of Cleburne v. Cleburne Living Ctr., 473 U.S. 432 (1985). In addition, the Supreme Court held over a century ago that "[i]nternational law is part of our law," indicating that U.S. courts can in certain situations apply international law when that body of law binds the United States. *See* The Paquete Habana, 175 U.S. 677, 700 (1900). Given the DPRC regulation's citation to decisions of the Supreme Court, it is at least arguable that the DPRC could rely on both principles of narrow tailoring under U.S. constitutional and statutory precedents and on international law values of necessity and proportionality.

The DPRC will issue a decision, which will also include any additional comments by the CLPO.[175]

One central question is whether the DPRC can conduct an iterative review of the kind described here. The FISC has been able to conduct iterative review, although progress has been incremental. The FISC's discussion of abuses in FBI querying practices regarding U.S. persons has identified this problem, formulated remedies, and assessed compliance with the court's decree.[176] There is an obvious link between the degree of FBI compliance with FISC orders and confidence in the DPRC as a robust source of review.[177] The DPRC's independence is a crucial point. The regulation expressly acknowledges the importance of "independent and impartial review."[178] Under the regulations, judges of the DPRC can only be fired for cause. A for-cause dismissal threshold provides a measure of independence, as does the regulation's guarantee that the DPRC will not be under the continuous supervision of the Attorney General.[179] Nevertheless, the DPRC's functioning is tied to the Attorney General's statutory authority to provide advice to the President.[180] Like other senior cabinet-level officials, the Attorney General serves at the President's pleasure.[181] In addition, the statutory sections cited in the regulation make clear the strong institutional tie between the President and the Attorney

175 Data Protection Review Court , 87 Fed. Reg. at 62305. Consistent with practice in the United Kingdom, the complainant will receive a bare-bones decision stating one of two outcomes: either the DPRC did not find any violation of law or the DPRC issued a ruling and required appropriate relief. *Id.*

176 *See* Memorandum Opinion and Order at 42–44, *In re* Section 702 2020 Certification (FISC Ct. Nov. 18, 2020), https://www.intel.gov/assets/documents/702%20Documents/declassified/20/2020_FISC%20Cert%20Opinion_10.19.2020.pdf [https://perma.cc/8FS4-WUNT] (citing FBI violations in conducting U.S.-person queries to section 702 database, while finding that the FBI was still implementing documentation and training requirements that the court had imposed previously); U.S. DEP'T OF JUST. & OFF. OF THE DIR. OF NAT'L INTEL., *supra* note 34, at 42–44 (noting reduction in volume of FBI queries and continued implementation of safeguards ordered by FISC).

177 The FBI's compliance record on U.S.-person queries has been a focus of commentators seeking heightened legal protections. *See* Goitein, *supra* note 159; Jeff Kosseff, *If Congress Wants to Protection Section 702, It Needs to Rein in the FBI*, LAWFARE (Feb. 9, 2023, 1:46 PM), https://www.lawfaremedia.org/article/if-congress-wants-protect-section-702-it-needs-rein-fbi [https://perma.cc/6HL2-BKWU]. The FISC in April 2023 found a substantial reduction in FBI compliance incidents involving querying practices. *See* Memorandum Opinion and Order at 82–88, *In re* DNI/AG 702(h) Certifications, Nos. 702(j)-23-01, 702(j)-23-02, 702(j)-23-03 (FISC Ct. April 11, 2023), https://www.intel.gov/assets/documents/702%20Documents/declassified/2023/FISC_2023_FISA_702_Certifications_Opinion_April11_2023.pdf [https://perma.cc/P2F9-RMMY].

178 Data Protection Review Court, 87 Fed. Reg. at 62304.

179 *Id.*

180 *Id.* (citing 28 U.S.C. §§ 511–512).

181 Seila Law LLC v. Consumer Fin. Prot. Bureau, 140 S. Ct. 2183, 2197–2200 (2020).

General.[182] The Attorney General, who provides opinions to the U.S. government through the Office of Legal Counsel, has an interest in maintaining a reputation for independent decisions.[183] However, the Attorney General's accountability to the President, including amenability to dismissal without cause, is constitutionally required.[184] Against this backdrop, ensuring the independence of the DPRC will be a constant challenge. It remains to be seen whether the DPRC can perform iterative reviews in this demanding environment.

A related question is whether the DPRC has the requisite degree of layered accountability. This should include input from a range of interested parties. It should also include whistleblower protections. The regulations do not expand on current intelligence-community whistleblower safeguards, which are insufficiently robust, since they provide only a narrow channel for conveying complaints, with no provision for public involvement.[185] Moreover, the regulation has no provision for involvement of interested parties, such as nongovernmental organizations (NGOs) focused on free speech and privacy. Organizations like the Electronic Frontier Foundation, the Center for Democracy and Technology, and the Brennan Center for Justice play a vital role in democratic checks on intelligence-community excesses. The U.S. intelligence community, in the wake of Edward Snowden's revelations, has become more open.[186] Moreover, the intelligence community has invited input from NGOs. The civil liberties and privacy officers at ODNI and the National Security Agency (NSA) have been instrumental in that process.[187] The ODNI CLPO will surely continue to cultivate exchanges with external interested parties, such as NGOs. However, a more formal role for NGOs would be useful in this process. It might entail additional measures to prevent unauthorized disclosure of sensitive information. But that would be a small price to pay for NGO involvement.

---

182 *See* 28 U.S.C. § 511 ("The Attorney General shall give . . . advice and opinion on questions of law when required by the President"); *id.* § 512 (providing that a cabinet secretary "may require the opinion of the Attorney General on questions of law" concerning the secretary's department).

183 *See* Trevor W. Morrison, *Stare Decisis in the Office of Legal Counsel*, 110 COLUM. L. REV. 1448, 1460–68 (2010).

184 *Seila Law*, 140 S. Ct. at 2197–200.

185 *See* Note, *Are Intelligence-Community Leakers Internationally Protected Whistleblowers or Simply "Whistling in the Dark"? Assessing the Protections Afforded to Intelligence-Community Whistleblowers Under International Law*, 67 CASE W. RES. L. REV. 897 (2017).

186 The government-sponsored website, *IC on the Record*, is an invaluable resource for scholars, advocates, and the public. *See ODNI Releases 24th Joint Assessment of Section 702 Compliance*, IC ON THE REC. (Dec. 21, 2022), https://icontherecord.tumblr.com/post/704277129317236736/odni-releases-24th-joint-assessment-of-section-702 [https://perma.cc/VL55-BULF].

187 See Schlanger, *supra* note 168, at 141–42 (noting role of NSA CLPO Rebecca Richards, while also cautioning that fostering compliance culture within agencies that participate in U.S. foreign surveillance is a challenging endeavor and that violations continue to occur).

As part of a commitment to layered accountability, the DPRC may be able to address functional trade-offs in surveillance and privacy. As noted above, one crucial trade-off is the conflict between curbing surveillance in the interest of privacy and gathering information about foreign powers and nonstate actors who use cyber means to gain access to sensitive personal information from website users, corporate customers, recipients of government benefits and services, and personnel in the public and private sectors.[188] Freedom from such intrusions is a valuable public good. All persons—save officials of foreign powers and private hackers responsible for the intrusions—benefit from living and working with reduced fear of unauthorized access to their private data. Yet, individual victims of cyber intrusions often lack the information or capacity to protect their own online information.

As with other public goods, government action can compensate for these information asymmetries, skewed agendas, and resource deficits. While in theory, individual victims could pool their knowledge and resources to create more robust protections, collective action problems, disparate agendas, and high transaction costs impede that effort.[189] Many consumers lack the awareness that cyber intrusions are a threat and confront other pressing problems, such as paying off a mortgage or ensuring a quality education for their children. Pursuit and deterrence of foreign threats is particularly difficult for individuals, while government has the information, technological capacity, and structural elements such as security agencies to accomplish this goal.

As part of layered accountability, both the ODNI CLPO and the DPRC could ask whether government was implementing this trade-off between surveillance and safety from foreign intrusions in a proportionate way. For example, evidence suggests that the government queries U.S. persons' data to determine the means, methods, and impact of foreign cyber intrusions.[190] Identifying victims of a cyber intrusion can facilitate alerting victims and understanding foreign intruders' motives and tradecraft. That information can aid in remedying the adverse effects of past cyber intrusions and deterring future attempts. U.S. collection of foreign intelligence might well include similar queries of foreign nationals outside the United States. The DPRC could assess whether the degree of intrusion on foreign national communications accounts was

188 *See* Pozen, *supra* note 30, at 229–32; *see also* U.S. CYBERSPACE SOLARIUM COMM'N, *supra* note 22 (detailing cyber threats).

189 U.S. CYBERSPACE SOLARIUM COMM'N, *supra* note 22; Derek E. Bambauer, *Ghost in the Network*, 162 U. PA. L. REV. 1011, 1031 (2014).

190 *See* Goitein, *supra* note 159.

proportional to the need to remedy and deter foreign cyber intrusions.[191] Surveillance by the United States might be insufficiently tailored to this goal, making that surveillance disproportionate. On the other hand, U.S. surveillance could be reasonably related to the need to deter foreign hackers. The CLPO and DPRC could help resolve this complex question.

Finally, the DPRC goes only part of the way toward establishing an institutionalized opposing voice. The Special Advocate that the DPRC will appoint may adopt some of the function of the amicus that the FISC has appointed in a range of cases.[192] The combination of lawyering skill and subject-matter knowledge will make the Special Advocate a formidable force in the DPRC's deliberations. However, the provisions for the Special Advocate frame that actor's function in narrow terms. The Special Advocate will play an episodic role, expressing views on a particular case. True, the regulations noted that the Special Advocate may provide information to the DPRC, which may involve a survey of developments to date. However, the Special Advocate's own knowledge may be limited by the episodic nature of their role. An expanded and robust public advocate would be better equipped to provide the institutional memory and counterweight that the DPRC needs.

B. *Data Breaches and Disinformation*

Stewardship would have a substantial impact on current practices regarding cybersecurity, data breaches, and online disinformation. To fully implement a stewardship approach, government could enlarge the jurisdiction of the Data Protection Review Court limned above. While a neutral term such as "data protection" can be helpful in identifying the tribunal, the name suggested in this Article, Algorithmic Rights Court, focuses on the design and implementation of machine-learning models.

---

191 *See* United States v. Hasbajrami, 945 F.3d 641, 670–73 (2d Cir. 2019) (holding that U.S.-person queries might constitute a search under the Fourth Amendment and remanding to the district court to determine whether querying was reasonable, given the context of the queries and safeguards that apply); United States v. Muhtorov, 20 F.4th 558, 602–06 (10th Cir. 2021) (applying balancing test from precedent and finding that querying of U.S. person information under section 702 is reasonable under the Fourth Amendment); Emily Berman, *When Database Queries are Fourth Amendment Searches*, 102 MINN. L. REV. 577, 629–37 (2017) (arguing that queries may constitute Fourth Amendment searches requiring heightened safeguards); Alan Z. Rozenshtein, *Digital Disease Surveillance*, 70 AM. U. L. REV. 1511, 1572–75 (2021) (arguing for more robust Fourth Amendment reasonableness standard for "special needs" searches that do not require a warrant); *cf.* Daphna Renan, *The Fourth Amendment as Administrative Governance*, 68 STAN. L. REV. 1039 (2016) (favoring an administrative law approach to surveillance programs such as section 702); Engstrom & Ho, *supra* note 42, at 827–45 (discussing virtues and risks of administrative law approach).

192 *See* Patel & Koreh, *supra* note 34.

The enlargement of DPRC jurisdiction and the name change to Algorithmic Rights Court establishes the ARC's more comprehensive focus. It also enhances the tribunal's ability to consider all relevant factors, including functional trade-offs between algorithmic efficiency and explainability. This Section addresses those concerns in the areas of data breaches and online disinformation.

1. Stewardship, Data Breaches, and Privacy

The ARC could address the adverse effects of data breaches through an enhanced focus on best information practices. The FTC has already displayed an iterative approach in the settlements that it has reached for over fifteen years with corporations whose lax practices resulted in serious data breaches.[193] The settlements have flowed from best practices such as requiring that employees use robust passwords. Best practices now also include multi-factor authentication, which requires users to demonstrate their identity and credentials through more than one gateway, such as a text prompt on a smartphone as well as a laptop log-in.[194] As best practices evolve, FTC action may hinge on other practices that experts believe promote privacy and reduce the risk of data breaches, including requiring that companies prepare a software bill of materials that describes the origin of each component used in a company's code.[195] The agreements between the FTC and companies have included monitoring mechanisms so that the FTC can assess ongoing compliance.[196]

FTC privacy rulemaking could provide even more precise guidance on cybersecurity practices. It could also address issues of targeted advertising and the privacy impact of harvesting users' personal data. In addition, it could address issues of insurance coverage, requiring insurers to provide adequate insurance against cyber risks.[197] Just as importantly, an ARC could break down the sectoral barriers that now balkanize

193 *See* Solove & Hartzog, *supra* note 9.

194 *See* Peter Swire, *The Portability and Other Required Transfers Impact Assessment (PORTIA): Assessing Competition, Privacy, Cybersecurity, and Other Considerations*, 6 GEO. L. TECH. REV. 57, 196–97 (2022).

195 *See* Exec. Order No. 14028, 86 Fed. Reg. 26633 (May 12, 2021).

196 *See* Solove & Hartzog, *supra* note 9.

197 *See* Koyejo-Isaac Idowu, Comment, *The Insurance Data Security Model Law: Strengthening Cybersecurity Insurer-Policyholder Relationships and Protecting Consumers*, 24 ROGER WILLIAMS U. L. REV. 115 (2019).

cybersecurity regulation in the United States.[198] Instead of a welter of disparate sectors, each with its own privacy rules, a comprehensive federal approach could impose uniform guidelines, tailoring those guidelines to the needs of specific business categories, such as power generation. A national regime of review and implementation would bridge the gaps that currently exists, in which some industries, such as health care, are regulated heavily—albeit imperfectly—while others are hardly regulated at all. Moreover, a comprehensive regime would allow information from each sector to influence others, as needed, without the arbitrary divisions that characterize the current sectoral framework.

Layered accountability in the data-breach context would also address functional trade-offs in a uniform rule for disclosure of data breaches to the public and government. Reporting requirements balance several factors: (1) regarding publicly traded corporations, the importance of prompt notice to investors; (2) the burden on reporting entities; and (3) the impact of public disclosure on collaboration between companies and government on the source and extent of the breach. This last factor is crucial. Public disclosure tips off investors, but also tips off cyber intruders that government is tracking them. That tip-off may allow a cyber intruder to cover its tracks, sabotaging an investigation before it has begun. Consider the SEC's proposed requirement that publicly traded corporations disclose material breaches within four business days.[199] As noted above, this proposed rule may be unduly rigid, limiting corporations' opportunities to work with government agencies, including the Department of Homeland Security and the FBI, on the source of the breach.[200] The balance here is delicate. Layered accountability would protect investors while preserving the efficacy of government responses to cyber intrusions.

Similarly, while the proposed SEC rule discusses the importance of disclosing cybersecurity-related corporate governance policies to investors, the proposed rule contains no express requirements regarding such policies.[201] Stewardship would require that cybersecurity

198 Robust state privacy laws like California's privacy legislation are valuable in raising standards but are no substitute for a comprehensive federal approach. *See* California Consumer Privacy Act of 2018, CAL. CIV. CODE § 1798.100 (West 2023).

199 Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure, 87 Fed. Reg. 16590, 16595 (proposed Mar. 23, 2022) (to be codified at 17 C.F.R. pts. 229, 232, 239, 240, 249).

200 *See* Sasha Hondagneu-Messner, Steve McInerney & Alan Charles Raul, *'Cyclops Blink' Shows Why the SEC's Cybersecurity Disclosure Rule Could Undermine the Nation's Cybersecurity*, LAWFARE (Aug. 30, 2022, 8:01 AM), https://www.lawfareblog.com/cyclops-blink-shows-why-secs-proposed-cybersecurity-disclosure-rule-could-undermine-nations [https://perma.cc/S78G-Y2QH].

201 *See* Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure, 87 Fed. Reg. at 16594.

governance meet certain baselines. Those baselines would recognize that an unduly prescriptive approach to governance can interfere with corporate innovation and state prerogatives. At the same time, a nebulous approach can permit a "check the box" mentality that changes little in the real world. A regulator practicing stewardship and a court reviewing the regulator's work would acknowledging these trade-offs by leaving corporations with concrete options, rather than taking a "one size fits all" approach.

As part of layered accountability, stewardship would also consider functional trade-offs between privacy and other values. For example, both European and U.S. officials wish to increase the portability of individual data.[202] Greater portability can open up more options for consumers: if consumers can transfer their data more efficiently between companies, they can more readily find the best companies for their needs. Portability can thus increase competition in the technology sector, breaking down the power of giant companies such as Google, Apple, and Facebook. However, portability also increases risks to privacy. Consider the transfer of anonymized data that is useful for aggregate datasets. The transfer of data heightens the risk that companies involved in the transfer will "re-identify" anonymized data with specific individuals.[203] More detailed disclosure duties and amenability to audits can reduce this risk. Layered accountability would include seeking consumer, corporate, and expert input on the relative weights that government should assign to portability and privacy. This in turn would enhance deliberation on the array of safeguards that will reduce risks to privacy while promoting portability's benefits for consumer choice.

Institutionalized opposition would also be helpful in the cybersecurity arena. While the SEC has generally been a vigilant regulator, gaps in government enforcement have left too many consumers and investors without effective protection. Moreover, even the SEC, because of scarcity of resources, legal uncertainty, and shifting political directions has on significant occasions been "late to the party" of investor protection in particular areas. The SEC did not act quickly enough to temper the effects of the subprime mortgage meltdown that caused the Great Recession of 2008.[204] More recently, the SEC failed to stem the excesses of the cryptocurrency industry that led to the collapse of FTX in late 2022. An institutional voice for consumers and investors at

202 *See* Swire, *supra* note 194.

203 *See id.* at 194–96.

204 *See* Jonathan Macey, Geoffrey Miller, Maureen O'Hara & Gabriel D. Rosenberg, *Helping Law Catch Up to Markets: Applying Broker-Dealer Law to Subprime Mortgages*, 34 J. CORP. L. 789, 804–05 (2009) (noting that players in subprime space included large publicly traded financial firms such as JPMorgan Chase).

the ARC could mitigate these tendencies. It would be proactive in calling the court's attention to troubling trends and have authority to seek remedies to protect the public.

2. Stewardship and the Dilemmas of Combating Online Disinformation

Stewardship's focus on iterative review, layered accountability, and institutionalized opposition can ease the challenge of combating online disinformation and harmful speech. The Meta OB's approach is instructive, although examination of the OB's work reveals gaps that stewardship would fill.

The Meta OB has crafted an iterative approach to review. Consider the OB's candid observation in the *Cross-Check* opinion that Meta was protecting "the wrong content" against automated errors.[205] This insight summed up Meta's track record of favoring the rich and famous, even at the risk of increasing false-negative errors. Each false negative represents potentially viral disinformation or harmful speech, such as incitements to violence against Myanmar's Rohingyas, that Meta should have taken down. Simultaneously, Meta has vastly increased false positives through erroneous takedowns of legitimate information or advocacy by marginalized groups.[206] The OB's requirement in the *Cross-Check* opinion that Meta provide data on compliance at regular intervals also set up an ongoing iterative process.[207]

However, in the layered accountability phase that assesses functional trade-offs, the OB's opinions reveal marked gaps. For example, the *Cross-Check* opinion asserts that machine learning methods often create false positives in the review of postings of marginalized groups.[208] This is a longstanding problem that the Board noted almost two years earlier in *Breast Cancer Symptoms and Nudity*.[209] The *Cross-Check* opinion does observe that Meta's algorithm's might be able to rank posts based on the likelihood of false positives.[210] Precise automated ranking is a useful innovation that would trigger its own benchmarking process. Meta could

---

205 *Cross-Check*, *supra* note 2, ¶ 108.

206 *Id.* ¶ 107.

207 *Id.* at 49.

208 *Id.* ¶¶ 158–160.

209 *Breast Cancer Symptoms and Nudity* at § 6. The difficulty of avoiding false positives in content moderation would counsel a broad interpretation of 47 U.S.C. § 230, a federal provision that shields social media companies from liability for the effects of posts by third parties. *See generally* Gonzalez v. Google LLC, 2 F.4th 871 (9th Cir. 2021), *vacated*, 143 S.Ct. 1191 (2023), *remanded to* 71 F.4th 1200 (9th Cir. 2023) (reading statutory grant of immunity broadly).

210 *Cross-Check*, *supra* note 2, ¶ 159.

undertake to design and deploy algorithms that would rank the probability that a given post was information by a marginalized group, based in part on words or phrases such as "critical perspective" or "subordination," or on a more holistic interpretation of each post's audio, textual, and video content. Having modified its algorithms in this way, Meta could then report to the Board on its progress.

Nevertheless, the two-year interval between *Breast Cancer Symptoms and Nudity* and *Gender Identity* constitutes a gap in layered accountability. The trade-off between false positives and negatives is a perennial issue in automated content moderation.[211] In ongoing assessments of such trade-offs, the optimal time interval between assessments should hinge on the need to address two factors: (1) downturns in overall accuracy caused by an algorithmic entity's backsliding, and (2) prospects for increased accuracy due to improved technology. In the dynamic world of machine learning, two years is an eternity.[212] The Board's lack of focused attention on algorithms allowed the major problem of false positives to fester. Admittedly, the Board was challenged by its own charter, which does not expressly place algorithms within the Board's jurisdiction and purports to preclude Board consideration of many Meta business matters.[213] However, the *Cross-Check* opinion shows that the Board is increasingly willing to call Meta to account for inequities that its algorithmic activity creates or compounds. The operation of Meta's content-moderation algorithms and the prospects for improvement are inextricably related to issues of equity and free expression that the Board has tackled for several years. The Board's path toward resolving this tension between false negatives and positives is strewn with obstacles. Nevertheless, deeper engagement would be beneficial, given the centrality of this issue.

Such challenges bolster the case for institutionalized opposition in content-moderation adjudication. Institutionalized opposition could highlight functional trade-offs and assess the prospects for positive change. Installing such an in-house critic may be a big ask for private industry, where an ostensibly neutral body such as the Meta OB encounters arduous challenges.[214] But the threatened loss of good will that

---

211 *See* Douek, *supra* note 5, at 550–51 (noting this issue and the difficulty of resolving it).

212 *See* Tejas N. Narechania, *Machine Learning as Natural Monopoly*, 107 IOWA L. REV. 1543, 1569–74 (2022) (discussing evolution of machine learning).

213 *See* Thomas E. Kadri, *Juridical Discourse for Platforms*, 136 HARV. L. REV. F. 163 (2022); Arun, *supra* note 160, at 245.

214 *See* Paul W. Grimm, Maura R. Grossman & Gordon V. Cormack, *Artificial Intelligence as Evidence*, 19 NW. J. TECH. & INTELL. PROP. 9, 38 n.128 (2021) (noting allegations that Google fired a persistent in-house researcher and critic, Timnit Gebru, who had written about bias and other adverse effects in Google's AI models).

helped drive the Meta OB's establishment as a neutral adjudicator could also prompt support for an opposing voice.

C. *Discrimination in Credit, Marketing, Employment, and Housing*

Stewardship would also have a role to play in reducing the discrimination caused by algorithmic activity. The brittleness of machine learning training sets, which often fail to include comprehensive context, has helped create this problem.[215] Bias in training sets, including inadequate representation of marginalized groups, has compounded these challenges.[216] Stewardship can address these concerns.

In iterative review, stewardship might start with an administrative agency or an internal tribunal like the Meta OB that could formulate and enforce benchmarking norms. A firm that screens for credit, employment, or other features would have to set standards.[217] For example, those standards would require that data that developers use to train AI models be generally representative of all groups. Consider facial recognition technology. If developers determined that their prior training data had not included sufficient images of women, trans people, or people of color, the developers would have to submit a plan for modifying their training data. These norms would also require periodic assessment of the impact of algorithmic activity on a range of groups, including groups framed by race, gender, ethnicity, and sexual orientation.

Benchmarking would require that firms and developers avoid backsliding. Just as firms identified problems, they would have to identify methods that worked and build on those successes. If training data samples became less inclusive, firms would have to document that issue and explain it.

In addition, a firm would have to consider the use of AI models that use algorithms to perform audits on other AI agents. These models use "post-processing" techniques to identify subgroups within a sample in which initial identification or prediction is inaccurate.[218] Post-processing, which computer scientists have called "multiaccuracy boost,"

---

215 *See* Strandburg, *supra* note 39; WHITE HOUSE, AI BILL OF RIGHTS, *supra* note 2 (discussing importance of combating discrimination in AI); *see also supra* notes 38–41 and accompanying text (discussing discrimination and AI).

216 *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016); Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459 (2019); Pauline D. Kim, *Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action*, 110 CALIF. L. REV. 1539 (2022); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019).

217 *See* NAT'L INST. OF STANDARDS & TECH., *supra* note 65; Coglianese & Hefter, *supra* note 39.

218 *See* Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389 (2019).

substantially increases the accuracy of the overall model. For example, using the boost model on FRT outputs that inaccurately classify Black women results in significant improvements in accuracy.[219] A firm might conclude that the boost method was unsuitable for its work. It would then consider other methods. Documenting that consideration would illustrate the firm's attention to the problem and encourage habits that produce solutions.

Part of that review would compare the performance of algorithms to traditional screening performed by humans. Since humans are prone to pervasive errors of inference and prediction, a comparative assessment of an algorithm's performance is material to the iterative review described here.[220] A particular algorithm might improve substantially on human performance. Or it might backslide from that benchmark. Only careful review can adequately address that issue.

An agency such as the FTC could monitor firms' performance and request remedies, if needed. If firms believed that the FTC was incorrect or its remedy was too broad, appeal would be available to federal circuit courts of appeal. As an alternative, an industry-wide trade association could set up a tribunal with rules similar to the Meta OB. A tribunal would have to exhibit the same iterative qualities that the Meta OB has displayed in cases such as *Breast Cancer Symptoms and Nudity*, *Gender Identity*, and *Cross-Check*, which apply benchmarking principles by documenting failures to live up to norms and requiring assessments to measure progress. To be robust, any decision-maker, be it an agency or a trade tribunal, would have to provide sufficient information about algorithmic activity to make its decisions intelligible to the public. That would entail greater information about technical issues than the Meta OB has provided to date, although the tribunal could frame the discussion at a level of generality that would preserve trade secrets.

---

219 *See* Michael P. Kim, Amirata Ghorbani & James Zou, *Multiaccuracy: Black-Box Post-Processing for Fairness in Classification*, AIES '19: PROC. 2019 AAAI/ACM CONF. ON AI, ETHICS, & SOC'Y 247, 248–49 (2019); *see also* Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman & Aram Galstyan, *A Survey on Bias and Fairness in Machine Learning*, 54 ACM COMPUTING SURV., July 2021, at 1 (discussing bias in machine learning and ways to address it); ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER, MORE INCLUSIVE FUTURE (2022) (discussing multiaccuracy and related boost methods); Gillis, *supra* note 10, at 1254–56 (discussing use of baselines derived from traditional credit screening to ascertain performance of algorithms).

220 *See* KAHNEMAN, SIBONY & SUNSTEIN, *supra* note 51, at 138–47 (discussing human beings' overconfidence in the quality and accuracy of their decisions, compared with the wide variations in human decisions about similarly situated cases in hiring and other contexts); Selmi, *supra* note 10; Charles A. Sullivan, *Employing AI*, 63 VILL. L. REV. 395 (2018); Coglianese & Lai, *supra* note 6; Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour & Kristian Lum, *Algorithmic Fairness: Choices, Assumptions, and Definitions*, 8 ANN. REV. STAT. & ITS APPLICATION 141 (2021).

In terms of layered accountability, such a decision-maker would require timely disclosure of standards, assessments, and remedial plans. The decision-maker would also solicit input from a broad range of interested parties, including scholars, practitioners, and consumer and advocacy groups.[221] Trade tribunals would have to follow Meta's lead and fund a feedback infrastructure. Engagement at this level would make for better decisions, in part by identifying faulty assumptions or neglected elements of context that developers might not fully appreciate.

An agency or trade tribunal would also have to establish institutionalized opposition. If industry argued that a given algorithmic activity was accurate and unbiased, an institutional voice would challenge that view, putting industry to its proof. That pushback would ensure that a tribunal received the sharpest possible framing of issues.

As part of this process, a tribunal would have to consider functional trade-offs. For example, prior to deploying a screening algorithm, industry would have to consider whether it could preserve the algorithm's accuracy while also increasing its explainability.[222] It would then need to document the nature and scope of that consideration, including machine learning tools that it evaluated. Social media firms like Meta that proposed new algorithmic tools for limiting harmful speech would have to address overenforcement that impinged on the free expression of marginalized groups. If industry assessed the accuracy of past algorithmic activity, it would also have to explain how it conducted that assessment in a manner that preserved the privacy of users and applicants. Approaches to these functional trade-offs would then become models for subsequent industry efforts.

This approach would not banish bias from AI applications. However, it would provide information about the problem and about tools to combat it. It would enable informed arguments, free from both alarmism and undue complacency. Moreover, it would promote sound habits within algorithmic entities. These achievements form a notable signature for stewardship.

CONCLUSION

From a modest start, the movement for accountable algorithms has achieved a consensus among government officials, scholars, advocates, and business executives. That consensus results from acknowledgement

---

221 *See* Michele Estrin Gilman, *Expanding Civil Rights to Combat Digital Discrimination on the Basis of Poverty*, 75 SMU L. REV. 571 (2022).

222 *See* Clement, Kemmerzell, Abdelaal & Amberg, *supra* note 39; Gohel, Singh & Mohanty, *supra* note 39.

of problems that algorithms have either caused or compounded, including widespread surveillance, data breaches and social media disinformation, and bias in applicant screening in credit, housing, employment, and government benefits. However, despite this emerging consensus, proposed solutions suffer from gaps and reflect continuing disagreement.

Gaps and ongoing controversies are evident in debates about the utility of various modes of regulation. For some, prohibition of certain uses of AI is the only answer. With respect to most AI applications, the majority of government officials, commentators, and advocates view prohibition as a blunt instrument that bars the benefits AI can bring in efficiency and relief from pervasive errors in human judgment. Instead, most recommend a panoply of curative measures, including standards, assessments, explanations, and procedural safeguards. However, at least in the United States, the conception and proposed implementation of these models of reform have been spotty and riddled with contradictions.

In the United States, part of the problem has been the challenge of overcoming the sectoral approach that has long characterized U.S. data protection efforts. Cybersecurity is the most obvious example of the sectoral approach, where safeguards for data vary widely across disparately regulated domains, with exacting standards for health care and critical infrastructure such as the power grid but little comprehensive regulation elsewhere. In addition, a sectoral approach has produced arbitrary divisions among data processing domains, with different standards for surveillance, online disinformation, and algorithmic screening.

Controversies and blind spots extend beyond this sectoral divide. For example, scholars have disagreed robustly on the virtues of adjudication, with some scholars discounting the value of prominent examples of adjudication, such as the Meta OB. These critiques have failed to address the pivot toward programmatic remedies in the Board's cases, and have thus missed the opportunity to pinpoint more precisely ongoing challenges that the Board confronts. In addition, many scholars urging algorithmic accountability have failed to fully grapple with functional trade-offs, such as the importance of tailored surveillance as a tool for combating cyber intrusions that impinge on consumers' privacy.

This Article has sought to fill gaps and enhance consistency across domains with a stewardship model. Stewardship requires iterative review, layered accountability, and institutionalized opposition. Iterative review sets benchmarks through adjudication and requires improvement over time, as the Meta OB has begun to do in cases such as *Cross-Check* with the problems of over- and underenforcement of Meta's content moderation policies. Layered accountability requires broad input from

stakeholders, including consumers, advocates, and experts. It also addresses functional trade-offs, such as the conflict between algorithms' accuracy and explainability. Lastly, institutionalized opposition sets up a constant counterweight to established narratives in government and the private sector. It thus reduces the risk of capture of regulators by algorithmic entities.

The stewardship model provides fresh insights on the prospects for the Biden administration's new Data Protection Review Court, as well as approaches to cybersecurity, online disinformation, and algorithmic discrimination. No reform model will wholly eliminate the adverse impacts of various AI applications. However, the stewardship model offers a consistent approach to reform efforts. That guidance represents a decisive step toward digital accountability.