TechREG CHRONICLE

MACHINE LEARNING

FEBRUARY 2023

CPI COMPETITION POLICY INTERNATIONAL

Competition Policy International, a What's Next Media and Analytics Company

LETTER FROM THE EDITOR

Dear Readers,

Consider the following:

> **Q**: "ChatGPT, write a brief editorial introduction to a set of law review articles on the legal implications of machine learning and language models."
>
> **A**: "As artificial intelligence ("AI") technologies become increasingly sophisticated, their impact on the legal landscape cannot be ignored. One of the most prominent AI applications is machine learning, which enables computer systems to learn and improve from experience without being explicitly programmed. In recent years, machine learning has facilitated the development of powerful language models that can generate human-like text.
>
> However, these language models raise a host of legal issues that require careful consideration. In this set of law review articles, we explore the legal implications of machine learning and language models. Our contributors analyze a range of issues, including intellectual property, data privacy, bias and discrimination, and ethical considerations.
>
> One key question is whether language generated by machine learning models can be protected by copyright. Other issues include the use of personal data to train these models, the potential for algorithmic bias and discrimination, and the ethical implications of using language models for sensitive tasks, such as criminal sentencing or medical diagnoses.
>
> As the use of language models becomes more widespread in both the public and private sectors, it is important that policymakers, lawyers, and scholars alike understand the legal implications of these powerful tools. Our contributors offer thoughtful analysis and recommendations for how the law can adapt to the challenges posed by machine learning and language models."

CPI COMPETITION POLICY® INTERNATIONAL

The foregoing is an unedited interaction with a well-known AI chatbot. As this exchange should make clear, machine learning, and, specifically, language models, are advancing at an alarming pace. Appropriately, the authors of the pieces in this edition of the Chronicle address the issues that, almost existentially, ChatGPT raises regarding itself.

Gary E. Marchant opens by noting that at the international level, no binding treaties or conventions to regulate AI are likely anytime soon, but organizations such as the OECD and UNESCO have developed non-binding recommendations that can help guide AI/ML governance by governments and industry. Major AI powers such as China and the European Union are putting in place legislative frameworks for AI with uncertain impacts and effectiveness, whereas the U.S. Congress has not enacted any substantive controls on AI/ML to date. Rather, various federal agencies have started producing guidance documents and recommendations, primarily focused on discouraging algorithm applications with biased or discriminatory impacts. Most governance of AI/ML consists of a variety of "soft law" programs. Given the central role these programs in AI/ML governance, it is important to make these programs more effective and credible.

Cary Coglianese expands on the heterogeneous nature of AI. The term refers to a vast array of algorithms that are put to varied uses, including transportation, medicine, social media, marketing, and others. Not only do they vary widely across their types and uses, but they evolve constantly. Due to the staggering heterogeneity of these algorithms, multiple regulatory agencies will be needed to regulate the use of machine learning, each within their own discrete area of specialization. Even specialized expert agencies, though, still face the challenge of heterogeneity and must approach their task of regulating machine learning with agility. Regulators should consider how to use machine-learning tools to enhance their ability to protect the public from the adverse effects of machine learning.

Further broadening the canvas, **Danni Yu & Benjamin Cedric Larsen** discuss how different AI regulatory regimes are emerging across Europe, the United States, China, and elsewhere. The author queries how these new regulatory regimes have implications for companies and their adoption of self-regulatory and compliance-based tools and practices. The article outlines how and where AI regulations emerge and how these, in some cases, seem to be on divergent paths.

Marco Almada takes a broader perspective, noting how the regulation of digital technologies, by necessity, draws from various regulatory techniques. He pleads in favor of "regulation by design," by which laws or regulations would specify requirements that software designers would need to follow when creating new systems. The paper examines the suitability of "regulation by design" approaches to the specific case of machine learning, arguing that such an approach is potentially useful, but would need to have a narrow scope of application. Drawing from EU law examples, the paper claims to show how "regulation by design" relies on the delegation of normative definitions and enforcement to software designers. Given the risks of such an approach, such delegation would only be effective if certain conditions are satisfied.

Wrapping up, **Heather Egan Sussman, Ian Adams & Nur Lalj** discuss the differing approaches to regulating AI and ML in Europe and at the federal and state levels in the United States. The article proposes best practices for building compliance. Finally, **Thomas Freeman & Aaron McKain** note the profound interaction between AI systems and privacy laws. Their article outlines how the legal system and society at large need to determine what information about individuals can be gathered and maintained and when and how that data can be used to judge individuals. It is essential to have thoughtful conversations about the core principles for digital law and ethics. Those conversations should involve broad, diverse, and interdisciplinary groups, which can consider factors such as biases in historical data, whether a given algorithm is being programmed or trained appropriately, and what type of decisions we are comfortable automating or trusting algorithms to make.

As always, many thanks to our great panel of authors.

Sincerely,
CPI Team

TABLE OF CONTENTS

COMPETITION POLICY® INTERNATIONAL

MACHINE
LEARNING

FEBRUARY 2023

SUMMARIES

REGULATING MACHINE LEARNING

By Gary E. Marchant

Artificial intelligence using machine-learning ("AI/ML") is already providing countless benefits to society, but is also presenting some risks and concerns that require governance. Yet, the rapid pace of AI/ML, the many diverse applications and industries across which it is being implemented, and the complexity of the technology itself challenge effective governance. At the international level, no binding treaties or conventions are likely anytime soon, but organizations such as the OECD and UNESCO have developed non-binding recommendations that can help guide AI/ML governance by governments and industry. Other major AI powers such as China and the European Union are putting in place legislative frameworks for AI with uncertain impacts and effectiveness, whereas the U.S. Congress has not enacted any substantive controls on AI/ML to date. Rather, various federal agencies have started producing guidance documents and recommendations, primarily focused on discouraging algorithm applications with biased or discriminatory impacts. Some state and local governments are also in the process of starting to adopt some restrictions on problematic AI/ML applications and uses. At this time, most governance of AI/ML consists of a variety of "soft law" programs. Given the central role these programs in AI/ML governance, it is important to make these programs more effective and credible.

REGULATING MACHINE LEARNING: THE CHALLENGE OF HETEROGENEITY

By Cary Coglianese

Machine learning, or artificial intelligence, refers to a vast array of different algorithms that are being put to highly varied uses, including in transportation, medicine, social media, marketing, and many other settings. Not only do machine-learning algorithms vary widely across their types and uses, but they are evolving constantly. Even the same algorithm can perform quite differently over time as it is fed new data. Due to the staggering heterogeneity of these algorithms, multiple regulatory agencies will be needed to regulate the use of machine learning, each within their own discrete area of specialization. Even these specialized expert agencies, though, will still face the challenge of heterogeneity and must approach their task of regulating machine learning with agility. They must build up their capacity in data sciences, deploy flexible strategies such as management-based regulation, and remain constantly vigilant. Regulators should also consider how they can use machine-learning tools themselves to enhance their ability to protect the public from the adverse effects of machine learning. Effective regulatory governance of machine learning should be possible, but it will depend on the constant pursuit of regulatory excellence.

EMERGING AI REGULATORY ECOSYSTEMS: IMPLICATIONS FOR BUSINESSES AND REGULATORS

By Danni Yu & Benjamin Cedric Larsen

Different AI regulatory regimes are currently emerging across Europe, the United States, China, and elsewhere. But what do these new regulatory regimes mean for companies and their adoption of self-regulatory and compliance-based tools and practices? This article outlines how and where AI regulations emerge and how these, in some cases, seem to be on divergent paths. Second, it discusses what this means for businesses and their global operations. Third, it comments on a way forward in the growing complexities of AI use and regulation, as it exists between soft law practices and emerging hard law measures.

REGULATING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

By Heather Egan Sussman, Ian Adams & Nur Lalji

Artificial Intelligence ("AI") and machine learning ("ML") have the potential to create breakthrough advances in a range of industries, but they also raise novel legal, ethical, and privacy questions that will likely define the next era of technological advancement. Over the last several years, there has been a flurry of AI- and ML-related regulations and guidance issued by international bodies, governments, and regulators seeking to mitigate the risks posed by AI and ML, especially when these technologies are used to make important decisions related to employment or healthcare. Given the proliferation of these technologies across various industries, more regulation is likely to come. Organizations with AI and ML-based products and services should understand and consider how existing laws apply to them, as well as how the changing regulatory landscape may impact their business plans going forward. In this article, we discuss the differing approaches to regulating AI and ML in Europe and at the federal and state levels in the United States and the best practices for building compliance.

PRINCIPLES OF DIGITAL LAW AND ETHICS
By Thomas Freeman & Dr. Aaron McKain

As more personal data is collected and more decisions that affect individuals are automated, individual rights are increasingly threatened. The legal system and society at large need to determine what information about individuals can be gathered and maintained and when and how that data can be used to judge individuals. It is essential that we have thoughtful conversations about the core principles for digital law and ethics. Those conversations should involve broad, diverse, and interdisciplinary groups, which can consider factors such as biases in historical data, whether an algorithm is being programmed or trained appropriately, and what type of decisions we are comfortable automating or trusting algorithms to make. The best safeguard of our digital rights will ultimately be engaging diverse teams that thoughtfully consider how their fellow humans are affected as they establish legal and ethical guardrails around emerging technology.

REGULATING MACHINE LEARNING BY DESIGN
By Marco Almada

The regulation of digital technologies around the world draws from various regulatory techniques. One such technique is regulation by design, in which regulation specify requirements that software designers must follow when creating any systems. This paper examines the suitability of regulation by design approaches to machine learning, arguing that they are potentially useful but have a narrow scope of application. Drawing from EU law examples, it shows how regulation by design relies on the delegation of normative definitions and enforcement to software designers, but such delegation is only effective if a few conditions are present. These conditions, however, are seldom met by applications of machine learning technologies in the real world, and so regulation by design cannot address many of the pressing concerns driving regulation. Nonetheless, by-design provisions can support regulation if applied to well-defined problems that lend themselves to clear expression in software code. Hence, regulation by design, within its proper limits, can be a powerful tool for regulators of machine learning technologies.

REGULATING MACHINE LEARNING

BY
GARY E. MARCHANT

Regents Professor and Faculty Director, Center for Law, Science & Innovation, Sandra Day O'Connor College of Law at Arizona State University.

Artificial intelligence ("AI") has surged in its applications, public awareness, and policy priority in recent years. Several technical advances have driven this surge, including faster computer processors, unprecedented availability of massive sets of data and images on the internet, rapidly improved capabilities in optical recognition, and greatly improved abilities of computers to understand and interact with written and verbal human speech, a skill known as natural language processing. Yet the most important factor driving AI forward has been the rise of machine learning ("ML"). In contrast to previous models of AI in which a human programmer codes a set of instructions for the AI to follow (rule-based AI), in ML the machine learns itself by processing data and incrementally learning from that data (data-based AI).

ML AI has already achieved many valuable benefits, with many more to come. But is has also generated some concerns, which must be effectively governed if we are to enjoy the full benefits of this technology.[2] This comment summarizes the challenges and opportunities of governing ML AI. Part I discusses unique issues and problems in governing ML. Part II addresses the international framework and status for AI governance. Part III summarizes U.S. government efforts to regulate AI to date. Finally, Part IV discusses a "soft law" alternative to traditional government regulation of AI.

01

GOVERNANCE CHALLENGES OF AI MACHINE LEARNING

The capabilities of ML turn out to far surpass those of earlier AI models, which explains the recent proliferation of AI usefulness across virtually every industry sector and human activity. But ML also presents some unique policy and governance challenges. For one, because AI systems require large sets of data to learn from, they have an almost insatiable need for data, including data that may present significant privacy concerns. Unlike earlier products, ML algorithms continue to learn and thus evolve throughout their lifespan, making obsolete regulatory approval systems based on a "once and done" government review.

Another complication with ML systems is that the data they are trained on is derived from actual human experience, which often reflects various types of societal bias. The ML algorithms will often replicate or even amplify the biases hidden in the training data, which can result in discrimination against under-privileged groups in applications such as criminal justice or hiring.[3] ML systems also do not follow pre-set human-created instructions, but rather are capable of making their own decisions as they learn, creating unique issues of who is accountable when a machine makes a decision. Finally, ML systems currently cannot explain their decisions, so their reasoning remains a black box.[4]

In addition to the substantive aspects of ML, the dynamic adoption of ML also creates governance challenges. ML applications are developing and evolving at a frantic pace, much faster than traditional regulatory systems can keep up, creating a pacing problem.[5] Moreover, even if new rules are enacted, they will quickly be out of date, and nations are understandably concerned about "freezing" in place their AI technology with outdated regulations in a highly competitive global economy. Another challenge is that AI is being applied across every industry in the economy, and spanning almost every regulatory agency, creating a formidable coordination problem.[6] AI also potentially presents a broad range of potential risks, going beyond health and safety risks traditionally regulated by governments to also include other concerns that agencies have less experience and delegated authority to regulate, such as privacy, bias, fairness, worker displacement, autonomy, lack of transparency, and more.[7] Finally, AI has international applications, making national regulation problematic.[8]

02

INTERNATIONAL FRAMEWORK FOR AI GOVERNANCE

While there has been much discussion about possible international regulatory instruments for AI, especially for lethal

2 Wendell Wallach & Gary Marchant, *Toward the Agile and Comprehensive International Governance of AI and Robotics,* 107 Proceedings of the IEEE 505, 505-06 (2019).

3 Nicol Turner Lee, Paul Resnick & Genie Barton, Algorithmic Bias Detection and Mitigation: Best Practices and Policies To Reduce Consumer Harms, Brookings Inst., May 22, 2019, available at https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/#footref-6.

4 Will Knight, *The Dark Secret at the Heart of AI*, Technology Review, April 11, 2017, available at https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/.

5 Gary E. Marchant, *The Growing Gap Between Emerging Technologies and the Law*, *in* The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem 19, 22–23 (Gary E. Marchant et al. eds., 2011.

6 Wallach & Marchant, *supra* note 2, at 505.

7 *Id.* at 505-06.

8 *Id.* at 506.

autonomous weapons, no international treaties or conventions on AI have been enacted. Various international organizations have adopted non-binding international guidelines on AI, including UNESCO[9] and the OECD.[10] Other international initiatives, such as the Global Partnership on AI, led by Canada and France,[11] have also considered international governance options, but nothing concrete has come of such efforts to date.

In the absence of any binding international AI regulation for the foreseeable future, many jurisdictions pursuing AI technology have also been developing their own regulatory frameworks. Most notable is the European Union ("EU"), which is actively developing a comprehensive regulatory program known as the AI Act,[12] anticipated to be completed in 2023 and to take effect in 2024. The draft EU AI Act takes a risk-based approach and applies different regulatory requirements to different tiers. The highest risk applications that present a central threat to fundamental rights are banned outright, high-risk applications are subject to conformity assessments, and lower risk applications rely on industry standards and other soft law measures.[13]

The third major AI power in addition to the U.S. and E.U. is China, which has promulgated a series of AI regulatory programs. Some of these requirements are unique to China, such as the requirement that recommendation algorithms must "vigorously disseminate positive energy," but others address more common ML governance challenges such as transparency and accountability.[14] On March 1, 2022, another major set of AI regulations took effect in China that among other things prevented companies from discriminating among users in price based on ML algorithms.[15] Many other countries such as Australia, Canada, the U.K., Japan, Singapore and others have adopted their own AI policy frameworks, but have generally not yet enacted enforceable requirements that apply to individual companies.

03

U.S. GOVERNMENT REGULATORY INITIATIVES

There have been several bills in the U.S. Congress to regulate AI, most notably the Algorithmic Accountability Act, the most recent iteration of which would mandate the Federal Trade Commission ("FTC") to require impact assessments for high-risk automated decision systems.[16] This proposed bill was not enacted, and although a similar bill is likely to be introduced in the new Congress, there is no evidence it will fare any better than previous versions. Absent a major accident or abuse, which is usually needed to trigger Congress to adopt new statutes, it unlikely that Congress will undertake major legislative change on AI anytime soon. Instead, the U.S. government is likely to approach AI in the same way it had governed other emerging technologies such as the internet, biotechnology and nanotechnology, relying primarily on existing regulatory agencies and statutes to apply oversight, supplemented by private governance initiatives. This results in a more decentralized, sector-specific, and incremental governance approach, quite distinct from the European approach of centralized, top-down control.[17]

U.S. government policy on AI through the Obama, Trump, and Biden (so far) administrations has consisted of a "light touch" sector-specific approach that has become gradually more proactive as AI technology and applications have advanced over the past decade.[18] The U.S. government first started identifying AI as a policy priority in the latter days of the Obama Administration, when a subcommittee on ML/AI was created by the White House Office of Science and Technology Policy ("OSTP") to coordinate government AI policy.

---

9 https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

10 https://oecd.ai/en/ai-principles.

11 Launch of the French-Canadian Initiative Global Partnership on Ai (GPAI) (June 15, 2020), available at https://ai-regulation.com/launch-of-the-french-canadian-initiative-global-partnership-on-ai-gpai/.

12 EUROPEAN COMMISSION, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, (2021), available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

13 *Id.*

14 See Matt Sheehan, China's New AI Governance Initiatives Shouldn't Be Ignored, Carnegie Endowment for International Peace, Jan. 4, 2022, available at https://carnegieendowment.org/2022/01/04/china-s-new-ai-governance-initiatives-shouldn-t-be-ignored-pub-86127.

15 Jennifer Conrad & Will Knight, *China Is About to Regulate AI – And the World is Watching*, Wired (Feb. 22, 2022), available at https://www.wired.com/story/china-regulate-ai-world-watching/.

16 Algorithmic Accountability Act of 2022, H.R. 6580, 117th Cong. (2021-22).

17 See Adam Thierer, *U.S. Artificial Intelligence Governance in the Obama-Trump Years*, 2 IEEE Trans. Tech. & Soc'y 175, 179 (2021).

18 *Id.* at 176.

The OSTP subcommittee held a series of public hearings across the country and issued reports, including one entitled *Preparing for the Future of AI*.[19] This report raised several concerns about the implementation of AI/ML, such as the potential for discrimination based on biased data used to train ML systems, but noted that experts agreed "that broad regulation of AI research or practice would be inadvisable at this time" and instead called for relying on existing statutory authority to address problems created by AI.[20]

The Trump Administration was somewhat more active on AI, but continued the "light touch" approach of his predecessor. President Trump issued Executive Order 13859 in February 2019 that emphasized the need for the U.S. to retain global leadership in AI.[21] While much of the Executive Order focused on enhancing investment and innovation in AI, on the regulatory side it called upon the National Institute of Standards and Technology ("NIST") to promote standard-setting on AI, and instructed the Office of Management and Budget ("OMB") to produce a memorandum on regulatory principles for AI that federal agencies should follow. That guidance memorandum was finalized in November 2020, and identified ten principles for regulation of AI, with an emphasis on ensuring safety, but also advising US regulatory agencies to consider "nonregulatory approaches for AI."[22]

Just as the Biden administration was about to take office, Congress passed the National Artificial Intelligence Initiative Act of 2020, which took effect on January 1, 2021.[23] This bipartisan statute created the National AI Initiative, which "provides an overarching framework to strengthen and coordinate AI research, development, demonstration, and education activities across all U.S. Departments and Agencies, in cooperation with academia, industry, non-profits, and civil society organizations."[24] The Initiative is structured around six "strategic pillars" – Innovation, Advancing Trustworthy AI, Education and Training, Infrastructure, Applications, and International Cooperation.[25] This Initiative created a framework for the incoming Biden administration to structure its AI activities.

> ***"The Trump Administration was somewhat more active on AI, but continued the "light touch" approach of his predecessor***

To date, the Biden administration has continued the sector-specific approach that relies on existing statutory authorities, with no proposals or efforts to establish comprehensive regulation of AI. However, many federal agencies have ramped up their focus on AI under the Biden presidency. Perhaps the highest profile activity was the promulgation of a "Blueprint for an AI Bill of Rights" by the OSTP in October 2022.[26] The Blueprint set forth five principles for responsible AI: (1) safe and effective systems; (2) algorithmic discrimination protections; (3) data privacy; (4) notice and explanation; and (5) human alternatives, consideration, and fallback.[27] The proposed Bill of Rights received mixed reviews, with one frequent criticism being that the document was "toothless."[28]

Several other agencies have started new AI guidance or enforcement initiatives for specific industry sectors, mostly driven by the potential for bias from ML systems. The FTC

19 Executive Office Of The President National Science And Technology Council Committee On Technology, Preparing For the Future of Artificial Intelligence (Oct. 2016), available at https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

20 *Id.* at 17.

21 President Donald Trump, Executive Order 13859: Maintaining American Leadership in Artificial Intelligence, 84 Fed. Reg. 3964 (Feb. 14, 2019).

22 Russell T. Vought, OMB Director, Guidance for Regulation of Artificial Intelligence Applications (Nov. 17, 2020), available at https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf.

23 See National Artificial Intelligence Initiative Act of 2020, §§ 5001 et seq., 2020 Defense Appropriations Act (Dec. 3, 2020), available at https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210.

24 National Artificial Intelligence Initiative (NAII), About NAII (undated), available at https://www.ai.gov/about/#NAII-NATIONAL-ARTIFICIAL-INTELLIGENCE-INITIATIVE.

25 *Id.*

26 OSTP, Blueprint For an AI Bill Of Rights: Making Automated Systems Work For the American People (Oct. 2022), available at https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

27 *Id.*

28 See, e.g. Khari Johnson, *Biden's AI Bill of Rights Is Toothless Against Big Tech,* Wired, Oct. 4, 2022, available at https://www.wired.com/story/bidens-ai-bill-of-rights-is-toothless-against-big-tech/.

has been at the forefront of these efforts. In April 2021, the FTC issued a statement notifying stakeholders that it intends to use its authority under the 1970 Fair Credit Reporting Act, the 1970 Equal Credit Opportunity Act, and section 5 of the FTC Act to ensure that AI systems are fair, transparent and truthful.[29] The FTC has used this authority to take enforcement action against a number of algorithmic AI products that violate its principles, including applying a new remedy of "algorithmic disgorgement" to require an offending company to destroy all records of the relevant algorithm.[30] Perhaps most significantly, the FTC published an advance notice of proposed rulemaking in August 2022 on possible new regulations "concerning the ways in which companies collect, aggregate, protect, use, analyze, and retain consumer data…."[31] Although this notice applied broadly to all types of commercial surveillance, it included a section specifically addressing automated decision-making systems (i.e. ML).[32]

The Food and Drug Administration ("FDA") has been particularly proactive in considering the impact of AI and ML for its regulatory programs. Many medical devices are using AI and ML – the FDA has already approved over 500 such devices.[33] One problem with the traditional FDA regulatory model for medical devices is it assumes that the products are static, and thus once approved, they will remain the same for their useful life. AI devices using ML are dynamic in that they continue to learn and improve even after FDA approval, which the existing FDA oversight approach does not accommodate or address. The FDA released a discussion paper and then a follow-up action plan to create a revised regulatory approval pathway for AI/ML systems given their unique dynamic nature.[34] FDA also explored the development of a software pre-certification program to allow more flexible approval of complex software programs such as those using AI/ML.[35] Unfortunately, the FDA determined that this model would not comport with its existing statutory authority and thus would not proceed further with the program,[36] a clear example of an outdated regulatory statute blocking an innovative governance approach.

The Food and Drug Administration ("FDA") has been particularly proactive in considering the impact of AI and ML for its regulatory programs

The Department of Transportation has also actively engaged the development of AI for autonomous vehicle driving systems, including publishing a series of major reports providing guidance for industry and state and local governments on the safe development of autonomous vehicles.[37] These reports primarily rely on private standards to ensure autonomous vehicle safety, but the agency has recently issued a request for comment on a governance framework for autonomous driving system safety.[38] The National Institute of Standards and Technology ("NIST)" has also been very active in interacting with private standard-setting efforts, by issuing a series of recommendations on topics such as explainable AI, AI bias, and risk management that can inform both standards-setting bodies and individual companies.[39] Other federal agencies are also taking action by issuing various types of guidance documents, including the Equal Employment Opportunity

29 FTC, *Aiming for Truth, Fairness, and Equity In Your Company's Use of AI* (April 19, 2021), available at https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

30 See Kate Kaye, *The FTC's New Enforcement Weapon Spells Death for Algorithms*, Protocol, March 14, 2022, available at https://www.protocol.com/policy/ftc-algorithm-destroy-data-privacy.

31 FTC*, Trade Regulation Rule on Commercial Surveillance and Data Security*, 87 Fed. Reg. 51273 (Aug. 22, 2022).

32 *Id.* at 51283-84.

33 FDA, Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices (Oct. 5, 2022), available at https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.

34 FDA, Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan (Jan. 2021), available at https://www.fda.gov/media/145022/download.

35 FDA, Developing Software Precertification Program: A Working Model, v 2.0, June 2018, available at https://www.fda.gov/media/113802/download.

36 FDA, The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings (Sept. 2022), available at https://www.fda.gov/media/161815/download.

37 DOT, USDOT Automated Vehicles Activities, available at https://www.transportation.gov/AV.

38 NHTSA, Framework for Automated Driving System Safety, 85 Fed. Reg. 78058 (Dec. 2, 2020).

39 NIST, Trustworthy and Responsible AI, available at https://www.nist.gov/programs-projects/trustworthy-and-responsible-ai.

Commission ("EEOC"),[40] the Department of Health and Human Services ("DHHS"),[41] the Consumer Product Safety Commission ("CPSC"),[42] the Department of the Treasury,[43] the Consumer Financial Protection Bureau ("CFPB"),[44] and the Federal Housing Finance Agency ("FHFA").[45]

In addition to these federal efforts, some state and local governments have also begun regulatory initiatives on AI/ML. At the state level, California has been most active, and is pursuing a number of regulatory measures for AI. The State has recently proposed amendments to its employment anti-discrimination laws that would impose liability on companies using AI tools that discriminate against protected groups.[46] California has also adopted a law that requires AI bots to disclose their non-human nature.[47] California's data privacy statutes, specifically the California Consumer Privacy Act of 2018 as amended by the California Privacy Rights Act of 2020, will apply to many AI applications using ML, since they will often use consumer data. California and several other states are in the early stages of trying to adopt other measures relating to AI, although many such initiatives have been unsuccessful in previous years.[48] At the local level, New York City is leading the way by enacting Local Law 144 that will require employers to conduct a bias audit before using any algorithm in the hiring process, and will require notification to job applicants before its use.[49] This law was originally scheduled to take effect on January 1, 203, but has now been delayed to April 15, 2023.[50]

In summary then, we have seen a significant ramping up of activity relating to AI by US regulatory agencies in the past couple years, primarily at the federal level but also at the state level, but this activity is limited to applying existing statutory authority to AI. Many of these statutes were enacted decades ago, long before the modern wave of AI/ML and there does not appear to be any momentum in Congress towards adopting comprehensive AI legislation. As such, U.S. government regulation of AI will likely remain limited for the foreseeable future, and various soft law initiatives, discussed in the next section, are likely to continue to play a central role in AI governance.

04

AI SOFT LAW

The light touch of AI regulation in the U.S. has been supplemented with soft law to fill the governance gaps, if not voids. Soft law are programs that set substantive expectations but which are not directly enforceable by governments.[51] Soft law comes in many different forms, including private standards, codes of conduct, best practices, statements of principles,

40 EEOC, *Draft Strategic Enforcement Plan*, 88 Fed. Reg. 1379, 1381 (Jan. 10, 2023).

41 DHHS*, Nondiscrimination in Health Programs and Activities*, 87 Fed. Reg. 47824 (Aug. 4, 2022).

42 CPSC, Artificial Intelligence and Machine Learning In Consumer Products (May 19, 2021), available at https://www.cpsc.gov/s3fs-public/Artificial-Intelligence-and-Machine-Learning-In-Consumer-Products.pdf.

43 Department of Treasury et al., *Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning*, 86 Fed. Reg. 16837 (March 31, 2021).

44 CFPB, *Consumer Financial Protection Circular 2022–03: Adverse Action Notification Requirements in Connection With Credit Decisions Based on Complex Algorithms*, 87 Fed. Reg. 35864 (June 14, 2022).

45 FHFA, Advisory Bulletin AB 2022-02: Artificial Intelligence/Machine Learning Risk Management (Feb. 10, 2022), available at https://www.fhfa.gov/SupervisionRegulation/AdvisoryBulletins/AdvisoryBulletinDocuments/Advisory-Bulletin-2022-02.pdf.

46 Fair Employment & Housing Council, Draft Modifications to Employment Regulations Regarding Automated-Decision Systems (March 15, 2022), available at https://calcivilrights.ca.gov/wp-content/uploads/sites/32/2022/03/AttachB-ModtoEmployRegAutomated-Decision-Systems.pdf.

47 California Code, Business and Professions Code - BPC § 17940 - last updated January 01, 2019 | https://codes.findlaw.com/ca/business-and-professions-code/bpc-sect-17940/.

48 National Conference of State Legislatures, Legislation Related to Artificial Intelligence (Aug. 26, 2022), available at https://www.ncsl.org/technology-and-communication/legislation-related-to-artificial-intelligence.

49 New York City Council, Law 2021/144, Local Law to Amend the Administrative Code of the City of New York, In Relation To Automated Employment Decision Tools (Dec. 11, 2021), available at https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9.

50 Ryan Golden, *NYC Delays Enforcement of AI in Hiring Law to April 2023*, HR Dive (Dec. 14, 2021), available at https://www.hrdive.com/news/nyc-ai-in-hiring-law-delayed-enforcement-april-2023/638793/.

51 Gary E. Marchant & Brad Allenby, *Soft Law: New Tools for Governing Emerging Technologies*, 73 Bull. Atomic Sci. 108, 108 (2017).

certification programs, voluntary programs, and private-public partnerships.[52] A variety of different types of organizations can promulgate soft law, including governmental bodies, industry groups, individual companies, non-governmental organizations, or any combination of the above.[53]

Soft law is the most prominent form of AI governance today, both in the United States and elsewhere. A recent empirical survey by Carlos Ignacio Gutierrez identified and characterized over 600 AI soft law programs that had been adopted by the end of 2019.[54] These soft law programs were extremely diverse, varying in what issues they addressed, the form of the soft law instrument, the type of organization that promulgated them, the geographical origin and reach of the program, and whether they included any implementation or enforcement provisions. One of the most surprising findings was that government entities were the most frequent participant in developing soft law programs, serving in more of a convening or coordination role rather than traditional coercive regulatory role.[55] Another significant finding was that only about one-third (31 percent) of the soft law programs analyzed publicly disclosed any type of implementation or enforcement provisions.[56]

Soft law is currently the dominant form of AI governance, and is likely to continue to be so for some time, but as the empirical study by Gutierrez shows, the AI soft law environment is complex and multi-layered. At the international level, organizations such as the OECD and UNESCO have promulgated principles or codes of ethics for responsible AI, which many organizations in the private and public sector attempt to integrate into their own practices. In addition, international standard setting bodies such as the ISO and IEEE are issuing private standards on responsible AI and AI governance. For example, the IEEE P7000 standards are a set of standards under development addressing various aspects of ethical AI.[57] IEEE is also developing a standard for governance of AI by entities that develop or use AI.[58] NIST is developing a series of documents to assist AI standard-setting bodies, or to assist companies directly in building their own AI governance programs, such as the recently released NIST framework for AI risk management.[59] A large variety of more focused AI soft law instruments have been produced by trade associations, professional societies, think tanks, non-governmental organizations, and individual companies.

In recent years there has been a "techlash" against technology companies as a result of incidents such as the Boeing crashes, Theranos' fraud, and data handling scandals such as Facebook's Cambridge Analytica debacle. This has translated into a backlash against self-regulatory and soft law approaches to technology governance. The lack of implementation and enforcement measures in the majority of AI soft law programs no doubt contributes to this unease. We can learn from the history of soft law for AI and other technologies that accountability and indirect enforcement mechanisms can make soft law more effective and credible, without losing the important benefits of soft law in terms of flexibility, agility and diversity.[60] Since soft law will be essential for the safe and responsible development of beneficial AI, making it successful should be a common goal. To paraphrase Winston Churchill, "[Soft law] is the worst form of govern[ance], except for all the others."

> ***Soft law is the most prominent form of AI governance today, both in the United States and elsewhere***

52 Gary Marchant, Lucille Tournas & Carlos Ignacio Gutierrez, *Governing Emerging Technologies Through Soft Law: Lessons for Artificial Intelligence- An Introduction*, 61 Jurimetrics 1, 5 (2020).

53 Kenneth W. Abbott, Gary E. Marchant & Elizabeth A. Corley, *Soft Law Oversight Mechanisms for Nanotechnology*, 52(3) Jurimetrics, The Journal of Law, Science, and Technology 279, 298-99 (2012).

54 Carlos I. Gutierrez & Gary Marchant, *A Global Perspective of Soft Law Programs for the Governance of Artificial Intelligence*, SSRN (May 28, 2021), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3855171.

55 *Id.* at 13-14.

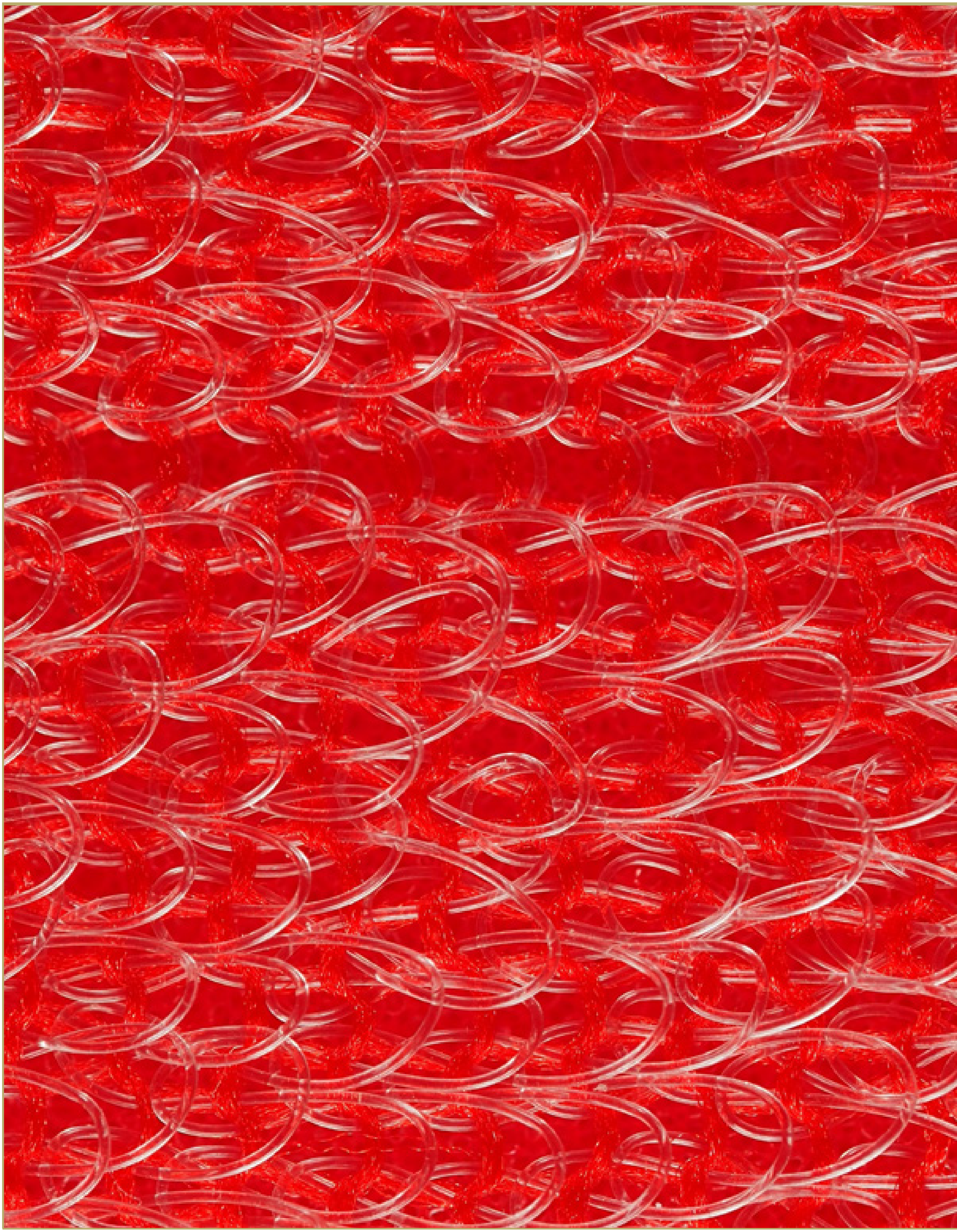
56 Carlos Ignacio Gutierrez, *Transitioning From Ideas to Action: Trends in the Enforcement of Soft Law for the Governance of Artificial Intelligence*, 2 IEEE Transactions on Technology and Society 210, 211 (2021).

57 IEEE P7000 Projects, available at https://ethicsstandards.org/p7000/.

58 IEEE P2863, available at https://sagroups.ieee.org/2863/.

59 NIST, AI Risk Management Framework, available at https://www.nist.gov/itl/ai-risk-management-framework.

60 Gary Marchant, Lucille Tournas & Carlos Ignacio Gutierrez, *Governing Emerging Technologies Through Soft Law: Lessons for Artificial Intelligence- An Introduction*, 61 Jurimetrics 1, 9-16 (2020); Gutierrez, *supra* note 56, at 211-15.



REGULATING MACHINE LEARNING: **THE CHALLENGE OF HETEROGENEITY**

BY
CARY COGLIANESE

Edward B. Shils Professor of Law and Professor of Political Science, and Director, Penn Program on Regulation, University of Pennsylvania.

Machine-learning algorithms increasingly drive technological advances that deliver valuable improvements for society and the economy. But these algorithms also raise important concerns. The way machine-learning algorithms work autonomously to find patterns in large datasets has given rise to fears of a world that will ultimately cede critical aspects of human control to the dictates of artificial intelligence. These fears seem only exacerbated by the intrinsic opacity surrounding how machine-learning algorithms achieve their results. To a greater degree than with other statistical tools, the outcomes generated by machine learning cannot be easily interpreted and explained, which can make it hard for the public to trust the fairness of products or processes powered by these algorithms.

For these reasons, the autonomous and opaque qualities of machine-learning algorithms make these digital tools both distinctive and a matter of public concern. But when it comes to *regulating* machine learning, a different quality of these algorithms matters most of all: their heterogeneity. The Merriam-Webster Dictionary defines "heterogeneity" as "the quality or state of consisting of dissimilar or diverse elements." Machine learning algorithms' heterogeneity will make all the difference in deciding when to regulate them, who should regulate them, and how to design regulations imposed on their development and use.

01

MACHINE LEARNING'S HETEROGENEITY

One of the most important sources of machine learning's heterogeneity derives from the highly diverse uses to which it is put. These uses could hardly vary more widely. Consider just a small sample of ways that different entities use machine-learning algorithms:

· Social media platforms use them to select and highlight content for users;
· Hospital radiology departments use them to detect cancer in patients;
· Credit card companies use them to identify potential fraudulent charges;
· Commercial airlines use them to operate aircraft with auto-piloting systems;
· Online retailers use them to make product recommendations to visitors to their websites; and
· Political campaigns use them in deciding where and how advertise.

Even within the same organizations, different machine-learning algorithms can perform different functions. An automobile manufacturer, for example, might use one type of machine-learning algorithm to automate certain on-road operations of their vehicles, while using other machine-learning algorithms as part of its manufacturing processes or for managing its supply chain and inventory.

In addition to their varied uses, machine-learning algorithms can themselves take many different forms and possess diverse qualities. These algorithms are often grouped into several main categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Within each category, the range of algorithms and their forms can be highly diverse. Naïve Bayesian models, decision trees, random forests, and neural networks are just a few types of supervised learning models.[2] Even within any single type, finer points about how each model generated by an algorithm is structured, not to mention differences in the data used to train it, can lead each application of machine learning almost to fall within a category of its own.

Despite the wide variation in algorithms, it also remains that the same machine-learning model can be put to different uses within a single organization. For example, Meta—the corporation that owns Facebook and Instagram—has noted that, even though its "image classification models are all designed to predict what's in a given image, they may be used differently in an integrity system that flags harmful content versus a recommender system used to show people posts they might be interested in."[3]

Added to the extreme variation in uses and designs of algorithms is the fact that, for many uses, multiple different algorithms are used in combination with each other to support automated systems. What may at times be referred to as "an" algorithm is often actually a suite or family of algorithms, integrated into an automated system or process in a manner designed to perform a specified function. Furthermore, these algorithms and their combinations are updated and changed over time, as new or refined algorithms are shown to do better. Today's ChatGPT, for example, runs on models that are markedly different than earlier language models, and it will only be updated, enhanced, and modified repeatedly in the years to come.

Finally, these changes in machine-learning models come on top of the fact that when the data processed by a learning algorithm changes, then so too can its performance. This means that, for some algorithms, their performance can be constantly evolving as they encounter and process new data.[4]

2 Differences of expert opinion even exist over what counts as machine learning, with some data scientists treating forms of what others see as standard regression analysis as a type of machine learning.

3 MetaAI, *System Cards, A New Resource for Understanding How AI Systems Work* (Feb. 23, 2022), https://ai.facebook.com/blog/system-cards-a-new-resource-for-understanding-how-ai-systems-work/.

4 See, e.g., Jessa Boubker, *When Medical Devices Have a Mind of Their Own: The Challenges of Regulating Artificial Intelligence*, 47 Am. J.L. & Med. 427, 434 (2021) (indicating that, if an algorithm is continuously learning, it "will not always be able to predict how a software is going to react in real-time based on new data").

In short, machine-learning algorithms place the definition of heterogeneity on steroids. These algorithms vary widely across different types and different uses at any given time — and they are highly dynamic, with their performance evolving over time. All this heterogeneity holds crucial implications for whether and how machine-learning algorithms should be regulated.

02 DECIDING TO REGULATE MACHINE LEARNING

The first question to ask, of course, is whether machine learning needs to be regulated at all.[5] Regulation is a tool designed to respond to and help solve social and economic problems. But by themselves, machine-learning algorithms are just mathematical constructs and create no social or economic problems.[6] If they were used only for intellectual pleasure—say, as a hobby pursued by a mathematically inclined subset of the population — then there would surely be no need to consider regulating them. Regulating machine learning becomes a topic of conversation only when it is used in ways that have tangible effects on people.

If machine learning is to be a candidate for regulation, then, it is because of *the uses* for which it gets employed. This is not unlike other physical machines. When other machines have had consequential effects on the public, they have come to be regulated. The National Highway Traffic Safety Administration ("NHTSA"), for example, long ago starting imposing regulatory standards on different parts of an automobile not because of something intrinsic about the parts themselves, but rather because of how they are used in vehicles and how those uses affect the safety of the vehicle. Machine-learning algorithms are much the same. They are or will become objects of regulation because of the systems in which they are situated and how they ultimately affect system outcomes in ways that touch people's lives and livelihoods.

Because machine-learning algorithms can be used in so many different ways, this means that the regulatory problems they can create will vary quite widely as well. Looking across a host of different uses of machine learning, it is possible to say that the potential problems cover the gamut of classic market failures that justify regulation. Machine-learning algorithms used as part of automated pricing systems by online retailers, for example, may contribute to anticompetitive behavior in the marketplace.[7] Machine-learning algorithms used in medical treatments and consumer products can contribute to the kind of information asymmetries that typically justify consumer protection regulation.[8] And any pedestrian put at an increased risk from a self-driving car should easily be able to see another obvious market failure—an externality—created by vehicles that operate autonomously using sensors and machine-learning algorithms.

Regulation is often justified by more than just these classic market failures. It can also be used, for example, as a tool for preventing injustices and protecting civil rights, such as when regulations aim to combat employment discrimination.[9] Grounds exist for regulating machine learning on this basis as well. When society's prevailing biases have been reflected in the design of machine-learning algorithms or in the data on which they are trained, these algorithms can end up reinforcing, if not even exacerbating, existing injustices.[10] Machine learning used as part of an employer's hiring process, for example, can thus create the problems

5 In posing the question in terms of whether to "regulate machine learning," I mean to distinguish it from the question of whether to impose antitrust regulation on the structural or other business decisions of firms that rely heavily on machine learning—namely, the so-called big tech firms. Deciding to impose regulatory scrutiny on mergers and acquisitions in the big tech space is not what I mean here by regulating machine learning. Only if machine-learning tools are themselves directly used to impede competition or concentrate market power would antitrust law become relevant for regulating machine learning in the sense I mean here.

6 This is putting to the side, of course, the fact that processing data using machine-learning algorithms can result in externalities from the production of energy needed to power the necessary computer hardware.

7 Cary Coglianese & Alicia Lai, *Antitrust by Algorithm,* Stan. Computational Antitrust, Vol. 2, no. 1, 2022, at 4.

8 *Cf. id.* at 18 (describing the difficulty in supporting algorithmic forecasts with intuitive explanations, which may run in some tension with consumer protection principles favoring disclosure and transparency).

9 See, e.g., Olatunde C.A. Johnson, Beyond the Private Attorney General: Equality Directives in American Law, 87 N.Y.U. L. Rev. 1339 (2012) (providing an overview of civil rights regulation in the United States).

10 See, e.g., Dorothy Roberts, *Digitizing the Carceral State,* 132 Harv. L. Rev. 1695, 1698 (2019) (reviewing Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (2018)); Sandra G. Mayson*, Bias in, Bias Out*, 128 Yale L.J. 2218 (2019).

that antidiscrimination regulation has been established to solve.[11]

Privacy is another civil rights concern that is often raised in the context of calls for regulation of machine learning. One worry centers on protecting the private information contained in the extensive data on which these algorithms draw — as well as ensuring individual notice of or consent to the use of such information. Still another concern arises from the ability of machine-learning algorithms to make accurate inferences about certain private characteristics that are not contained in the data themselves. Yet another concern centers on how machine-learning algorithms can make possible the use of facial recognition and other tools that can track individuals' whereabouts and contribute to fears of a "surveillance state."[12]

And then there are a host of other public policy concerns surrounding machine-learning algorithms that lie at the heart of many conversations about regulating artificial intelligence.[13] The availability of ChatGPT, for example, has raised new questions about what artificial intelligence means for education.[14] Social media platforms use machine-learning algorithms to push content to users in ways that accentuate conflict, keep users distracted, or make them crave more time on their smart phones.[15] Digital tools driven by machine-learning algorithms can also generate new artwork from existing works, raising questions about ownership rights and rules about appropriation.[16] These tools can be used perniciously too, such as by facilitating new opportunities for fraud through deep fakes.[17] Pernicious actors can also use artificial intelligence to propagate cyberattacks that threaten both digital and physical assets.[18]

As should be evident, the heterogeneous uses for machine-learning algorithms lead to a variety of regulatory concerns. It is surely axiomatic to observe that when the types of regulatory problems vary, regulation itself must vary as well to fit the nature of the problem. At the very least, regulation must be designed in a way that accommodates variation in uses and either targets diverse problems or provides appropriate incentives for regulated entities to find and address those problems.[19]

03

WHO SHOULD REGULATE MACHINE LEARNING?

Before turning to how regulation might be designed to accommodate machine learning's heterogeneity, a prior question arises about what type of institution should regulate machine learning, whenever that regulation is justified.

With respect to other technologies and their regulatory problems, the need for regulation to be adapted to fit different circumstances has led governments to establish different regulatory bodies, each targeting a circumscribed range of problems. The problems created by anticompetitive behavior, after all, are different than those created by industrial pollution, which are in turn different than the problems of unsafe and ineffective consumer products. As a result, antitrust regulatory institutions exist to target anticompetitive behavior; environmental regulatory bodies specialize in reducing pollution; and drug and consumer safety regulators

11 Jeffrey Dastin, *Amazon Scraps Secret Ai Recruiting Tool That Showed Bias Against Women*, Reuters (Oct. 10, 2018, 7:04 PM), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

12 A number of jurisdictions have prohibited law enforcement agencies from using facial recognition tools. *See* Cary Coglianese & Kat Hefter, *From Negative to Positive Algorithm Rights*, 30 Wm. & Mary Bill Rts J. 883, 886 n.15 (2022).

13 *Id*. at 886-893.

14 Kalley Huang, *Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach*, N.Y. Times (Jan. 16, 2023), https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html.

15 Barbara Ortutay & David Klepper, *Facebook Whistleblower Testifies: Five Highlights*, Assoc. Press (Oct. 5, 2021), https://apnews.com/article/facebook-frances-haugen-congress-testimony-af86188337d25b179153b973754b71a4. *See generally* Tim Wu, The Attention Merchants: The Epic Scramble to Get Inside Our Heads (2016).

16 Elizabeth Penava, *AI Art Is in Legal Greyscale*, Regul. Rev. (Jan. 24, 2023), https://www.theregreview.org/2023/01/24/penava-ai-art-is-in-legal-greyscale/.

17 Todd C. Helmus, RAND Corp., Artificial Intelligence, Deepfakes, and Disinformation: A Primer (2022).

18 Blessing Guembe, Ambrose Azeta, Sanjay Misra, Victor Chukwudi Osamor, Luis Fernandez-Sanz & Vera Pospelova, *The Emerging Threat of AI-Driven Cyber Attacks: A Review*, 36 Applied A.I. 1 (2022).

19 For a related discussion, see Cary Coglianese, *Regulating New Tech: Problems, Pathways, and People*, TechReg Chron., Dec. 2021, at 65-73.

aim to protect consumers from unsafe products. A single firm will need to comply with the regulations of several distinct regulators with respect to different facets of its operations and market behavior.

These different, specialized regulatory bodies have the advantage over a general legislature in that they can draw upon the specialized knowledge needed to address the different types of problems, their origins in different industries, and their effects on different subsets of the population. This is not to say that, even within their specializations, regulators do not confront heterogeneity. On the contrary, antitrust regulators are usually tasked with looking across all sectors of the economy for different ways businesses might engage in anticompetitive behavior. Environmental regulators are commonly tasked with regulating a variety of types of pollution, such as to the air, water, and land, and from a myriad of different businesses, large and small. Even regulatory bodies with relatively narrow targets — such as the U.S. Nuclear Regulatory Commission, which targets a single industry for the important but still circumscribed problem of nuclear safety[20] — will face some degree of heterogeneity in the different sources of risks and different scenarios that must be accounted for if regulation is to be effective. Nevertheless, because of the value of specialized expertise, nuclear regulators exist to look at nuclear safety and are not responsible for, say, ensuring the safety and soundness of banks. This is why, as a prescriptive matter, environmental regulators do not also seek to combat anticompetitive market conduct, and antitrust regulators are not responsible for addressing pollution problems.

It may be tempting to conclude that machine-learning algorithms are like nuclear power plants and that they need their own regulator. Recently, U.S. Representative Ted Lieu, for example, has argued that "[w]hat we need is a dedicated agency to regulate A.I."[21] Certainly, machine-learning algorithms do require specialized skills to understand how they work and how they can go awry. Regulating machine-learning algorithms' impact on any segment of society or the economy will require sophisticated knowledge about artificial intelligence. But because the regulatory problems that machine-learning algorithms are associated with can be so varied—and often so closely connected to long-standing regulatory problems that already have dedicated regulatory institutions—it is unrealistic to expect that any single regulator could ever sufficiently regulate all the problematic aspects of machine learning. Regulating algorithmic stock market trading will necessarily require great expertise about financial markets. A similar need for substantive expertise will apply when regulating the effects of machine-learning algorithms on the safety of medical devices, the operation of automobiles, and the pricing behavior of firms. No dedicated AI regulatory agency could possibly possess all of the additional related technical knowledge and capacity needed to regulate algorithms' many uses.

It may be tempting to conclude that machine-learning algorithms are like nuclear power plants and that they need their own regulator

Given the many ways that machine-learning algorithms are intertwined with different problems, many of which are already addressed by existing regulatory bodies, it is not surprising that these existing regulators have so far taken the lead in responding to potential problems related to machine learning. Within the Department of Transportation, for example, NHTSA has issued regulatory guidance for automobile manufacturers on safety assessments for autonomous vehicle technology.[22] It ordered these manufacturers to file reports on crashes involving their autonomous vehicles.[23] NHTSA also recently prodded Tesla to recall more than 350,000 of its vehicles over safety concerns related to its driver assistance software.[24]

Separately, the U.S. Food and Drug Administration (FDA) has developed an action plan for addressing the use of ma-

20 *About NRC*, U.S. Nuclear Regul. Comm'n, https://www.nrc.gov/about-nrc.html (last visited Feb. 4, 2023).

21 Ted Lieu, *I'm a Congressman Who Codes. A.I. Freaks Me Out*., N.Y. Times (Jan. 23, 2023), https://www.nytimes.com/2023/01/23/opinion/ted-lieu-ai-chatgpt-congress.html.

22 U.S. Dep't Transp. Nat'l Highway Traffic Safety Admin., *Federal Automated Vehicles Policy* (Sept. 2016), https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/av_policy_guidance_pdf.pdf.

23 First Amended Standing General Order, U.S. Dep't Transp. Nat'l Highway Traffic Safety Admin., Incident Reporting for Automated Driving Systems (ADS) and Level 2 Advanced Driver Assistance Systems (ADAS), Order No. 2021-01 (August 2021), https://www.nhtsa.gov/sites/nhtsa.gov/files/2021-08/First_Amended_SGO_2021_01_Final.pdf.

24 Neal E. Boudette, *Tesla to Recall 362,000 Cars With Its "Full Self Driving" System*, N.Y. Times (Feb. 16, 2023), https://www.nytimes.com/2023/02/16/business/tesla-recall-full-self-driving.html.

chine learning in medical devices, announcing it will treat them under a separate category for innovative devices.[25] In 2020, FDA approved the first AI-based cardiac ultrasound software under this alternative track.[26]

As existing regulatory bodies go forward to address AI-related problems within their domains, they will certainly need to develop further their data science expertise. It is not inconceivable that they could benefit from a centralized expert body that can provide guidance and support. Already, the National Institute of Standards and Technology (NIST) within the U.S. Department of Commerce has issued a generalized risk management framework for artificial intelligence that could be of value if customized to fit the needs of other more specialized regulatory settings.[27] NIST's framework joins other similar documents issued by other federal entities — such as the U.S. Government Accountability Office,[28] the White House Office of Science and Technology,[29] and the Administrative Conference of the United States[30] — that articulate general principles to follow when using machine-learning tools. The federal government has also established an AI Center of Excellence within the General Services Administration.[31]

Nevertheless, as helpful as these general, cross-cutting initiatives may be, existing regulators still need to build up their own capacity to understand and regulate AI tools, given how intertwined they can be with so many longstanding regulatory problems. Admittedly, even with sufficient capacity within existing agencies, some kinds of new problems will fall through the cracks. Ill effects from social media platforms' use of algorithms, for example, have so far have elided serious governmental oversight. Nevertheless, rather than hoping that a new omnibus AI regulatory body can swoop in to save the day by regulating all uses of machine learning, policymakers would do well to look instead to empower existing centers of regulatory expertise. Where gaps or overlaps exist in current regulatory authority, policymakers can then work to fill those gaps or work out any conflicting jurisdictions. Gaps could be filled either by creating new regulatory bodies focused on unattended problems or by assigning those new problems to existing regulators with relevant expertise.

04

HOW TO REGULATE MACHINE LEARNING

No matter which institutions take responsibility for regulating machine learning, they will still confront heterogeneity. Even within a specified industry and even with respect to some identical uses of machine learning, heterogeneity will remain because both the algorithms themselves and the data they use vary so widely. Moreover, the algorithms and the automated systems of which they are a part are changing over time. As a result, even within specialized domains, regulators will need to pursue measures that take into account the varied and dynamic nature of these algorithms.

For this reason, it is impossible to specify a tidy, one-size-fits-all formula for how regulators should approach their task of regulating machine learning. But at a broad brush, it is possible to say that regulators will need to approach their work with agility, flexibility, and vigilance.

1. Regulate with agility. Regulators will need to be active and adaptive. Regulation of machine learning cannot be approached as a matter of finding the "right" rule and then

25 U.S. Food & Drug Admin., *Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan* (Sept. 22, 2021), https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices; U.S. Food & Drug Admin., *Clinical Decision Support Software Guidance for Industry and Food and Drug Administration Staff* (Sept. 28, 2022), https://www.fda.gov/media/109618/download.

26 Press Release, U.S. Food & Drug Admin., *FDA Authorizes Marketing of First Cardiac Ultrasound Software That Uses Artificial Intelligence to Guide User* (Feb. 7, 2020), https://www.fda.gov/news-events/press-announcements/fda-authorizes-marketing-first-cardiac-ultrasound-software-uses-artificial-intelligence-guide-user.

27 Nat'l Inst. of Standards & Tech. (NIST), Artificial Intelligence Risk Management Framework (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

28 U.S. Gov't Accountability Off., GAO-21-519SP, Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities (June 2021), https://www.gao.gov/assets/gao-21-519sp.pdf.

29 White House Off. Of Sci. & Tech. Pol'y, Blueprint For An Ai Bill Of Rights: Making Automated Systems Work For The American People, https://www.whitehouse.gov/ostp/ai-bill-of-rights.

30 Admin. Conf. of the U.S., Administrative Conference Statement #20: Agency Use of Artificial Intelligence, 86 Fed. Reg. 6616, 6616 n.1 (Jan. 22, 2021).

31 Gen. Servs. Admin., Accelerate Adoption of Artificial Intelligence to Discover Insights at Machine Speed, https://coe.gsa.gov/docs/2020/AIServiceCatalog.pdf.

moving on simply to enforcing that rule. Instead, regulators need to think of their work as incremental and constantly provisional. When the world that regulators seek to regulate keeps changing, the last thing regulators can do is remain static.

To regulate machine learning with agility, regulators need to build up their capacity to keep pace with changes in industry.[32] This requires building up a regulator's internal technological infrastructure and human capital with expertise in data sciences. It also means finding ways to engage with and gather information from industry.[33] Industry, after all, will be best-positioned to know the most about their algorithms and how they are used. Regulators cannot avoid active engagement with industry if they are to adopt smart approaches to regulation.

> ***No matter which institutions take responsibility for regulating machine learning, they will still confront heterogeneity.***

Of course, in seeking to engage with industry, regulators should never lose sight of their distinctive role as protectors of public value. To be sure, the public does gain from technological innovation in the private sector and regulation that unduly impedes innovation should be avoided. But regulators also should avoid embracing a perspective that values innovation for its own sake. They should not take their eyes off of the risks and other regulatory problems that innovations might bring.[34] Private firms will see some of these problems too, but if regulation is needed, that is because the firms lack the socially optimal incentives to ferret out and redress these problems, especially when the solutions are costly.

2. Deploy flexible rules. Machine learning's heterogeneity will make flexible rules strong candidates for adoption. A one-size-fits-all "prescriptive" or "specification" standard will not make sense, as that would necessitate the regulator telling firms exactly how to design, train, and use their algorithms. Regulators will almost surely never have sufficient capacity to regulate with such specificity.

An obvious alternative would be for the regulator to adopt performance standards that specify outcomes to be achieved (or avoided) but then give regulated firms the flexibility to decide how to proceed as long as they meet (or avoid) the outcome in the regulatory standard.[35] As appealing as performance standards may be, they necessitate that the regulator will be able to specify the desired outcome in a clear, monitorable fashion—and then have the capacity to do the actual monitoring.[36] Sometimes that might be the case, such as when machine learning is embedded in a larger system that can be observed independently and subjected to sufficient testing and monitoring. But in many cases it will be unlikely that regulators can develop sufficiently clear, monitorable performance tests for algorithms themselves.

When standard-setting organizations around the world have adopted voluntary performance guidelines for algorithms, they have tended to do so by articulating general performance *principles* calling for algorithms to yield outcomes that are "fair," "safe," "explainable," and so forth.[37] Although these principles-based approaches may be helpful in offering general guidance to industry, they are far from operational. It remains to be seen whether and how regulators could articulate with greater precision outcome values such as fairness and explainability.[38] Even with safety, one must surely ask: Exactly how safe is safe enough? Absent an ability to specify outcome values in measurable and

32 Cary Coglianese, *Optimizing Regulation for an Optimizing Economy*, 4 U. Pa. J.L. & Pub. Affs. 1, 2 (2018).

33 Cary Coglianese, Richard Zeckhauser & Edward Parson, *Seeking Truth for Power: Informational Strategy and Regulatory Policy Making*, 89 Minn. L. Rev. 277, 278-79 (2004).

34 Cary Coglianese, *Regulatory Vigilance in a Changing World*, Regul. Rev. (Feb. 25, 2019), https://www.theregreview.org/2019/02/25/coglianese-innovation-regulatory-vigilance/.

35 Cary Coglianese, Jennifer Nash & Todd Olmstead, *Performance-Based Regulation: Prospects and Limitations in Health, Safety, and Environmental Regulation*, 55 Admin. L. Rev. 705 (2003).

36 Cary Coglianese, *The Limits of Performance-Based Regulation*, 50 U. Mich. J.L. Reform 525 (2017).

37 Gary E. Marchant, Lucille Tournas & Carlos Ignacio Gutierrez, *Governing Emerging Technologies Through Soft Law: Lessons for Artificial Intelligence*, 61 Jurimetrics J. 1, 5-6 (Fall 2020).

38 For a discussion of principles-based regulation in other contexts, see Julia Black, *Forms and Paradoxes of Principles-Based Regulation*, 3 Cap. Markets L.J. 425 (2008); Cristie L. Ford, *New Governance, Compliance, and Principles-Based Securities Regulation*, 45 Am. Bus. L.J. 1 (2008). For treatment in the context of artificial intelligence, see Julia Black & Andrew Murray, *Regulating AI and Machine Learning: Setting the Regulatory Agenda*, 10 Eur. J. L. & Tech. 1 (2019).

monitorable terms, it is hard to see how regulators could rely on a performance-based approach to the regulation of machine learning.

In situations where neither a one-size-fits-all prescriptive rule nor a performance-based rule seem likely to work, regulators have turned to an alternative regulatory strategy called *management-based* regulation.[39] Under a management-based approach, the regulator requires the firm to engage in systemic managerial activities that seek to identify problems and then create internal responses to correct them. This approach has been widely applied to address other regulatory problems where heterogeneity dominates, such as food safety and chemical facility security. In these situations, the sources of the underlying regulatory problem are highly diverse and dynamic. The management-based approach typically calls for a regulated entity to develop a management plan, monitor for potential risks, produce internal procedures and trainings to address those risks, and maintain documentation on the operation of the firm's management system. Sometimes these regulations also require firms to subject their management systems to third-party auditing and certification.

> ***"In situations where neither a one-size-fits-all prescriptive rule nor a performance-based rule seem likely to work, regulators have turned to an alternative regulatory strategy called management-based regulation***

Management-based regulation will be an obvious option to consider for machine learning. This regulatory option does not demand that the regulator have the same level of knowledge as regulated firms themselves, nor does it require that the regulator be able to specify and measure all the relevant outcomes. It also gives firms considerable flexibility and thereby accommodates heterogeneity across firms and over time.

Unsurprisingly, many emerging soft law standards for machine learning are taking a management-based approach. The voluntary framework that NIST recently issued to improve the trustworthiness of machine-learning applications, for example, bears all the hallmarks of a management-based approach. Specifically, it calls for firms to develop "structures, systems, processes, and teams" for "[a]nticipating, assessing, and otherwise addressing potential sources of negative risks" and to put in place "rigorous software testing and performance assessment methodologies," "[s]ystematic documentation practices," and "plans for prioritizing risk and regular monitoring and improvement."[40]

Although the NIST framework is not mandatory, similar approaches are starting to emerge in regulations or proposed regulations in various parts of the world. Canada, for example, has imposed a requirement that its own federal government agencies conduct algorithmic impact assessments, quality assurance auditing, and various documentation measures before launching algorithmic systems that substitute for human decision-makers.[41] A proposed European Union regulation would impose similar impact assessment and auditing requirements on both public and private sector machine-learning systems.[42] These auditing and impact assessment requirements are management-based. They do not impose any specific prescriptions for the design and use of algorithms nor what outcomes they achieve — but they do direct firms to undertake a series of risk management steps.

In other contexts, management-based regulations have sometimes required firms to disclose publicly their plans and audit results. Mandatory disclosure is another likely option for the future regulation of machine-learning algorithms. Already, big-tech firms are starting to develop their own semi-standardized means of disclosing information about their uses of machine learning as well as the basic properties of the algorithms and the data on which they are

39 Cary Coglianese & David Lazer, *Management-Based Regulation: Prescribing Private Management to Achieve Public Goals*, 37 L. & Soc. Rev. 691 (2003); Cary Coglianese, *Management-Based Regulation: Implications for Public Policy*, *in* Risk and Regulatory Policy: Improving the Governance of Risk (Gregory Bounds & Nikolai Malyshev, eds., 2010); Cary Coglianese & Shana Starobin, *Management-Based Regulation*, *in* Policy Instruments in Environmental Law 292 (Kenneth R. Richards & Josephine van Zeben, eds., 2020).

40 NIST, *supra* note 28.

41 Government of Canada, Directive on Automated Decision-Making (2021), https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592.

42 European Commission, Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (2021), https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

trained and deployed.[43] These voluntary disclosure efforts — what are currently known as "model cards" — could provide a template in the future for mandatory disclosure of information about machine-learning algorithms.[44] For the same reasons that performance-based standards are unlikely to prove viable as a regulatory strategy, it is unlikely that any disclosure regulation could demand a unified outcome metric to be applied to all algorithms and all use cases.[45] But any firm that has an internal management process supportive of the responsible use of artificial intelligence will necessarily generate some common types of information that could be disclosed.[46] The disclosure of information from firms' management of their algorithms would go some distance toward addressing concerns about machine learning's opacity as well as providing consumers and the public better assurance that firms are testing, validating, and deploying machine learning in a responsible manner.[47]

3. Remain vigilant. Research in other regulatory domains shows that management-based regulation can lead firms to reduce risks.[48] But as much as management-based regulation has been demonstrated to work in other contexts and is conceptually well-suited for regulating machine learning, it is hardly a panacea. The evidence for the long-term efficacy of this strategy remains less clear and worries exist that managerial rigor and steadfastness by firms can atrophy over time. The possibility exists that, even if firms subjected to AI impact assessment and auditing requirements take their required risk management responsibilities seriously at first, these management-based requirements can become rote paperwork exercises over time.[49] It is crucial that regulators build the capacity to assess the quality of firms' management efforts and that regulators sustain rigor in their oversight of their management-based regulatory regime.

Vigilance is also needed simply because of the rapid pace of change. Machine learning's future is a dynamic one and regulators need to equip themselves to make smart decisions in a changing environment. This means regulators must remain engaged with the industry they are overseeing and continue learning constantly. Regulators will make mistakes—they always have. But the key will be to try to minimize the consequences of those mistakes and, most of all, to learn from failures. Responsible regulation, like the responsible use of AI, requires vision, attentiveness, and the capacity to learn and adapt. If regulation of machine learning is to succeed, it must be viewed as an ongoing pursuit of continuous improvement.

05

REGULATING MACHINE LEARNING WITH MACHINE LEARNING?

A final aspect of the regulation of machine learning should not be overlooked: using machine learning to regulate machine learning. Algorithms, after all, are not merely tools for private

43 Vasi Philomin & Peter Hallinan, *Introducing AWS AI Service Cards: A New Resource to Enhance Transparency and Advance Responsible AI* (Nov. 30, 2022), https://aws.amazon.com/blogs/machine-learning/introducing-aws-ai-service-cards-a-new-resource-to-enhance-transparency-and-advance-responsible-ai/; *The Value of a Shared Understanding of AI Models*, Google Cloud, https://modelcards.withgoogle.com/about (last visited Feb. 16, 2023); Meta AI, System Cards, A New Resource for Understanding How AI Systems Work (Feb. 23, 2022), https://ai.facebook.com/blog/system-cards-a-new-resource-for-understanding-how-ai-systems-work/.

44 Margaret Mitchell et al., *Model Cards for Model Reporting* 221 (Jan. 14, 2019) (paper prepared for FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency), https://dl.acm.org/doi/10.1145/3287560.3287596 ("Model cards provide a way to inform users about what machine learning systems can and cannot do, the types of errors they make, and additional steps that could create more fair and inclusive outcomes with the technology.").

45 *See supra* note 37 and accompanying text. Model cards, on the other hand, "are designed to be flexible in both scope and specificity in order to accommodate the wide variety of machine learning model types and potential use cases." *Id*. at 228.

46 *Cf*. *Service Cards and ML Governance with Michael Kearns* (January 2, 2023), https://twimlai.com/podcast/twimlai/service-cards-and-ml-governance/ (discussing the quantitative technical assessments and extensive internal reviews that underlie AWS service cards and noting that "a lot of work went into these cards").

47 *Cf*. Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 Admin. L. Rev. 1, 49-55 (2019) (discussing emerging technical advances that can enhance machine learning's transparency).

48 See Lori S. Bennear, *Are Management-based Regulations Effective? Evidence from State Pollution Prevention Programs,* 26 J. Pol'y Analysis & Mgmt. 327 (2007); Travis Minor & Matt Parrett, *The Economic Impact of the Food and Drug Administration's Final Juice HACCP Rule*, 68 Food Pol'y 206 (2017).

49 See, e.g., Cary Coglianese & Jennifer Nash, *Compliance Management Systems: Do They Make a Difference?*, *in* Cambridge Handbook of Compliance 571 (D. Daniel Sokol & Benjamin van Rooij, eds., 2021); Garry C. Gray & Susan S. Silbey, *Governing Inside the Organization: Interpreting Regulation and Compliance*, 96 Amer. J. Soc. 120 (2014).

sector firms seeking to innovate and enhance value. Regulators can also look to machine-learning algorithms as tools for improving their own performance.[50] At present, some regulators use them to identify firms that are likely in violation of applicable rules. Rather than sending out auditors or inspectors at random, and thereby using limited oversight resources to monitor firms that will be in compliance, regulators can vastly improve the detection of violators by using machine learning to decide how to target their limited resources.[51]

This same approach could be used by regulators when allocating limited resources to oversee firms' compliance with machine-learning regulation. With so many different uses for machine learning, and the prospect of vast numbers of firms using this digital technology, regulators will have to be smart about how to allocate their oversight resources. This may include using natural language processing algorithms to identify firms with inadequate risk management plans. It may include using algorithms to select firms for regulatory auditing that are most likely to be treating required management-based planning in a pro forma fashion. The kind of vigilance that regulators will need to maintain will require that regulators themselves use the most sophisticated tools in their arsenals.

> ***This same approach could be used by regulators when allocating limited resources to oversee firms' compliance with machine-learning regulation***

The time may also come when regulators develop automated regulatory tools that match the speed and heterogeneity of private sector machine learning with the speed and heterogeneity of regulatory machine learning. When businesses rely on machine-learning tools to make subtle but anticompetitive pricing decisions in real time, for example, antitrust regulators might do well to use machine-learning tools to detect these collusive pricing patterns at the same speed.[52] When high-speed algorithms facilitate ever-so-slight but profitable forms of stock market manipulation, securities regulators would likely do well to use similarly sophisticated algorithms to discover that manipulation.[53] Over time, regulators' own algorithms might even be used as part of larger automated systems that can detect and algorithmically punish at the same time.

Perhaps the idea of regulatory robots seems a bit fanciful, but it is already becoming a reality, even if in seemingly banal ways. Automated regulatory systems already are already being used in one of the most familiar venues of daily life: the roadway. Several cities around the United States have installed automated rule-makers and rule-enforcers on their streets and highways to optimize traffic flow.[54] These digital traffic light systems rely on sensors and machine-learning algorithms to determine when signals turn red and green. Other jurisdictions have installed automated systems on highways that can detect vehicles traveling at excessive speeds and then send tickets to the vehicles' owners.[55]

It is not hard to imagine a future in which machine-learning systems that operate self-driving cars are integrated into automated systems of traffic control and management, making the regulation of the nation's roadways run entirely on machine learning. Nor is it difficult to envision a world in which many other activities and business practices are regulated by automated systems driven by machine-learning algorithms.[56]

Admittedly, the regulatory tasks involved in detecting vehicle speed and changing traffic lights may seem simple compared with the tasks regulators face in overseeing all the myriad uses of machine learning. And technology will not erase the regulatory challenges created by machine learning's heterogeneity. But the existence of even crude automated regulatory systems today on the nation's roadways offers a vision of a future in which at least some private sector uses of machine-learning algorithms will be overseen by regulatory systems driven themselves by machine-learning algorithms.[57]

50 Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision-Making in the Machine Learning Era*, 105 Geo. L. J. 1147 (2017); Cary Coglianese, *Algorithmic Regulation: Machine Learning as Governance Tool*, *in* The Algorithmic Society: Power, Knowledge and Technology in the Age of Algorithms 35 (Marc Schuilenburg & Rik Peeters, eds., 2021).

51 Cary Coglianese & Alicia Lai, *Algorithm vs. Algorithm*, 72 Duke L.J. 1281, 1311 (2021).

52 Coglianese & Lai, *supra* note 8.

53 Coglianese & Lehr, *supra* note 52.

54 Cary Coglianese & Lavi M. Ben Dor, *AI in Adjudication and Administration*, 86 Brook. L. Rev. 791, 824-25 (2021).

55 Coglianese & Hefter, *supra* note 13.

56 Coglianese & Lehr, *supra* note 52; Cary Coglianese & Alicia Lai, *Assessing Automated Administration*, *in* Oxford Handbook of AI Governance (Justin Bullock et al., eds., forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4084844.

57 Cary Coglianese, *Moving Toward Personalized Law*, U. Chi. L. Rev. Online (2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4051776.

06

MEETING THE CHALLENGE OF HETEROGENEITY

Regulating machine-learning algorithms sensibly will not be easy. Their complexity, self-learning autonomy, and opacity create reasons for, as well as challenges to, sound regulation. But it is machine learning's *heterogeneity* that poses regulators' greatest challenge of all. These algorithms' varied forms, multiple uses, and dynamic properties make most conventional regulatory strategies obsolete. The tradition of a regulatory body that establishes and then enforces rigid, general commands will not fit well in a world of rapidly evolving, highly varied digital tools.

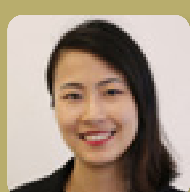
> *Regulating machine learning well must draw upon the expertise of multiple regulatory institutions that can target machine learning's multiple uses*

Regulating machine learning well must draw upon the expertise of multiple regulatory institutions that can target machine learning's multiple uses. These specialized regulators will need to deploy flexible regulatory instruments, such as management-based regulation, and use smart oversight strategies, such as by using algorithmic tools for prioritizing resources.

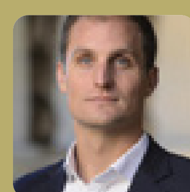
In the end, effective governance in a world driven by heterogeneous algorithmic machines will depend on sophisticated decision-making and top-level performance by human institutions tasked with regulatory oversight. Regulating machine learning well will demand the utmost levels of vigilance and excellence by regulatory officials as they practice their craft.[58]

58 Achieving Regulatory Excellence (Cary Coglianese, ed., 2017); Malcolm K. Sparrow, The Regulatory Craft: Controlling Risks, Solving Problems & Managing Compliance (2000).

EMERGING AI REGULATORY ECOSYSTEMS: **IMPLICATIONS FOR BUSINESSES AND REGULATORS**



BY
DANNI YU



&
BENJAMIN CEDRIC LARSEN

Project Fellow, Artificial Intelligence and Machine Learning, World Economic Forum, Consultant, Boston Consulting Group, and AI Lead, Centre for the Fourth Industrial Revolution, World Economic Forum, respectively.

Different AI regulatory regimes are currently emerging across Europe, the United States, China, and elsewhere. But what do these new regulatory regimes mean for companies and their adoption of self-regulatory and compliance-based tools and practices? This article outlines how and where AI regulations emerge and how these, in some cases, seem to be on divergent paths. Second, it discusses what this means for businesses and their global operations. Third, it comments on a way forward in the growing complexities of AI use and regulation, as it exists between soft law practices and emerging hard law measures.

01

AI GOVERNANCE CONCEPTUALIZED

Two distinct but connected forms of AI governance are currently emerging. One is soft law governance, which functions as self-regulation based on non-legislative policy instruments. This group includes private sector firms issuing principles, guidelines, and internal audits and assessment frameworks for developing ethical AI. Actionable mechanisms by the private sector usually focus on developing concrete technical solutions, including the development of internal audits, standards, or explicit normative encoding.[2] Soft law governance also entails multi-stakeholder organizations such as The Partnership on AI, international organizations such as the World Economic Forum, standard-setting bodies such as the ISO/IEC,[3] CEN/CENELEC,[4] NIST,[5] and interest organizations such as the Association for Computing Machinery ("ACM"), among others. This means that soft-law governance and associated mechanisms are essential in setting the default for how AI technologies are governed.

Hard law measures, on the other hand, entail laws and legally binding regulations that define permitted or prohibited conduct. Regulatory approaches generally refer to legal compliance, the issuing of standards-related certificates, or the creation or adaptation of laws and regulations that target AI systems.[6] Policymakers are currently contemplating several approaches to regulating AI, which broadly can be categorized across AI-specific regulations (e.g. EU AI Act), data-related regulations (e.g. GDPR, CCPA, COPPA), existing laws and legislation (e.g. antitrust and anti-discrimination law), and domain or sector-specific regulations (e.g. HIPAA and SR 11-7).

02

EMERGING REGULATORY LANDSCAPES

According to the OECD AI Policy Observatory, which tracks 69 countries and territories, these have already released more than 200 initiatives targeting AI governance and regulation. Initiatives are aimed at different areas such as antitrust concerns, interoperability standards, risk mitigation -hereunder consumer and social protection, the delivery of public services, and the protection of public values.[7]

While many countries have implemented national AI strategies, not all countries and territories take the same approach to AI governance and regulation. Different approaches are connected to a country's existing institutions, including culture and value systems and economic considerations, e.g. regarding innovation. Before understanding what this means for businesses and their international operations, a few examples of emerging AI regulations are highlighted below.

In many ways, **the European Union ("EU")** has been a frontrunner in data and AI regulation. The EU's AI Act ("AIA"),[8] which is expected to gradually go into effect starting in 2024, establishes a horizontal set of rules for developing and using AI-driven products, services, and systems within the EU. The Act is modeled on a risk-based approach where AI systems that pose unacceptable risks are entirely banned, while high-risk systems will be subject to conformity assessments, including independent audits and new forms of oversight and control.[9] Limited risk systems are subject to transparency obligations, and little or no risk systems remain unaffected by the EU AI Act. The EU has also

---

2 AI Ethics Impact Group. (2020). From Principles to Practice - An interdisciplinary framework to operationalise AI ethics. VDE Association for Electrical Electronic & Information Technologies e.V., Bertelsmann Stiftung, 1–56. https://doi.org/10.11586/2020013.

3 "ISO/IEC JTC 1/SC 42 - Artificial Intelligence." Accessed January 25, 2023. https://www.iso.org/committee/6794475.html.

4 "CEN and CENELEC Launched a New Joint TC on Artificial Intelligence." CEN-CENELEC. March 03, 2021. https://www.cencenelec.eu/news-and-events/news/2021/briefnews/2021-03-03-new-joint-tc-on-artificial-intelligence.

5 "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." 2023. https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.pdf.

6 Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2.

7 OECD.AI (2021), powered by EC/OECD (2021), database of national AI policies, accessed on 4/01/2023. https://oecd.ai/en/dashboards/policy-instruments/Emerging_technology_regulation.

8 EUR-lex Access to European Union law, accessed on 4/01/2023. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex-%3A52021PC0206.

9 European Commission. "Regulatory framework proposal on artificial intelligence." https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

proposed an AI Liability Directive, which targets harmonization of national liability rules for AI.[10]

In the **United Kingdom**, the government released a proposal for regulating the use of AI technologies in June 2022, which focuses on a "light touch" sectoral approach where guidance, voluntary measures, and sandbox environments are encouraged as a means to assess and test AI technologies before they are marketed. The proposal is meant to reflect a less centralized approach than the EU AI Act.[11]

In **Canada**, the Directive on Automated Decision-Making came into effect in April 2019 to ensure that the government's use of AI to make administrative decisions is compatible with core administrative values.[12] Canada's Artificial Intelligence and Data Act ("AIDA") was introduced in June of 2022 and would be the first law in the country to regulate the use of AI systems if approved. The objective of AIDA is to establish common requirements across Canada for the design, development, and deployment of artificial intelligence technologies that are consistent with national values and international standards.[13]

The United States' approach to artificial intelligence is more fragmented and characterized by the idea that companies, in general, must remain in control of industrial development and governance-related criteria.[14] In terms of AI regulation, the U.S. Algorithmic Accountability Act,[15] a horizontal AI regulation, was reintroduced in 2022. Should the Act be passed, it would require companies that develop, sell, and use automated systems to be subject to new rules on when and how AI systems are used.[16] It would require organizations to perform impact assessments of automated decision-making systems ("ADS") before deployment and augmented decision-making processes after deployment. This approach mirrors the conformity assessments and post-market monitoring plans mandated by the EU AI Act. In the absence of national legislation, some states and cities have started implementing their own regulations, such as The California Consumer Privacy Act ("CCPA") and New York City's Law on Automated Employment Decision Tools (Local Law 144). Local Law 144 stipulates that any automated hiring system used on or after January 1, 2023, in NYC must undergo a bias audit consisting of an impartial evaluation by an independent auditor, including testing to assess the potential disparate impact on some groups.[17]

> *In many ways, the European Union ("EU") has been a frontrunner in data and AI regulation*

China's approach to AI legislation is evolving rapidly and is heavily based on central government guidance.[18] China, for example, oversees recommender engines through the "Internet Information Service Algorithmic Recommendation Management Provisions,"[19] which went into effect in

10 European Commission, 28 September 2022, Brussels. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_5807.

11 Zhang, Cynthia O'Donoghue, Sarah O'Brien & Yunzhe. "UK Government Announces Its Proposals for Regulating AI." Technology Law Dispatch. September 2, 2022. https://www.technologylawdispatch.com/2022/09/privacy-data-protection/uk-government-announces-its-proposals-for-regulating-ai/#:~:text=On%2018%20July%202022%2C%20the.

12 Government of Canada. Directive on Automated Decision-Making. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592.

13 "Government of Canada's Artificial Intelligence and Data Act: Brief Overview." 2022. https://www.osler.com/en/resources/regulations/2022/government-of-canada-s-artificial-intelligence-and-data-act-brief-overview.

14 Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M. & Floridi, L. "Artificial Intelligence and the 'Good Society': the US, EU, and UK approach." Science and Engineering Ethics 24, no. 2: 505–528. https://doi.org/10.1007/s11948-017-9901-7.

15 https://www.congress.gov/bill/117th-congress/house-bill/6580/text#:~:text=To%20direct%20the%20Federal%20Trade,Algorithmic%20Accountability%20Act%20of%202022%E2%80%9D.

16 Vought, R. "Guidance for Regulation of Artificial Intelligence Applications Introduction." Executive Office of the President, Office Of Management and Budget. November 17, 2020. https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf.

17 Crowell. New York City Issues Proposed Regulations on Law Governing Automated Employment Decision Tools. October 14, 2022. https://www.crowell.com/NewsEvents/AlertsNewsletters/all/New-York-City-Issues-Proposed-Regulations-on-Law-Governing-Automated-Employment-Decision-Tools#:~:text=October%2014%2C%202022&text=Local%20Law%20144%2C%20which%20is,the%20use%20of%20such%20tool.

18 Larsen, B. C. (2022). Governing Artificial Intelligence: Lessons from the United States and China. Copenhagen Business School [Phd]. PhD Series No. 29.2022. https://research.cbs.dk/en/publications/governing-artificial-intelligence-lessons-from-the-united-states-.

19 Rogier C, Graham W. & Helen T. "Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022." DigiChina. Stanford University, January 10, 2022. https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/.

March 2022, the first regulation of its kind worldwide. The law gives users new rights, including the ability to opt-out of using recommendation algorithms and delete user data. The regulation goes further, however, with its content moderation provisions, which require private companies to actively promote "positive" information that follows the official line of the Communist Party.[20] Regarding generative AI, the Cyberspace Administration of China implemented regulations on AI-generated image, audio, and text-generation software, so-called synthetic media, on January 10, 2023, also marking the first regulation of its kind globally.[21]

In **Singapore**, A.I. Verify[22] was introduced in May of 2022 as the world's first AI Governance Testing Framework and Toolkit for companies who want to demonstrate responsible AI ("RAI") in an objective and verifiable manner. The toolkit, which remains voluntary, provides a governance testing framework that verifies the performance of an AI system against the developer's claims - with respect to internationally accepted AI ethics principles.[23]

Many other countries currently devise AI-related regulations. The Philippines, for example, enacted regulations on spreading false news in 2021.[24] In Brazil**,** a December 2022 proposal outlines a risk-based approach to AI regulation which includes specifying new rights for individuals affected by AI systems.[25] In India, the Ministry of Electronics and Information Technology ("MeitY") is considering Niti Aayog's proposed Responsible #AIForAll10 to be incorporated into India's AI mission,[26] and MeitY has also proposed new privacy legislation, the Digital Personal Data Protection Act, 2022.[27]

While there are too many national AI regulations to recite here, it serves the point that these are materializing across a variety of countries and contexts. It is likely that governments' disparate approaches to AI application and regulation could have varying consequences for businesses in terms of the perceived costs of compliance, which could result in diverging organizational practices.

03

BUSINESSES TAKE THE LEAD ON SELF-GOVERNANCE

As the regulatory landscape slowly evolves, companies increasingly take the lead on self-governance to ensure their development and use of AI systems comply with incoming regulations across regions of operation.

Early adopters of AI-related self-governance come from various sectors such as technology, media, and telecom ("TMT"), financial services, healthcare, and consumer goods. As AI is widely used in these sectors, some companies have adopted a global best practices approach to AI governance.

The first step in this approach relies on creating a list of principles that demonstrate the business' commitment to responsible AI. These principles are usually created by

20 Huld, A. China's Sweeping Recommendation Algorithm Regulations in Effect from March 1. China Briefing. January 6, 2022. https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations-effect-march-1-2022/.

21 Hao, Karen. n.d. "China, a Pioneer in Regulating Algorithms, Turns Its Focus to Deepfakes." The Wall Street Journal. January 8, 2023. https://www.wsj.com/articles/china-a-pioneer-in-regulating-algorithms-turns-its-focus-to-deepfakes-11673149283.

22 "Singapore's A.I.Verify Builds Trust through Transparency." OECD.ai. Accessed January 25, 2023. https://oecd.ai/en/wonk/singapore-ai-verify.

23 "Singapore Launches World's First AI Testing Framework and Toolkit to Promote Transparency; Invites Companies to Pilot and Contribute to International Standards Development." Infocomm Media Development Authority. Accessed January 25, 2023. https://www.imda.gov.sg/content-and-news/press-releases-and-speeches/press-releases/2022/singapore-launches-worlds-first-ai-testing-framework-and-toolkit-to-promote-transparency-invites-companies-to-pilot-and-contribute-to-international-standards-development.

24 Seventeenth Congress of the Republic of the Philippines. June 17, 2021. / http://legacy.senate.gov.ph/lisdata/2624822593!.pdf.

25 Iapp. (2022) Brazil's AI commission to deliver final report. December 2, 2022. https://iapp.org/news/a/brazils-ai-commission-to-deliver-final-report/.

26 https://www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdf.

27 Iapp (2022) India's Digital Personal Data Protection Bill 2022: Does it overhaul the former PDPB? https://iapp.org/news/a/indias-digital-personal-data-protection-bill-2022-does-it-overhaul-the-former-pdpb/.

the company's senior leadership and are aligned with the company's core values and culture. Microsoft,[28] Google,[29] Amazon,[30] Meta,[31] HSBC,[32] AstraZeneca,[33] Novartis,[34] and H&M,[35] among others, have also publicly shared their responsible AI principles. Fairness, transparency, privacy, explainability, safety, controllability, and human-centeredness are among the most common themes and are generally in line with the OECD's AI Principles.[36]

> *The first step in this approach relies on creating a list of principles that demonstrate the business' commitment to responsible AI*

While AI principles is a good starting point, successful implementation rests on developing a cross-organization AI governance structure. One common approach is to have decision-making and oversight responsibilities at a centralized level, for example, in a hub or Center of Excellence ("CoE"). In this model, a board of senior business and functional leaders are responsible for decisions on AI, including for creating and enacting associated governance mechanisms. To operationalize AI governance, the hub or CoE usually assembles a group of technical and subject matter experts tasked with increasing awareness and literacy, e.g. on sensitive use cases, while developing processes, tools, and best practices linked to responsible AI.

An example of this structure can be found in Microsoft. The Microsoft senior leadership team is the final decision maker accountable for the company's direction on responsible AI and steers the company's commitments to AI principles, values, and human rights. A committee called AETHER, made up of expert working groups, provides advice to the senior leadership and practitioners on questions, challenges, and opportunities linked to the development and use of AI.[37] Their decisions are subsequently enacted by the Office of Responsible AI, which serves as a hub working with stakeholders across the company to define governance mechanisms and establish new best practices.[38]

While the above structure is effective for AI governance in some companies, it is not a one-size-fits-all solution. Companies must choose their AI governance model based on their culture, organizational structure, and existing governance model. For example, a company with highly autonomous business units may decentralize decision-making for individual use cases while creating a Center of Excellence to provide expertise and best practices across business units.

Despite differences in governance models, the global best practices approach usually features a group, hub, or CoE that embodies the following capabilities:

- Understanding of the company's values, culture, and operations.
- Multi-disciplinary expertise on the topics of data & AI, risks, compliance, legal, public policy, and any sector- and business-specific knowledge relevant to key AI use cases.
- Up-to-date knowledge of the RAI landscape, including regulations and best practices.

28 "Responsible AI Principles from Microsoft." Microsoft. Accessed January 25, 2023. https://www.microsoft.com/en-us/ai/responsible-ai.

29 "Our Principles." Google AI. Accessed January 25, 2023. https://ai.google/principle.

30 "Responsible use of artificial intelligence and machine learning." Amazon. Accessed January 25, 2023. https://aws.amazon.com/machine-learning/responsible-machine-learning.

31 "Facebook's Five Pillars of Responsible AI." Meta AI. June 22, 2021. https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/.

32 "HSBC's Principles for the Ethical Use of Data and AI." Accessed January 25, 2023. https://www.hsbc.com/-/files/hsbc/our-approach/risk-and-responsibility/pdfs/220308-hsbc-principles-for-the-ethical-use-of-data-and-ai.pdf.

33 "Astrazeneca Data and AI Ethics." Accessed January 25, 2023. https://www.astrazeneca.com/sustainability/ethics-and-transparency/data-and-ai-ethics.html.

34 "Our commitment to ethical and responsible use of AI." Accessed January 25, 2023. https://www.novartis.com/about/strategy/data-and-digital/artificial-intelligence/our-commitment-ethical-and-responsible-use-ai.

35 "Responsible AI, Is Better AI." H&M Group, June 17, 2021. https://hmgroup.com/our-stories/responsible-ai-is-better-ai/.

36 "The OECD Artificial Intelligence (AI) Principles" OECD.AI. Accessed January 25, 2023. https://oecd.ai/en/ai-principles.

37 Green, Brian, Daniel Lim, and Emily Ratté. "Responsible Use of Technology: The Microsoft Case Study." World Economic Forum. February 2021. https://www.weforum.org/whitepapers/responsible-use-of-technology-the-microsoft-case-study/#:~:text=The%20World%20Economic%20Forum%20Responsible,technology%20product%20design%20and%20development.

38 "Putting principles into practice: How we approach responsible AI at Microsoft." Microsoft AI. Accessed January 25, 2023. https://www.microsoft.com/cms/api/am/binary/RE4pKH5.

- Sponsorship from the top management and ability to navigate the organizational structure to roll out communications, cultural change, and upskilling.
- And, for companies that wish to take a lead role in responsible AI – R&D capabilities devoted to developing new frameworks and solutions.

This group / hub / CoE can, for example, facilitate the risk classification of AI systems, monitor high-risk AI use cases, create resources such as guidelines and tools for responsible assessment, development, and deployment of AI. Furthermore, this group can also collaborate with external actors, such as policymakers and researchers, that are devoted to shaping new laws and regulations around AI technology.

> ***While the above structure is effective for AI governance in some companies, it is not a one-size-fits-all solution***

As an industry example, J.P. Morgan created the Explainable AI Center of Excellence to research the explainability and fairness of AI systems. The center aims to develop new techniques, tools, and frameworks that make AI/ML models more explainable and fair to advance the company's AI vision.[39]

By setting up a rigorous self-governance approach to responsible AI, these first-mover companies aim to not only comply with the legal standards across the regions they operate in – but also stay ahead of them. This avoids a patchwork approach in dealing with compliance and risk in the evolving regulatory landscape. By demonstrating sufficient and advanced self-governance practices, companies are better positioned to promote public and private collaboration on AI governance, for example, in support of more flexible regulatory arrangements.

When it comes to sector-specific regulations, these tend to differ considerably across countries and regions, calling for a more targeted approach to compliance for a company with global operations. The company will need to understand varying jurisdictions and decide on a potential local path diverging from the global best practices approach. For example, China's Internet Security Law and National Intelligence Law could require companies to share data with the Chinese government if requested,[40] which could conflict with a global best practices approach.

To address such conflicts, companies may opt for a customized/localized approach, adopting separate regional operations and governance structures to meet local regulatory requirements. This approach is currently embraced by many Chinese tech companies with large international customer bases to reconcile geopolitical implications and diverging regulatory requirements.

For instance, Bytedance carved out TikTok as a standalone business that operates independently from its Chinese counterpart Douyin.[41] Despite having almost the same user interfaces, TikTok, and Douyin are allegedly separated in terms of user data and operations. The implications of the separation go beyond data and are directly linked to China's specific vision for socio-technological governance, which, among others, requires social media companies to promote "positive" content aligned with the Communist Party's values. Consequently, social media companies operating in China must adopt content monitoring and moderation protocols that differ from requirements placed on social media platforms in other countries.

The rise of digital sovereignty, defined as a nation's ability to control and affect domestic information infrastructure,[42] is another challenge for companies, which compels a regional customization approach. For instance, Xiaomi, a Chinese consumer electronics company, moved its international user data and cloud services out of China to comply with data protection regulations in other markets.[43] Furthermore, Xiaomi developed different phone operating systems for the Chinese and international markets and built a version specifically for India after more than 100 Chinese apps and ser-

39 "Explainable AI Center of Excellence." J.P. Morgan. Accessed January 25, 2023. https://www.jpmorgan.com/technology/artificial-intelligence/initiatives/explainable-ai-center-of-excellence.

40 Cimpanu, Catalin. "China's Cybersecurity Law Update Lets State Agencies 'Pen-Test' Local Companies." ZDNET, Feb. 8, 2019. https://www.zdnet.com/article/chinas-cybersecurity-law-update-lets-state-agencies-pen-test-local-companies/.

41 Feng, Coco. "ByteDance carves out TikTok as world's most valuable technology unicorn finds way to satisfy US-China regulatory demands." South China Morning Post, November 2, 2021. https://www.scmp.com/tech/article/3154537/bytedance-carve-out-tiktok-worlds-sole-hectocorn-splits-six-units-delineating.

42 Larsen, B. C. (2022). "The Geopolitics of AI and the Rise of Digital Sovereignty." Brookings, December 8, 2022. https://www.brookings.edu/research/the-geopolitics-of-ai-and-the-rise-of-digital-sovereignty/.

43 "Xiaomi Moving International User Data and Cloud Services out of Beijing." ZDNET. Accessed January 25, 2023. https://www.zdnet.com/article/xiaomi-moving-international-user-data-and-cloud-services-out-of-beijing/.

vices were banned in India.[44] Examples can also be found in American companies operating in China. To meet Chinese regulations, companies such as Apple, AWS, and Microsoft have all partnered with local Chinese entities, which is a legal requirement that needs to be fulfilled for them to provide their data center services in the country.[45]

If geo-political tensions in the digital space keep intensifying and regulatory requirements diverge, we may see more multinational businesses customize, separate, or, in some cases, even shut down entire business units to be compliant. In particular, the diverging governance approaches indicate increasing differences in socio-technological values among these countries, and alignment with all these values at the same time could be increasingly difficult. Hence, this type of decision goes beyond sheer regulatory considerations and reflects on a company's core values. One prominent example is Google's exit from the Chinese market in 2010 due to increased Internet censorship in China, along with regular cyber-attack concerns.[46]

While many companies with global operations have adopted a best practices approach, sometimes with regional characteristics, this approach is not feasible for many small and medium-sized enterprises ("SME"). A local recruiting agency operating only in New York City, likely neither has the resources nor the incentives to keep track of the highest global standards surrounding AI but still needs to meet local legal requirements, for example, on the use of automated employment decision tools. For many SMEs, a local approach to AI governance allows them to comply with regional and sector-specific regulations in a cost-effective way.

However, even for businesses that choose a local approach, there may still be significant costs associated with compliance. At a minimum, companies must establish an oversight process and sometimes work with external auditors. This process involves building entirely new capabilities that most local businesses currently do not have, such as understanding and evaluating the technical and social implications of the algorithmic systems and tools that are used.

Last but not least, startups play a crucial role in creating new AI innovations. While most startups (except for the ones in the RegTech space) will not devote many resources to AI governance e.g. due to cost constraints, it is vital that they have basic checks in place to ensure their innovations are responsible. One possible way is to incorporate existing oversight tools such as Model Cards.[47] Another incentive can be guidelines and checklists, for example, provided by investors to ensure the legality of the start-up's products and long-term viability of its business model.

04

EMERGING RESPONSIBLE AI ECOSYSTEMS

As companies move from AI principles to adopt self-governance practices and new organizational processes, sometimes linked to external audits and services, they increasingly fill the institutional vacuum of trailing AI regulations. However, as discussed at the beginning of this article, a growing plethora of legislation is slowly emerging globally. In many cases, these support the advancement of an entirely new ecosystem of third-party auditors, assessment bodies, and services at the intersection of soft- and hard law measures.

In the case of the European Union, the EU AI Act delineates one vision for what an AI auditing ecosystem could look like.[48] The system would need two core components: First, a clear organizational structure for assigning responsibilities to private companies, government agencies, and supranational organizations would need to be established, along with delineating accountability for different types of system failures. Second, these actors all need access to effective auditing tools and expert knowledge to ensure that high-risk systems are safe and in compliance with the EU AI Act.[49]

44 Wright, Arol. "Xiaomi Is Rebuilding MIUI for India without Any of Its Banned Apps." XDA Developers. August 8, 2020. https://www.xda-developers.com/xiaomi-rebuilding-miui-for-india-without-banned-apps/.

45 Swinhoe, Dan. 2021. "Apple Officially Opens Data Center in China." DCD, May 28, 2021. https://www.datacenterdynamics.com/en/news/apple-officially-opens-data-center-in-china/.

46 Dahiya, Rekha. "Google's Exit from China - a Case Study." Delhi Business Review, Vol. 11, No. 2 (July - December 2010). https://www.delhibusinessreview.org/V_11n2/v11n2case-study.pdf.

47 "Model Card." Google. Accessed January 25, 2023. https://modelcards.withgoogle.com/about.

48 "The European Commission's Artificial Intelligence Act Highlights the Need for an Effective AI Assurance Ecosystem - Centre for Data Ethics and Innovation Blog." CDEI. May 11, 2021. https://cdei.blog.gov.uk/2021/05/11/the-european-commissions-artificial-intelligence-act-highlights-the-need-for-an-effective-ai-assurance-ecosystem/.

49 Mökander, J., Axente, M., Casolari, F. et al. Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation. Minds & Machines 32, 241–268 (2022). https://doi.org/10.1007/s11023-021-09577-4.

Several private sector startups have been moving into the AI governance space and provide a range of services that are specifically linked to optimizing AI governance across enterprises. Companies such as Fiddler[50] and Vera,[51] for example, ask clients to provide access to their models, code, and data, potentially allowing them to adjust model features and find more equitable outcomes. This process can be accompanied by an algorithmic impact assessment that could be provided to third-party auditors and regulators. Credo AI[52] helps companies manage AI risk through a unified platform that standardizes AI governance efforts across an organization, and TruEra[53] similarly provides a platform for explaining and monitoring AI models according to quality and reliability.

> *In the case of the European Union, the EU AI Act delineates one vision for what an AI auditing ecosystem could look like*

Traditional consulting companies are also creating new services to assess AI. EY, for example, sells a service that turns responses to questions about AI systems into a score that quantifies risks.[54] BCG X created Rate.AI, a web-based self-administered tool to assess AI projects and benchmark companies across seven dimensions of responsible AI.[55] Accenture[56] provides an algorithmic assessment process that checks for disparities in potential outcomes of AI systems and monitors for future problems once a model is deployed. BSR[57] does human rights assessments without auditing for bias or accuracy of the algorithm itself.

For now, it remains clear that on the public side of the regulatory equation, the necessary know-how of putting words into practice is lagging, and the public sector has, in many cases, not yet built the necessary institutional infrastructure to operationalize new policies. This is also true for the underlying standards where these are intended as governance mechanisms.[58]

NYC Local Law 144 is a case in point. While the law went into effect on January 1st, 2023, enforcement has been postponed to April 14, 2023. New York City's Department of Consumer and Worker Protection will use this time to provide additional guidance on how companies can comply with the law before the new enforcement date.[59]

To ensure regulatory oversight in the case of the EU, the European Commission has proposed setting up a governance structure that spans both Union and national levels. At a Union level, a "European Artificial Intelligence Board" is intended to be established to collect and share best practices among member states and to issue recommendations on uniform administrative practices. At the national level, member states will be required to appoint a competent agency to oversee the application and execution of the AI Act. This structure has similarities to the self-governance model in the private sector, as the role of the European Artificial Intelligence Board related to recommending and operationalizing best practices, is comparable to the functions of a corporate AI hub / CoE.

Going forward, the idea of creating AI Centers of Excellence ("CoE") is therefore not only applicable to private sector organizations but also to the public sector. Establishing public and private AI-focused CoEs could prove to be a critical step in (1) strengthening and (2) harmonizing approaches to AI governance and regulation, both nationally and also at the international level.

50 https://www.fiddler.ai/ Accessed January 25, 2023.

51 https://www.askvera.io/ Accessed January 25, 2023.

52 https://www.credo.ai/ Accessed January 25, 2023.

53 https://truera.com/ Accessed January 25, 2023.

54 "EY Trusted AI Platform." Accessed January 25, 2023. https://www.ey.com/en_uk/consulting/trusted-ai-platform.

55 Duranton, Sylvain, Mills, Steven. "Responsible AI: Leading by Example." Medium. February 3, 2021. https://medium.com/bcggamma/responsible-ai-leading-by-example-c25a8a0a98ea.

56 "Responsible AI | AI Ethics & Governance." Accessed January 25, 2023. https://www.accenture.com/us-en/services/applied-intelligence/ai-ethics-governance.

57 https://www.bsr.org/.

58 "The EU's AI Act Is Barreling toward AI Standards That Do Not Exist." Lawfare. January 12, 2023. https://www.lawfareblog.com/eus-ai-act-barreling-toward-ai-standards-do-not-exist#:~:text=The%20EU.

59 "New York City Proposes Regulations to Clarify Requirements for Using Automated Employment Decision Tools." JD Supra. September 26, 2022. https://www.jdsupra.com/legalnews/new-york-city-proposes-regulations-to-3740630/.

One promising avenue toward building common capacity in the public sector could be creating an AI and Regulation Common Capacity Hub ("ARCCH").[60] To act as a trusted partner for regulatory bodies, the Hub could have its home at a politically independent institution, established as a Center of Excellence in AI, drawing on multidisciplinary knowledge and expertise from across the national and international research community. The Hub would also act as an interface for regulators to interact with relevant stakeholders, including other regulators, industry, and civil society.[61] It would serve as an important source of expertise, especially for companies with fewer resources and technical expertise to draw from to understand and address risks posed by AI. Singapore's A.I Verify is a good example of a publicly provided tool that promotes transparency and trust in AI products and services through voluntary adoption and disclosure by companies.[62] Additionally, a national hub or CoE could provide regulatory sandboxes that businesses could use to test their AI innovations, and it could work with sector-specific CoEs to advise on the interactions between horizontal AI- and sector-specific regulations.

When establishing a public sector AI hub / CoE, it is important to clarify its roles and interactions with other public agencies. In the UK, for example, a new Hub or AI CoE could interface with the Digital Regulation Cooperation Forum ("DRCF") in cross-regulator collaboration, providing knowledge and expertise on AI regulations while liaising with the Office for Artificial Intelligence (OAI) to get the latest strategy updates and ensure a pro-innovation governance approach. The Hub could also collaborate with the Centre for Data Ethics and Innovation (CDEI) in best practices for operationalizing data and AI policies and collect and curate research e.g. conducted by the Alan Turing Institute (ATI) to improve its policies and recommendations.

While national AI Hubs / Centers of Excellence would be enabled to work with the private sector, they would also be able to work with national and supranational AI CoEs, such as the European Artificial Intelligence Board and the OECD's AI Policy Observatory, for example. Over time, this networked approach to AI governance could form a new institutional arena for debating potential issues and areas of alignment between private sector practices and the growing complexities of emerging regulatory regimes.

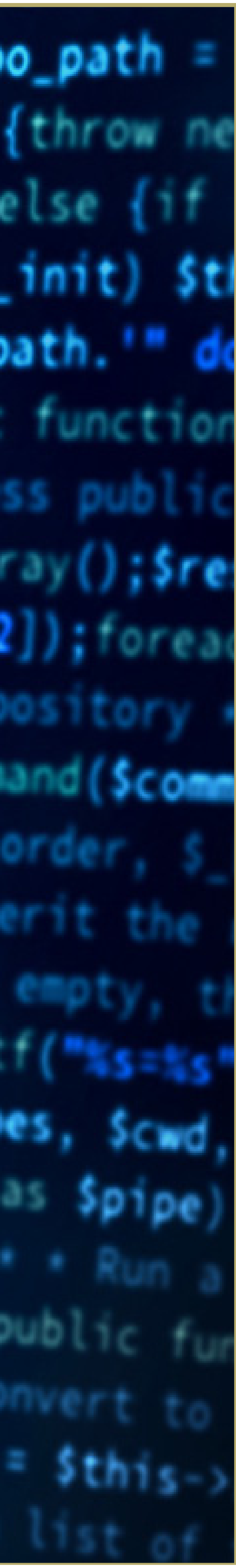
> ***One promising avenue toward building common capacity in the public sector could be creating an AI and Regulation Common Capacity Hub ("ARCCH")***

60 Aitken, M., Leslie, D., Ostmann, F., Pratt, J., Margetts, H., & Dorobantu, C. "Common Regulatory Capacity for AI." The Alan Turing Institute. 2022. https://doi.org/10.5281/zenodo.6838946.

61 *Ibid*.

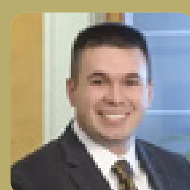
62 "Singapore's A.I.Verify Builds Trust through Transparency." OECD.ai. Accessed January 25, 2023. https://oecd.ai/en/wonk/singapore-ai-verify.

path."/config");if ($parse_ini['bare']) {$this->rep
$repo_path;if ($_init) {$this->run('init');}} else
xception('"'.$repo_path.'" is not a directory');}}
dir($repo_path);$this->repo_path = $repo_path;if ($
directory');}} else {throw new Exception('"'.$repo_
ectory) * * @access public * @return string */publi
h."/.git";}/** * Tests if git is installed * * @acce
y('pipe', 'w'),2 => array('pipe', 'w'),);$pipes = a
nts($pipes[1]);$stderr = stream_get_contents($pipes[
($status != 127);}/** * Run a command in the git re
to run * @return string */protected function run_co
pes = array();/* Depending on the value of variables
ith * putenv, and call proc_open with env=null to inh
ust those * variables afterwards. * * If $_ENV is not
LL;foreach($this->envopts as $k => $v) {putenv(sprin
;$resource = proc_open($command, $descriptorspec, $pi
derr = stream_get_contents($pipes[2]);foreach ($pipes
atus) throw new Exception($stderr);return $stdout;}
ic * @param string command to run * @return string
mmand);}/** * Runs a 'git status' call * * Accept a c
string */public function status($html = false) {$msg
return $msg;}/** * Runs a `git add` call * * Accepts
lic function add($files = "*") {if (is_array(
`git rm` call * * Accep



REGULATING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

BY HEATHER EGAN SUSSMAN



& IAN ADAMS

& NUR LALJI

Heather Egan Sussman is a partner at Orrick, Herrington & Sutcliffe LLP, and head of the firm's Strategic Advisory and Government Enforcement Business Unit. Ian Adams is a public policy attorney at Orrick, advising clients on matters at the intersection of law, business and public policy. Nur Lalji is an associate in ORrick's Cyber, Privacy & Data innovation Practice Group.

01 INTRODUCTION

AI and ML are considered some of the most important technological developments in recent years, and their use across a myriad of industries has exploded over the past decade. A 2022 survey by NewVantage Partners found that nearly 92 percent of executives said their organizations were increasing investments in data and AI systems and 26 percent of companies already have AI systems in widespread production.

Perhaps what is most compelling about AI and ML from a business perspective is its potential to make organizations more efficient and data driven in their decision-making. Specifically, the use of ML algorithms makes it possible for organizations to ingest huge amounts of data, identify patterns, and create rules that enable the machine learning model to make automated decisions and provide an output to the organization that otherwise may have been either too time or capital intensive. However, there is real risk that the power of AI and ML may be misused. In order to mitigate the potential for harm at the hands of AI and ML systems, there is increasing pressure for regulatory oversight. Over the last two years, policymakers and regulators, from international bodies to municipal governments, have begun to focus on the potential for AI applications to cause harm. The increasing drum beat of regulation of AI and ML on both sides of the Atlantic makes clear that the global race to regulate AI and ML has begun in earnest.

Compliance (and noncompliance) with these regulations may have a steep cost for businesses. AI and ML touch on many aspects of the regulatory tapestry in the U.S. and abroad — privacy, security, employment, civil rights, regulation of BigTech, and beyond. The potential for large fines, lawsuits, and regulatory investigations makes it essential for organizations to build a risk and governance strategy that explicitly accounts for AL and ML-related activities. In fact, it may be necessary to consider structural modifications within firms to identify an individual or cross-functional committee to take responsibility over AI and ML compliance.

In this article, we (i) identify some of the novel legal, ethical, and privacy issues that AI and ML present; (ii) evaluate the differing approaches to regulating AI and ML in Europe and at the federal and state levels in the United States; and (iii) discuss considerations for building an effective risk management and governance strategy.

02
LEGAL, ETHICAL, AND PRIVACY CONCERNS

AI and ML systems present novel legal, ethical, and privacy challenges. For example, these systems can produce unintentionally biased outputs based on bias inherent in the data they ingest or the algorithms that processes the underlying data. This can produce discriminatory or otherwise negative outcomes. Additionally, due to the large troves of data these systems require, the use of AI and ML is also often at odds with privacy and consumer protection principles. We discuss each of these in turn.

A. Bias in AI

In 2018, researchers Joy Buolumwini and Timnit Gebru exposed the inherent biases in the facial recognition models across several major technology companies. According to their study, *Gender Shades*, these companies' facial recognition technologies were significantly more likely to misidentify women and individuals with darker skin tones.[2] These disproportionate error rates were reportedly produced, in part, because of the training data fed to the model — which was predominantly white and male.[3] This study, and others like it, show the often unintended but discriminatory consequences of AI and ML systems that are not carefully reviewed by diverse and cross-disciplinary teams of engineers, data scientists, compliance professionals, and lawyers that are tasked with considering the ethical use of AI and ML.

Moreover, although AI and ML has often been touted as a neutral solution, often, the inverse is true — without human intervention, these models may reflect back historic biases that had previously gone undetected. Amazon, for example found that the algorithms it developed for hiring were disproportionately disadvantaging women. This was reportedly because the algorithms were trained on resumes submitted to Amazon in the previous ten years, which disproportionately "came from men, a reflection of male dominance across the tech industry."[4] One can easily see how companies utilizing AI and ML in their hiring processes may unintentionally produce similar effects if there are not adequate safeguards in place to review the underlying data and the algorithm and remove inherent biases.

2 Joy Buolamwini & Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, Proceedings of Machine Learning Research 81:1–15 (2018).

3 *Id.* at 3.

4 Jeffrey Dastin, Amazon scraps secret AI recruiting tool that showed bias against women, Reuters (October 10, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

These examples showcase how important it is for organizations to audit their training set data and algorithmic outputs to account for unintentional results, such as the incomplete or inaccurate representation of a particular group or a legally protected class. It is worth noting that mitigating biases of this type has been the impetus for several regulatory proposals, many of which center around audit requirements that would result in the proliferation of disparate impact tests — an outcome many companies are likely to find problematic.[5] However, companies that pro-actively endeavor to address such issues and seek to promote transparency at a high level in how their AI and ML systems operate, may inoculate themselves from the worst scrutiny.

B. Privacy and AI

There is an inherent tension between privacy and AI. Privacy laws generally promote the concept of "data minimization," which stands on the principle that organizations should limit their collection of personal information to only that which is directly relevant and necessary to accomplish the purpose for which the personal information was collected for in the first place. From a consumer protection standpoint, principles of lawfulness, fairness, and transparency are also key, meaning that individuals should be provided with information and afforded meaningful choices with regards to how their personal information is collected and used. AI and ML systems, however, need to be trained with large amounts of data, and they improve as more data is fed to them. This friction has led some business to obfuscate how they use personal information to train their AI and ML models, with consumers only learning about this data usage after the fact.[6]

Despite these inherent challenges, only 44% of executives said their organizations have well-established policies and practices to support data responsibility and AI ethics.[7] Nonetheless, the legal landscape surrounding AI and ML has changed dramatically in recent years, and new laws seek to protect consumers from these legal and ethical harms.

03
LEGAL LANDSCAPE

In the spring of 2021, the European Commission (the "Commission") published its highly anticipated communication and "Proposal for a Regulation laying down harmonized rules on artificial intelligence" (the "EU AI Regulation").[8] The EU AI Regulation was released just days after the Federal Trade Commission (the "FTC") published a blog post entitled "Aiming for truth, fairness, and equity in your company's use of AI" (the "2021 FTC Memo").[9] Additionally, there have been a flurry of AI and ML-related action from U.S. regulatory agencies, state governments vis-à-vis privacy laws, and U.S. city governments relating to the use of AI and ML for employment decisions.[10]

A. Europe

The European Commission proposed the EU AI Regulation in the spring of 2021 to harmonize AI rules across the continent. The EU AI Regulation takes a risk-based approach to controls on using AI and ML systems, depending on the intended purpose of the system. The EU AI Act proposes a sliding scale of rules based on risk that would classify different AI and ML applications as unacceptable, high, limited, or minimal risks.[11]

5 See, e.g. U.S. EEOC, Artificial Intelligence and Algorithmic Fairness Initiative, https://www.eeoc.gov/ai (Last accessed Jan. 23, 2023); New York City Council, Automated Employment Decision Legislation, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search; Algorithmic Accountability Act of 2022, congress.gov/bill/117th-congress/house-bill/6580/text; American Data Privacy and Protection Act of 2022, https://www.congress.gov/bill/117th-congress/house-bill/8152/text.

6 See, e.g. Alex Hern, TechScape: Clearview AI was fined £7.5m for brazenly harvesting your data – does it care?, The Guardian (May 25, 2022), https://www.theguardian.com/technology/2022/may/25/techscape-clearview-ai-facial-recognition-fine.

7 Tam Habert, *Regulations Ahead on AI*, SHRM (April 2, 2022), https://www.shrm.org/hr-today/news/all-things-work/pages/regulations-ahead-on-artificial-intelligence.aspx.

8 European Commission, *Proposal for a Regulation laying down harmonized rules on artificial intelligence (April 21, 2021)*, digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

9 Press Release, Federal Trade Commission, Aiming for truth, fairness, and equity in your company's use of AI (April 19, 2021), https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

10 AI and ML are dependent on huge data sets which can include personal information, including sensitive personal information. Consequently, privacy laws have become a primary means to address the risks inherent in relying on AI to make decisions that have legal and social consequences such as loan approvals or employment decisions.

11 Press Release, Orrick, The New EU Approach to the Regulation of Artificial Intelligence (May 7, 2021), https://www.orrick.com/en/Insights/2021/05/The-New-EU-Approach-to-the-Regulation-of-Artificial-Intelligence.

The EU AI Regulation is intended to have extraterritorial effect and establishes the European Artificial Intelligence Board that will have significant authority to levy "dissuasive" fines for noncompliance of up to 6% of annual global turnover for certain breaches, as well as the power to order AI and ML systems to be withdrawn from the market. The inclusion of these GDPR-like penalties shows that the EU is serious about regulating the burgeoning AI and ML industry. The EU AI Regulation applies across all sectors (public and private) to "ensure a level playing field." On December 6, 2022, the European Council adopted its common position on the Artificial Intelligence Act.[12] The adoption of this approach enables the Council to enter negotiations with the European Parliament once the European Parliament adopts a position on the proposed regulation. Negotiations are expected to be complex with thousands of amendments already proposed by political groups in the European Parliament.

The EU AI Regulation will become law once both the European Commission and the European Parliament agree on a common version of the text and will enter into force 24 months after that date, though some provisions may apply sooner. If enacted, the regulation would have significant consequences for organizations that develop, sell, or use AI or ML systems. Those consequences include the introduction of legal obligations and a monitoring and enforcement regime with hefty penalties for non-compliance. Specifically, organizations will be required to register standalone high-risk AI or ML systems, such as remote biometric identification systems, in an EU database. Potential fines for noncompliance range from 2-6% of a company's annual revenue. The regulation has striking similarities to the General Data Protection Regulation, or GDPR, which already carries implications for AI as Article 22[13] prohibits decisions based on solely automated processes that produce legal consequences or similar effects for individuals unless the user has explicitly consented, or the AI or ML system meets other requirements.

The proposed EU AI Regulation will have a significant impact on any organization that operates anywhere in Europe or targets the European market. It is likely that the regulation of AI will follow a path similar to the evolution of data privacy regulations where the sweeping regulations that start in the EU cause other jurisdictions to follow that lead. In the United States, a patchwork of local, state, and federal regulations, guidance, and frameworks have already emerged in the wake of the EU AI Regulation and do not appear to be losing steam.

> *The proposed EU AI Regulation will have a significant impact on any organization that operates anywhere in Europe or targets the European market*

B. United States

Unlike the comprehensive framework proposed in Europe, regulatory guidelines have generally been proposed on an agency-by-agency basis in the United States, as well as regulation at the state and local levels.

1. National AI Initiative Act

In January 2021, the National AI Initiative Act (the "U.S. AI Act")[14] became law creating the National AI Initiative that provides "an overarching framework to strengthen and coordinate AI research, development, demonstration, and education activities across all U.S. Departments and Agencies." The U.S. AI Act created new offices and task forces aimed at implementing a national AI strategy, implicating a multitude of U.S. administrative agencies including the FTC, Department of Defense, Department of Agriculture, Department of Education, and the Department of Health and Human Services.

2. Algorithmic Accountability Act of 2022

The Algorithmic Accountability Act (the "AAA")[15] of 2022 was introduced on February 3, 2022, by Sen. Ron Wyden, Sen. Cory Booker, and Rep. Yvette Clark. The bill is likely to be reintroduced in a substantially similar form in the new Congress and would require large technology companies across the states to perform a bias impact assessment of any automated decision-making system that

12 Press Release, Council of the EU, Artificial Intelligence Act: Council Calls for Promoting Safe AI that Respects Fundamental Rights (December 6, 2022), https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/#:~:text=The%20Council%20has%20adopted%20its,fundamental%20rights%20and%20Union%20values.

13 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal information and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) GDPR Article 22, https://gdpr-info.eu/art-22-gdpr/.

14 National AI Initiative Act of 2020, https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210.

15 Algorithmic Accountability Act of 2022, congress.gov/bill/117th-congress/house-bill/6580/text.

makes critical decisions in a variety of sectors, including employment, financial services, healthcare, housing, and legal services. The Act's scope is potentially far reaching as it defines "automated decision system" to include "any system, software, or process (including one derived from ML, statistics, or other data processing or artificial intelligence techniques and excluding passive computing infrastructure) that uses computation, the result of which serves as a basis for a decision or judgment." Notably, significant sections of the Act as introduced in 2022 were incorporated into the "three corners" privacy bill (known as the ADPPA) that will function as the basis for future efforts to develop a national digital privacy standard.[16]

3. Department of Commerce

A flurry of AI-related activity has emanated from the Department of Commerce, including a move towards a risk-management framework. Congress has directed the National Institute of Standards and Technology, part of the Commerce Department, to develop "a voluntary risk management framework for trustworthy AI systems." That framework may greatly influence how organizations approach AI-related risks, including avoiding bias and promoting accuracy, privacy, and security.

In September 2021, the Department of Commerce established the National Artificial Intelligence Advisory Committee[17] to advise the president and federal agencies. It will offer recommendations on the "state of U.S. AI competitiveness, the state of science around AI, issues related to the AI workforce" and how AI can enhance opportunities for underrepresented populations, among other topics. Given its responsibilities and engagement with AI, the Department of Commerce appears poised to play a central role in the federal approach to AI regulation.

4. Federal Trade Commission

The FTC has also made it clear that it will use its power under Section 5 of the FTC Act, the Fair Credit Reporting Act, and the Equal Credit Opportunity Act to help ensure AI is used truthfully, fairly, and equitably in the United States.

In September 2021, the Department of Commerce established the National Artificial Intelligence Advisory Committee to advise the president and federal agencies

The 2021 FTC Memo discussed above made clear that the FTC will marshal its resources to pursue the use of biased algorithms. The FTC provided a roadmap for its compliance expectations and organizations should "keep in mind that if you don't hold yourself accountable, the FTC may do it." Among other things, organizations should:

· Rely on inclusive data sets: "Companies should think about ways to improve their data set, design their model to account for data gaps, and — in light of any shortcomings — limit where or how they use the model."
· Test an algorithm before use and periodically afterwards "to make sure that it doesn't discriminate based on race, gender, or other protected class."
· Be truthful about how they use customers' data and don't exaggerate an algorithm's abilities.
· Embrace transparency and independence.[18]

In June 2022, the FTC indicated that it plans to submit an Advanced Notice of Preliminary Rulemaking to "ensure that algorithmic decision-making does not result in harmful discrimination."[19] Also in June 2022, the FTC issued a report to Congress discussing how AI may be used to combat online harms, ranging from scams, deep fakes, and opioid sales.[20] However, the report sought to strike a balance and noted that AI is also susceptible to producing biased and discriminatory outcomes.

16 American Data Privacy and Protection Act of 2022, https://www.congress.gov/bill/117th-congress/house-bill/8152/text.

17 Press Release, Department of Commerce Establishes National Artificial Intelligence Advisory Committee (September 8, 2021), https://www.commerce.gov/news/press-releases/2021/09/department-commerce-establishes-national-artificial-intelligence.

18 Press Release, Federal Trade Commission, Aiming for truth, fairness, and equity in your company's use of AI (April 19, 2021), https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

19 FTC Plans to Submit an Advanced Notice of Preliminary Rulemaking ("ANPRM") "under section 18 of the FTC Act to curb lax security practices, limit privacy abuses, and ensure that algorithmic decision-making does not result in unlawful discrimination." https://www.reginfo.gov/public/do/eAgendaViewRule?pubId=202204&RIN=3084-AB69.

20 Federal Trade Commission Report to Congress, *Combatting Online Harms Through Innovation* (June 16, 2022), https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf.

5. The White House

The E.U.-U.S. Trade and Technology Council has committed[21] to develop "AI systems that are innovative and trustworthy and that respect universal human rights and shared democratic values." The council also plans to discuss "measurement and evaluation tools. . . to assess the technical requirements for trustworthy AI" and study the technology's impact on the labor market.

In November 2021, the White House Office of Science and Technology Policy solicited engagement[22] from stakeholders across industries in an effort to develop a "Bill of Rights for an Automated Society." Such a Bill of Rights could cover topics like AI's role in the criminal justice system, equal opportunities, consumer rights, and the healthcare system.

6. Food and Drug Administration

The Food and Drug Administration (the "FDA") issued Artificial Intelligence/Machine Learning Based Software as a Medical Device ("SaMD") Action Plan to outline its proposed steps for creating a regulatory framework "that would allow for modifications to be made from real-world learning and adaptation, while ensuring that the safety and effectiveness of the software as a medical device are maintained."[23] The Action Plan outlines how the agency intends to oversee development and use of the software in the SaMD context.

7. National Security Commission and Government Accountability Office ("GAO")

The National Security Commission on Artificial Intelligence submitted its final report to Congress in 2021. It recommends the government take domestic actions to protect privacy, civil rights, and civil liberties in its AI deployment. It notes that a lack of public trust in AI from a privacy or civil rights/civil liberties standpoint will undermine the deployment of AI to promote U.S. intelligence, homeland security, and law enforcement. The report advocates for public sector leadership to promote trustworthy AI, which will likely affect how AI is deployed and regulated in the private sector.

Also in 2021, the GAO identified practices to help ensure accountability and responsible AI use by federal agencies. The report identified four key focus areas:

· Organization and algorithmic governance
· System performance
· Documenting and analyzing data to develop and operate an AI system
· Continuous monitoring and assessment to ensure reliability and relevance over time.

> ***"Also in 2021, the GAO identified practices to help ensure accountability and responsible AI use by federal agencies***

8. EEOC

In May 2022, the U.S. Equal Employment Opportunity Commission (the "EEOC") released a guidance[24] warning to U.S. companies that their use of algorithmic decision-making tools to assess job applicants and employees could violate the Americans with Disabilities Act by intentionally or unintentionally screening out individuals with disabilities when utilizing algorithms in the hiring process.

9. NIST

The National Institute of Standards and Technology ("NIST"), which falls under the U.S. Department of Commerce, is currently engaging with stakeholders to develop "a voluntary risk management framework for trustworthy AI systems."[25] Additionally, in September 2021, NIST released a paper describing its Principles on Explainable AI.[26] Under these non-binding principles, AI algorithms should:

21 Press Release, U.S.-EU Trade and Technology Council Inaugural Joint Statement (September 29, 2021), https://www.whitehouse.gov/briefing-room/statements-releases/2021/09/29/u-s-eu-trade-and-technology-council-inaugural-joint-statement/.

22 Press Release, Join the Effort to Create A Bill of Rights for an Automated Society (November 10, 2021), whitehouse.gov/ostp/news-updates/2021/11/10/join-the-effort-to-create-a-bill-of-rights-for-an-automated-society/.

23 Press Release, Artificial Intelligence and Machine Learning in Software as a Medical Device (January 2021), https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.

24 Press Release, The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees (May 12, 2022), https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.

25 NIST AI RISK MANAGEMENT FRAMEWORK (August 18, 2021), https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.pdf.

26 NIST Four Principles of Explainable Artificial Intelligence (September 29, 2021), https://www.nist.gov/publications/four-principles-explainable-artificial-intelligence.

- · Have accompanying evidence or reason(s) for all outputs;
- · Be understandable to individual users;
- · Correctly represent how system generates the output;
- · Have confidence in output and only operate in the conditions for which it was designed.

10. <u>State Privacy Laws</u>

Because AI and ML are dependent on huge data sets which may include personal information, U.S. state privacy laws have become one of the means to mitigate risk. These laws, which include the California Privacy Rights Act (the "CPRA"), the Colorado Privacy Act (the "CPA"), the Virginia Consumer Data Protection Act (the "VCDPA"), and the Connecticut Data Privacy Act (the "CTDPA"), seek to put the consumer in control of their personal information and ensure that AI is used in a responsible manner. However, they also create new obligations for organizations to assess and potentially comply with, and also are structured in a way that poses unique challenges in the AI and ML context.

(i) Obligations for AI and ML Systems Use

Each of the state privacy laws grant consumers rights regarding opting out of the processing of their personal information for purposes of profiling and create requirements that impact automated decision-making. Though the definitions of automated decision-making and profiling differ slightly across the state privacy laws, profiling generally refers to an organization attempting to evaluate personal aspects of a data subject via the processing of their personal information. Relatedly, automated decision-making refers to an organization either (i) acting upon profiling to make a decision by automated means without human intervention or with limited human intervention or (ii) establishing an automated system that renders a decision based directly on information provided by a data subject (such as an age gate that would prevent anyone under a certain age from being able to participate in a program or apply for a position). Several of the state privacy laws also require data controllers to conduct a data protection impact assessment (a "DPIA") for processing activities that present a "heightened risk of harm" to a consumer.

(ii) Challenges with the privacy law framework

The state privacy laws generally split businesses up into two categories: entities that control the ways in which consumers' personal information is collected, used, and disclosed (i.e. that act as a "controller") and entities that assist these businesses and act as a "service provider" or "processor" on their behalf. Acting as a controller or service provider/processor come with varying obligations, risks, and benefits that organizations must consider. While conventional businesses may more naturally fit into one category or the other, businesses that use AI and ML systems may have a difficult time classifying themselves in accordance with these definitions.

> ***Because AI and ML are dependent on huge data sets which may include personal information, U.S. state privacy laws have become one of the means to mitigate risk***

For example, a SaaS-based vendor that uses AI and ML systems as part of their product offerings may generally consider themselves to be a service provider, but want to use the data they collect from their customers to improve their machine learning model. Where the data they receive from customers includes personal information, this provides a challenge under the state privacy law frameworks, as service providers are generally prohibited from using personal information for their own purposes.[27] Although certain privacy laws include exceptions to this requirement, such as permitting service providers to use personal information purely for their own internal purposes, use of personal information to train and improve a machine learning model does not clearly fit within this exception, as that data may be combined with other datasets and used for the benefit of other customers.[28] Practices such as deidentifying and aggregating personal information may solve part of this problem, it may not provide a workable solution for all vendors, such as where their model may be predicated on the use of the personal information to provide the service. In addition, customers themselves may be hesitant to allow the personal information they provide to a vendor to be used to enhance the vendor's machine learning model, even where such information is deidentified. As such, businesses that use AI and ML will need to think carefully about how they classify themselves under the privacy law framework and develop their compliance strategy with this classification in mind.

All of the state privacy laws come into effect at varying points in 2023. Accordingly, the compliance obligations and the potential for increased regulatory scrutiny in the U.S. will increase significantly in the coming year. For more infor-

27 See, e.g. CCPA Draft Regulations, § 7051(a).

28 CCPA Draft Regulations, § 7050(a)(3).

mation about how state privacy laws will affect AI, see our Orrick's Insight, "New State Privacy Laws Zero in on AI."[29]

11. New York City's Biometric Data Protection Law

On July 9, 2021, the New York City Biometric Identifier Information Law (the "NY Biometric Act")[30] went into effect. The NY Biometric Act applies to the collection and processing of "biometric identifier information," which is defined as "physiological or biological characteristic that is used by or on behalf of a commercial establishment, singly or in combination, to identify, or assist in identifying, an individual." The NY Biometric Act identifies a retina or iris scan, a fingerprint or voiceprint, and a scan of hand or face geometry as examples of biometric identifier information. The NY Biometric Act only applies to a "commercial establishment," defined as a place of entertainment, a retail store, or a food and drink establishment.

There are two primary legal requirements: (i) commercial establishments that collect, retain, or share a customer's biometric identifier information must disclose these activities "by placing a clear and conspicuous sign near all… customer entrances notifying customers in plain, simple language." The NY Biometric Act does not require commercial establishments to obtain any type of written consent from consumers either before or even after their biometric data is collected; and (ii) commercial establishments cannot "sell, lease, trade, share in exchange for anything of value or otherwise profit from the transaction of biometric identifier information."

The NY Biometric Act includes a private right of action under which individuals can recover damages of $500 per violation for an establishment's failure to post a conspicuous notice, $500 for each negligent violation of the ban on the sale or sharing of biometric data, and $5,000 for each intentional or reckless violation of the ban on selling or sharing biometric identifier information.

Additionally, New York City has passed the first law[31] in the United States that will require employers to conduct audits of automated decision-making tools used to evaluate job candidates or employees (the "New York City Automated Employment Decision Law"). The law, which took effect in January 2023, calls for audits of tools that automatically screen job candidates. Failure to comply with the law may result in the imposition of civil penalties of up to $500 for a first violation and each additional violation occurring on the same day as the initial violation, and between $500 and $1,500 for each subsequent violation. The law specifies that "[e]ach day on which an automated employment decision tool is used" in violation of the provision requiring a bias audit "shall give rise to a separate violation." Additionally, the failure to provide any of the required notices constitutes a separate violation.

While the New York City Automated Employment Decision Law applies only to employers in New York City, it's likely to have a much broader impact as large companies that hire employees in New York will likely be forced to update their hiring systems across the board to meet the floor established by the legislation.

04

NEXT STEPS: WHAT SHOULD ORGANIZATIONS DO?

The regulation of AI and ML will continue to be a rapidly developing area of law. To mitigate the risk of legal liability and "future proof" compliance efforts, organizations are wise to build a compliance framework that focuses on predictability and transparency, as well as continuous auditing, refining, and monitoring with programmatic modification of AI and ML systems as appropriate. This can include:

- Crafting policies and procedures to create a compliance-by-design program that promotes AI innovation while ensuring transparency and explicability. In practice, this may involve the development of first-order principles that inform more granular practical guidance, including methods for human intervention where appropriate.
- Instituting cross-disciplinary teams of engineers, compliance and legal professionals, relevant executives, HR professionals, and members from across the organization to recognize problematic applications of AI and ML and cure such applications in a responsible and efficient manner.
- Developing privacy-forward solutions to data usage where possible, such as:
    - Deidentify and anonymize personal data when possible;
    - Create methods for removal of personalized information from machine learning

29 Press Release, Orrick, New State Privacy Laws Zero in on AI (August 11, 2022), https://www.orrick.com/en/Insights/2022/08/New-State-Privacy-Laws-Zero-in-on-AI.

30 New York City Biometric Identifier Information Law, https://codelibrary.amlegal.com/codes/newyorkcity/latest/NYCadmin/0-0-0-42626.

31 New York City Council, Automated Employment Decision Legislation, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search.

models upon request;

- o Enact notice and consent frameworks for the use of individuals' sensitive personal information within the machine learning context.

· Implementing rigorous testing and review practices designed to identify, analyze, and address patterns and outcomes, all focused on continuous improvement.

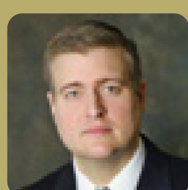
· Documenting these processes to comply with regulators who may seek further information.

Taking these steps will not only help to future proof compliance efforts as AI and ML regulation continues to develop over the coming months and years, but it will also provide the evidence to show regulators, investors, and the public alike that responsible AI and ML is a top priority. Orrick has assembled further resources about steps organizations can take to maximize the benefits of AI while minimizing regulatory risk.[32]

> *The regulation of AI and ML will continue to be a rapidly developing area of law*

32 Press Release, Orrick, AI Tips: 10 Steps to Future-Proof Your Artificial Intelligence Regulatory Strategy, (July 1, 2021), https://www.orrick.com/en/Insights/2021/07/AI-Tips-10-Steps-to-Future-Proof-Your-Artificial-Intelligence-Regulatory-Strategy.

PRINCIPLES OF DIGITAL LAW AND ETHICS



BY
THOMAS FREEMAN

&
DR. AARON MCKAIN

Thomas J. Freeman, JD, MS, MBA, is a Business Law Professor at Creighton University, a practicing attorney, and the Director of Research and Policy Analysis at the Institute for Digital Humanity. Aaron McKain, Ph.D., M.S.L. is the Director of the School of English and Communication Arts and Associate Professor of Digital Media at North Central University, as well as the Executive Director of the Institute for Digital Humanity.

The authors would like to thank: Samson Hall is a Research Associate for the Institute for Digital Humanity. Joseph Masi, Michael Bartz, William Harper, Kylie Karsky, Madeleine Britto, C.J. Braccia, Emily Atamov, Ava Verzani, and Clare Hart are students at Creighton University and contributed to this article.

We live in an era of Big Data, where most information about us is known or knowable. Gone are the days when privacy consisted of the ability to choose what we shared with the world. We spend more and more of our lives online. Our online activities, movements, purchases, and communications are tracked, cataloged, and used to judge and influence us. There are almost no rules for this Brave New World we live in, which permits our government(s) and private companies to engage in all manner of questionable practices that would have been unthinkable – violating state and federal law as well as industry and civic norms – only a decade ago.

If we do not change course quickly, hard-won privacy and civil rights can be lost, possibly forever. There is a desperate need for legal regulation and ethical guardrails for the digital world. An emerging industry is sprouting up to fix the messes caused by corporations and governments vacuuming and monetizing personal data. However, there is a conspicuous lack of careful attention to some basic and fundamental questions of digital ethics and law. What types of information about people should be allowed to be gathered about patients, students, defendants, and citizens? How should the governments and corporations that gather it be allowed to use it? When and how should data about individuals be used to make decisions about them that affect their abilities to secure housing, medical care, employment, parole, or probation? When algorithms or other automated processes are designed to make decisions about people, what safeguards are necessary to ensure those decisions are accurate and free from bias?

Digital ethics doesn't pertain to one piece of technology or another. It's an ecosystem that demands a public referendum on what personal information should be private, how automated decisions should be made, what information should be censored (and by whom), and what it means to be a citizen or person in criminal, employment, educational, financial and health care contexts. Programmers, policymakers, teachers, advocates, and lawyers struggle to adequately address these issues. Legislators and courts are asking for help because they cannot keep up with rapidly evolving technology. The digital ethics community has few proven holistic solutions accepted across industries, education levels, and academic disciplines. There is a danger that ideological divisions could become entrenched and block effective bipartisan coalitions and solutions.

As a guide to addressing these concerns, this article will attempt to lay out the core principles for digital law and ethics, gleaned from both the author team's research and their experience with their technology ethics initiative: The Institute for Digital Humanity. Started in 2018, the IDH is a bipartisan and multi-faith digital rights think tank that works to secure the rights of everyone. And by following some basic principles of digital ethics and constitutional law, the IDH has found a "cheat code" that should be of interest to any organization that is serious about reclaiming civil rights. We begin this article by outlining some overarching principles of digital ethics and constitutional law. Then we turn to two specific examples – privacy and algorithmic decision-making – to show how these principles apply.

01 DIGITAL LAW AND ETHICS

Although it seems overwhelming, the issues regarding how to regulate this new digital world can be distilled into two fundamental questions: 1) what information about us can be gathered and retained, and 2) how can it be used to judge us? With those guidelines in mind, we have developed a list of core digital rights principles.

A. Core Principle #1: Digital Ethics and Law Issues Require a Holistic Approach

Privacy and AI decision-making are often the central focus of AI ethics reform. But *all* digital ethics issues must be dealt with holistically and based on consistent principles. While the case studies in this article deal explicitly with AI decision-making and intrusions into privacy that would have been unethical or illegal in the pre-digital age, these methodologies and principles have been developed – and are compatible with – the two other and irretrievably interconnected, primary digital threats to our civil rights and democratic values:

- *Disinformation/Misinformation:* Who decides what information misleads or is false and should be censored as a result? What rules or guidelines should be used to make those determinations? All of us are uncomfortable with the fake news and conspiracy theories we see online. But who do we trust to identify those and determine whether we should be allowed to view them and judge them for ourselves?

- What are the rules of behavioral advertising and political micro-targeting?

- *Rebalancing Free Speech Versus Hate Speech: How can we determine when to suppress online speech? When does speech become so hateful that it must be censored? Who should be empowered to make those decisions? What standards should be used for making those decisions?*

B. Core Principle #2: Give Everyone a Seat at the Table: Digital DEI

Digital ethics affects all of us. This is an enormous and diverse world. Every person in it has a stake in how key digital civil rights issues – from privacy to algorithmic discrimination to disinformation to free speech versus online hate speech – are determined. Every political, cultural, or religious faction can veto any tech solution. The teams composed to design, assess, and evaluate algorithms must be truly diverse, based on the presence of those with different races, genders, ages, disability statuses, etc., as well as of thought leaders from various industries, professions, academic fields, backgrounds, and worldviews. A trusted leader from a particular minority community can explain how members of that community might be affected by or react to a product or service. Historians, philosophers, attorneys, and industry leaders can bring unique insights about how a data set is biased or a practice might be illegal or impracticable. If humans are to be weighed, measured, and judged by algorithms, those algorithms should at least be intelligently and thoughtfully designed.

C. Core Principle #3: Interdisciplinary and Peer-Reviewed Methods

Digital ethics, by its nature, is an interdisciplinary field. The practices of effectively designing and evaluating digital tools and policies will require diverse groups of people drawn from different industries and academic fields. Lawyers, ethicists, philosophers, historians, writers, and artists must be included in those conversations, as they can all bring different perspectives. The methods by which they assess questions like what types of data collection should be permitted or when an algorithm should be allowed to judge a person should be peer-reviewed to ensure they work as intended. Too many unintended problems caused by unregulated technology occurred because the experts from various disciplines, and with diverse life experiences were not consulted before implementation. It would help if you had a methodology — and here, the IDH uses narrative theory, but there are others – that are accepted and valued (and considered unbiased) by multiple professional and academic communities.

D. Core Principle #4: Teachable and Understandable to Everyone

Our lives are increasingly lived online. Our resumes are stored on sites like LinkedIn. Our thoughts are collected on applications like Facebook, Instagram, and TikTok. Algorithms are increasingly making decisions about us. These systems of data collection are almost inescapable. Each interaction with the world is increasingly monitored, cataloged, and used for algorithmic assessment and/or prediction. It is, therefore, vital that individuals understand how and why they are being judged. And principles of digital ethics – while complex enough to be of use to lawyers, legislators, policymakers, and programmers – need to be simple enough that anyone can understand them to both know and express their rights.

02

RETHINKING THE RIGHT TO PRIVACY

How the world views us is increasingly a function of how we conduct ourselves online. We view news stories, advertisements, and other online content based on who tech companies think we are. How do you carve out a "private" space for your identity when "how" you present yourself (via social media, search engines, browser clicks, and purchases) is radically re-contextualized and algorithmically calculated by – to name just a few prominent examples – future schools, employers, retail companies, political advertisers, and police departments? The state of U.S. privacy law remains in flux, with states such as California trying to go it alone with laws like the California Consumer Privacy Act ("CCPA"). To date, the United States has yet to pass any comprehensive laws regarding privacy similar to Europe's General Data Protection Regulation ("GDPR").

There is a common misconception – regrettably shared even within the digital ethics community – that privacy no longer exists. This is both dangerous and patently incorrect in the context of post-digital Constitutional law in the United States. The right to privacy still exists, but in the era of Big Data needs to be reconceptualized: It isn't the right to privacy that has disappeared. It's the traditional view of *privacy as secrecy* that no longer works. As Justice Alito argued in *U.S. v. Jones*, if "an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties," and almost all information about us is known or discernible in the post-digital era, then privacy, as a legally protected concept, would, wrongly, cease to meaningfully exist. ("I would not assume that all information voluntarily disclosed to some member of the public for a limited purpose is, for that reason alone, disentitled to Fourth Amendment protection."[2]

More importantly, Alito recognizes that the nature of privacy as a protected Constitutional right turns on – both philosophically and legally – what the public "believes" its rights to post-digital privacy are, both affirmatively and via the traditional violations of privacy that they are – every day in the post-digital era – acquiescing to. Every judgment we make – whether to use that CVS card on our keychain or to turn on our GPS – is unavoidably larger than itself and an *epideictic* declaration of our values.

The modern right to privacy will need to shift from a singular focus on "what is shared" – by acknowledging the reality that most information about us has already been collected or can be ascertained based on statistical guesswork – to a

2 *U.S. v. Jones*, 565 U.S. __ (2012), slip op., Sotomayor concurrence at 5.

symbiotic legal/cultural paradigm that asks "whether, when, and how" information about us can be used to form judgments and make decisions about and against an individual: How was the information about the individual obtained? How old was the person at the time of the alleged misdeed? What were the circumstances surrounding it? As we are increasingly deprived of the ability to maintain secrets or personal information, we must become more thoughtful and intentional about how people are judged and when they should be forgiven. Even after perhaps the most shunning in literary history, Hester Prynne earned back her community's respect.[3] The modern world will likewise have to think about the concept of forgiveness.

> ***There is a common misconception – regrettably shared even within the digital ethics community – that privacy no longer exists***

The erosion of privacy has been a slow-moving crisis for decades. A more holistic approach to privacy is needed, which envisions it as an ecosystem of public, private, and professional decisions about when, where, and why data should be, legally and ethically, allowed to be used as evidence when making judgments and interpretations about particular people in particular contexts.

A. Privacy Value #1: Transparency and Knowing Consent

The targets of any data collection should have to consent to it. That consent must be premised on full and fair disclosure about the data collection. Knowing consent requires that those who collect our data disclose what they are gathering and doing with it. Those whose data is collected should have the right to withdraw their consent and request that their data not be used in a certain way or deleted entirely.

B. Privacy Value #2: Evidence Exclusion Rules

When do we allow data about someone to be used as evidence against them? In a criminal context, evidence that is the product of an unreasonable search or seizure cannot be used against a defendant. But this principle – an "evidence exclusion rule" – is also a key means to methodologies of a critical privacy value in all data contexts (financial, medical, educational, etc.): *Interpretive restraint*. Or, to put it in less legal terms: "Evidence exclusion" means a determination that, despite the known availability of potential interpretive evidence (which, in the digital era, includes everything from emails to social media posts to search engine histories), an organization, government agency, or community has chosen (for ethical, practical, legal, and/or political reasons) to exclude and ignore this data when making a judgment, analysis, or prediction about this particular person or group of people.

C. Privacy Value #3: Forgiveness and Grace:

The notion of a statute of limitations on shunning, shaming, or cancellation should also be explored. In our view, we must create a clear and unifying version of the right to be forgotten in the United States and perhaps the world.[4] We have all read stories about someone making a poor decision or social media gaffe. In one recent example, an 18-year-old cheerleader was "canceled" and forced to withdraw from college due to her unfortunate use of a racial epithet years before when she was 15 years old. For the rest of her life, anytime anyone performs a web search for her name, the first page of search results will recount a dumb and embarrassing thing she did as a child.

Forgiveness and grace need to be a cornerstone of our social ethos. As previously discussed, our collective data is out there and is there to stay. We have seen the effects of cancellation already, but as more of us live more of our lives online, our society will face this problem on a much larger scale.

Generation Z is the first generation to have social media as a part of their whole lives. Social media companies alike have been tracking, collecting, and utilizing personalized data they started collecting from users when they were children. These individuals are now applying for colleges and jobs and entering adulthood. This means the call for forgiveness and grace must be even stronger. We have a generation who has spent much of their lives on social media without a second thought. The information they shared, even casually and thoughtlessly, can come back to haunt. As a society, we must ask ourselves if we want to shun, shame, and/or cancel individuals when they might have changed, but their data has followed them forever.

As a society, we must build a collective post-digital ethos around privacy. These new standards and community norms must include such quaint but critical notions as understanding, forgiveness, and grace. That should start with some form of right to be forgotten, where an individual can ask search engines to delete stories about their past, so long as they do not concern criminal behavior or a matter of ongoing public interest.

---

3 Hawthorne, N. (1850). The Scarlet Letter. Boston, MA: Ticknor and Fields.

4 Everything you need to know about the "Right to be forgotten," GDPR-EU, https://gdpr.eu/right-to-be-forgotten/?cn-reloaded=1.

03

RE-PROGRAMMING AND REFORMING ALGORITHMIC DECISION MAKING

An increasing number of the decisions once made by humans are now made by algorithms, which are automated processes used by computers. Algorithms are also increasingly used by prospective employers, landlords, businesses, health providers, police, schools, and government agencies to determine whether individuals are worthy of jobs, housing, health care, education, parole, or probation. The laws we have put in place to guarantee civil rights are premised on human actors making decisions that affect people. In post-digital America, regulations and Constitutional precedent are still in the early stages of determining how to enforce civil rights laws on computer processes.

People too often treat algorithms like calculators and their decisions like solutions to math problems. Algorithms are step-by-step sequences of instructions we direct computers to use. When a machine is tasked with something objective, like adding two numbers, we can reliably trust and use the answers it produces. However, in the post-digital world, machines are often tasked with complex decision-making that cannot be reduced to binary code (i.e. algorithmic error rates); they fail to understand the "intersectional" nature of our fellow citizen's identities (i.e. algorithmic discrimination); unjustly use prior data to re-reinforce old stereotypes and systemic disadvantages; or fail to holistically judge and evaluate an individual and their circumstances (as employees, defendants, patients, and suspects).

While numerous organizations and influencers now provide "solutions" to algorithmic decision-making – including the White House's AI Bill of Rights – these solutions often simply "re-program" the ideological divisions already blocking meaningful reform. More troubling, by focusing narrowly on particular types of discrimination (i.e. Race and gender), current "solutions" further marginalize other protected categories of identity. Even worse, they inadvertently program in a "separate but equal" digital society that legal scholar Margaret Hu has rightfully called "Algorithmic Jim Crow."

We utilize the following core methodological principles as a bipartisan and interdisciplinary solution that works with any AI decision-making process.

A. AI Decision-Making Principle #1: Transparency and Consent

Those who use algorithms to make decisions should be required to notify those affected by those decisions that 1) algorithms are judging them and 2) how those algorithmic judgments are rendered. Transparency and consent also trigger these sub-concerns, values, and questions about data sets, machine learning, and burden shifting.

• ***Data Sets****:* What data sets are used to train algorithms? Are they trained on large, diverse, and representative data sets? Have those datasets been evaluated thoughtfully and measured for historical biases? Are they regularly audited to ensure they are making decisions fairly and in a manner free from bias? Do the parties using them understand why and how they make decisions? Can their decisions be explained and replicated? When technology is not designed with thoughtfulness and intentionality about how it can affect different groups of people, the results can be disastrous.

• ***Machine Learning****:* The processes of how an algorithm is designed and learns, what techniques it uses for training and validation, how it makes decisions, how it is audited and evaluated, and whether it is working as intended must be fully transparent and explainable. Those affected by its findings, the legal community at large, government regulators, etc., must be able to "see" and understand how and why the algorithm makes decisions so those decisions can be evaluated for legality and fairness.

• ***Burden Shifting****:* Our legal system is designed to evaluate how humans make decisions regarding other humans. The individual who believes an algorithm judged them unfairly should not have to bear the burden of proof, which could be quite costly. Instead, the parties using algorithms to evaluate individuals and determine their qualifications should be forced to explain how the algorithms work and unpack and justify their decisions. For example, suppose an algorithm is used during job interviews for a position and prefers one candidate over another. In that case, the company using the algorithm must have an affirmative duty to explain how it works and why it formed that opinion.

Once these transparency questions are answered, digital ethics and law demand a more holistic approach to how AI renders decisions. This brings us to three more principles for meaningful AI reform: Error rates, bias and discrimination, and human oversight when technology is making life-altering decisions about people.

B. AI Decision-Making Principle #2: Error Rates

A quick sampling of recent algorithmic injustice instances highlights this growing problem. In terms of algorithmic error rates, Bank of America was recently fined 225 million dollars and ordered to offer redress, which could amount to millions more for a fraud detection algorithm that lacked human oversight.[5] Predictive policing algorithms are proving to be inherently flawed and rely on historical crime data which replicate discriminatory police practices and reinforces over-policing of communities of color and is not guaranteed to predict a crime.[6] Individuals like Robert McDaniels found themselves on the Chicago Police Department heat list [even with no history of violence]. McDaniels became the subject of police harassment and constant surveillance, which led to him being shot twice.[7]

> *A quick sampling of recent algorithmic injustice instances highlights this growing problem*

C. AI Decision-Making Principle #3: Bias and Discrimination

A key reason algorithms exhibit error rates, and biases is that they are trained on flawed datasets. Algorithms are asked to predict the future based on the past. It should be no surprise that many racial, gender, and other biases are built into historical datasets. These biased datasets are also called coding bias and, if left uncorrected, reinforce decades of marginalization and discrimination.

The U.S. Department of Justice recently settled a case against Meta Platforms (formerly Facebook) for allowing features in its advertising business to discriminate against groups protected by federal civil rights laws.[8] The Correctional Offender Management Profiling for Alternative Sanctions ("COMPAS") algorithm predicts the likelihood of a criminal defendant's recidivism.[9] COMPAS predicted twice as many false positives for recidivism for black offenders (45 percent) than for white offenders (23 percent). Facial recognition programs are used to make employment decisions and identify criminal suspects, despite often struggling to "see" darker skin.[10] Algorithms used to find more "high quality" or "successful" job candidates can look for "more of the same," replicating a company's biased historical hiring practices and overlooking qualified candidates who belong to historically marginalized groups.[11]

In situations with a higher-than-normal risk that an algorithmic assessment might be incorrect and/or biased, consideration should be given to restricting the use of algorithms and insisting on human decision makers. For example, a facial recognition program that judges a person's personality based on their facial expressions might judge someone who is non-neurotypical harshly. In such cases, human assessors who can consider such factors would likely be more appropriate.

D. AI Decision-Making Principle #4: Human Oversight When Machines Judge People

In our view, algorithmic decision-making has yet to advance to the level where it should be completely autonomous and unaccountable. The explosion of systems that determine everything from who gets a job to who gains access to housing or medical care or is granted parole or probation is very concerning. Algorithms offer us increased convenience and efficiency: They can enable companies to review massive amounts of data far more quickly than humans could. But we should ask ourselves, should we ever allow a machine, no matter how alike in human consciousness, to be able to be free of human oversight? It is imperative that diverse groups of humans, with careful deliberations, thoughtfulness, and intentionality, ensure that algorithmic judgments are made correctly, in a manner free from bias, and with due respect to privacy when life-changing decisions – economic, medical, educational, legal, professional, and/or financial – about our fellow humans are at stake.

5 Jenna McNamee, CFPB fines Bank of America for faulty unemployment benefits fraud detection, Jul 18, 2022, https://www.insiderintelligence.com/content/cfpb-bank-of-america-faulty-fraud-detection.

6 Pitfalls of Predictive Policing: An Ethical Analysis Viterbi Conversations in Ethics Volume 6 Issue 1 17 February 2022; see also Predictive Policing Explained Brennen Center for Justice, Tim Lau, 1 April 2020.

7 Heat Listed - Chicago PD automated Policing Got a Man Shot Twice , The Verge - Matt Stroud, 24 May 2021

8 Jenna McNamee, CFPB fines Bank of America for faulty unemployment benefit fraud detection, Jul 18, 2022, https://www.insiderintelligence.com/content/cfpb-bank-of-america-faulty-fraud-detection.

9 Terence Shin, Real-life Examples of Discriminating Artificial Intelligence, Towards Data Science, Jun 4, 2020, https://towardsdatascience.com/real-life-examples-of-discriminating-artificial-intelligence-cae395a90070.

10 Reuters, Thomas, Black and Asian faces misidentified more often by facial recognition software, Dec 20, 2019, https://www.cbc.ca/news/science/facial-recognition-race-1.5403899.

11 Bogen, Miranda, All the Ways Hiring Algorithms Can Introduce Bias, May 6, 2019, https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias.

04
CONCLUSION

The era of Big Data requires a rethinking of legal and ethical principles. If we are to have a right to privacy, the parameters of it must be based on how data can be used. If we are to be judged by machines, the algorithms that make them and the judgments they make will need to be carefully monitored. And the best safeguard of our digital rights will ultimately be engaging diverse teams that thoughtfully consider how their fellow humans are affected as they establish guardrails around emerging technology.

This article highlights only a few of the ways that blockchain networks and Web3 applications may open new ways to approach antitrust analysis for zero-price goods

REGULATING MACHINE LEARNING **BY DESIGN**

BY
MARCO ALMADA

Researcher, EUI Department of Law. Email: Marco.Almada@eui.eu. The author's work on this paper was partially funded by a Fundación Carolina doctoral grant. The author would also like to thank Felipe Roquete for comments in related work, which helped shape this paper.

01 INTRODUCTION

Machine learning ("ML") is a new frontier for regulation. Little more than a decade ago, ML-based technologies were a niche concern even in the field of technology regulation, as the field of artificial intelligence ("AI") lingered on at a low point of investment. Move forward a decade, and the situation could not be more different. The risks and opportunities associated with AI technologies have become a problem not only in domains typically associated with digital technologies, such as privacy or intellectual property, but they now permeate the most varied dimensions of social life. For example, the early months of 2023 have seen intense debates about the impact of large language models such as ChatGPT. Will these systems facilitate the spread of online misinformation? Did the creators of these systems breach intellectual property rights as they assembled the massive datasets powering them? Are the capabilities of these systems enough to transform the work of lawyers — or even replace them altogether? To answer these and other questions, regulators need to engage with the technical aspects of ML technologies.

When it comes to the governance of the technical side of ML and other digital technologies, regulators worldwide are increasingly reliant on Regulation by Design ("RbD"). In general lines, RbD operates by incorporating legal requirements into software design:[2] the law specifies requirements that a computer system must meet, and the designers of any computer system subject to that law must choose the technical arrangements that ensure the system always meets those requirements. Provisions laying down RbD requirements are common in data protection law,[3] and, since ML systems require huge data sets for their training, these requirements encompass most applications of ML directed at natural persons. However, by-design provisions are not restricted to the field of personal data. In the EU, the Digital Markets Act ("DMA") encourages those undertakings designated as gatekeepers to adopt by-design measures to foster fairness and market contestability,[4] and the proposed AI regulation ("AI Act")[5] imposes various design measures to be adopted by AI systems deemed to pose a high risk to fundamental rights and other public interests. Whenever such provisions are adopted, the technical decisions made in constructing and deploying ML systems become directly relevant for regulatory compliance. Therefore, the use of RbD amounts to a technology-sensitive approach to regulation.

This paper argues that RbD offers a powerful tool for regulators, but one with a narrow scope of application. The following section characterizes RbD as a modality of meta-regulation, in which the designers of ML systems are required to give effect to legal requirements through code. After this high-level overview, Section III argues that RbD provisions can lay down standards and offer guidance to designers as they seek to implement legal requirements in ML. But every regulatory tool comes with its drawbacks, and Section IV shows that the limits of RbD render it unsuitable to address many of the pressing challenges driving AI regulation, even being harmful to some of these regulatory goals. However, one should not throw the baby out with the bathwater, so the paper concludes with a discussion of the proper scope for RbD provisions in the regulation of ML systems.

02

REGULATION BY DESIGN AS META-REGULATION

A key aspect of RbD approaches is that they afford considerable discretion to software designers.[6] Consider Article 25 GDPR, titled *data protection by design and by default*. According to Article 25(1) GDPR, any data processing must be accompanied by technical and organizational measures that ensure personal data is processed in conformity with the data protection principles enshrined in the regulation. While the principles are determined by legislation, their interpretation in a specific data processing context, the choice of which measures will be adopted, and the implementation of these measures all remain in the hands of the actors who determine the purposes and means of data processing.[7]

2 Pieter Van Cleynenbreugel, *EU By-Design Regulation in the Algorithmic Society: A Promising Way Forward or Constitutional Nightmare in the Making?*, *in* Constitutional Challenges in the Algorithmic Society 202, 202 (Amnon Reichman et al. eds., 2021).

3 See, e.g. Article 25 of the EU General Data Protection Regulation ("GDPR": *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ 2016 L 119/1), Articles 46, § 2, and 49 of the Brazilian Data Protection Law (*Lei 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais (LGPD))*, (2018)), and Article 10 of the Council of Europe's modernized Convention 108 on personal data.

4 See Recital 65 of Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance), OJ 2022 L 265/1 (2022).

5 European Commission, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (2021) arts. 10-15.

6 That is, the ones responsible for the various activities involved in the creation of a system, such as the definition of overall goals and requirements for that system, the stipulation of software architectures, and the actual programming.

7 In GDPR parlance, the "controllers" of the personal data being processed: Lina Jasmontaite et al., *Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR*, 4 Eur. Data Prot. L. Rev. 168 (2018).

The effectiveness of these principles will, accordingly, hinge on the technical (and, in this case, organizational) decisions made by designers as they pursue compliance with the legal requirement.

Designer discretion is, at least in part, a consequence of the broad scope of some RbD provisions. In the GDPR example above, the category "data protection principles" covers various legal interests. Since it is unlikely that any particular measure will succeed in protecting all such interests in all circumstances, the regulation does not specify any action that must be adopted in all cases.[8] But some forms of RbD are much narrower. For example, Article 12(1) of the proposed EU AI Act stipulates that the design and development of high-risk AI systems must create the conditions for automated logging of events during the system's operation. The remaining paragraphs of Article 12 AI present various requirements that must be met by any acceptable solution for logging: any high-risk AI system that does not provide the capabilities listed there is not in compliance with the AI Act. Still, designers are free to adopt any technical approach to produce the event logs as long as they meet these requirements.

This is not to say RbD is always a "hollow" approach to regulation,[9] devoid of any enforceable substance. Courts and data protection authorities can rely on those criteria when adjudicating particular cases[10] and thus subject designer decisions to external scrutiny. Furthermore, technical decisions can be compared to external benchmarks, such as those provided by administrative guidelines,[11] technical standards, and certification procedures. By-design provisions might not be sufficient to eliminate all designer discretion, but they offer constraints to its exercise. In doing so, they set up a meta-regulatory regime, in which the state delegates to designers the specification of norms in the contexts in which a system is meant to operate while establishing mechanisms to supervise the use of delegated power.

"A key aspect of RbD approaches is that they afford considerable discretion to software designers

The meta-regulatory character of RbD follows from its targeting of design decisions. When designers make technical decisions that respond to a legal requirement, they effectively hardcode an interpretation of that requirement as a rule — or, more likely, a set of rules — in a computer system. Some of these rules bind the future behavior of system users: for example, an eBook might be accompanied by Digital Rights Management mechanisms that prevent unauthorized users from reading its contents.[12] But the encoded rules might also affect third parties without direct interaction with the system, as is the case when a public sector authority relies on an ML system to assess fraud risks.[13] In both cases, the impact of legal requirements on the outcome will be mediated by how these requirements influence system outputs. So, by regulating system design practices, RbD governs the role of designers in giving force to regulation.

Scholarship on meta-regulation points out that such approaches tend to emerge when the state lacks direct regulatory capability.[14] In the case of ML technologies, the capability gap stems from several factors. First, and perhaps foremost, the sheer variety of potential ML applications prevents regulators from addressing in-depth the risks associated with every single use context.[15] Second, the technological complexity of ML systems means that their regulation requires considerable resources and technical

8 It does, however, provide criteria that must be considered when determining the measures to be adopted, such as the risks posed by the processing, the state of the art, and the cost of implementing such measures.

9 Aurelia Tamò-Larrieux, Designing for Privacy and its Legal Framework: Data Protection by Design and Default for the Internet of Things 209 (2018).

10 For an overview of administrative cases on Article 25 GDPR, see Marco Almada, Juliano Maranhão & Giovanni Sartor, *Article 25. Data protection by design and by default*, *in* General Data Protection Regulation. Article-by-article commentary (Indra Spiecker gen. Döhmann et al. eds., 2023).

11 See, e.g. EDPB, *Guidelines 4/2019 on Article 25 on Data Protection by Design and by Default*, (2020).

12 On computer systems as a source of rules for uses, see Laurence Diver, Digisprudence: code as law rebooted (2021).

13 See, e.g. David Hadwick & Shimeng Lan, *Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany*, 13 World Tax Journal (2021).

14 Peter Grabosky, *Meta-regulation*, *in* Regulatory Theory: Foundations and applications 149, 155 (Peter Drahos ed., 1st ed. 2017).

15 See, e.g. the claims that AI is a general-purpose technology: Manuel Trajtenberg, *Artificial Intelligence as the Next GPT: A Political-Economy Perspective*, *in* The Economics of Artificial Intelligence: An Agenda (Ajay Agrawal, Joshua Gans, & Avi Goldfarb eds., 2019).

know-how, which are not always available to regulators.[16] Using RbD as a meta-regulatory strategy thus allows public regulators to tap into the resources and domain-specific knowledge available to the actors that design ML systems and still rein in their regulatory power. The following section examines some of the contributions RbD approaches can give to ML regulation.

03 REGULATORY INTERVENTIONS IN MACHINE LEARNING DESIGN

RbD is, by necessity, a context-sensitive practice. Certain technical solutions might be adequate for some problems and not for others: for example, approaches that minimize the use of personal data are widely promoted as conducive to the protection of privacy, but they might create obstacles to the kind of statistical analyses needed to detect algorithmic discrimination.[17] Still, all ML systems share a few traits: they rely on large data sets for training and operation, are opaque to untrained observers and technical experts, and rely on a somewhat narrow set of technical approaches and technological infrastructure.[18] Therefore, some kinds of design requirements will likely benefit various regulatory goals.

One of the primary objectives of ML regulation is to avoid, or at least mitigate, the harms produced by algorithms. In recent years, various such harms have been identified. Some of these harms have been detected early in software adoption, as is the case of the recent news on AI-powered search delivering wrongful results.[19] In other cases, the harmful impact of algorithmic systems can be more difficult to detect: in the Dutch Childcare Benefits scandal, a risk assessment system produced outputs that led to discriminatory enforcement of anti-fraud mechanisms against minoritized groups.[20] RbD approaches can contribute to avoiding such incidents by forcing designers to address *ex ante* known risks stemming from their systems.

The contribution of RbD to the quality of ML outputs comes from its binding force. If designers are mandated to achieve certain goals, or even required to use certain technical approaches, the legal obligation can lead them to use techniques that would otherwise be seen as too complex or expensive. For example, statistical metrics such as conditional demographic disparity can be used to detect whether an ML system discriminates against protected groups,[21] and accuracy benchmarks can be used to evaluate whether the system actually delivers the promised results.[22] Compliance with well-designed RbD rules can lead to adopting ML systems that produce higher-quality outputs.

Output quality is not the only goal that design practices can foster. In fact, it has been argued that promoting increases in quality metrics, such as accuracy and completeness of training data sets, often happens at the expense of other human values, such as privacy and dignity.[23] To protect these values, RbD approaches can mandate the

16 This issue is particularly salient in developing countries: Cecil Abungu, *Algorithmic Decision-Making and Discrimination in Developing Countries*, 13 Case W. Res. J.L. Tech. & Internet 39 (2022). However, even developed countries can struggle to cultivate and retain technical expertise, especially when research in ML is concentrated in a few corporate actors: Daniel Zhang et al., *The AI Index 2021 Annual Report*, (2021).

17 Marvin van Bekkum & Frederik Zuiderveen Borgesius, *Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?*, 48 Comput. Law Secur. Rev 105770 (2023).

18 For an overview of the technical convergence in AI for non-technical audiences, see Matthias Gallé, *Foundation Models in AI: what impact for policies and law?* (2022), https://www.eui.eu/Documents/Research/Clusters/Techcluster-memos/20220530-Foundation-Models-in-AI-memo.pdf (last visited Sep 29, 2022).

19 Chloe Xiang, *Bing's ChatGPT-Powered Search Has a Misinformation Problem*, Vice (2023), https://www.vice.com/en/article/3ad3ey/bings-chatgpt-powered-search-has-a-misinformation-problem (last visited Feb 20, 2023).

20 Amnesty International, *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* (2021).

21 Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI*, 41 Comput. Law Secur. Rev 105567 (2021).

22 On the various ways they may fail to do so, see Inioluwa Deborah Raji et al., *The Fallacy of AI Functionality*, *in* 2022 ACM Conference on Fairness, Accountability, and Transparency 959 (2022).

23 Paul Ohm, *Throttling machine learning*, *in* Life and the Law in the Era of Data-Driven Agency 214 (Mireille Hildebrandt & Kieron O'Hara eds., 2020).

use of certain technical practices that foster them even if at the expense of some accuracy. For example, many of the most potent ML systems, such as ChatGPT, are unfathomably big and complex. Still, for some tasks, it might be possible to achieve similar results — or, at least, good enough performance — with simpler techniques that are more amenable to human scrutiny.[24] In that case, a requirement imposing the use of interpretable ML techniques, or the adoption of explainable AI techniques that reduce the complexity of the larger systems,[25] can bolster human oversight at a reduced cost to the system's fitness to purpose. Requirements such as these supply designers with solutions to value conflicts they would otherwise need to solve, thus easing compliance with general regulatory requirements.

By-design regulation is not limited to promoting a specific set of values, but it can direct designers towards values they would otherwise disregard or treat as secondary. For example, much of contemporary software development is guided by so-called agile methodologies, in which the continuing evolution of software systems is prioritized over the comprehensiveness of documentation.[26] Requirements such as the AI Act's demand for up-to-date documentation of various aspects of high-risk AI systems[27] counteract this tendency by forcing designers to evaluate whether their documentation-light approach is enough to meet legal requirements and, if not, drawing up additional documents.

> *By-design regulation is not limited to promoting a specific set of values, but it can direct designers towards values they would otherwise disregard or treat as secondary*

The examples above show that RbD plays a triple role in shaping the regulatory activities of designers: it supplies those actors with quality standards, specifies solutions to potential value conflicts, and shifts the weight they must confer to the values considered. In doing so, RbD approaches can rely on practices that apply to all sorts of digital systems, such as those governing data protection by design. But, to address the specific challenges of ML systems, regulators will need to look into the technical arrangements that power those systems. Otherwise, they might fail to spot how the mechanisms through which ML can produce harmful or otherwise undesirable effects, as well as potential technical fixes for these problems.

04

THE LIMITS OF MACHINE LEARNING REGULATION BY DESIGN

Despite its relative novelty, RbD has been extensively critiqued in technology regulation scholarship. Among other important points, it has been argued that the delegation of regulatory power to designers can suffer from legitimacy issues,[28] that software code cannot properly capture the ambiguity and value judgments that are inherently involved in legal interpretation,[29] and that rules hardcoded in software are difficult to change if the initial implementation was wrong or if the regulatory requirement changes after initial implementation.[30] These issues constrain the efficiency of by-design approaches not just in ML systems, but in any kind of digital system.[31] Until they are solved, if they are at all solvable, it follows that RbD approaches are at their

24 See, e.g. section 6 of Madalina Busuioc, Deirdre Curtin & Marco Almada, *Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act*, Eur. Law Open FirstView (2022).

25 Adrien Bibal et al., *Legal requirements on explainability in machine learning*, 29 Artif Intell Law 149 (2021).

26 Thomas Hoeren & Stefan Pinelli, *Agile programming – Introduction and current legal challenges*, 34 Comput. Law Secur. Rev 1131 (2018).

27 Article 11 AI Act.

28 Diver, *supra* note 12.

29 Mireille Hildebrandt, *The adaptive nature of text-driven law*, 1 CRCL (2022).

30 Lyria Bennett Moses & Monika Zalnieriute, *Law and Technology in the Dimension of Time*, *in* Time, Law, and Change: An Interdisciplinary Study 303, 317 (Sofia Ranchordás & Yaniv Roznai eds., 2020).

31 Indeed, some of these critiques are older than the current wave of ML technologies prompting calls for regulation: Bert-Jaap Koops & Ronald Leenes, *Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law*, 28 Int. Rev. Law, Comput. Technol. 159 (2014).

most effective when they are deployed to implement requirements that are socially perceived as legitimate, which can be expressed in terms of "if–then" statements that can be implemented in computer code, and that are amenable to change in the future.

These conditions rarely hold in real-world ML applications. When it comes to legitimacy, the use of ML systems has a mixed track record: recommender systems are part of everyday experiences in social media, but the automation of sensitive tasks such as grading high school leaving exams has been met with protests and other forms of political contestation.[32] Furthermore, these systems produce their outputs by applying statistical models to their input model, and this reliance on statistics introduces a degree of uncertainty into any decisions relying on ML techniques. As a result, a change to regulation that seems relatively small for a human observer might require considerable rework if it is to be implemented in a ML system.[33]

> ***Despite its relative novelty, RbD has been extensively critiqued in technology regulation scholarship***

Beyond these conceptual challenges, the economics of ML introduce additional constraints to RbD. The construction of ML systems requires large amounts of data and computing resources,[34] which are not accessible to most private actors — and even to many public actors. To access those capabilities, most designers rely on ML-as-a-service solutions, hiring timeshares in platforms offered by large providers such as Amazon or Google. Such arrangements not only incentivize market concentration, but they create a regulatory conundrum: the actors who use ML-as-a-service lack the power to effect change into the large systems they rely on, but the providers of the general-purpose systems offered as a service lack the context-specific knowledge they need to address the risks associated with each possible application of their tools. Design requirements for ML systems walk a thin line between imposing impossible obligations to designers and creating obligations that are too general to have any binding content.

In light of the issues presented above, regulators should be wary of turning RbD into a pillar of their regulatory strategies. If designers cannot implement measures that contribute to the overall goals of the strategies but are nonetheless obliged to do *something* to fulfil a legal requirement, they might be pushed towards a Procrustean solution: pursing the regulatory goals only to the extent said goals can be expressed in terms of design requirements. For example, it has been argued that the EU AI Act's approach of protecting fundamental rights through product safety standards overlooks systemic violations of rights and dignity harms that cannot be described in terms of quantified risk.[35] Applying RbD approaches in contexts they are not suited to handle may lead to the construction of ML systems that produce undesirable side-effects or even undermine the goals that drive RbD in the first place.

32 See, among others, Geoffrey Mead & Barbara Barbosa Neves, *Contested delegation: Understanding critical public responses to algorithmic decision-making in the UK and Australia*, Sociol. Rev. 00380261221105380 (2022).

33 For a broad comparison between rules in code and roles in machine-learning systems, see Reuben Binns, *Analogies and Disanalogies Between Machine-Driven and Human-Driven Legal Judgement*, 1 CRCL (2022).

34 Andrew Lohn & Micah Musser, *AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?* (2022).

35 See, among others, Marco Almada & Nicolas Petit, *The EU AI Act: Between Product Safety and Fundamental Rights*, (2022), https://papers.ssrn.com/abstract=4308072 (last visited Dec 21, 2022); Nathalie Smuha et al., *A Response to the European Commission's Proposal for an Artificial Intelligent Act*, 64 (2021).

Still, RbD provisions can be part of a well-calculated delegation strategy. Some technical goals and solutions are pretty much universal, and so their stipulation by design would face little opposition. Consider the human oversight requirements from Article 14 AI Act, which respond to widespread calls for tools that support human control over high-risk ML systems. RbD is also unlike to raise further issues if it is applied to a well-defined problem, as is the case of the logging requirement discussed in Section II. Finally, RbD provisions can also be useful if they remove some of the barriers to the effectiveness of other RbD provisions outlined above. For example, a requirement that ML systems must be designed with modularity and long-term maintenance in mind would reduce the costs involved in adapting software to cope with changes in the regulations it must comply to. The effectiveness of any such measures must, evidently, be evaluated in light of the context in which a particular ML system is used and of the techniques available for their implementation. But, if seen as a supporting tool rather than a full-blown approach to regulation, RbD can help regulators in addressing the complexities of ML.

> ***In light of the issues presented above, regulators should be wary of turning RbD into a pillar of their regulatory strategies***

WHAT'S **NEXT**

For March 2023, we will feature a TechREG Chronicle focused on issues related to **Cybersecurity**.

ANNOUNCEMENTS

CPI TechREG CHRONICLES April 2023

For April 2023, we will feature a TechREG Chronicle focused on issues related to ***Ex Ante* Regulations**.

Contributions to the TechREG Chronicle are about 2,500 – 4,000 words long. They should be lightly cited and not be written as long law-review articles with many in-depth footnotes. As with all CPI publications, articles for the CPI TechREG Chronicle should be written clearly and with the reader always in mind.

Interested authors should send their contributions to Sam Sadden (**ssadden@competitionpolicyinternational.com**) with the subject line "TechREG Chronicle," a short bio and picture(s) of the author(s).

The CPI Editorial Team will evaluate all submissions and will publish the best papers. Authors can submit papers in any topic related to competition and regulation, however, priority will be given to articles addressing the abovementioned topics. Co-authors are always welcome.

ABOUT **US**

Since 2006, **Competition Policy International** ("CPI") has provided comprehensive resources and continuing education for the global antitrust and competition policy community. Created and managed by leaders in the competition policy community, CPI and CPI TV deliver timely commentary and analysis on antitrust and global competition policy matters through a variety of events, media, and applications.

As of October 2021, CPI forms part of **What's Next Media & Analytics Company** and has teamed up with **PYMNTS**, a global leader for data, news, and insights on innovation in payments and the platforms powering the connected economy.

This partnership will reinforce both CPI's and PYMNTS' coverage of technology regulation, as jurisdictions worldwide tackle the regulation of digital businesses across the connected economy, including questions pertaining to BigTech, FinTech, crypto, healthcare, social media, AI, privacy, and more.

Our partnership is timely. The antitrust world is evolving, and new, specific rules are being developed to regulate the so-called "digital economy." A new wave of regulation will increasingly displace traditional antitrust laws insofar as they apply to certain classes of businesses, including payments, online commerce, and the management of social media and search.

This insight is reflected in the launch of the **TechREG** Chronicle, which brings all these aspects together — combining the strengths and expertise of both CPI and PYMNTS.

Continue reading CPI as we expand the scope of analysis and discussions beyond antitrust-related issues to include Tech Reg news and information, and we are excited for you, our readers, to join us on this journey.

Scan to Stay Connected!

Scan here to subscribe to CPI's **FREE** daily newsletter.

CPI
SUBSCRIPTIONS

CPI reaches more than **35,000 readers** in over **150 countries** every day. Our online library houses over **23,000 papers**, articles and interviews.

Visit **competitionpolicyinternational.com** today to see our available plans and join CPI's global community of antitrust experts.

CPI COMPETITION POLICY® INTERNATIONAL