Non-Decisional Statement by the National AI Advisory Commission (NAIAC)
Working Group on Regulation and Executive Action

Rationales, Mechanisms, and Challenges to Regulating AI:

A Concise Guide and Explanation

Disclaimer: This non-decisional document has been drafted by Members of the National Artificial Intelligence Advisory Committee (NAIAC) for the purposes of explaining concepts and does not offer formal recommendations. The opinions discussed in this document do not represent the views of the full Committee, and should not be considered a recommendation by the NAIAC.

Calls for AI regulation have been wide-ranging. This concise guide provides an explanation of the varied rationales for potentially regulating AI,[1] the main types of regulatory interventions, and some of the distinct challenges that AI poses to effective regulation.[2]

I. Why Regulate AI?

The potential benefits of AI are considerable. As the first-year report by NAIAC noted, "AI is one of the most powerful and transformative technologies of our time" and has the potential to "address society's most pressing challenges."[3] But different applications of AI can pose vastly different types of risk, at different levels of severity and on different timescales, depending on the technology and context of deployment.[4] Addressing these risks requires reasoning about both

[1] Regulations are rules that executive branch departments and agencies issue, subject to rulemaking authority delegated by Congress. Legislation is statutory law passed by Congress. *E.g.*, United States Senate, "Laws and Regulations," n.d., https://www.senate.gov/reference/reference_index_subjects/Laws_and_Regulations_vrd.htm [https://perma.cc/FZ89-54K9]].

[2] NAIAC's statutory obligation is to advise the White House and National AI Initiative Office, but this document may be more broadly informative for parties considering AI regulation.

[3] NAIAC, "National Artificial Intelligence Advisory Committee (NAIAC) Year 1," May 2023, https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf [https://perma.cc/PG5V-9M63].

[4] "For example, an AI system used to recommend TV shows to consumers may pose little risk of harm whereas an AI system that screens job applications can have an enormous impact on a person's economic opportunity." *See, e.g.*, Christina Montgomery, Francesca Rossi, and Joshua New, "A Policymaker's Guide to Foundation Models," IBM Newsroom, May 1, 2023, https://perma.cc/3STC-3NRC. There is much discussion and disagreement about what AI use cases are "high-risk," particularly given the proposed EU AI Act sets different regulatory requirements for "high-risk" AI. For examples of this discussion, *see*, *e.g.,* "Regulatory Framework Proposal on Artificial Intelligence" (European Commission, n.d.), https://perma.cc/NYE2-GGCU; "The EU AI Act's Risk-Based Approach: High-Risk Systems and What They Mean for Users" (Holistic AI, November 2022), https://perma.cc/5BAA-ULCB; Khari Johnson, "The Fight to Define When AI Is 'High Risk," *Wired*, September 1, 2021, https://perma.cc/Z3CF-9GPU; Alex Engler, "The EU and U.S. Diverge on AI Regulation: A Transatlantic Comparison and Steps to Alignment" (Brookings Institution, April 25, 2023), https://perma.cc/27WK-2GQG; Gabriella Shea and Sabine Neschke, "Defining High-Risk, High-Reward AI, Bipartisan Policy Center" (Bipartisan Policy Center, April 6, 2023), https://perma.cc/3KCA-QAKE. The IEEE's AI Impact Use Cases Initiative is

potential benefits and harms along the entire technology lifecycle, including design, development, deployment and monitoring.[5] Regulation can play an important role in ensuring that the advancement of technology achieves outcomes that are beneficial to society by, for example, incentivizing certain behaviors or establishing and enforcing appropriate guardrails.[6] Designing an effective regulatory intervention requires balancing, and gaining clarity about, the distinct and marginal benefits and risks, relative to appropriate baselines as opposed to in the absolute.[7] In the debate about AI regulation, there are many conceptions of harm that are often commingled. We refrain from assessing their validity, interrelationships, relative importance, or immediacy, but some of the main concerns articulated include:

> Poor Performance. Like any technology, use of AI systems will result in a range of performance outcomes, including potentially unacceptable errors.[8] At the systems-level,

attempting to curate concrete AI use cases to support the EU's risk-level classifications. "The IEEE AI Impact Use Cases Initiative" (IEEE SA), accessed June 19, 2023, https://perma.cc/7VQN-MCMF. ISO and IEC also published, through its subcommittee on AI, a report delineating 142 AI use cases. *See, e.g.*, Antoinette Price, "IEC and ISO Publish over 130 Emerging AI Use Cases," *E-Tech* (blog), May 17, 2021, https://perma.cc/RBX5-ZLQ7.

[5] "The balance we establish in addressing these two divergent AI realities — fully harnessing its benefits while also effectively addressing its challenges and risks — will significantly impact our future. If navigated appropriately, the U.S. government can ensure that AI creates greater opportunities, providing economic and societal benefits for a broader cross section of the population. However, if navigated poorly, AI will further widen the opportunity gap, and trustworthy AI for all may become an unrealized aspiration." NAIAC, "National Artificial Intelligence Advisory Committee (NAIAC) Year 1," May 2023, 7, https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf [https://perma.cc/PG5V-9M63].

[6] *See, e.g.*, "M-21-06, Memorandum for the Heads of Executive Departments and Agencies" (Executive Office of the White House, Office of Management and Budget, November 17, 2020); Lawrence O. Gostin, "Legal and Public Policy Interventions to Advance the Population's Health" (National Library of Medicine, 2000), https://www.ncbi.nlm.nih.gov/books/NBK222835/; Erin Simpson and Adam Conner, "How To Regulate Tech: A Technology Policy Framework for Online Services" (Center for American Progress, November 16, 2021), https://www.americanprogress.org/article/how-to-regulate-tech-a-technology-policy-framework-for-online-services/.

[7] Conventionally, regulation addresses market failures and each of these harms can be conceived of as such. For instance, environmental costs can be conceived of as either (a) negative externalities of training large models or (b) resulting from transaction costs that make bargaining around the externality difficult. Balancing benefits and costs has been central to the U.S. regulatory system and affirmed across administrations. See Executive Order 12,291, Federal Regulation, 46 Fed. Reg. 13,193 (1981); Executive Order 12,866, Regulatory Planning and Review, 58 Fed. Reg. 51,735 (1993); Executive Order 13,563, Improving Regulation and Regulatory Review (2011), 76 Fed. Reg. 3821 (2011). The NAIAC has similarly also recommended that, "The Administration should require an approach that protects against these risks while allowing the benefits of values-based AI services to accrue to the public." NAIAC, "National Artificial Intelligence Advisory Committee (NAIAC) Year 1," May 2023, https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf [https://perma.cc/PG5V-9M63]. The idea of matching regulatory intervention to harms is a mainstay of regulatory analysis. *See, e.g.*, Stephen Breyer, *Regulation and Its Reform* (Harvard University Press, 1982).

[8] AI can err, or fail to perform its intended task, in a variety of ways. For example, AI can be inaccurate because of labeling errors in test sets, see Curtis Northcutt, Anish Athalye, and Jonas Mueller, "Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks," November 7, 2021, https://doi.org/10.48550/arXiv.2103.14749; Will Knight, "The Foundations of AI Are Riddled With Errors," *Wired*, March 31, 2021, https://perma.cc/EG2D-VB9Y. AI can also fail to function as intended where the AI has been assigned an impossible task and as a result of engineering, post-deployment, and communication failures. *See,* Inioluwa Deborah Raji et al., "The Fallacy of AI Functionality," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22 (New York, NY, USA: Association for Computing Machinery, 2022), 959–72, https://doi.org/10.1145/3531146.3533158. AI can also suffer from "label choice bias" where the AI's actual target or

AI errors can also be particularly difficult to uncover and manage[9] and AI can introduce greater uncertainty about performance outcomes.[10] This can cause harm, particularly in high-risk fields such as criminal justice,[11] housing,[12] finance,[13] and employment.[14]

Bias. AI systems can create, perpetuate, or exacerbate and scale bias against particular demographic groups, including through algorithmic discrimination.[15] Although AI systems (e.g., data, algorithms, and their interactions) can be easier to audit than manual systems under certain conditions,[16]AI systems can also pose unique bias risks that are not present in manual systems or human decision-making (e.g., biased misidentification by facial recognition technologies can introduce bias at scale).[17]

---

prediction is a suboptimal proxy variable for the actual target or prediction of interest. Ziad Obermeyer et al., "Algorithmic Bias Playbook" (https://perma.cc/T9JE-QAZH, June 2021), 2, https://perma.cc/T9JE-QAZH. For example, medical algorithms may prioritize healthy patients over sick patients because the algorithm uses health care costs as a proxy for illness. Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (2019): 447–53.

[9] At the systems level, failures in AI systems can be especially surprising and difficult to uncover. For example, with the addition of a small amount of targeted noise, model classifications can swing from one extreme to another. *See, e.g.*, Christian Szegedy et al., "Intriguing Properties of Neural Networks," *ArXiv Preprint ArXiv:1312.6199*, 2013; Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard, "Deepfool: A Simple and Accurate Method to Fool Deep Neural Networks," 2016, 2574–82. For more, see the literature on adversarial attacks. Kyle Wiggers, "Adversarial Attacks in Machine Learning: What They Are and How to Stop Them," *VentureBeat*, May 21, 2021, https://perma.cc/J9SG-VTF4; Samuel G Finlayson et al., "Adversarial Attacks on Medical Machine Learning," *Science* 363, no. 6433 (2019): 1287–89.

[10] Unlike rule-based systems, outputs from neural AI systems may be unpredictable in ways that affects system trustworthiness. There is also a line of research that leverages rules to steer the behavior of neural networks—*see, e.g.*, Sungyong Seo et al., "Controlling Neural Networks with Rule Representations," *Advances in Neural Information Processing Systems* 34 (2021): 11196–207.

[11] Logan Kugler, "AI Judges and Juries," *Communications of the ACM* 61, no. 12 (2018): 19–21.

[12] Wonyoung So et al., "Beyond Fairness: Reparative Algorithms to Address Historical Injustices of Housing Discrimination in the US," 2022, 988–1004.

[13] I Elizabeth Kumar, Keegan E Hines, and John P Dickerson, "Equalizing Credit Opportunity in Algorithms: Aligning Algorithmic Fairness Research with Us Fair Lending Regulation," 2022, 357–68.

[14] Manish Raghavan et al., "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices," 2020, 469–81.

[15] *See, e.g.*, Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, ed. Sorelle A. Friedler and Christo Wilson, vol. 81, Proceedings of Machine Learning Research (New York, NY, USA: PMLR, 2018), 77–91, http://proceedings.mlr.press/v81/buolamwini18a.html; Haoran Zhang et al., "Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings," in *Proceedings of the ACM Conference on Health, Inference, and Learning* (ACM CHIL '20: ACM Conference on Health, Inference, and Learning, Toronto Ontario Canada: ACM, 2020), 110–20, https://doi.org/10.1145/3368555.3384448.

[16] Jon Kleinberg et al., "Discrimination in the Age of Algorithms," *Journal of Legal Analysis* 10 (April 22, 2019): 113–74. Secrecy surrounding AI systems can also make them opaque and complicate an understanding of potential bias and due process. Danielle Keats Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Wash. L. Rev.* 89 (2014): 1.

[17] Stephanie Fontenot, "Study Outlines What Creates Racial Bias in Facial Recognition Technology," December 4, 2020, https://news.utdallas.edu/science-technology/racial-bias-facial-recognition-2020/; Brianna Rauenzahn, Jamison Chung, and Aaron Kaufman, "Facing Bias in Facial Recognition Technology," *The Regulatory Review*

Privacy. AI systems can erode privacy protections by ingesting and integrating massive volumes of data to train large models.[18] Such models can, for instance, inadvertently memorize training data, regurgitating sensitive information such as home addresses or social security numbers in response to prompts.[19]

Labor Displacement, Job Quality, and Workers Rights. AI systems can displace workers through the automation of tasks, reduce job quality (e.g., through deskilling), harm workers' rights and autonomy (e.g., through workplace surveillance), and / or violate labor and employment law.[20]

Environmental Costs. AI systems may have greenhouse gas and water footprints, rely on chips that use rare earth metals, or have other environmental impacts, and, as a result, harm the environment.[21]

(blog), March 20, 2021, https://www.theregreview.org/2021/03/20/saturday-seminar-facing-bias-in-facial-recognition-technology/; Clare Garvie, *The Perpetual Line-up: Unregulated Police Face Recognition in America* (Georgetown Law, Center on Privacy & Technology, 2016).

[18] AI poses a number of privacy concerns, see Cameron Kerry, "Protecting Privacy in an AI-Driven World" (Brookings Institution, February 10, 2020), https://perma.cc/WL78-PDAE. These concerns also arise from data persistence, data repurposing, and data spillover. Catherine Tucker, "Privacy, Algorithms, and Artificial Intelligence," in *The Economics of Artificial Intelligence: An Agenda* (National Bureau of Economic Research, 2019), 423–37, http://www.nber.org/chapters/c14011. Recent work has also demonstrated the capacity for diffusion models to leak training data. *See, e.g.*, Nicholas Carlini et al., "Extracting Training Data from Diffusion Models" (arXiv, January 30, 2023), https://doi.org/10.48550/arXiv.2301.13188. More broadly, ML models may leak information about their training data, including information about sensitive individual health records. *See* Reza Shokri et al., "Membership Inference Attacks Against Machine Learning Models," in *2017 IEEE Symposium on Security and Privacy (SP)*, 2017, 3–18, https://doi.org/10.1109/SP.2017.41.

[19] Nicholas Carlini et al., "The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks," in *28th USENIX Security Symposium (USENIX Security 19)*, 2019, 267–84; Ethan Perez et al., "Red Teaming Language Models with Language Models," *ArXiv Preprint ArXiv:2202.03286*, 2022.

[20] For a discussion on technology labor displacement, see Daron Acemoglu and Pascual Restrepo, "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives* 33, no. 2 (May 2019): 3–30, https://doi.org/10.1257/jep.33.2.3. OECD has an AI in Work, Innovation, Productivity and Skills (AI-WIPS) research program, where the OECD with support from Germany is analyzing and publishing reports on the impact of AI on the labor market, labor skills, and broader social policy. For an overview of the research program and an example report, see "OECD Programme on AI in Work, Innovation, Productivity and Skills," *OECD.AI Policy Observatory* (blog), n.d., https://perma.cc/HPC6-R5AB; Anna Milanez, "The Impact of AI on the Workplace: Evidence from OECD Case Studies of AI Implementation" (OECD, March 27, 2023), https://doi.org/10.1787/2247ce58-en. The Global Partnership on AI also has a Working Group on the Future of Work that is analyzing how the deployment of AI can affect workers, the workplace, and the broader workforce. For an overview of the working group and its reports, see "Working Group on the Future of Work," *GPAI* (blog), n.d., https://gpai.ai/projects/future-of-work/. For a case study of AI surveillance in the trucking industry, see Karen Levy, *Data Driven: Truckers, Technology, and the New Workplace Surveillance* (Princeton University Press, 2022).

[21] *See, e.g.*, Emma Strubell, Ananya Ganesh, and Andrew McCallum, "Energy and Policy Considerations for Modern Deep Learning Research," vol. 34, 2020, 13693–96; Nestor Maslej et al., "The AI Index 2023 Annual Report" (Stanford University: Institute for Human-Centered AI, April 2023), 73; Pengfei Li et al., "Making AI Less 'Thirsty': Uncovering and Addressing the Secret Water Footprint of AI Models," April 6, 2023, https://arxiv.org/pdf/2304.03271.pdf; Elsabet Jones and Baylee Easterday, "Artificial Intelligence's Environmental Costs and Promise," *Council on Foreign Relations* (blog), June 28, 2022, https://www.cfr.org/blog/artificial-intelligences-environmental-costs-and-promise.

Cybersecurity. AI systems may create or enable novel forms of cybersecurity risks, such as risks for identity theft or data breaches, and adversarial attacks.[22]

Geopolitical competition. AI systems can affect America's geopolitical competitive advantage and national security interests (e.g., through AI-enabled and autonomous weapons, "dual use" technologies, and supply chain risks).[23]

Democratic Erosion. AI systems could lead to erosion of trust and undermine democratic institutions, due, for instance, to AI-enhanced misinformation or surveillance.[24]

Existential Risk. AI systems that move towards artificial general intelligence but are not aligned with human values could increase long-term existential risk to humanity.[25]

Regulatory interventions are most effective when matched to address the underlying problem.[26] If the question is one of environmental costs, for instance, a typical intervention might be to *tax* energy-intensive computing (to incentivize parties to internalize the pollution cost). But if the concern is about inaccurate AI, fixing such errors could require *more* compute access to interrogate and assess systems, which a tax could undermine. Similarly, if the concern is about existential risk, an intervention might focus on *restricting* access generally to compute (to limit

[22] For more details on the cybersecurity risks of AI systems, see Anne Johnson and Emily Grumbling, eds., *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop* (Washington, D.C.: National Academies Press, 2019), https://doi.org/10.17226/25488. Examples include deepfakes and adversarial attacks. For more details on adversarial AI, see Micah Musser et al., "Adversarial Machine Learning and Cybersecurity: Risks, Challenges, and Legal Implications" (Center for Security and Emerging Technology, April 2023), https://doi.org/10.51593/2022CA003; Kyle Wiggers, "Adversarial Attacks in Machine Learning: What They Are and How to Stop Them," *VentureBeat*, May 21, 2021, https://perma.cc/J9SG-VTF4. As another resource, see the discussion of the "secure and resilient" characteristic within National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," January 2023, 15, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [https://perma.cc/LX34-FGZU].

[23] *See, e.g.*, National Security Commission on Artificial Intelligence, "Final Report," 2021.

[24] Rachelle Faust, "The Global Struggle Over AI Surveillance: Emerging Trends and Democratic Responses" (National Endowment for Democracy, June 7, 2022); Maria Pawalec, "Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions" (Digital Society, 2022), doi: 10.1007/s44206-022-00010-6.; Jackson Cote, "Deepfakes and Fake News Pose a Growing Threat to Democracy, Experts Warn," *Northeastern Global News*, April 1, 2022, https://perma.cc/THQ8-Z53C. For a discussion of the implications of AI deepfake technologies on democracy, *see, e.g.*, Danielle Citron and Robert Chesney, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *California Law Review* 107, no. 6 (December 1, 2019): 1753. For a technical exploration of how to protect world leaders and democracy from deepfake imposters, see Matyáš Boháček and Hany Farid, "Protecting World Leaders against Deep Fakes Using Facial, Gestural, and Vocal Mannerisms," *Proceedings of the National Academy of Sciences* 119, no. 48 (November 29, 2022): e2216035119, https://doi.org/10.1073/pnas.2216035119.

[25] "The possibility of an intelligence explosion is often cited as the main source of risk to humanity from AI because it would give us so little time to solve the control problem." Stuart J. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* ([London] New York NY: Penguin Books, 2020).

[26] Stephen Breyer, *Regulation and Its Reform* (Harvard University Press, 1982). As a corollary, a dominant perspective – adopted in NAIAC's recommendation endorsing the NIST AI RMF – is that regulatory interventions should also be tied to level of risk. NAIAC, "National Artificial Intelligence Advisory Committee (NAIAC) Year 1," May 2023, https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf [https://perma.cc/PG5V-9M63].

the development of, for example, "foundation models"[27]), while concerns about national competitiveness might call for *expanded* access domestically.

The current debate can be confusing because proponents of regulation are not always clear or precise about which harm their proposed regulatory mechanism purports to address.[28] The relative importance of harms – and magnitude of harms relative to baseline systems – can be fiercely contested. Ideally, policymakers should be able to make evidence-informed decisions about the relative gravity of distinct harms[29], which could also involve a consideration of how harms are forecasted[30] and mitigation strategies are operationalized.[31]

II. Types of Regulatory Interventions

"Regulation" comes in many forms.[32] To provide some clarity, we spell out some of the most commonly contemplated regulatory interventions, which may be considered against the status quo of common law liability and existing laws.[33] We identify these interventions below, without addressing the separate question of which types of AI systems each intervention may regulate.

> Registration. Certain parties could be required to register with a governmental body, or to register details of major compute-intensive training runs with a government body.[34]

---

[27] Jeanne Casusi, "What Is a Foundation Model? An Explainer for Non-Experts," Institute for Human-Centered AI, Stanford University, May 10, 2023, https://perma.cc/8VSE-L6XA.

[28] Many executive orders direct agencies to explicitly provide rational weighing of costs and benefits in regulation, *see infra* note 6. Of course, Congress is not required to provide explicit reasoning or discuss tradeoffs in legislation.

[29] For example OMB M-21-06 discusses the need to tailor regulation and consider the costs and benefits: "While narrowly tailored and evidence based regulations that address specific and identifiable risks could provide an enabling environment for U.S. companies to maintain global competitiveness, agencies must avoid a precautionary approach that holds AI systems to an impossibly high standard such that society cannot enjoy their benefits and that could undermine America's position as the global leader in AI innovation. Where AI entails risk, agencies should consider the potential benefits and costs of employing AI, as compared to the systems AI has been designed to complement or replace."

[30] For more on forecasting and AI harms, *see, e.g.*, Miles Brundage et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," *ArXiv Preprint ArXiv:1802.07228*, 2018.

[31] One example of a challenge in operationalizing harm mitigation strategies lies in collecting demographic data. *See, e.g.*, McKane Andrus et al., "What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21 (New York, NY, USA: Association for Computing Machinery, 2021), 249–60, https://doi.org/10.1145/3442188.3445888.

[32] *See supra note* 2. *See infra note* 6 for a related discussion of regulation.

[33] The common law is a kind of regulatory system as well. Tort liability, for instance, may adapt to recognize harms of AI systems. But AI harms may not fit well within the tort system, which relies on private parties and imposes liability ex post. *See, e.g.*, Andrew D Selbst, "Negligence and AI's Human Users," *BUL Rev.* 100 (2020): 1315. "By inserting a layer of inscrutable, unintuitive, and statistically derived code in between a human decisionmaker and the consequences of her decisions, AI disrupts our typical understanding of responsibility for choices gone wrong."

[34] Registration is also used in other regulations. For example, the FDA has criteria that would trigger the requirement to register with the agency. See FDA, "Registration and Listing," U.S. Food and Drug Administration, February 2, 2023, https://www.fda.gov/industry/fda-basics-industry/registration-and-listing [https://perma.cc/3PGD-

Proposed amendments to the EU AI Act, for instance, would require the registration of high-risk AI systems and foundation models with the European Commission.[35]

Licensing. A licensing scheme allows only parties that have met certain qualifications (e.g., educational, professional, training qualifications) to engage in specific activities.[36] Recent proposals have called for governmental licensing of certain types of advanced AI systems, infrastructure, or "frontier models."[37] A variant of this is to have certain AI-production activities, such as training runs above a certain computational scale, to require a license.

Premarket Approval. Some regulatory schemes require governmental approval before a product enters the market. The Food and Drug Administration, for instance, has issued approvals for over 500 medical devices that involve AI.[38] Premarket approval regimes may also include post-market monitoring, which may lead to recalls or other corrective action.[39] Other regulatory schemes rely exclusively on post-market recalls instead of premarket approval.

Disclosure. Many regulations mandate disclosures by parties.[40] Executive Order 13,960 and the Advancing American AI Act (incorporated into FY 2023 NDAA), for instance,

---

PZY8]. The SEC also lists registration requirements under the Exchange Act: SEC, "Exchange Act Reporting and Registration," U.S. Securities and Exchange Commission, April 6, 2023, https://www.sec.gov/education/smallbusiness/goingpublic/exchangeactreporting [https://perma.cc/44LH-MYMM].

[35] European Parliament, "Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts" (2023), https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf [https://perma.cc/7GZC-935T].

[36] Licensing is typically distinguished from certification. A license is required to operate, while a certification may be optional. For an example, the Secure and Fair Enforcement for Mortgage Licensing Act of 2008 limits mortgage loan originators to those who are licensed. "Secure and Fair Enforcement for Mortgage Licensing (SAFE) Act Examination Procedures," *CFPB* (blog), October 1, 2012, https://www.consumerfinance.gov/compliance/supervision-examinations/secure-and-fair-enforcement-for-mortgage-licensing-safe-act-examination-procedures/.

[37] Luke Muelhauser, "12 Tentative Ideas for US AI Policy," *Open Philanthropy* (blog), April 17, 2023, https://www.openphilanthropy.org/research/12-tentative-ideas-for-us-ai-policy/.

[38] Eric Wu et al., "How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals," *Nature Medicine* 27, no. 4 (2021): 582–84.

[39] "Recalls, Corrections and Removals (Devices)," *U.S. Food & Drug Administration* (blog), September 29, 2020, https://www.fda.gov/medical-devices/postmarket-requirements-devices/recalls-corrections-and-removals-devices; "Postmarket Requirements (Devices)," *U.S. Food & Drug Administration* (blog), September 27, 2018, https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/postmarket-requirements-devices; "522 Postmarket Surveillance Studies Program," *U.S. Food & Drug Administration* (blog), October 6, 2022, https://www.fda.gov/medical-devices/postmarket-requirements-devices/522-postmarket-surveillance-studies-program; "Postmarket Surveillance Under Section 522 of the Federal Food, Drug, and Cosmetic Act: Guidance for Industry and Food and Drug Administration Staff" (U.S. Food & Drug Administration, October 7, 2022), https://www.fda.gov/media/81015/download.

[40] In the securities context, one example is the mandated disclosure of public companies about financial information. *See, e.g.*, SEC, "Filings and Forms," U.S. Securities and Exchange Commission, January 9, 2017, https://www.sec.gov/edgar [https://perma.cc/VKK2-ZTHZ]; Wex Definitions Team, "Regulation S-K," Legal

require federal agencies to disclose AI use case inventories.[41] Proposed amendments to the EU AI Act also require the publication of summaries of copyrighted data used in training for generative models.[42] Other common disclosure and transparency measures lie in mechanisms like model cards,[43] AI Factsheets,[44] and notice requirements about data collection or the use of AI.[45]

Rulemaking and Enforcement. Many regulatory bodies set standards of conduct through (notice-and-comment) rulemaking and enforcement.[46] The Federal Trade Commission, for instance, has set consumer privacy and security standards through such vehicles.

Auditing / Inspection. Many regulatory schemes involve audits or inspections of parties for compliance with standards. New York City, for instance, mandated bias audits of automated employment decision tools.[47] These interventions may require the development of international standards on AI,[48] the establishment of bodies who can accredit AI auditors,[49] and risk assessments.[50]

Information Institute, January 2022, https://www.law.cornell.edu/wex/regulation_s-k [https://perma.cc/7P6P-JWR7].

[41] "Executive Order 13960, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government," 85 FR 78939 § (2020), https://perma.cc/9N3C-GGU5; "James M. Inhofe National Defense Authorization Act for Fiscal Year 2023," Pub. L. No. 117–263 (2022).

[42] European Parliament, Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts.

[43] Margaret Mitchell et al., "Model Cards for Model Reporting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, 220–29.

[44] Matthew Arnold et al., "FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity," *IBM Journal of Research and Development* 63, no. 4/5 (2019): 6–1.

[45] The Illinois Artificial Intelligence Video Interview Act, for instance, requires employers to notify applicants of the use of AI to analyze their video interviews. Illinois State Legislature, "Artificial Intelligence Video Interview Act" (2020), https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68 [https://perma.cc/HD5W-UDQM].

[46] *See, e.g.*, rulemaking and enforcement in sectoral agencies such as the EPA and NHTSA.

[47]Laurie Combo et al., "A Local Law to Amend the Administrative Code of the City of New York, in Relation to Automated Employment Decision Tools," Pub. L. No. 2021/144 (2021), https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9 [https://perma.cc/6B98-79LR].

[48] At the same time, standards relating to AI already put forth by the ISO and the OECD among other international bodies have been important for policy harmonization efforts. *See, e.g.*, NIST's crosswalks to OECD recommendations and ISO standards. See NIST, "Crosswalks to the NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0)," AI Risk Management Framework, February 21, 2023, https://www.nist.gov/itl/ai-risk-management-framework/crosswalks-nist-artificial-intelligence-risk-management-framework [https://perma.cc/X6T5-FEWM].

[49] For example, see the Public Company Accounting Oversight Board (PCAOB) which oversees certain financial audits. "About," *PCAOB* (blog), n.d., https://pcaobus.org/about.

[50] "Recommended Practices for Safety and Health Programs: Hazard Identification and Assessment" (Occupational Safety and Health Administration, n.d.), https://www.osha.gov/safety-management/hazard-Identification; "Risk Assessment: What Is It, and What Does It Have to Do with My Food?," *U.S. Food & Drug Administration* (blog),

Ban and Export Controls. Regulations can also ban products. Numerous localities in the United States, for instance, have banned the use of facial recognition technology for police usage.[51] Export controls can also be used to prevent the proliferation of certain technologies, including dual-use technologies, that serve a clear military purpose.[52]

Tax. Some regulatory schemes tax activity that may have an external cost. Some have proposed taxing AI systems, for instance, that displace workers, and using the tax revenue to offset displacement costs (e.g., by retraining).

Subsidies. Some regulatory schemes promote activities by providing subsidies. The CHIPS Incentives program, for instance, provides federal subsidies to encourage semiconductor manufacturing in the United States.

Technical Assistance. Some government agencies improve practices by providing assistance for regulated entities to come into compliance. The Americans with Disabilities Act, for instance, requires the Department of Justice to provide technical assistance to businesses, local governments, and individuals.[53]

In addition to regulatory interventions that are binding, many proposals also emphasize voluntary commitments, self-regulation, and standards.[54] The AI Risk Management Framework developed by the National Institute of Standards and Technology, for instance, is a voluntary risk-based

February 17, 2022, https://www.fda.gov/food/buy-store-serve-safe-food/risk-assessment-what-it-and-what-does-it-have-do-my-food.

[51]Somerville City Council, "Ban on Facial Recognition Technology" (2019), https://s3.documentcloud.org/documents/6210431/CITY-OF-SOMERVILLE-ORDINANCE-NUMBER-2019-16.pdf [https://perma.cc/6EZ4-9GYJ]; Shannon Van Sant and Richard Gonzales, "San Francsico Approves Ban On Government's Use of Facial Recognition Technology," May 14, 2019, https://www.npr.org/2019/05/14/723193785/san-francisco-considers-ban-on-governments-use-of-facial-recognition-technology [https://perma.cc/QB9U-WEXP].

[52] "U.S. Export Controls" (International Trade Administration, n.d.), https://www.trade.gov/us-export-controls; Hannah Kelley, "Dual-Use Technology and U.S. Export Controls," *Center for a New American Security* (blog), June 15, 2023, https://www.cnas.org/publications/reports/dual-use-technology-and-u-s-export-controls; Kelley.

[53] "Guidance & Resource Materials," *U.S. Department of Justice* (blog), n.d., https://www.ada.gov/resources/?filters=.

[54] For a discussion of soft law's role in AI regulation, see John Villasenor, "Soft Law as a Complement to AI Regulation," July 31, 2020, https://www.brookings.edu/research/soft-law-as-a-complement-to-ai-regulation/ [https://perma.cc/AW43-4NAR]. For a discussion of the role of soft law for emerging technologies more broadly, see Ryan Hagemann, Jennifer Huddleston Skees, and Adam Thierer, "Soft Law for Hard Problems: The Governance of Emerging Technologies in an Uncertain Future," *Colo. Tech. LJ* 17 (2018): 37.

standard.[55] Other such examples include regulatory sandboxes,[56] certification schemes,[57] self-attestation,[58] voluntary disclosures,[59] codes of ethical conduct,[60] international standards,[61] or company commitments.[62] Human intervention is also commonly discussed in recent regulatory proposals and risk management guidelines.[63]

These interventions are not exclusive and many schemes in practice mix, match, and combine interventions.

III. Distinct Regulatory Challenges with AI

Effectively regulating AI involves distinct challenges.

---

[55] See National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0).", pg. 2: "The Framework is intended to be voluntary, rights-preserving, non-sector-specific, and use-case agnostic, providing flexibility to organizations of all sizes and in all sectors and throughout society to implement the approaches in the Framework."

[56] Regulatory sandboxes are a controlled and supervised environment where companies can test and deploy new products, services and business models under the oversight of regulatory authorities. For instance, the EU has started piloting a regulatory sandbox in Spain: European Commission, "First Regulatory Sandbox on Artificial Intelligence Presented," June 27, 2022, https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented [https://perma.cc/ULQ2-GBJ7].

[57] *See, e.g.*, Peter Cihon et al., "AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries," *IEEE Transactions on Technology and Society* 2, no. 4 (December 2021): 200–209, https://doi.org/10.1109/TTS.2021.3077595.

[58] *See, e.g.*, the use of self-attestation to certify compliance in security: National Institute of Standards and Technology, "Secure Software Development Framework (SSDF) Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities," February 2022, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218.pdf [https://perma.cc/8ZJE-CS2G].

[59] *See, e.g.*, voluntary disclosures in the form of environmental reports. Google, "Environmental Report," 2022, https://www.gstatic.com/gumdrop/sustainability/google-2022-environmental-report.pdf [https://perma.cc/TJE7-ELHU]; Meta, "2021 Sustainability Report," 2021, https://sustainability.fb.com/wp-content/uploads/2022/06/Meta-2021-Sustainability-Report.pdf [https://perma.cc/7W5M-ZJFB].

[60] *See, e.g.*, Facebook AI, "Facebook's Five Pillars of Responsible AI," June 22, 2021, https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/ [https://perma.cc/T3RN-EHAH]; Google AI, "Our Principles," n.d., https://ai.google/responsibility/principles/ [https://perma.cc/8FJV-J72U]; Microsoft, "Code of Conduct for Azure OpenAI Service," March 12, 2023, https://learn.microsoft.com/en-us/legal/cognitive-services/openai/code-of-conduct [https://perma.cc/LR9H-AGKK].

[61] *See, e.g.*, ISO/IEC, "ISO/IEC 23894:2023 Information Technology — Artificial Intelligence — Guidance on Risk Management," ISO, 2023, https://www.iso.org/standard/77304.html.

[62] *See, e.g.*, Antony Cook, "Announcing Microsoft AI's Customer Commitments," June 8, 2023, https://blogs.microsoft.com/blog/2023/06/08/announcing-microsofts-ai-customer-commitments/ [https://perma.cc/93D8-S8BQ]; YouTube, "Our Commitments," n.d., https://www.youtube.com/howyoutubeworks/ [https://perma.cc/EBH3-6CRG].

[63] *See, e.g.*, Microsoft, "Governing AI: Blueprint for the Future," May 25, 2023, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw [https://perma.cc/V37L-37B7]; For considerations that should be in place during human-AI interaction, see Appendix C of National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," January 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [https://perma.cc/LX34-FGZU].

First, some forms of AI are already subject to regulation under existing laws and regulations, where important lessons might be learned. The FDA regulates AI medical devices;[64] HUD addresses algorithmic bias in housing;[65] the CFPB regulates AI used in consumer financial products;[66] the CPSC protects safety in consumer products;[67] the FTC regulates advertising claims and enforces consumer protection;[68] and DOT oversees self-driving cars,[69] to name a few.[70] Novel forms of regulation should be based on an assessment of where existing statutory and regulatory authorities fall short, either because of (a) the rise of AI in unregulated domains or (b) gaps in authorities within existing domains. NAIAC, for instance, recommended determining whether certain authorities be granted to the DOJ Civil Rights Division to reduce gaps in its capacity to protect civil rights with AI systems.[71] If a new AI agency is contemplated,[72] a critical question will be about the relationship between existing agencies that have deeper domain expertise and the new agency that would ostensibly have greater AI expertise.[73]

---

[64] Eric Wu et al., "How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals," *Nature Medicine* 27, no. 4 (2021): 582–84.

[65] https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known

[66] Consumer Financial Protection Bureau, "CFPB Acts to Protect the Public from Black-Box Credit Models Using Complex Algorithms," May 26, 2022, https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms/ [https://perma.cc/X2DM-MMUZ].

[67] Consumer Product Safety Commission, "Artificial Intelligence and Machine Learning In Consumer Products," May 19, 2021, https://www.cpsc.gov/s3fs-public/Artificial%20Intelligence%20and%20Machine%20Learning%20In%20Consumer%20Products.pdf.

[68] Michael Atleson, "Keep Your AI Claims in Check," *Federal Trade Commission* (blog), February 27, 2023, https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check [https://perma.cc/BQ9J-NAWJ].

[69] U.S. Department of Transportation, "U.S. Department of Transportation Releases Automated Vehicles Comprehensive Plan," January 11, 2021, https://www.transportation.gov/briefing-room/us-department-transportation-releases-automated-vehicles-comprehensive-plan [https://perma.cc/PQV3-ENLJ].

[70] AI developers have also seen large settlement amounts from data protection laws–such as the EU GDPR and state privacy statutes in the U.S. These laws regulate data, which constitute input to AI models. *See, e.g.*, European Data Protection Board, "The French SA Fines Clearview AI EUR 20 Million," October 20, 2022, https://edpb.europa.eu/news/national-news/2022/french-sa-fines-clearview-ai-eur-20-million_en [https://perma.cc/9V47-SS7V]; Woodrow Hartzog, "BIPA: The Most Important Biometric Privacy Law in the US?," *Regulating Biometrics: Global Approaches and Urgent Questions, Ed. Amba Kak (AI Now 2020)*, 2020, 96–103; Rui-Jie Yew and Alice Xiang, "Regulating Facial Processing Technologies: Tensions Between Legal and Technical Considerations in the Application of Illinois BIPA" (2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT, 2022), 1017–27.

[71] NAIAC, "National Artificial Intelligence Advisory Committee (NAIAC) Year 1," May 2023, https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf [https://perma.cc/PG5V-9M63].

[72] Andrew Tutt, "An FDA for Algorithms," Administrative Law Review 69, no. 1 (2017): 83–123.

[73] Special Competitive Studies Project, "AI Governance Authority Options Memo."

Second, any regulatory scheme specifically addressing AI must define AI.[74] Yet there are considerable uncertainties about how to implement a definition in practice.[75] Some risks with foundation models are similar to risks of other AI systems, but foundation models can both amplify existing risks and pose new ones. For example, if one is primarily concerned about existential risk, regulatory proposals would perhaps be targeted towards foundation models.[76] But if one takes the position that it is necessary to address existing harms of AI systems, regulatory proposals could be based on the regulation of use and application. In addition, regulatory schemes must define when AI systems trigger requirements. For example, foundation models have widely varying capabilities[77] and are commonly subject to "fine-tuning" and it is not clear whether each instance of fine-tuning should trigger regulatory requirements afresh. FDA, for instance, has proposed to deal with AI software changes without triggering pre-approval with Predetermined Change Control Plans.

Third, any regulatory scheme armed with definitions of the AI systems it seeks to regulate then needs to devise a method for assessing for the presence (or absence) of the harms or failures the regulation is aimed at. This means any regulatory scheme specific to AI must be able to measure the property it is seeking to regulate.[78] This is a challenging task because many regulatory

---

[74] We note that definitions of AI created by NIST AI RMF and the OECD have found significant impact among many stakeholders. *See, e.g.*, references to OECD definitions in National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," January 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [https://perma.cc/LX34-FGZU]; We also note that incumbent regulatory schemes and implementing standards or guidance (e.g., the EEOC's four-fifths or 80% rule) can operate across the board, and be applied to AI, without defining AI. *See, e.g.*, https://www.eeoc.gov/select-issues-assessing-adverse-impact-software-algorithms-and-artificial-intelligence-used. U.S. Equal Employment Opportunity Commission, "Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964," n.d., https://www.eeoc.gov/select-issues-assessing-adverse-impact-software-algorithms-and-artificial-intelligence-used [https://perma.cc/QW88-NZR4].

[75] Some of the complexities with implementing a definition of AI in practice come from ambiguity in the differences in the breadth of AI definitions and carve-outs for certain uses of the technology. *See, e.g.*, Christie Lawrence, Isaac Cui, and Daniel E. Ho, "Implementation Challenges to Three Pillars of America's AI Strategy" (Stanford RegLab, December 2022); David Freeman Engstrom et al., "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies," *NYU School of Law, Public Law Research Paper*, no. 20–54 (February 1, 2020), http://dx.doi.org/10.2139/ssrn.3551505.

[76] Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models" (arXiv, July 12, 2022), http://arxiv.org/abs/2108.07258.

[77] Although capabilities differ across foundation models, understanding precisely how particular training mechanisms produce those capabilities (e.g., model architecture and size, dataset composition and size, etc.) is an area of scientific research.

[78] McKane Andrus et al., "What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21 (New York, NY, USA: Association for Computing Machinery, 2021), 249–60, https://doi.org/10.1145/3442188.3445888. Arushi Gupta et al., "The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in US Government," in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 2023, 492–505. Agarwal, Sushant, "Trade-Offs between Fairness, Interpretability, and Privacy in Machine Learning" (Master's Thesis, UWSpace, 2020), http://hdl.handle.net/10012/15861.

schemes seem to be able to measure normative and/or qualitative properties of AI systems, such as the perceived "fairness" of an AI system in a given context,[79] or its potential for generating "harmful" content with regard to specific demographic or ideological groups.[80] All these schemes require the technical tools to be able to measure and assess a system in relation to the qualities being studied, yet relatively few consensus measures and assessments exist to achieve this. Even a systematic framework of relevant measures and assessments could prove useful, for healthy scrutiny and comparison.[81]

Fourth, regulatory schemes struggle with allocating responsibilities across parties. Who should be covered by regulation? This issue is particularly challenging with the rise of general purpose foundation models. Licensing, for instance, would restrict which parties can develop foundation models, but decreased access could then in turn limit R&D at the frontiers of AI and reduce who is able to develop, use, adapt, or assess such models.[82] A White House Report noted that licensing can benefit consumers by setting higher standards, but can also protect incumbent industries and be unduly restrictive.[83] Allocating responsibilities between developers and deployers is particularly important given the fact that general purpose AI may be released by developers with no particular intended use, and then put to wide range of downstream uses by deployers. Vendor liability is also a contested issue. Effective regulatory approaches will need to account for the different roles of different parties.[84]

---

[79] Sam Corbett-Davies and Sharad Goel, "The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning," *ArXiv:1808.00023 [Cs]*, August 14, 2018, http://arxiv.org/abs/1808.00023; Alice Xiang and Inioluwa Deborah Raji, "On the Legal Compatibility of Fairness Definitions," *Workshop on Human-Centric Machine Learning at the 33rd Conference on Neural Information Processing Systems*, 2019; Daniel E. Ho and Alice Xiang, "Affirmative Algorithms: The Legal Grounds for Fairness as Awareness," *University of Chicago Law Review Online*, 2020.

[80] Peter Henderson et al., "Pile of Law: Learning Responsible Data Filtering from the Law and a 256gb Open-Source Legal Dataset," *Advances in Neural Information Processing Systems* 35 (2022): 29217–34.

[81] Related projects in the development of metrics and evaluations are ongoing NIST. National Institute of Standards and Technology, "NIST AI Measurement and Evaluation Projects," 2022, https://www.nist.gov/programs-projects/ai-measurement-and-evaluation/nist-ai-measurement-and-evaluation-projects [https://perma.cc/69XG-EJEN].

[82] The costs of licensing, for instance, could restrict access to foundation models.

[83] White House, "Occupational Licensing: A Framework for Policymakers," *Report Prepared by the Department of the Treasury Office of Economic Policy, the Council of Economic Advisers and the Department of Labor* 7 (2015), https://obamawhitehouse.archives.gov/sites/default/files/docs/licensing_report_final_nonembargo.pdf [https://perma.cc/JD2E-Q9XF].

[84] To articulate expectations for use and liability, licenses have been contemplated for ML datasets. *See, e.g.*, Misha Benjamin et al., "Towards Standardization of Data Licenses: The Montreal Data License," *ArXiv Preprint ArXiv:1903.12262*, 2019. However, implemented licenses have faced a lack of compliance and enforcement. *See, e.g.*, Kenny Peng, Arunesh Mathur, and Arvind Narayanan, "Mitigating Dataset Harms Requires Stewardship: Lessons from 1000 Papers," *ArXiv Preprint ArXiv:2108.02922*, 2021. As AI systems increasingly function as general purpose technology, licenses for model use rather than dataset use are considered in, for example, Danish Contractor et al., "Behavioral Use Licensing for Responsible AI," in *2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, 778–88. Responsibility could also be clarified for model development and deployment through contracting around liability. If liability is placed squarely on the deployer, deployers may want assurances/indemnification/warranties from the developer that incentivize the developer to address risks or not make

Fifth, numerous regulatory schemes have focused on tailoring requirements based on the risk level of AI use cases. Such schemes have commonly pointed to particular domains (e.g., employment, finance, housing, law enforcement, health) as indicative of high-risk. But risk can vary dramatically *within* a domain. The same language model, for instance, can be used to automatically make employment decisions (high-risk) or simply to assist job applicants to identify relevant positions without restricting any options (lower risk).[85] Similarly, risk assessments become more challenging with the shift toward general purpose foundation models, where downstream use cases are wide-ranging and uncertain at the time of development. Making risk determinations itself can be challenging but is important in developing risk-based regulatory proposals.[86]

Sixth, AI systems are rapidly evolving. Current calls for regulation are in part inspired by the rise of large language models, but regulation must be flexible to adapt to tomorrow's technology. Disclosure may become much more feasible as technical standards such as watermarking emerge (and government could potentially help to foster the creation of such standards).[87] Impact assessments, standard setting, and audits can be challenging with general purpose technology and may be more feasible by focusing on domain-specific implementation and use cases.[88] Requiring developers to incorporate safeguards for all possible downstream use cases could mean that a much smaller number of parties are able to meet such requirements, further concentrating industry. For example, the costs of incorporating mandatory safeguards for model outputs, such as through hard-coded constraints or reinforcement learning with human feedback, could make developing them infeasible for smaller groups and academic researchers.[89] Regulations targeting

---

the model available for certain risky purpose. For a discussion of AI liability and contracting considerations broadly, s*ee, e.g.*, Dan Felz, Wim Nauwelaerts, Paul Greaves, and Josh Fox, "ChatGPT & Generative AI: Everything You Need to Know," April 14, 2023, https://www.law.com/2023/04/14/chatgpt-generative-ai-everything-you-need-to-know/?slreturn=20230619012204; Matthew F. Ferraro, Natalie Li, Haixia Lin, and Louis W. Tompros, Ten Legal and Business Risks of Chatbots and Generative AI, February 28, 2023, Tech Policy Press, https://techpolicy.press/ten-legal-and-business-risks-of-chatbots-and-generative-ai/

[85] Of course, even in this setting, certain forms of job assistance can be higher risk. Additionally, lower risk categories of applications may still have vulnerabilities that carry high risk. For example, the use of cookies in personalized advertising could pose privacy risks related to law enforcement.

[86] Special Competitive Studies Project, "AI Governance Authority Options Memo," June 2023, https://www.scsp.ai/wp-content/uploads/2023/06/AI-Regulatory-Options-Memo.pdf [https://perma.cc/6ZF7-N2R9]. The premise of risk regulation itself can also be contested. Margot E. Kaminski, "Regulating the Risks of AI," *Boston University Law Review* 103 (2023), https://doi.org/10.2139/ssrn.4195066.

[87] See, John Kirchenbauer et al., "A Watermark for Large Language Models," *ICML*, 2023. for an example of output watermarking for language models. At the same time, inventing new watermarking mechanisms and bypassing these mechanisms can become a cat-and-mouse game. See Peter Henderson, "Should the United States or the European Union Follow China's Lead and Require Watermarks for Generative AI?," *Georgetown Journal of International Affairs* (blog), May 24, 2023, https://gjia.georgetown.edu/2023/05/24/should-the-united-states-or-the-european-union-follow-chinas-lead-and-require-watermarks-for-generative-ai/ [https://perma.cc/C5XU-EPRD].

[88] Inioluwa Deborah Raji et al., "Outsider Oversight: Designing a Third Party Audit Ecosystem for Ai Governance," in Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 2022, 557–71.

[89] *See, e.g.*, Ryan Lowe and Jan Leike, "Aligning Language Models to Follow Instructions," January 27, 2022, https://openai.com/research/instruction-following [https://perma.cc/YG3B-FZBN]; Or Sharir, Barak Peleg, and Yoav Shoham, "The Cost of Training Nlp Models: A Concise Overview," *ArXiv Preprint ArXiv:2004.08900*, 2020.

AI development may also include wide-ranging requirements around, for example, system safeguards, data, and algorithm development, while regulations of AI deployment may be context or use-specific. NAIAC has supported the creation of the National AI Research Resource to broaden data and computing access for researchers and small businesses to understand the implications of such models.[90]

Seventh, regulation requires technical and sociotechnical expertise in government. Crafting effective audits, standards, disclosures, and enforcements will not be possible absent government expertise in AI systems – including engineering, machine learning, data ethics, and sociotechnical expertise. Yet such expertise is currently lacking within the public sector.[91] NAIAC has recommended a range of mechanisms to foster sociotechnical research and to improve the talent pipeline into government, such as the Digital Service Academic Compact, the US Digital Service Academy, short-term rotations (PIFs, IPAs), and training programs of the existing federal workforce.[92]

Last, the AI innovation ecosystem has historically been driven by many open-source and open science innovations. The concentration of AI innovation, driven by larger data and large-scale computing resources, has begun to challenge such open-source principles in practice. Different regulatory goals can likewise heighten the tension around openness.[93] A focus on existential risk, for instance, could lead some to advocate for a more closed ecosystem, while a focus on errors and biases could lead some to advocate for greater transparency and access.[94] Such choices, however, are not necessarily binary: release and access policies exist on a spectrum[95] and regulatory interventions may be matched to the model and contemplated use case.[96]

---

[90] NAIAC, "National Artificial Intelligence Advisory Committee (NAIAC) Year 1," May 2023, 43, https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf [https://perma.cc/PG5V-9M63].

[91] National Security Commission on Artificial Intelligence, "Final Report," 2021; Engstrom et al., Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies.

[92] NAIAC, "National Artificial Intelligence Advisory Committee (NAIAC) Year 1," May 2023, 37–39, https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf [https://perma.cc/PG5V-9M63]; NAIAC, 44–51.

[93] Alex Engler, "The EU's Attempt to Regulate Open-Source AI Is Counterproductive," *Brookings* (blog), August 24, 2022, https://www.brookings.edu/blog/techtank/2022/08/24/the-eus-attempt-to-regulate-open-source-ai-is-counterproductive/ [https://perma.cc/3HRY-4B48].

[94] Inioluwa Deborah Raji and Joy Buolamwini, "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial Ai Products," in *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 2019, 429–35; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015); Danielle Keats Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Wash. L. Rev.* 89 (2014): 1.

[95] Irene Solaiman, "The Gradient of Generative AI Release: Methods and Considerations," in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 2023, 111–22.

[96]Percy Liang et al., "The Time Is Now to Develop Community Norms for the Release of Foundation Models," May 17, 2022, https://hai.stanford.edu/news/time-now-develop-community-norms-release-foundation-models [https://perma.cc/JKY3-YRJW].

Working Group Members

Amanda Ballantyne, Jack Clark, Victoria Espinel (*Co-chair*), Janet Haven, Daniel E. Ho (*Co-chair*), Ashley Llorens, Frank Pasquale, Christina Montgomery, Navrina Singh

Acknowledgements

Thank you to the following individuals for their research and support in creating this document: Rui-Jie Yew and Christie Lawrence. All contributions made by non-Members have been performed under the supervision of a NAIAC Member.