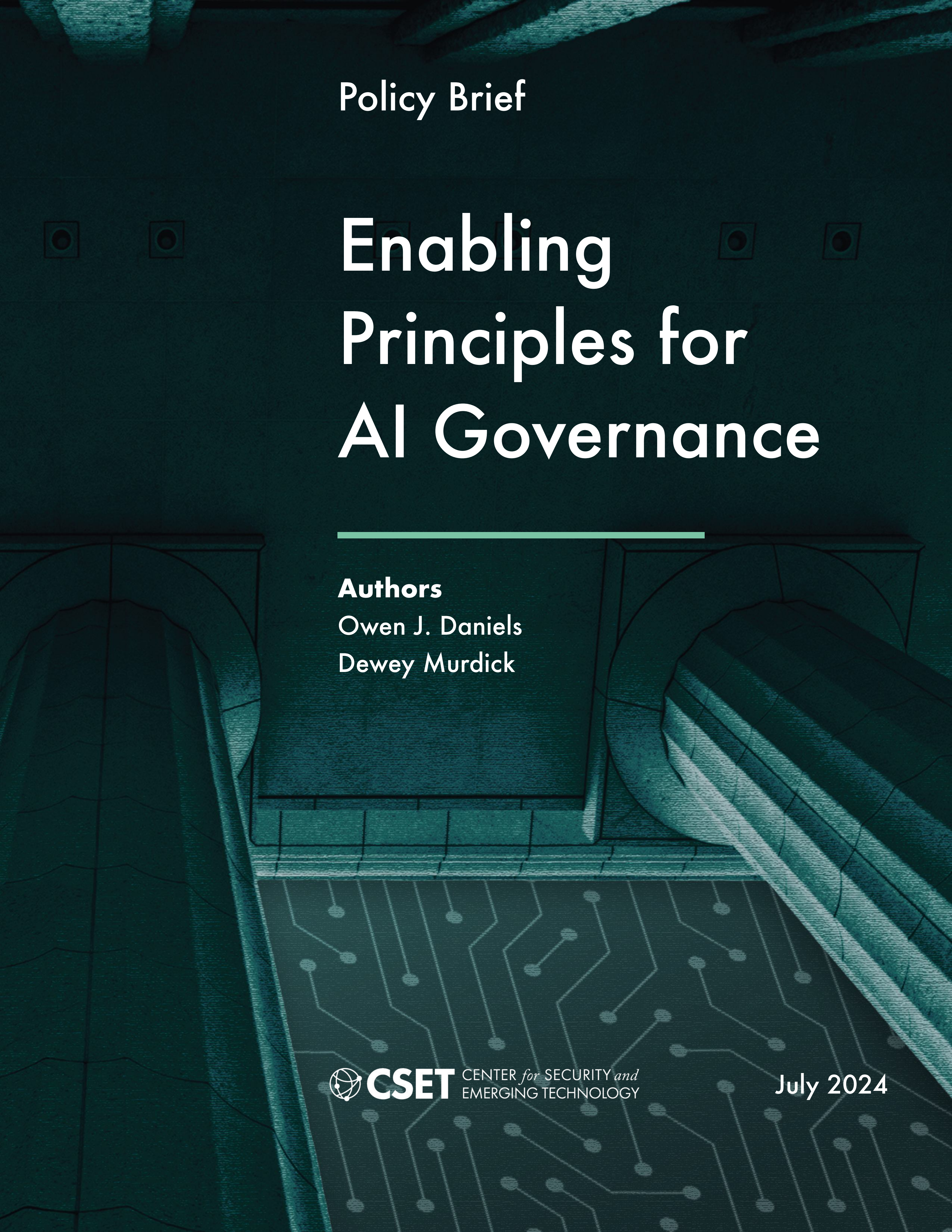

Policy Brief

Enabling Principles for AI Governance

Authors
Owen J. Daniels
Dewey Murdick

CSET CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

July 2024

Introduction

The question of how to govern artificial intelligence (AI) is rightfully top of mind for U.S. lawmakers and policymakers alike. Strides in the development of high-powered large language models (LLMs) like ChatGPT/GPT-4o, Claude, Gemini, and Microsoft Copilot have demonstrated the potentially transformative impact that AI could have on society, replete with opportunities and risks. At the same time, international partners in Europe and competitors like China are taking their own steps toward AI governance.[1] In the United States and abroad, public analyses and speculation about AI's potential impact generally lie along a spectrum ranging from utopian at one end—AI as enormously beneficial for society—to dystopian on the other—an existential risk that could lead to the end of humanity—and many nuanced positions in between.

LLMs grabbed public attention in 2023 and sparked concern about AI risks, but other models and applications, such as prediction models, natural language processing (NLP) tools, and autonomous navigation systems, could also lead to myriad harms and benefits today. Challenges include discriminatory model outputs based on bad or skewed input data, risks from AI-enabled military weapon systems, as well as accidents with AI-enabled autonomous systems.

Given AI's multifaceted potential, in the United States, a flexible approach to AI governance offers the most likely path to success. The different development trajectories, risks, and harms from various AI systems make the prospect of a one-size-fits-all regulatory approach implausible, if not impossible. Regulators should begin to build strength through the heavy lifting of addressing today's challenges. Even if early regulatory efforts need to be revised regularly, the cycle of repetition and feedback will lead to improved muscle memory, crucial to governing more advanced future systems whose risks are not yet well understood.

President Biden's October 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, as well as proposed bipartisan AI regulatory frameworks, have provided useful starting points for establishing a comprehensive approach to AI governance in the United States.[2] These stand atop existing statements and policies by federal agencies like the U.S. Department of Justice, the Federal Trade Commission, as well as the U.S. Equal Employment Opportunity Commission, among others.[3]

In order for future AI governance efforts to prove most effective, we offer three principles for U.S. policymakers to follow. We have drawn these thematic principles from across CSET's wide body of original, in-depth research, as well as granular findings and specific recommendations on different aspects of AI, which we cite throughout this report. They are:

1. **Know the terrain of AI risk and harm:** Use incident tracking and horizon-scanning across industry, academia, and the government to understand the extent of AI risks and harms; gather supporting data to inform governance efforts and manage risk.

2. **Prepare humans to capitalize on AI:** Develop AI literacy among policymakers and the public to be aware of AI opportunities, risks, and harms while employing AI applications effectively, responsibly, and lawfully.

3. **Preserve adaptability and agility:** Develop policies that can be updated and adapted as AI evolves, avoiding onerous regulations or regulations that become obsolete with technological progress; ensure that legislation does not allow incumbent AI firms to crowd out new competitors through regulatory capture.

These principles are interlinked and self-reinforcing: continually updating the understanding of the AI landscape will help lawmakers remain agile and responsive to the latest advancements, and inform evolving risk calculations and consensus.

1. Know the terrain of AI risk and harm

As AI adoption progresses, supporting data will be necessary to better understand the types, and extent of, various public and societal risks and harms. U.S. regulators should prioritize collecting information on AI incidents to inform policymaking and take necessary corrective measures, while preserving the technology's benefits and not stifling innovation. Ideally, an effective, multipronged approach to AI governance would mix incident reporting, evaluation science, and intelligence collection.

Capture data on AI harms through incident reporting. AI systems should be tested rigorously before deployment, including with each update, but they may be prone to drift or failure in environments dissimilar to their testing conditions and can behave in ways unforeseen by system developers.[4] Malicious actors can also use AI to cause intentional harm, for instance using generative AI to perpetuate fraud by creating deepfake images or videos.[5] In conceptualizing harm on the spectrum of minimal to existential risk, lawmakers can consider harm exposure in four buckets: 1) demonstrated harms; 2) probable harms involving known risks in deployed AI systems; 3) implied harms, where studies could uncover new weaknesses in deployed systems;

and 4) speculative harms, including existential risks.[6] These four risk-based buckets provide structure to different harms that regulators can use in AI governance.

Incident collection would entail collecting data from accidents and events where AI systems caused harm, relying on mandatory, voluntary, and citizen reporting of risks and harms.[7] A public incident reporting system would not cover military or intelligence AI incidents, and there could be a separate channel for reporting sensitive AI incidents, protected within secure enclaves. Mandatory and voluntary reporting would likely need to be overseen by federal agencies with clear regulatory roles and distance from AI developers, such as the Federal Aviation Administration or the Securities and Exchange Commission.[8] Citizen reporting could be collected either as part of a governmental complaint reporting system or for public consumption by nongovernmental organizations like the UL Research Institutes, the Organization for Economic Cooperation and Development, or even a news media outlet. Initially, incident reporting could prioritize incidents that generate tangible harms and shift political will, including fatalities, major property damage, or child safety. CSET research has explored the pros and cons of these risk collection approaches.[9]

Knowledge garnered through incident reporting would help achieve several goals.

First, it could help improve public awareness around existing real-world AI risks and harms. With clearer insights into today's most pressing AI challenges, regulators and legislators can better shape laws and address liability issues of public interest.

Second, as patterns of AI incidents develop across different industries, regulators may be able to prioritize certain AI governance actions based on the prevalence of certain harms. For example, regulators might create risk-based requirements for certain AI systems to undergo retesting and recertification if and when iterative improvements are made to models, similar to how the U.S. Food and Drug Administration subjects high-risk medical devices like pacemakers to continuous evaluation.[10] Incident collection would provide regulators with more granular data to better identify new or more serious harms and to rapidly devise robust responses.[11]

Third, developing an incident reporting system is a concrete bureaucratic step that could beget more government action to address AI harms. It would require determining where a mandatory and voluntary reporting incident collection body would sit within the U.S. government, along with the criteria for different reporting requirements. It would also require an action plan and implementation process to stand it up, and the establishment of a decision-making process for budgeting and resource allocation. The process of establishing this body would generate momentum and build muscle memory that carries over to work on thornier AI governance questions.

Finally, incident reporting could help build U.S. leadership in AI governance globally. Building a strong exemplar of an incident monitoring and reporting center could facilitate collaboration, exchanges, and best-practice sharing with other nations. Incubating international cooperation could make the United States more aware and better prepared to address AI harms that may be more prevalent in other parts of the world, and help build a common foundation with other countries to monitor and spread awareness of shared AI risks.

Invest in evaluation and measurement methods to strengthen our understanding of cutting-edge AI systems. The science of measuring the properties of AI systems, especially the capabilities of foundation models that can be adapted for many different downstream tasks, is currently in early development. Investment is needed to advance basic research into how to evaluate AI models and systems, and to develop standardized methods and tool kits that AI developers and regulators can use. Policymakers' creation of appropriate governance mechanisms for AI depends on their ability to understand what AI systems can and cannot do, and how these systems rate on trustworthiness properties such as robustness, fairness, and security. The establishment of the U.S. Artificial Intelligence Safety Institute within the National Institute of Standards and Technology is a promising step in this direction, though it may currently lack sufficient resourcing to accomplish the tasks it has been set under the 2023 AI executive order and other policy guidance.

Build a robust horizon scanning capability to monitor new and emerging AI developments, both domestically and internationally. Alongside incident collection, maintaining information awareness and avoiding technological surprise (unexpectedly discovering that competitors have developed advanced capabilities) will allow U.S. legislators and regulators to be adaptive in addressing risks and potential harms.[12] Horizon scanning capabilities would be relevant for a range of agencies and bodies, and could take on unique relevant focus areas.

For instance, an open-source technical monitoring center would be instrumental for the United States. It could help the U.S. intelligence community and other federal agencies by establishing a core capability to track progress in various AI fields throughout commercial industry, academia, and government. This would not only keep the community well-informed but also enhance the integration of open-source knowledge with classified sources, thereby improving the overall intelligence gathering and interpretation process—particularly focused outside of the United States.[13] For intelligence community agencies, this monitoring would likely focus on specific technology that augments military systems; agencies outside the intelligence community might focus their horizon scanning on AI applications that could have a significant (though less clearly defined) impact on the economic competitiveness and societal well-being of the United States. Scanning the horizon for new and emerging capabilities can help to ensure that regulators are prepared to handle emerging challenges from abroad. This could be valuable amid competition with China or other authoritarian states that develop capabilities with negative implications for democratic societies, such as AI for mass surveillance or for generating and spreading political disinformation. Robust U.S. horizon-scanning capabilities could improve policymakers' responsiveness to the latest threats across AI fields and applications.[14]

2. Prepare humans to capitalize on AI

AI is ultimately a tool, and like other tools, familiarity with its strengths and limitations is critical to its effective use. Without adequately educated and trained human users, society will struggle to realize AI's potential safely and securely. This section presents several points for how regulators and policymakers can prepare the human side of the equation for emerging AI policy challenges.

Develop AI literacy among policymakers. AI literacy for policymakers is key to effectively understanding and governing risks from AI. At a minimum, policymakers should understand different types of AI models at a basic level. They should also grasp AI's present strengths and limitations for certain tasks, recognize AI models' outputs, and acknowledge the technical and societal risks from factors like bias or data issues. Policymakers should be keenly aware of the ways that AI systems can be imperfect and prone to unexpected, sometimes strange failures, often with limited transparency or explainability. They will need to understand in what contexts using certain AI models is suitable and how machine inputs may bias human decision-making. Grounding in these and other details of AI systems will be important for understanding how new AI differs from current models and for anticipating new regulatory challenges.[15] Developing training and curricula for those in policy positions could help build AI literacy today, while investing in AI education would benefit the policymakers of tomorrow and society in general.[16]

Develop AI literacy among the public. Building public AI literacy, beginning as early as possible and continuing throughout adulthood, can help citizens grasp the opportunities, risks, and harms posed by AI to society. For instance, AI literacy can help workers across fields where intelligent systems are already starting to be applied—ranging from industrial manufacturing to healthcare and finance—to better understand the limitations of systems that help them perform their jobs. Knowing when to rely on the outputs of AI systems or to exercise skepticism, particularly in decision-making contexts, will be important. Alerting workers in other fields to the possibility of upskilling programs and accreditations could create employment opportunities beyond the cutting-edge of AI in competencies like computer and information science. AI literacy will be key to participation in the economy of the future for both workers and consumers. Promoting AI literacy could also help the public use outputs from systems like LLMs appropriately to boost productivity and grasp where risks of plagiarism or copyright infringement might exist. The United States could look to countries that have attempted to implement their own public AI literacy programs, such as Finland, for best practices and lessons learned in trying to provide citizens with digital skills.[17]

More broadly, alerting the public to the risks of convincing AI-generated disinformation, including text, images, videos, and other multimedia that could manipulate public opinion, could help citizens remain alert to risks from artificial content.[18] This could be a first line of defense against nefarious attempts by malicious actors to use AI to harm democratic processes and societies. AI developers should also be alert to and versed in the risks of harm that integrating their models into different products could create.

3. Preserve adaptability and agility

Finally, given the dynamic nature of AI research, development, deployment, and adoption, policymakers must be able to incorporate new knowledge into governance efforts. Allowing space to iteratively build and update policies as technology changes and incorporating learning into policy formulation could make AI governance more flexible and effective.

Consider where existing processes and authorities can already help govern AI if certain implementation gaps are addressed. AI is likely to require some new types of regulations and novel policy solutions, but not all regulations for AI will need to be cut from whole cloth. Using existing regulations offers the benefits of speed and familiarity to lawmakers, as well as the ability to fall back on previously delineated authorities among federal agencies (compared to the need to litigate overlapping authorities between existing agencies and newly created AI governance agencies). Policymakers will need to differentiate between truly novel and comparatively familiar questions that AI systems may raise. There are harms that existing protections, such as the Federal Trade Commission Act and the Civil Rights Act of 1964, might already cover when it comes to issues like copyright infringement or discrimination. Other AI applications mix corporate activity, product development, and commercialization in familiar ways that are already covered by protections by bodies like the Federal Trade Commission or the U.S. Food and Drug Administration.[19]

For effective AI governance, policymakers must identify where gaps exist in legal structures and authorities, as well as areas where implementation infrastructure could be lacking. Where applicable legislation does already exist, it will be important to consider where agencies require new resources for analyzing AI systems and applications, such as relevant expertise, sandboxes, and other assessment tools. Given AI's wide-ranging applications and their tendency to get at points of tension in current practices and procedures, new guidance and implementing statutes may be necessary to ensure that existing laws are effective. In some cases, the regulators that enforce these laws may be able to address some of the challenges posed by AI, but they may be reluctant to do so based on resource constraints, lack of precedent with a new technology, or the need to overcome procedural hurdles. Examining where procedural changes or additional resources can unlock the potential for existing laws to be applied to AI may allow lawmakers to move more quickly in addressing harms with regulation, rather than tailoring bespoke solutions to AI problems.

Where it is less clear that existing regulatory or legal frameworks apply, regulators should consider how to develop frameworks that are flexible and can be adapted to incorporate new information. The National Institute of Science and Technology's *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* is a compelling example of a policy document designed to be adapted based on new information and knowledge.[20] The United States can also draw on its mix of state and federal regulations to aggregate data and information and explore the suitability of flexible, experimental governance approaches.[21]

Remain open to future AI capabilities that may evolve in new, unanticipated, and unexpected ways. AI models and applications are diverse, and not all technological progress will be identical. Policymakers should remain open to the possibility that future AI advancements will not rely on the same factors that enabled recent progress. For example, much of the progress in LLM development was driven by a mix of algorithmic improvements and increases in computing power, achieved at great cost, over roughly the past decade.[22] Companies may use more compute to fill the increasing demand for LLM-based products and to continue to innovate in the near term, at an increasingly high cost. That said, it is possible that meaningful future advancement may come not just from research achieved with massive compute, but also from algorithmic innovation or improvements in data processing that require smaller amounts to advance the state of the art.[23] Indeed, CSET research suggests that growth in the amount of compute used to train large models appears to be slowing.[24] Policymakers should be aware of new trends—through connection to information sources like open-source collection, incident reporting, and horizon scanning—and be prepared to effectively regulate to mitigate the risks and capitalize on the opportunities inherent in new AI models.

Lawmakers should consider the costs and tradeoffs involved when planning AI governance approaches. Estimating the labor and resourcing required to implement various governance regimes is an essential step in selecting a feasible strategy. For example, consider regulatory capture, which occurs when a regulatory agency, created to act in the public's interest, instead advances the commercial or special interests of the industry it is charged with regulating, often resulting in policies and decisions that favor the regulated entities rather than the public. Congress should welcome not only input from AI companies as legislators develop regulatory policy, but also their cooperation in regulatory enforcement. Industry can help identify the latest trends in AI development, including nascent risks and harms, and it has a large, highly-skilled workforce whose knowledge the government can draw on.[25] However, lawmakers should keep in mind that companies are not disinterested parties and have their own visions for how to gain and cement advantageous market positions.[26] Regulatory capture presents similar risks in AI as in other industries.[27] However, avoiding it is likely to require the maintenance of a large, skilled government workforce capable of tasks like assessing risks and harms from AI models, and performing analysis and testing. This is likely to be both difficult to attain and costly. While the government could limit such costs by adopting governance models that shift responsibility for testing and risk mitigation onto firms, allowing major AI firms to entrench regulatory positions could permit firms to develop standards that benefit their development models at the expense of others.

Depending on the scope of effort involved, if lawmakers seek to eliminate certain AI risks, they may be more willing to devote costly resources to develop a high-intensity, government-first approach that avoids regulatory capture. If risk minimization is sufficient, avoiding regulatory capture may be less of a priority. Keeping these trade-offs in mind will be key going forward.

Conclusion

AI governance shapes how humans develop and use AI in ways that reflect their societal values. By adhering to the principles outlined in this brief—understanding AI incidents, closely monitoring tech advancement, fostering AI literacy, and maintaining regulatory flexibility—the United States can lead in responsible AI development. This approach will help safeguard important societal values, promote innovation, and navigate the dynamic landscape of AI advancements. These enabling principles offer a roadmap for crafting agile, informed policies that can keep pace with technological progress and ensure AI benefits society as a whole. The next step is for leaders, policymakers, and regulators to craft governance oversight that allows innovation to progress under watchful supervision and in an atmosphere of accountability.

Authors

Owen J. Daniels is the Andrew W. Marshall fellow at CSET.

Dewey Murdick is the executive director of CSET.

Acknowledgments

For helpful feedback and suggestions, the authors would like to thank Igor Mikolic-Torreira, Jack Shanahan, Miriam Vogel, Jill Crisman, Helen Toner, Zachary Arnold, Heather Frase, and Jack Corrigan. They would also like to thank Jahnavi Mukul and Shelton Fitch for editorial support.

Document Identifier: doi: 10.51593/20240029

Endnotes

[1] For Europe, the EU AI Act is the preeminent piece of AI regulation. See: Adam Satariano, "E.U. Agrees on Landmark Artificial Intelligence Rules," *The New York Times*, December 8, 2023, https://www.nytimes.com/2023/12/08/technology/eu-ai-act-regulation.html; and Mia Hoffmann, "The EU AI Act: A Primer," Center for Security and Emerging Technology, September 26, 2023, https://cset.georgetown.edu/article/the-eu-ai-act-a-primer/. See also, "The EU Artificial Intelligence Act: Up-to-Date Developments and Analyses of the EU AI Act." EU Artificial Intelligence Act, 2024, https://artificialintelligenceact.eu/. For China, see, for example, the CSET translation of, "Regulations for the Promotion of the Development of the Artificial Intelligence Industry in Shanghai Municipality," the Standing Committee of the 15th Shanghai Municipal People's, originally published September 23, 2022, https://cset.georgetown.edu/publication/regulations-for-the-promotion-of-the-development-of-the-artificial-intelligence-industry-in-shanghai-municipality/.

[2] The White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," October 30, 2023, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/; Office of Management and Budget, Executive Office of the President, OMB Memorandum M-24-10, "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence" (2024), https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf. In September 2023, U.S. Senators Richard Blumenthal (D-CT) and Josh Hawley (R-MO), chair and ranking member of the U.S. Senate Subcommittee on Privacy, Technology, and the Law, introduced a Bipartisan Framework for U.S. AI Act. The framework calls for establishing a licensing regime administered by an independent oversight body; ensuring legal accountability for harms caused by AI; promoting transparency; protecting consumers and children; and defending national security amid international competition. See, Senator Richard Blumenthal and Senator Josh Hawley, "Bipartisan Framework for U.S. AI Act," Senate Subcommittee on Privacy, Technology, and the Law, September 07, 2023, https://www.blumenthal.senate.gov/imo/media/doc/09072023bipartisanaiframework.pdf. Senator Chuck Schumer (D-NY) introduced a high-level SAFE Innovation Framework around AI. The acronym SAFE in Senator Schumer's framework stands for Security, Accountability, Foundations (in democratic values), and Explain (i.e., "determine what information the federal government needs from AI developers and deployers to be a better steward of the public good, and what information the public needs to know about an AI system, data, or content. See, Senator Chuck Schumer, "Safe Innovation Framework," 2023, https://www.democrats.senate.gov/imo/media/doc/schumer_ai_framework.pdf.

[3] Department of Labor, "Joint Statement on Enforcement of Civil Rights, Fair Competition, Consumer Protection, and Equal Opportunity Laws in Automated Systems," April 2024, https://www.dol.gov/sites/dolgov/files/OFCCP/pdf/Joint-Statement-on-AI.pdf.

[4] For example, certain facial recognition systems performed poorly in identifying individuals with darker skin complexions. "Incident 484: US CBP App's Failure to Detect Black Faces Reportedly Blocked Asylum Applications," AI Incident Database, January 18, 2023, https://incidentdatabase.ai/cite/484/#r2803.

[5] “Incident 88: ‘Jewish Baby Strollers’ Provided Anti-Semitic Google Images, Allegedly Resulting from Hate Speech Campaign,” AI Incident Database, August 15, 2017, https://incidentdatabase.ai/cite/88/#r2183.

[6] Heather Frase and Owen Daniels, “Understanding AI Harms: An Overview,” Center for Security and Emerging Technology, August 11, 2023, https://cset.georgetown.edu/article/understanding-ai-harms-an-overview/; Mia Hoffmann and Heather Frase, “Adding Structure to AI Harm An Introduction to CSET's AI Harm Framework,” Center for Security and Emerging Technology, July 2023, https://cset.georgetown.edu/publication/adding-structure-to-ai-harm/.

[7] Heather Frase and Ren Bin Lee Dixon, “AI Incident Collection: An Observational Study of the Great AI Experiment,” Center for Security and Emerging Technology, September 18, 2023, https://cset.georgetown.edu/wp-content/uploads/20230044-AI-Incident-Collection_-An-Explainer.pdf.

[8] Jack Corrigan, Owen J. Daniels et al., “Governing AI with Existing Authorities: A Case Study in Commercial Aviation,” Center for Security and Emerging Technology (forthcoming).

[9] Frase and Dixon, “AI Incident Collection: An Observational Study of the Great AI Experiment.”

[10] Mina Narayanan, Alexandra Seymour, Heather Frase, and Karson Elmgren, “Repurposing the Wheel: Lessons for AI Standards,” Center for Security and Emerging Technology, November 2023, https://cset.georgetown.edu/wp-content/uploads/20230021-Repurposing-the-Wheel-Final-11.29.2023-1.pdf.

[11] Helen Toner, Jessica Ji, John Bansemer et al., “Skating to Where the Puck is Going: Anticipating and Managing Risks from Frontier AI Systems,” Center for Security and Emerging Technology and Google DeepMind, October 2023, https://cset.georgetown.edu/wp-content/uploads/Frontier-AI-Roundtable-Paper-Final-2023CA004-v2.pdf.

[12] Mark F. Cancian, “Technological Surprise,”*Avoiding Coping with Surprise in Great Power Conflicts*, Center for Strategic and International Studies (CSIS), February 2018. http://www.jstor.org/stable/resrep22428.9.

[13] Tarun Chhabra, William Hannas, Dewey Murdick, and Anna Puglisi, “Open-Source Intelligence for S&T Analysis,” Center for Security and Emerging Technology, September 2020, https://cset.georgetown.edu/publication/open-source-intelligence-for-st-analysis/.

[14] Dewey Murdick, “For a Senate Homeland Security and Governmental Affairs Subcommittee on Emerging Threats and Spending Oversight hearing on Advanced Technology: Examining Threats to National Security,” Center for Security and Emerging Technology, September 19, 2023, https://cset.georgetown.edu/wp-content/uploads/2023-09-19-Emerging-Threats-and-Spending-Oversight-Subcommittee-Written-Testimony-v1.5.pdf; Dewey Murdick, “Advanced Technology: Examining Threats to National Security,” Center for Security and Emerging Technology, September 19,

2023, https://cset.georgetown.edu/publication/advanced-technology-examining-threats-to-national-security/.

[15] Toner, Ji, Bansemer, et al, “Skating to Where the Puck is Going: Anticipating and Managing Risks from Frontier AI Systems,” https://cset.georgetown.edu/wp-content/uploads/Frontier-AI-Roundtable-Paper-Final-2023CA004-v2.pdf.

[16] Diana Gehlhaus, Luke Koslosky, Kayla Goode and Claire Perkins, “U.S. AI Workforce: Policy Recommendations,” Center for Security and Emerging Technology, October 2021, https://cset.georgetown.edu/wp-content/uploads/CSET-U.S.-AI-Workforce-Policy-Recommendations.pdf.

[17] Tarmo Virki, “Finland Seeks to Teach 1% of All Europeans Basics on AI,” Reuters, January 2020, https://www.reuters.com/article/idUSKBN1YE1CT/.

[18] Josh A. Goldstein and Andrew Lohn, “Deepfakes, Elections, and Shrinking the Liar’s Dividend,” Brennan Center for Justice, January 23, 2024, https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend; Josh A Goldstein, Jason Chao, Shelby Grossman, Alex Stamos and Michael Tomz, “How persuasive is AI-generated propaganda?” *PNAS Nexus* 3, no. 2, (February 2024), https://academic.oup.com/pnasnexus/article/3/2/pgae034/7610937?login=false.

[19] Zachary Arnold and Micah Musser, “The Next Frontier in AI Regulation Is Procedure,” Lawfare, August 10, 2023, https://www.lawfaremedia.org/article/the-next-frontier-in-ai-regulation-is-procedure.

[20] National Institute of Standards and Technology, “Artificial Intelligence Risk Management Framework,” U.S. Department of Commerce, January 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[21] David A. Wolfe, “Experimental Governance: Conceptual Approaches and Practical Cases,” OECD, https://www.oecd.org/cfe/regionaldevelopment/Wolfe(2018)ExperimentalGovernanceConceptualApproaches.pdf.

[22] Anson Ho, Tamay Besiroglu, and Ege Erdil et al., “Algorithmic progress in language models,” arXiv preprint arXiv:2403.05812 (2024), https://arxiv.org/abs/2403.05812.

[23] Ibid.; Micah Musser, Rebecca Gelles, Ronnie Kinoshita et al., “The Main Resource is the Human: A Survey of AI Researchers on the Importance of Compute,” Center for Security and Emerging Technology, April 2023, https://cset.georgetown.edu/publication/the-main-resource-is-the-human/; Husanjot Chahal Helen Toner, and Ilya Rahkovsky, “Small Data’s Big AI Potential,” Center for Security and Emerging Technology, September 2021, https://cset.georgetown.edu/wp-content/uploads/CSET-Small-Datas-Big-AI-Potential-1.pdf.

[24] This assessment was based on download data from Github, an online code repository, and subsequent analysis by CSET. Andrew J. Lohn, “Scaling AI Cost and Performance of AI at the Leading Edge,” Center for Security and Emerging Technology, December 2023, https://cset.georgetown.edu/wp-

content/uploads/Scaling-AI-Cost-and-Performance-of-AI-at-the-Leading-Edge.pdf; Andrew J. Lohn and Micah Musser, "AI and Compute," Center for Security and Emerging Technology, January 2022, https://cset.georgetown.edu/wp-content/uploads/AI-and-Compute-How-Much-Longer-Can-Computing-Power-Drive-Artificial-Intelligence-Progress_v2.pdf.

[25] Toner, Ji, Bansemer et al., "Skating to Where the Puck is Going: Anticipating and Managing Risks from Frontier AI Systems," https://cset.georgetown.edu/wp-content/uploads/Frontier-AI-Roundtable-Paper-Final-2023CA004-v2.pdf.

[26] Owen Tucker-Smith, "Congress wants to regulate AI. Big Tech is eager to help," *Los Angeles Times*, July 5, 2023, https://www.latimes.com/politics/story/2023-07-05/ai-congress-regulation-lobbying.

[27] For example, in the aviation industry, the Federal Aviation Administration has increasingly come to rely on aircraft designers and manufacturers to certify aircraft, raising questions about safety and quality assurance. See, "Cockpit Automation, Flight Systems Complexity, and Aircraft Certification: Background and Issues for Congress," Congressional Research Service, October 3, 2019, https://www.everycrsreport.com/files/20191003_R45939_6765d75d1e7ad16986dffe2ad018f48a6777a230.pdf.