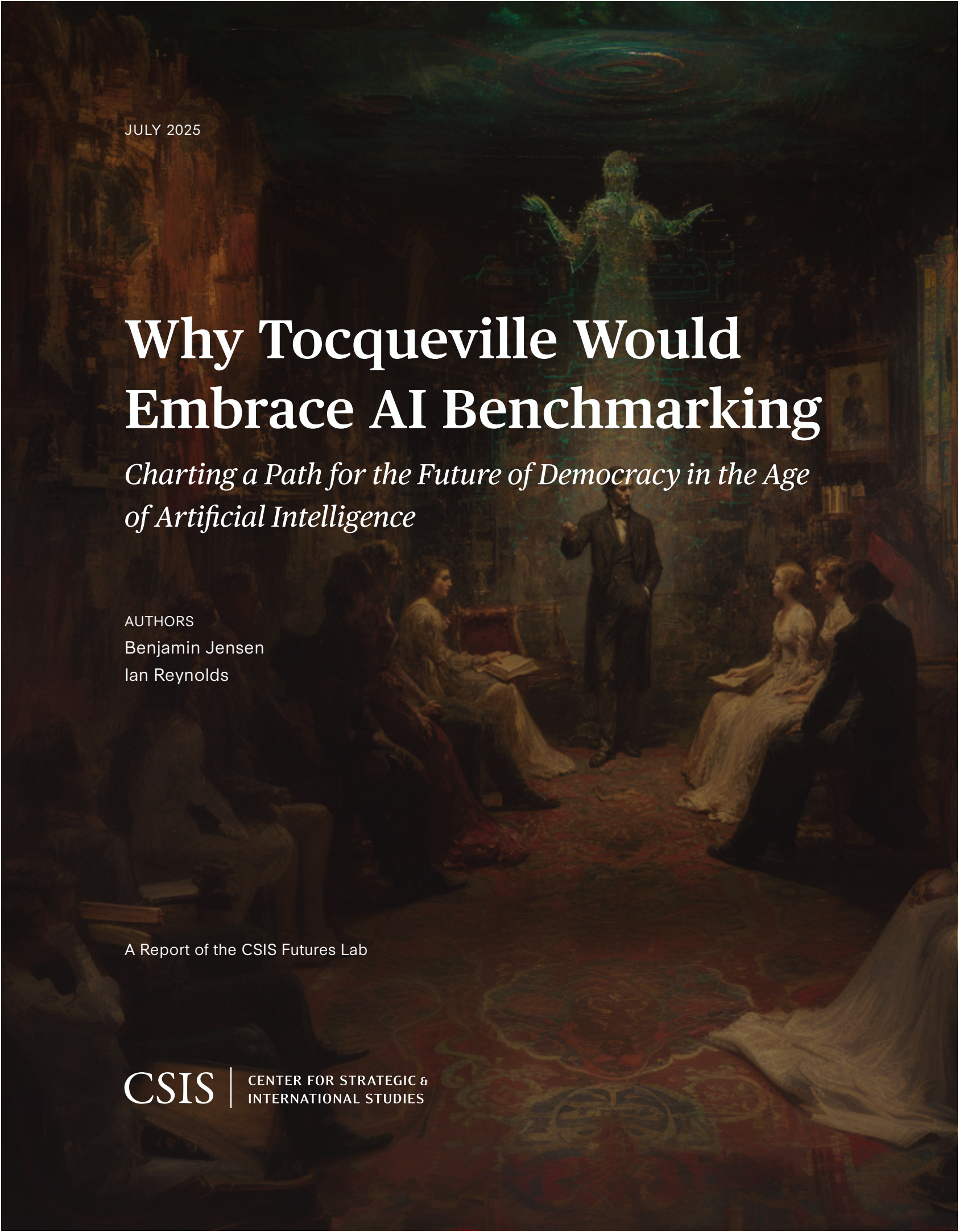


JULY 2025

Why Tocqueville Would Embrace AI Benchmarking

Charting a Path for the Future of Democracy in the Age of Artificial Intelligence

AUTHORS
Benjamin Jensen
Ian Reynolds

A Report of the CSIS Futures Lab

CSIS | CENTER FOR STRATEGIC & INTERNATIONAL STUDIES

JULY 2025

Why Tocqueville Would Embrace AI Benchmarking

Charting a Path for the Future of Democracy in the Age of Artificial Intelligence

AUTHORS
Benjamin Jensen
Ian Reynolds

A Report of the CSIS Futures Lab

CSIS | CENTER FOR STRATEGIC & INTERNATIONAL STUDIES

About CSIS

The Center for Strategic and International Studies (CSIS) is a bipartisan, nonprofit policy research organization dedicated to advancing practical ideas to address the world's greatest challenges.

Thomas J. Pritzker was named chairman of the CSIS Board of Trustees in 2015, succeeding former U.S. Senator Sam Nunn (D-GA). Founded in 1962, CSIS is led by John J. Hamre, who has served as president and chief executive officer since 2000.

CSIS's purpose is to define the future of national security. We are guided by a distinct set of values—nonpartisanship, independent thought, innovative thinking, cross-disciplinary scholarship, integrity and professionalism, and talent development. CSIS's values work in concert toward the goal of making real-world impact.

CSIS scholars bring their policy expertise, judgment, and robust networks to their research, analysis, and recommendations. We organize conferences, publish, lecture, and make media appearances that aim to increase the knowledge, awareness, and salience of policy issues with relevant stakeholders and the interested public.

CSIS has impact when our research helps to inform the decisionmaking of key policymakers and the thinking of key influencers. We work toward a vision of a safer and more prosperous world.

CSIS does not take specific policy positions; accordingly, all views expressed herein should be understood to be solely those of the author(s).

Center for Strategic & International Studies
1616 Rhode Island Avenue, NW
Washington, DC 20036
202-887-0200 | www.csis.org

Acknowledgments

This report is made possible by general support to CSIS. No direct sponsorship contributed to this report.

Contents

Executive Summary

Drawing inspiration from the work of Alexis de Tocqueville, CSIS Futures Lab sees an emergent role for artificial intelligence (AI) benchmarking—independent studies that test and evaluate AI performance in domain-specific tasks—as a new mechanism for ensuring accountability in a free society. Through adapting Tocqueville's concept of association, AI agents can be held accountable, thus ensuring a more open and transparent society. For Tocqueville, association was the "mother science" of life in a democracy that illuminated moral truths, fostered public virtue, and guarded against social isolation. In a world of algorithmic reasoning, association is no longer confined to human debates in the public square: It has taken on a technical dimension that involves testing and evaluating AI foundation models to help the body politic understand their strengths and weaknesses as well as push firms to fine-tune their models to meet specific user needs. This process must be independent and transparent to ensure the free exchange of ideas in a democracy.

As a result, AI benchmarking is a central task in a free society that embraces AI technology. Furthermore, to preserve its independence, the process should be open and free of direct government or business sector control. True association is bottom-up and maximizes transparency and accountability. In an era where individual actors—whether analysts or citizens—are increasingly reliant on opaque agentic workflows, collective action through transparent, domain-specific model evaluation is essential. Without such associations to mediate the influence of AI in national security and foreign policy, decisionmaking risks becoming centralized, brittle, and divorced from democratic oversight. Worse still, the polarization that defines online discourse in the modern

United States will be overrun by narrow, self-interested factions convinced of the rightness of their causes by never-ending feedback loops of tailored information.

Seen in this light, there is a mix of public and private sector action needed to create a world where technology and democracy are co-constituted and thrive, a new town square where algorithms help mitigate—as opposed to reinforce—bias and create the possibility of open dialogue and civic association. First, the public sector needs to pursue legislation that encourages private sector, independent model benchmarking. This push should include a mix of funding for efforts at nonprofits, including universities, and exploring which congressional committee has the authorities to call for routine hearings on model benchmarks. Second, legislators should encourage collaboration between civil society and AI firms through experimenting with legal incentives. Third, and most importantly, the United States' foundations need to come together to fund multiple benchmarking initiatives to preserve the independence of this modern form of association and accountability. To ensure its role in holding algorithms accountable, benchmarking should be free of both government and industry interests.

Introduction

It is 20XX. A defense analyst struggles to keep up with the array of incoming information on adversary troop movements near a contested border region. A combination of autonomous systems, infantry units, and air assets are already in the area. Moreover, some reports suggest that the adversary may be preparing precision long-range strike capabilities targeting the contested border, should a conflict break out. Pressured to demonstrate their resolve by an array of computational propaganda bots amplifying extreme nationalist views, political leaders on both sides have implied that they will not back down in the dispute. However, the analyst has been made aware that backdoor negotiations are underway in an attempt to avoid further crisis escalation. Circumstances are unfolding rapidly, uncertainty is high, and defense leaders are demanding a recommended course of action from the analyst immediately.

The analyst is equipped with a decision support agent leveraging live intelligence feeds, other relevant information, and a large language model (LLM)-enabled natural language chat interface. Based on the current information and its underlying training data, the decision support system recommends that an escalatory policy, in which use of force occurs, is the best course of action. The analyst, however, is hesitant. Her gut says to avoid escalatory behavior, but the decision support system is pushing for the exact opposite. Her superiors are increasingly demanding of her recommendation due to tight operational timelines, but she is unsure of the exact reasoning behind the decision support system's recommendation to use force. Moreover, she is not abreast of the data and training processes the system has undergone. While the final decision on what to recommend does lay in her hands, she is worried that her decision is being nudged toward a course of action by a complicated system she does not entirely understand.

Luckily in this case, the situation was simply a crisis simulation helping benchmark a new decision support agent within an AI-enabled military command system.[1] However, the tensions that are a part of this fictional scenario are increasingly very real. And they extend beyond national security to economics, energy policy, and even online discourses that shape how people mobilize in their communities and exercise their right to vote.[2]

Both citizens and government institutions around the world are experimenting with integrating AI agents into their daily lives, including in decisionmaking contexts.[3] AI agents that perceive the environment, make decisions, and recommend actions are increasingly ubiquitous, further blurring the line between human and machine.[4] As a consequence, it is likely that, in the near future, defense and foreign policy analysts, as well as city councils and even individual voters, will be faced with the question of whether to trust AI agents in critical circumstances.

The resulting situation leaves society at large with a choice. Citizens and institutions could passively rely on opaque and complicated systems that harness AI agents to augment decisionmaking. Alternatively, a free people could endeavor to achieve an active approach to broad and robust civic engagement with such systems to better understand, evaluate, and shape human-machine interactions in the future.

This report argues for the second option and suggests that it is unwise to leave either the public or the private sector to be the sole arbitrators for evaluating the use of AI agents, especially when such agents inform political and national security decisions. A fundamental element to implementing this approach will be successfully developing methodologies for continuous and robust benchmarking and evaluation of AI foundation models and derivative agents in contexts where ground truth may not exist, information is subject to abrupt change, and uncertainty reigns. Moreover, such approaches must be transparent, plural, and feature shared responsibility across a broad spectrum of U.S. society. No one party, no one government agency, and no single corporate cabal should control the ability of free people to exchange ideas and shape their society.

Processes of benchmarking and evaluation can be developed and practiced not only as abstract technical or regulatory tasks, but also—and primarily—as a form of civil association essential to preserving democratic judgment in an era of increasing digital dependency. That is the core proposition of this report and its recommendations for ensuring that the benchmarking process is free from government or industry control and influence.

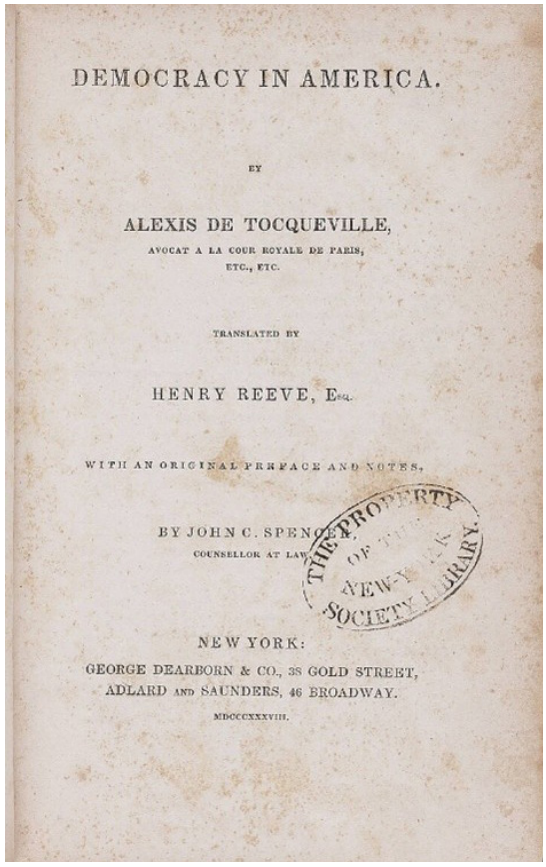
To make its case, this report revisits Tocqueville's arguments about U.S. society from his work *Democracy in America*, as well as other writings in democratic theory, political philosophy, and governance, suggesting that such perspectives can inform an approach on model benchmarking and evaluation that incorporates civil society and leverages the associative potential of different forms of actors across the United States to build a robust and accountable space for agentic AI to flourish.

Tocqueville's Warning and the Civil Society Imperative

Though first published in the nineteenth century, Tocqueville's *Democracy in America* holds contemporary relevance for thinking about how to ensure democratic accountability in the era of agentic AI. Two related factors are most pertinent to this report: (1) the possible detrimental effects of isolation, and (2) the corrective and beneficial consequences of associations. These factors are relevant in the context of AI agents in that delegating social action to AI could increase isolation and reduce human agency in governance decisions absent proper benchmarking and fine-tuning. However, robust collective human involvement within processes of agentic AI development and evaluation hold the potential to make the integration of agentic AI into society an associational process, thus reducing isolation rather than exacerbating it.

In his study, Tocqueville expressed concern regarding isolation in democratic life. Because democracies emphasize equality among citizens, each person is free to pursue their own interests and desires. The result is a generally individualistic social structure. Individualism, in the context of agentic AI, further incentivizes the risk of abdicating governance decisions to technology as citizens see to their own individual lives and daily priorities. The consequence could include leaving technological development and implementation to private companies and government actors with little-to-no public input.

For Tocqueville, despite citizen equality, left to their own devices, equity and individualism can also lead to isolation—and isolation can lead to weakness and the threat of despotism. As he argued, "the vices fostered by tyranny are exactly those supported by equality. These two things are complementary and mutually supportive, with fatal results."[5] Moreover, apart from the risk of


DEMOCRACY IN AMERICA.

BY

ALEXIS DE TOCQUEVILLE,

AVOCAT A LA COUR ROYALE DE PARIS,
ETC., ETC.

TRANSLATED BY

HENRY REEVE, Esq.

WITH AN ORIGINAL PREFACE AND NOTES,

BY JOHN C. SPENCER,
COUNSELLOR AT LAW.

NEW YORK:
GEORGE DEARBORN & CO., 38 GOLD STREET,
ADLARD AND SAUNDERS, 46 BROADWAY.
MDCCCXXXVIII.

Photo: Beinecke Rare Book and Mnuscript Library, Yale University/Wikipedia

tyranny, Tocqueville wrote that isolation leads to impotence in that alone, "citizens can achieve almost nothing."[6] As AI agents are increasingly deployed across public and private life, further isolation could easily lead to a public that simply–even unconsciously–accepts a version of AI's integration into practices of governance that challenge democratic principles and limit human capacity to exercise political agency.

For Tocqueville, the counterweight to the consequences of isolation and its threat to democratic governance was the formation of civil associations between citizens to take on problems of mutual concern, generate social cohesion, and foster care for others within an otherwise individualistic context.[7] The goals of associations can range from building local community centers to variable commercial interests, among many others, in which citizens realize and pursue common bonds. Tocqueville emphasized the critical nature of associations, writing "in democratic countries, the knowledge of how to form associations is the mother of all knowledge since the success of all others depends on it."[8] Thus, the fear is a surrender to a despotic state if associations, and their various positive effects, fail to form. Accordingly, within the context of agentic AI, associations can serve as a key check against the narrow interests of society's most powerful actors and even foster collective democratic practices guiding the technology's development and deployment in a fashion that has diffuse benefits across the U.S. public.

Absent the corrective forces of associations, broader democratic society risks surrendering to an antisocial, nondemocratic version of algorithmic governance.

Tocqueville's discussion of the risk of isolation and the need for association creates a framework for thinking about how to balance the efficiency gains of AI agents with the desire to maintain a free society. In a range of contexts, AI agents are replacing expert judgments and determining important life outcomes, including in the domains of medicine, bank loans, and even governance decisions, among others.[9] If broader society passively–and uncritically–accepts algorithmically shaped decisionmaking procedures, it risks surrendering to an antidemocratic version of

algorithmic governance, or a way of "social ordering" that incorporates algorithmic procedures into decisionmaking in which these computational systems may opaquely shape government processes and life outcomes.[10] As one assessment suggests, "advances in machine-learning—or artificial intelligence—portend a future in which many governmental decisions will no longer be made by people, but [rather] by computer-processed algorithms," presenting an "emerging threat" to liberty and democracy.[11] To map this on to Tocqueville's framework, instead of surrendering to the pull of tyrannical government, absent the corrective forces of associations, broader democratic society risks surrendering to an antisocial, nondemocratic version of algorithmic governance.

Importantly, scholars have shown that civil society can act as a pillar of accountability and an additional form of checks and balances within governance structures. For this reason, a brief review of social scientific work on the relationship between civil society and governance, as well as the possible impact of technology, is a useful exercise. While the modern version of the term "civil society" has been deployed at least as far back as the late eighteenth century, more contemporary expressions of the term, particularly in the work of U.S. political scientist Robert Putnam, have had broader impact in popular discussions of modern democracy.[12] As U.S. sociologist Larry Diamond argued in 1994, "no phenomenon has more vividly captured the imagination of democratic scholars, observers, and activists alike than 'civil society.'"[13] For the purposes of this discussion, following Larry Diamond's definition, civil society refers to sets of organized social groups that members voluntarily join, create, and support, and that are generally autonomous from the state.[14]

Scholars have attempted to demonstrate the important role that civil society can play in shaping political outcomes. For example, Putnam's work on civil society in Italy argues that robust networks of civic engagement hold a range of benefits for social cooperation, including increasing defection costs, fostering norms of reciprocity, and demonstrating the benefits of community collaboration.[15] In turn, he suggests that this form of civic activity—and the subsequent development of social capital—supports high-functioning democratic institutions. Putnam expands his argument beyond Italy. Drawing from Tocqueville, Putnam argues, "Tocqueville was right: democratic government is strengthened, not weakened, when it faces vigorous civil society."[16] As summed up by sociologist Sydney Tarrow, Putnam's basic argument is that "for where there is no social capital … democracy cannot flourish."[17]

Other scholars also emphasize the important links between civil association, democracy, and the spread of liberal norms and values. The study of nongovernmental organizations (NGOs) is a prominent example of the relationship between governance and civil society organizations. Scholars focusing on international politics have argued, for instance, that NGOs work as key norm entrepreneurs for the expansion of liberal human rights regimes.[18] Moreover, NGOs can constrain government action by applying public pressure through advocacy (i.e., naming and shaming), increasing the normative costs for pursuing policies that result in things such as human rights abuses.[19]

Apart from internationally oriented NGOs, researchers contest that, under the right conditions—such as sufficient levels of free press and political competition—robust and active civil society organizations can reduce government corruption and increase accountability.[20] Moreover, civil

society organizations can participate in processes of public oversight and information sharing that can increase the transparency and accountability of governing institutions.[21]

The picture, however, is not always so clear, and the impact of technology, particularly digital technology, on civil associations and democracy remains murky. Spanish social theorist Manuel Castells, for example, notes that technological changes have introduced a "networked society" that has fundamentally altered communicative practices and transformed "space and time in the human experience."[22] As such, technological changes have altered the basic ways in which individuals and groups in societies are linked together. Scholars express different perspectives on how technological factors will shape the role and structure of civil society. Some, such as technology expert Lee Rainie and sociologist Barry Wellman, suspect that such technological changes, particularly in the form of information and communication technologies (ICTs), improve the prospects of robust socialization and community building.[23] Others argue that ICTs offer the prospect of more impactful activism in restricted political environments due to ICTs' capacity to "facilitate open and inclusive participation."[24] Still other scholars express concern over the role that technology will play in social organizations and democratic structures. Returning to Castells, he suggests that one consequence of the networked society could be a destruction of organizations and the "delegitimating" of institutions, leading to the prospect of social alienation.[25] Moreover, research finds that, in some contexts, social media can decrease satisfaction with democracy, and scholars studying authoritarianism have argued that digital technologies can increase state capacity for repression of democratic freedoms.[26]

Of additional importance is the co-constitutive relationship between science, technology, and society.[27] All too commonly, technology is treated simply as a variable (or tool) that impacts the social world in a direct way.[28] Thus, in terms of civil society, technologies such as the internet will either help or hurt such organizations. Yet, the relationship is far more dynamic and dependent "on a complex pattern of interactions" between the social and technological.[29] The majority of peoples' interactions with each other and their environment are increasingly mediated through technology in general, and AI agents in particular. Humans may have made AI, but AI agents shape how people interact with their world.

A few examples from the daily practices of scientific laboratories as well as broader implementations of technology in public policy domains will help to illustrate this point. Social theorists Bruno Latour and Steve Woolgar demonstrate that even within the most pure scientific location—i.e., the laboratory—social factors still shape and direct scientific findings.[30] Others, such as political theorist Langdon Winner, discuss how technologies are not neutral artifacts, but, in fact, can have crucial implications for power relationships and political outcomes within broader society.[31] The intentional construction of bridges in Long Island, New York, to be too short for buses to pass under illustrates the role technology can play in power (and political) relationships. The goal in limiting the availability of public transportation in this way was to restrict poorer, typically minority, populations from accessing beaches and public areas served by the roads, as these groups commonly relied on public transit over more expensive modes of transportation, such as the automobile.[32] Moreover, technological discoveries are not divorced from the social

structures of power in which they are produced, and broader social imaginations of technology can shape the direction of technological developments.[33] The upshot of these wider effects is that there are critical choices to be made in technological design, implementation, and execution that, as those decisions are shaped by broader social structures and sequences of interactions, also have far-reaching social and political impacts. As Winner writes, "technologies are ways of building order in our world."[34] Thus, when applied to the question of the relationship between AI—specifically agentic AI decisionmaking—and civil society, associational politics can play a crucial role in building socio-technical orders that favor democratic outcomes and processes. However, this will not happen automatically. In simple terms, it is imperative to proactively create a relationship between AI and society that enables democracy and responds effectively to the political interests of everyday people.

The relationship between robust civil association and democracy is further complicated by empirical work that responded to Putnam's initial arguments suggesting robust civil society was the tonic for democratic governance. For example, a range of studies have documented cases in which civil society organizations have had detrimental effects on democratic outcomes. U.S. political scientist Sheri Berman's work on civil society in the German Weimar period, for example, illustrates how the Nazi party was able to co-opt aspects of robust German civil organizations in a fashion that supported the rise of fascism.[35] Further research demonstrates similar dynamics in Spain and Italy during the end of the nineteenth and into the twentieth century.[36] Consequently, contextual and historical factors matter in how strong civic associations shape democratic outcomes. While Putnam did emphasize this fact, he was perhaps wrong in then generalizing the relationship of civic association, social capital, and democracy to other cases so deterministically.[37] Yet, while we might not be able to isolate a consistent democratic effect of associations across time and space, this does not mean that civic associations and civil society cannot serve as practical tools in a larger tool kit for addressing broader issues of accountability and transparency related to artificial intelligence. Here we can draw lessons from the political theory of American pragmatism, which emphasizes social problem-solving oriented around joint public goals.[38]

Pragmatism highlights the "diversity of perspectives that different individuals and organizations bring to the definition of the problems, and to the generation of possible solutions."[39] As U.S. political scientist Christopher Ansell notes, "pragmatism is usefully described as a philosophy of evolutionary learning. It emphasizes the ability of both individuals and communities to improve their knowledge and problem-solving capacity over time through continuous inquiry, reflection, deliberation, and experimentation."[40] While pragmatism is dynamic in its broader philosophical commitments, a constant thread, particularly when connected to public-oriented governance, is an "emphasis on the open-ended process of refining values and knowledge."[41] In addition, this line of theory suggests the critical nature of learning through encountering tangible problems that require resolution, sometimes referred to as a "problem situation," a context in which actors must creatively and experimentally find resolutions to new dilemmas.[42] In fact, some scholars have emphasized how pragmatist philosophy can link practical, solution-oriented approaches to "grand problems" with high levels of complexity and uncertainty by offering a "situated, distributed, and processual approach to problem solving."[43]

Civil society groups have been shown to be important component parts of governance due to their capacity to unify citizen interests and improve the accountability and transparency of governing organizations.

This deliberative, action-oriented, experimental view can be supported by associational politics emphasized by Tocqueville, as well as by more contemporary lessons from scholarship on civil society. Consequently, it's worth synthesizing some of the above discussion into more concise takeaways. Fundamentally, from Tocqueville to modern research on democracy, under the right conditions, civil society groups have been shown to be important component parts of governance due to their capacity to unify citizen interests and improve the accountability and transparency of governing organizations. As a caution, however, research also demonstrates that civic organizations are not simply vacant objects of social good; they need specific social forces oriented toward democratic action. To draw on U.S. philosopher John Dewey, democracy is not something that perpetuates itself automatically.[44]

Accordingly, civil associations need to be positioned around democratic goals and social action. One way is to orient organizations toward joint problems facing large cross-sections of democratic society, including the case of governing and evaluating AI systems. Deliberation and communication will be fundamental to this form of associative politics, as through these processes actors can realize joint interests and civil associations can vocalize and propose creative solutions to communal problems. While technology can complicate these relationships, experimental problem-solving can create a link between society and technology that nudges the development of AI toward accountability, transparency, and democracy. Critically, if people, firms, and government entities are going to use AI agents to inform decisions, of paramount importance is transparency and accountability related to the information and technological systems used to make those decisions.

Model Benchmarking as a New Art of Association

New forms of association can play important roles in combating the consequences of a passively accepted version of algorithmic governance, and its subsequent detrimental impact on democratic society. One such connection is in the practice of benchmarking and evaluating AI models. Though this report leverages the empirical domain of national security as an example because it is the research team's area of expertise, this argument applies to other areas of governance and society more broadly speaking.

To begin this conversation, it's worth reviewing what benchmarking is. Benchmarks are datasets designed to evaluate model performance on a specific set of tasks.[45] The processes of benchmarking and evaluating models have become increasingly important as models are deployed in a range of situations that have real-world consequences. Successful benchmarking and evaluation of models not only allow for tracking model improvements but can also identify risks of models that do not perform at an adequate level for the desired use-case.[46]

For example, benchmarks have been developed to test model performance on tasks such as knowledge recall, quantitative reasoning, and other academic tasks.[47] Other benchmarks focus on harmful social biases with respect to gender or race.[48] Additionally, the research team at the CSIS Futures Lab has developed a benchmark, and associated methodology, for tracking model preferences with respect to critical foreign policy decisions in contexts such as crisis escalation scenarios.[49]

Successful evaluation processes can be made more robust through associative practices that unite diverse teams of researchers, policymakers, and civil society actors to steer technological

development away from opaque technological systems and toward a form of social organization in which AI public literacy is high, and a wide range of players have a say in the form of technology that is deployed in the public domain. In more tangible terms, this will require ongoing processes of domain-specific data creation and the testing and evaluation of models before and after deployment in contextually relevant scenarios. Key to the practice of benchmarking is the fact that, as Raji et al. argue, "the imagined artifact of a 'general' benchmark does not actually exist … presenting any single dataset in this way is ultimately dangerous and deceptive."[50] Moreover, issues of construct validity– "the degree to which a test or measurement tool accurately measures the construct it intends to measure"–and certain private sector actors successfully gaming benchmark results can problematically skew the reality of model performance.[51] Such conclusions point to the critical nature of having a wide range of experts and actors in the evaluative process of benchmarking highly specialized, yet broadly critical, domains.

Furthermore, such processes must be overseen by organizations with the public's interest at heart. Here, the information environment will be fundamental. Returning to Tocqueville, information sharing is critical for successful association to occur. Only through making interests and opinions of individuals available for consumption in the public domain can joint interests be realized and successful associations forged. In Tocqueville's era, local newspapers, as well as townhalls, were the key vector for information transfer.[52] While a broader range of communicative technologies exists today, a core lesson is clear–without communicating facts of interest to broader publics, the risk of isolation increases. When applied to AI, and specifically the task of benchmarking and evaluation, this means transparency and broad-based communication of technological risks and uses are essential to ensuring the public's interest. Binding interested actors into a dense, yet transparent, web of association has the potential to shape AI's use in governance away from the consequences of "dead hand" algorithmic governance.

Absent a form of Tocquevillian association related to AI development, the risk is that social passivity gives way to co-option and control of this general-purpose technology to private interests, ceding technological development to a narrower set of goals driven by a select few. Robust associative politics centered around the technology of AI, specifically with respect to evaluation and benchmarking, increase the probability of a strong democratic process working in coordination with efforts of integrating and developing new technologies. This is important as research illustrates that socio-technical relationships structure human agency and can have path-dependent effects as certain relational structures become stabilized, a process in science and technology studies frequently called closure.[53] Because the contemporary era is one in which socio-technical relations with respect to AI are still somewhat flexible, broader U.S. society has its greatest opportunity to structure a relationship more favorable to public–democratic–interests.

A General Review of U.S. AI Policy

Any attempt to initiate a process of associational benchmarking must embed itself within ongoing policy developments in AI. The last three presidential administrations have proposed a range of approaches to both governing AI and incentivizing its integration across the private and public

sectors. A review of these developments will provide a basis for building out the associative model of benchmarking governance discussed below. In 2019, Trump signed an executive order (EO) entitled "Maintaining American Leadership in Artificial Intelligence."[54] This order had a number of goals, including providing a foundation for AI innovation; integrating AI across federal agencies; fostering collaboration between government, the private sector, and academia; and developing technical standards, with the National Institute of Science and Technology (NIST) acting as the coordinating agency. The order sought to balance an innovation-friendly context for AI with the need to protect civil liberties and U.S. values. While the EO established a set of AI-related objectives, some criticized it for lacking details on both funding and practical implementation.[55] That said, later in 2019, the Trump administration released the National AI R&D Strategic Plan: 2019, which outlined a strategy of federal investment into the research and development of AI. The plan included eight specific goals supporting the earlier EO: making long-term investments in AI research; better understanding how AI and humans can work together; addressing ethical and societal implications of the technology; advancing the security and safety of AI; developing data sets for technology development; evaluating AI through standards and benchmarks; seeing to workforce implications; and fostering public-private partnerships.[56]

While the first Trump administration's efforts related to AI policy mention the need to establish fair, transparent, accountable, and ethical AI, the Biden administration furthered such efforts through the release of what the administration termed a "Blueprint for an AI Bill of Rights" in 2022.[57] This framework underscored the threats advances in AI posed to democratic processes and the rights of U.S. citizens, specifically related to bias and privacy. To attempt to address such issues, the Biden administration, led by the Office of Science and Technology Policy, identified five principles guiding "the design, use, and deployment of automated systems."[58] These principles were that AI systems should be safe and effective, that AI systems should not propagate biases or discrimination, that there must be protections from abusive data practices, that users should be aware when they are interacting with an automated system, and finally, that there should be clear human alternatives and fallbacks for any problems users have when interacting with AI-enabled systems.

This served as a platform for the eventual signing of the Biden administration's 2023 EO "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."[59] The 2023 EO focused on ensuring that the United States remained the leader in "safe, ethical, and responsible" AI use and, to that end, directed federal agencies to establish processes "managing dual-use AI models, implementing rigorous testing protocols for high-risk AI systems, enforcing accountability measures, safeguarding civil rights, and promoting transparency across the AI lifecycle."[60] An additional key event occurring alongside the Biden EO was the establishment of the U.S. AI Safety Institute (U.S. AISI), a part of NIST, which was tasked with testing models, conducting red-teaming, and attempting to identify risks from increasingly advanced systems, including those with national security implications.[61] Moreover, following guidance from the initial EO, in 2024, the administration released a memo focused on integrating AI into the national security enterprise in a fashion that both leveraged technological advances while also protecting human and civil rights.[62] The Biden administration's efforts continued through the final days of his presidency, with a second

EO released in January of 2025, focusing on AI infrastructure and directing federal agencies to find eligible lands that could be used for "frontier AI" data centers.[63]

Following Trump's election to a second term, the new administration quickly sought to stamp its authority on the direction of U.S. AI policy. The Trump administration quickly revoked elements of the Biden administration's executive action, suggesting that the latter had imposed "onerous and unnecessary government control over the development of AI."[64] The Trump team replaced Biden's executive guidance with an EO entitled "Removing Barriers to American Leadership in Artificial Intelligence." This late-January 2025 EO positioned the Trump administration, at least in terms of rhetoric, as far more innovation friendly than its predecessors. Furthermore, the EO provided guidance that relevant agency heads, as well as various presidential technology advisers, should review actions described within the Biden EO for revision or rescindment.[65] The 2025 Trump EO, while short and relatively scant on details, directed the Office of Management and Budget (OMB) to release additional guidance on AI implementation throughout the federal government.

Following this directive, in the spring of 2025, the OMB released two memos further laying out the Trump administration's approach to AI policy.[66] In general terms, the memos contain guidance for federal agencies to speed up AI innovation and adoption in relevant use cases while also improving public trust in the technology. Moreover, and in contrast to initial indications from the Trump administration, the OMB memos shared many elements with previous Biden policies. One significant change in the policy, however, is the combining of issues related to "rights" and "safety" impacting AI into a single category called "high-impact AI."[67] That said, as described by a Brookings report, "the continuity [between administrations] is more striking than the change" as both administrations' approaches appear similar on issues such as technology acquisition and the promotion of AI use throughout government.[68] Yet, a June 2025 change may illustrate the Trump team's desire to appear more innovation friendly with respect to AI: The AISA has been renamed the Center for AI Standards and Innovation (CAISI).[69] According to the CAISI, the organization will act as the "primary point of contact" for industry members in the U.S. government to assist with testing, securing, and implementing AI systems.[70]

Thus, as indicated by the above review, AI policy in the United States is progressing. However, civil society and robust benchmarking practices must serve as checks and balances to ensure that the technology functions as intended for broader society. Positioning the United States as a leading AI superpower should be accompanied less by regulation than by a robust benchmarking effort that brings civil society into the process.

The Case of AI in Foreign Policy Decisionmaking

To make this more tangible, the following section relies on an empirical discussion of the Futures Lab's efforts at benchmarking model preferences in foreign policy decisionmaking contexts as well as a review of literature in the field of international relations focusing on the relationship between civil society and foreign policy decisionmaking. Research in international relations has long explored the links between democratic regime type and foreign policy outcomes. For example, democratic peace theory suggests that democracies, whether as a result of institutional structures

or normative elements, tend to not go to war with one another.[71] Other scholars have illustrated that, due to fear of electoral punishment, public opinion related to foreign policy can shape leader preferences, particularly when issue salience is high.[72] However, research has also illustrated that the type of information environment—specifically how information flows from political elites to publics—is critical to the extent to which publics can constrain decisionmakers.[73]

That said, advances in digital technology have led some to argue that the constraining effects of democratic publics on elite decisionmakers may be eroded as constituents are further "fragmented and siloed" so that voters support their leaders, regardless of the leaders' policy decisions.[74] Thus, drawing on the field of international relations, research illustrates that mechanisms for constraining foreign policy decisionmakers can be created within democracies; however, information environments, particularly in the context of the internet, can challenge these processes of constraint. To put this into simple terms, while there are ways to ensure that democratic publics can influence foreign policy decisions, digital technologies present novel challenges to that relationship.

Importantly, advances in AI have the potential to further erode feedback mechanisms between democratic publics and foreign policy experts as the technology is integrated into foreign policy decisionmaking. The Futures Lab's findings related to its benchmarking of LLMs in foreign policy decision contexts can assist in illustrating this point.[75] Results from the Critical Foreign Policy Decisions (CFPD) benchmark demonstrate that some LLMs are significantly more escalatory in decision scenarios when compared to other models. For example, models such as DeepSeekV2, Qwen2 72B, and Llama 3.1 8B Instruct all tend to prefer recommending the use of force when compared to other LLMs included in the study. These escalatory preferences are particularly salient in scenarios where models are prompted to recommend courses of action for democratic countries, like the United States and United Kingdom, versus autocratic countries, such as China and Russia.

Imagine a scenario in which models are integrated into decision pathways (and are not continuously and robustly evaluated for their risk profiles in a transparent fashion) and nudge decisionmakers toward escalation, in the context of, for example, a Taiwan Strait crisis. While it is highly unlikely that LLMs would ever make such a critical decision alone, if models are leveraged to generate possible courses of action, and they are primed to favor the use of force, decisionmakers could be unknowingly constrained in looking for more peaceful pathways to crisis resolution. Moreover, consider a situation in which private organizations and government institutions are left to evaluate model performance, without any oversight. Absent robust transparent processes of evaluating models, such circumstances could lead to pathological outcomes as private companies seek to overpromise on system performance for financial gains and government institutions endeavor to avoid political consequences from deploying unreliable systems. For example, scholars have noted that private companies may be incentivized to keep critical information, such as code, in machine learning algorithms private both to increase profit and to avoid regulatory roadblocks.[76] Moreover, bureaucratic organizations can, at times, have issues with accountability due to political calculations and the pursuit of narrow institutional interests.[77]

Much like Tocqueville's warning about the adverse effects of isolation and passive obedience, broader civil society, in a nonassociative form and restricted from involvement in practices

of evaluating AI agents, may be left to believe claims that systems work reliably and in an unbiased, rational fashion. In such a situation, the public may lack the capabilities to go behind the curtain and access data in a transparent and accountable way. Or, to draw on Tocqueville, "a tyrant is relaxed enough to forgive his subjects for failing to love him, provided they do not love one another."[78]

Expanding on Tocqueville's analysis, the solution will involve tightly binding together civil society actors (ranging from local organizations to mainstream think tanks, research organizations, and beyond) into the processes of benchmarking and evaluating models for use cases relevant to their interests. While regulations, in terms of incentivizing government and private organizations to participate in the process, will be an element of such a vision, a more critical factor will be establishing a culture of civil involvement in influencing how the technology of AI will shape governance in the United States. This applies to cases as diverse as foreign policy to issues of local government.[79] As the following section will outline, a civil culture dedicated to open and transparent forms of AI development is more likely to shape technological—and social—developments in a direction more conducive to robust democratic governance in a context in which AI permeates many decisionmaking domains.

Civil Society in the Age of AI: Indirect Impacts and Democratic Renewal

Tocqueville argued that "as soon as communal affairs are treated as belonging to all, every man realizes that he is not as separate from his fellows as he first imagined."[80] This broad social realization is key to driving the development of associative action and the building up of a plural set of actors in U.S. society that are involved in the iterative and transparent evaluation of models. Associative behavior, however, must be cultivated. That said, its impacts can be recursive in form. Not only do associations mitigate the threats of isolation, but they also train citizens to be engaged in issues of joint concern and to hold a civic mindset.[81] To draw on John Dewey, such practices must become "a way of life," or as Tocqueville put it, "an imperceptible influence of habit."[82] In other words, a collective effort of Americans to ensure technology works for the average citizen must be encouraged and normalized.

In his work on U.S. civil society, Robert Putnam asked, "is technology … driving a wedge between our individual interests and our collective interests?"[83] This is a critical question, and while technology certainly can have this effect—see for example research illustrating that social media can contribute to problematic political polarization—there must be room for an alternative path.[84]

Benchmarking thus can serve as a practical vector for cultivating an associative, civic mindset related to the technology of AI and its relationship with governance to achieve meaningful results. In other words, benchmarking holds the potential to become a critical site of association in which collaborative spaces are generated for policy experimentation matched with public oversight through robust and dynamic evaluation methodologies. Simply put, benchmarking is a tool to ensure the public is aware of the impact that AI is having on their everyday lives.

For example, if AI is going to be involved in making decisions on how certain social benefits are provided, there is no reason why interested members or representatives of the community should not be involved in the evaluative process of system development and be kept abreast of what evaluation results mean for the AI-enabled decisions that could shape life outcomes. This claim is not new; some researchers have already committed to a vision of community-oriented AI development in which AI systems are cocreated with local communities.[85] This vision will undoubtably require a substantial measure of civic responsibility. Therefore, the benefits for communities of interest should not only be made clear; they must also be tangible. AI-enabled systems must feature accountable and transparent feedback loops that are responsive when algorithmic decision systems are failing. Absent such processes, broader democratic society is likely to fail the "tragic double bind" of governing algorithms while "governing by algorithms," as technology researchers Maciej Kuziemski and Gianluca Misuraca put it.[86] In such processes, large private tech organizations cannot be ignored. Yet, it is possible that, despite economic interests and with the right incentives, companies may prove to be amenable partners. Large AI firms have already set up industry organizations to share best practices in model assessment and agreed to a set of voluntary commitments regarding model performance in public domains, although issues of transparency and accountability remain.[87] Mapping market incentives to civic responsibility will undoubtably be a future challenge in generating an open, plural, and transparent evaluation environment.

The results of evaluations cannot simply be meaningless reports left to rot in untouched data repositories.

That said, there is potential for domain-specific, diverse associational methodologies of evaluation to emerge from below. As suggested by literature focusing on the role of civil society, such associations can serve as an injection of collective democratic spirit into the adoption of AI. These practices must be focused on collaboratively ensuring that governance-related decisions involving AI are continuously evaluated in a fashion that demystifies technological outputs and that ensures that citizens are not driven to simply "accept" algorithmic outputs out of either a lack of understanding or the absence of social structures that allow them to meaningfully contest seemingly unjust or incorrect algorithmic decisions. Moreover, people must feel that they receive some practical benefit of working within such associational models, implying that the socio-technical governance structures that are developed need to actually work for people at all levels of government and must respond to their feedback, ideas, and concerns. The results of evaluations cannot simply be meaningless reports left to rot in untouched data repositories. Importantly, research has demonstrated that there are consequential benefits both in terms of social trust and in terms of economic development for communities with high levels of civic engagement. There are thus clear advantages to achieving this collaborative ideal through AI benchmarking and to cementing benchmarking's role in empowering new forms of civic association.

Democratic AI Governance: Lessons from Tocqueville and Today

In many mainstream conversations related to AI, scaling is presented as the fundamental solution to advancing the technology. In technical terms, scaling can refer to increasing model parameters, the amount of training data, or computing capabilities to achieve performance benefits.[88] Scaling models in such a fashion takes significant financial resources, and, as a result, has contributed to the centralization of AI development in the hands of private organizations that already have the capacity to collect data, develop larger models, and afford significant computing resources. As a result of this understanding of scaling, the technical capabilities, model development, and overall direction of AI development are concentrated. This form of socio-technical relationship, and its political and economic results, presents a challenge to a more democratic vision for how AI can be integrated into governance. It is also misaligned with Tocqueville's preference for local initiative and community leadership to address issues of public concern.[89] More democratic-focused efforts surrounding AI will require local participation and association that can assist in generating the robust—and democratically oriented—civil society that this report's approach highlights.

CSIS Futures Lab researchers are not the first to recognize this issue. Civil society organizations have already called for more distributed models of AI research and deployment.[90] Legislative activity in the United States, such as the CREATE AI Act, is attempting to democratize access to computing resources.[91] Moreover, a robust community of open-source AI development is highly active, illustrated by organizations such as Hugging Face, which hosts a repository of open-source models available for public use.[92] Within open-source AI development, some level of model weights, code, and model parameters are available to the public. Proponents of an open-source approach

suggest that it is far more likely to democratize access to the technology and, furthermore, is more in line with scientific practices of transparency and the evaluation of research findings. In addition, within the research community, scholars have begun to pose the question: "How can scientists co-create AI systems with local communities to address context-specific concerns?"[93]

The picture that emerges centers on a range of actors, operating within a regulatory environment that incentivizes participatory action and a democratization of access, coming together to address and experiment with solving practical governance problems that are both posed by AI and, potentially, solved through the targeted use of the same technology. Such a view is supported by modern legal theorists suggesting that a diverse set of actors with clear, vocal interests can contribute to localized policy innovation.[94] Moreover, this methodology emphasizes the associational and civic-minded approach that scholars such as James and Deborah Fallows have revealed through their writings on the United States.[95] Thus, experimentation on, and subsequent technical evaluation of, how AI will interact with democratic governance cannot simply be a top-down driven process but instead must pursue associational pluralism.

While the domains of international relations and foreign policy may seem too abstract for such an associative politics related to general AI development, and the evaluation process specifically, this does not have to be the case. Much mainstream international relations research—particularly from the 1970s and 1980s—treated states as like units, or billiard balls which interacted to either cooperate or compete.[96] Complicating the picture, more recent work, such as that cited above on the relationship between democracy and foreign policy, as well as other research in international affairs emphasizing notions of "human security"—i.e., the desire to shift the lens of security away from states and toward the needs of individuals—has expanded the aperture for what, and who, counts when it comes to international affairs.[97] Moreover, there exists a diverse array of actors with a stake in the field, ranging from citizen organizations, universities, think tanks, NGOs, and others. The point is that even the domain of international relations—at times called the realm of high politics—has an existing grammar and set of actors with which to discuss and experiment with the issues addressed in this paper.

At its core, this paper advocates for a set of AI governance and model evaluation procedures that are multi-scalar, participatory, and grounded in civil society—not simply rooted in top-down regulatory regimes that, although well meaning, may fail to understand the complexity of certain local or domain-specific problems, or that are corporate controlled, and thus driven by market incentives rather than through robust civil involvement. This effort should be independent of government institutions and corporate interests and should be nonpartisan in form. The goal is to promote the maximum exchange of information by holding providers that power agentic systems accountable. Illustrating problematic tendencies or biases can pressure firms to correct and improve model performance, thereby creating a feedback system that incentivizes open information exchange. This argument is not without basis, as research on corporate social responsibility has demonstrated that, in some circumstances, firms respond to public pressures and update business practices.[98]

Toward an Associational Model of AI Benchmarking Governance

Drawing from the prior sections on the critical nature of civic associations in forming more publicly accountable, transparent, and effective practices of AI benchmarking, this section develops a model for associational practices related to AI governance, specifically focusing on AI evaluation and benchmarking. Many proposed AI governance models are broad based and offer wholistic views of the governance environment. They do so by incorporating elements or layers that are as general as categories such as society, technical factors, and ethics. Though these concepts are useful for identifying key considerations for successful AI governance, without more direct conceptualization, they are less practical in terms of generating practices in specific contexts.[99] The broad scope of ethical AI is a good example, as scholarly reviews of the emerging AI governance literature have noted that some discussions of AI ethics struggle to identify the details of real-world implementation, possibly limiting ethical AI's practical effectiveness.[100] Yet, problematically, some proposed governance approaches fail to discuss ethical dimensions at all, including fairness, transparency, and trust.[101] In fact, some reviews illustrate a general failure to broadly operationalize processes related to AI governance.[102] Thus, focusing on tangible practices within governing structures can help to illustrate what broader categories, such as ethics, mean for practical implementation purposes.

Some researchers address the specificity of organizational governance, offering important insight as to how organizations, such as private firms, play a role in governance structures.[103] However, due to their focus on the level of the organization, these approaches can miss the critical set of players that must be involved in successful governance beyond how individual organizations set up and implement their own AI-related standards.[104] For example, levels of AI governance touch on a range of players (frequently referred to as stakeholders in the governance literature) as diverse as small teams within organizations, large international governing bodies, and even individuals impacted by the technology.[105] Moreover, a diverse array of actors has been instrumental in proposing various models and principals of AI governance, including the U.S. National Institute for Standards and Technology (NIST) AI Risk Framework, the EU AI Act, private organization frameworks from the likes of Microsoft and Google, and civil society groups, among many others.[106]

As such, not only are there multiple levels in which AI governance occurs, but there are also multiple domains of practice in which AI governance must be implemented.[107] As some have noted, governance solutions must be implemented at all stages of the AI development lifecycle, from model development to deployment.[108] This includes factors in which benchmarking plays a role, for example, those related to testing, evaluation verification, and validation.[109]

To make this discussion more practical, the following section focuses specifically on the critical area of benchmarking and its possible role in AI governance, while also acknowledging the complex array of actors involved in any truly associational model of governance. The fundamental focus here is grounding benchmarking in transparency and accountability based on domain-specific, expert-crafted data, whether the context of the evaluation process be benchmarking model reasoning capabilities on tasks such as foreign policy decisions or evaluating model performance on local community implementation of AI to assist in providing efficient services. It is through

increasing transparency and accountability that U.S. citizens will be able to better decide how (and if) the technology is working to improve their lives in meaningful ways.

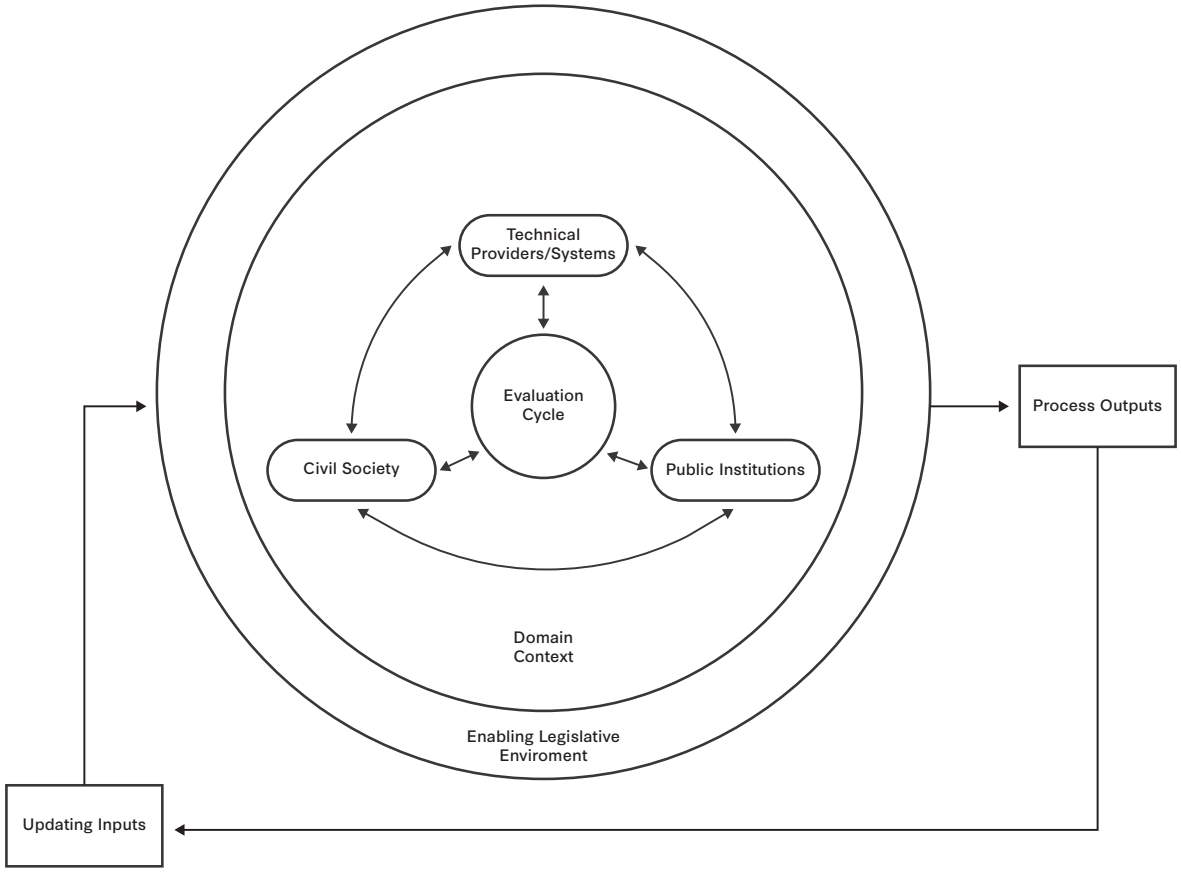
Rather than merely suggest that critical factors such as transparency are important in AI governance, this report posits a model for implementation that puts transparency and accountability into practice, thus contributing to calls from researchers to ensure that there is a "how for every what" involved in the governance of AI and ML technologies.[110] In other words, this report attempts to translate abstract principals into specific practices, roles, and benefits. Thus, the proposal outlined here is both micro-focused in that this model isolates one aspect of AI governance (evaluation and benchmarking) and *macro-focused* as it applies to a general range of domains from local community organizations to the governmental level. Simply put, it is a model that, while focused chiefly on benchmarking, can apply to many real-world use cases.

Figure 1 depicts this model. It starts by describing the constituent elements located at the central part of the model and then laying out their interactions—the most fundamental piece of the associational model. Three component parts contribute directly to the proposed model's main purpose—a robust AI evaluation and benchmarking cycle—which is further described below. Table 1 summarizes the roles and functions of the various elements involved in this cycle, as well as the benefits each element receives from participation in the process.

Civil Society. The first component part, as emphasized most directly in the above discussion, consists of civil society actors. With respect to a technology like AI that is largely a general-purpose technology, actors comprising this constituent part of the model are intentionally diverse.[111] For example, civil society actors with interests in the evaluation and benchmarking of AI could range from think tanks to universities and to local community organizations concerned with mitigating adverse impacts of poor performing AI on local services.

Within this associational model, civil society serves three key roles. The first is through the expression of interested community actors and organized groups in shaping technological performance. Due to the likely wide impacts of AI integration (from computer vision to language models), there are obvious incentives for civil society groups to become engaged in signaling their interests regarding model performance. Groups interested in AI evaluation also provide inputs related to how civil society organizations see the technology as benefiting their specific communities, along with the worries and concerns the same groups have regarding possible technological abuse and undesirable performance. As a result of this relationship, public institutions and technology companies receive important inputs on what broader social groups want—and do not want—from AI-enabled technology. These lines of communication between civil society and government and technology developers will be fundamental to ensuring that technology firms and government are aware of how civil society organizations imagine AI integrating into their daily lives.

Figure 1: Associational Benchmarking Concept



Source: CSIS Futures Lab.

Second, because civil society organizations can operate outside the incentive structures of private businesses and public institutions, they are able to provide important checks and balances by demanding transparency and accountability from other constituent parts of evaluation processes. Of course, transparency and accountability do not come from external demands alone, thus requiring important legislative enablers, but they can provide pressures for governments and private organizations to act or change behavior due to, for example, invoking reputational concerns or illustrating broad social desires for action.[112] Additionally, civil society organizations can serve as key interlocutors for publishing evaluation metrics, writing public-facing reports on benchmarking studies, and conducting reviews of evaluation results related to real-world use cases. In basic terms, civil society can push the development and deployment of AI in a direction that serves broader social interests.

Third, as discussed above, valid benchmarking datasets and evaluation processes require coherent, domain-specific knowledge that can be updated and refined as contexts change. In many circumstances, civil society organizations can assist in providing domain-relevant expertise and knowledge that typically cannot be replicated within governing institutions and private technology providers. Moreover, the organizations can serve as key partners in generating specific metrics and operationalizations

of domain-specific knowledge that can be integrated into benchmarking procedures. Without valid metrics, the construct validity of any evaluation process will be dubious. For instance, academic and scientific organizations retain a great deal of critical systematic knowledge relevant to real-world AI use cases that, absent transparent dialogue, cannot be obtained in a manner workable for designing model evaluations. The basic point is that expert-created data and evaluations will improve the quality of AI models and better demonstrate their strengths and weaknesses. To facilitate access to this data, government institutions, such as the Library of Congress, could serve as publicly accessibly data repositories for high-quality, curated, pooled data training that public organizations can access for free.

Importantly, as discussed in greater detail below, government must invest in broader AI literacy to improve the feedback mechanisms between civil society and other constituent parts of the model. Not all relevant civil society organizations with interest in participating in associative processes related to model benchmarking will have the specific skill set to make necessary contributions from the onset. If the U.S. public does not have a basic level of AI literacy, as well as access to resources to experiment with AI, civil society will be a less effective form of checks and balances.

Technical Providers and Systems. The second component part comprises technical providers and systems. As with civil society addressed above, this component can be multidimensional, including, first off, the critical actor set of private technology organizations. These organizations include the major AI developers: OpenAI, Anthropic, Meta, XAI, Google, and others. These private businesses are in the unique position of having the financial and compute resources, as well as the technical expertise, to build, train, and offer public-facing AI products. Here, the commonly discussed "black box" of AI- and ML-related technologies poses the highest risk, as these actors at times have incentives to restrict outsider access to algorithms, model weights, and the like. There is also generally a wide gap between technical knowledge within these organizations compared to the general public. The consequence is high levels of opacity due to technical literacy differences between communities.[113] Moreover, these companies typically offer application programing interfaces (APIs) that allow businesses and individuals to integrate high-performing models into their own workflows, which then may be converted into consumer-facing products, making technical providers key actors in broader business (and other organization) processes in which AI is increasingly deployed.[114] Yet technology companies are not the only actor that could qualify within this constituent part of the broader model. Open-source approaches to AI development allow for individuals and institutions to host models and to train them to their own specific purposes without access to a particular corporation's API, thus complicating the direct roles of technology provider and system developer. The basic point is that developers of AI products must be involved in governance processes for them to be effective.

Within evaluation processes, the role of these groups is critical as, in general, companies will likely continue to retain the highest degree of combined technical expertise in crafting model evaluations along with the related financial resources. Moreover, from an operational perspective, owing to their technical expertise, the technical providers and systems are important in offering guidance on converting local domain and institutional knowledge into a data structure on which a benchmark can be created and model performance measured. Due to the aforementioned motivations driving many private companies comprising this element of the model, the enabling legislative environment,

discussed in greater detail below, must structure incentives in a fashion that drives benefits for technology providers toward open and accountable participation in the benchmarking and evaluation processes. This effort at associational benchmarking will be most effective if friction between technology providers, government, and civil society is minimized through the right legal incentive structure.

Public Institutions. The third component part of the associational model presented here is public institutions. This includes organizations ranging from federal departments that are part of the executive branch to local governments that may be interested in implementing AI tools in their own daily workflows or service provision.[115] Although organizations can pursue their own bureaucratic political interests, the ideal function of public institutions, at least in a democratic context, is to implement policy in service of the public.[116] What these institutions contribute in the context of this governance model is that they are likely integrating AI into service provision and thereby impacting governing practices and policy implementation in a range of domains. In fact, recent executive guidance has directed federal departments to develop institutional policies and procedures for implementing AI.[117] The involvement of public institutions is unlikely to decrease. As such, these institutions will serve as a major touchpoint between the public and AI's real-world use cases in circumstances ranging from obtaining government benefits to courts making decisions on the likelihood of recidivism.[118] Moreover, such institutions retain significant public data on which institutional models are likely to be trained (and privacy concerns made acute) as well as specific expertise related to the institution's particular purview (for example, the Environmental Protection Agency on environmental issues or the National Institutes of Health on issues related to public health). As such, they will be critical in developing relevant benchmarks and metrics associated with public service performance.

Finally, public institutions, such as NIST, are critical in creating regulatory and standard-setting apparatuses in many scientific and technical contexts from which private organizations will derive their own institutional policies. In this space, CAISI (previously the U.S. AISI), as part of NIST, will play a key role as it operates as a standards-setting body and works with industry members to set voluntary commitments on model evaluation and security.[119] CAISI has already been an active participant in model evaluations. For example, in late 2024, alongside the UK AI Safety Institute, CAISI conducted a pre-deployment evaluation of Open AI's o1 model, focusing specifically on cyber, biological, and software and AI development.[120] Moreover, in early 2025, CAISI released a technical report related to initial evaluation results on how AI agents risk being subject to "high jacking," in which malicious actors inject unwanted instructions into AI agent workflows.[121] NIST and CAISA must continue to play a meaningful part within this benchmarking governance model.

Consequently, in terms of the associational evaluation model presented here, public institutions play a fundamental role. First, they are key targets of transparency and accountability feedback loops that are driven by civil society and civic association. If effective AI governance is to be implemented, it is up to such institutions to act on and respond to benchmarking results in meaningful ways by updating training data, workflows, and other relevant practices in response to feedback. Second, such organizations are likely to act as key data brokers in their specific domains as they work with private companies to train and deploy models on institutional use cases.

In simple terms, without the commitment of public institutions, effective implementation and governance of AI agents will remain out of reach.

Table 1: Summary of Model Elements

Element	Role or Function	Benefit
Technical Providers and Systems	▪ High level of technical expertise ▪ Experience in benchmarking and evaluation ▪ Access to financial and computing resources	▪ Better-performing technology in multiple contexts in which providers do not have expertise ▪ Higher trust in tech products, from local actors to public institutions
Public Institutions	▪ Major touchpoint between AI use case and public ▪ Significant repositories of data and expertise ▪ Important in standards development	▪ More successful tech integration ▪ Higher public trust in AI-enabled services ▪ Increased accountability and transparency from tech providers
Civil Society	▪ Expression of public interest in how technology works ▪ Checks and balances on benchmarking process; mechanism for transparency and accountability ▪ Provide domain- or context-specific expertise	▪ Meaningful stake in technology development ▪ Agency over when and how AI is used ▪ Increased AI literacy

Evaluation cycle. The three constituent parts discussed above are linked within the associational evaluation cycle, with each element providing feedback and incentives to the other component parts. Links between these component parts are representative of both communication and accountability. Responsiveness to impacted communities is fundamental to this associational model.[122] This section will briefly outline this recursive relationship along three parameters: (1) benchmark creation; (2) implementation, analysis, and dissemination of results; and (3) real-world responsiveness.

First, benchmark creation requires critical inputs from all constituent parts of the evaluation cycle. Civil society groups, for example, can be instrumental in expressing local interests, and, in certain circumstances (particularly in the case of think tanks or universities where specified knowledge is available), creating scenario-specific data on which to evaluate model performance. Thus, along this parameter, information flows to both technical providers and public institutions (as indicated by the arrows in Figure 1). Civil society groups transfer information in the forms of

expressing a civil association's specific interest when it comes to model performance, assisting in generating data for a benchmark, and providing input on which metrics account for a fair and realistic measure of performance on the relevant task. For benchmark creation, technical providers and systems offer important technical knowledge on how to design benchmarks valid for domain-specific data structures that they can share with partners in civil society and public institutions. Moreover, they can provide feedback on how technological progress in AI, largely driven by private firms, may shape requirements for updating relevant benchmarking practices. Finally, public institutions will be key in both listening to and engaging with relevant civic groups interested in involvement in benchmarking. Moreover, due to domain expertise and data access, public institutions will be crucial in generating contextual knowledge to include in benchmarking and, working recursively with interested civil society groups, in developing relevant metrics on which to assess evaluation results. Proper metrics thus will be the outcome of deliberative processes between stakeholders and based on domain-specific knowledge. Importantly, without the participation of multiple stakeholders in benchmarking, the results of benchmarking studies may lack real-world applicability.

Second, with respect to implementation, analysis, and information dissemination, civil society interacts with the other constituent parts largely in terms of interpreting benchmarking results according to their own local or domain-specific knowledge, as well as by acting as a key driver of transparency and accountability. Within the context of associational benchmarking, civil society organizations should participate in public advocacy and serve as essential providers of data and metrics to the public, particularly through creating and disseminating accessible information relevant to both their own association members and to the broader interested public. Moreover, because civil society operates outside of the for-profit incentives of private corporations and the bureaucratic pressures of public institutions, civil society organizations (particularly universities and think tanks) can offer objective analyses of evaluation results and their implications. During implementation and analysis, technology providers and systems typically should offer compute, model hosting, and other related technical skills to make evaluation processes methodologically consistent and relevant for the most advanced AI technologies (whether in computer vision, natural language processing, or others). Finally, public institutions within this parameter should focus on analyzing results specific to institutional goals based on agreed upon metrics. Along these lines, they should provide access to information to civil society organizations on evaluation results to ensure transparency and accountability to public interests. As with benchmark creation, meaningful participation of all stakeholders is crucial. This means matching local expertise with technical experience in interpreting benchmarking results and government actors maintaining awareness of how results link to their institutional goals.

Finally, as emphasized above, benchmarks should not be seen as simple technical resources, but rather as a set of socio-technical relations. As such, linking evaluation results to real-world processes of responsiveness by updating social and institutional practices as well as data resources is fundamental to this associational model functioning outside of theoretical terms. Thus, civil society actors must provide indicators to technical systems and providers, as well as public institutions, regarding how downstream AI applications are shaping their specific interests. In addition, such

groups must be prepared to apply pressure to public institutions and technical providers who may fail to commit to transparent practices and who lack accountability. When applicable, technology providers and systems must be prepared to share how their models perform on benchmarks relevant to a broad range of civic actors. In the long run, the technology providers and systems will benefit from group feedback on how their products perform in a range of contexts, thereby allowing them to make product improvements and increase user trust. Lastly, public institutions need to be responsive to updating and curating relevant institutional data while also retaining enough institutional and technical flexibility to adapt practices according to the results of evaluation. This is critical, as maintaining a stagnant approach within the highly fluid context of governance will likely lead to undesirable technological performance, and thereby worse public services. Notably, it is perfectly possible that evaluation results may demonstrate technological limitations, thus illustrating that certain use cases are not appropriate, or remain too high risk, for AI integration into institutional practices at this stage. Even with all the technical best practices imaginable, if evaluation results do not result in required change, AI will likely fail to have a positive impact on broader society.

Apart from the constituent elements of the central evaluation cycle, two constraining and enabling conditions feature in the model: the domain and the legislative environment. While each will vary based on the context in which benchmarking is needed, both will necessarily exist in some form. Moreover, both generate the type of data that will be needed for evaluation purposes and for the incentive structure under which key actors operate.

Domain. All benchmarking and evaluation processes, particularly those in contexts relevant to governing, must be shaped by domain-specific considerations. These specific considerations should orient and guide the entire benchmarking and evaluation cycle discussed above, including the technical processes of implementing evaluations; different use cases will require unique considerations as well. Moreover, the domain will influence what sort of organizations (e.g., environmental, legal, and human rights organizations) participate in the evaluation processes of AI technologies. Data curation and subsequent model evaluation should include a range of domain-specific experts, depending on the desired AI use case. In addition, this should involve leveraging civil society associations, among other possible pools of experts, in the entire evaluation lifecycle. Ignoring this critical element risks evaluation results not properly corresponding to the real-world contexts in which the technology may be used, thus seriously threatening a benchmark's construct validity, or put simply, its applicability to the real world.

Legislative Environment. All elements of the model are embedded within the legislative environment and, therefore, this environment must serve as an enabler of the associative model proposed here. While legislation and regulation are unlikely to solve all governance issues related to AI, they can create incentive structures that give value for participation to a broad range of stakeholders. Moreover, legislation can enable civic groups to have a real impact on model benchmarking and, thus, on how models are deployed. This includes supporting efforts aimed at increasing AI literacy among civil associations in local government contexts in order to ensure more robust and effective benchmarking processes and creating funding pools to enable expert civil society organizations to commit labor and other resources to creating robust evaluation cycles.

Moreover, a truly enabling legislative environment must manage power differentials between organizations that may be involved in benchmarking processes. Technological development and deployment take place within large social power structures, and the gap between the financial resources of exsisting AI companies and, for instance, locally oriented civil society groups will be vast.[123] Thus, it is important to create an environment in which civil associations, the heart of this model of governance, do not get steamrolled by more powerful interests, compromising the validity of evaluative processes and their benefits. Within this model, truly robust benchmarking can only take place if civil organizations' concerns are heard and acted upon rather than ignored in favor of more powerful corporate or political interests.

The final element of the model centers on the components that emphasize the process outputs and the requirements of updating inputs.

Process outputs. The evaluation cycle, per the stated goals of transparency and accountability, will result in multiple outputs, including publicly available metrics and reports that indicate model performance on domain-specific tasks. However, process outputs must also lead to iterative updating of institutional practices that align AI workflows and use cases with evaluation results. Moreover, new technical methodologies in domain-specific evaluation are likely to emerge through this process, and civil society organizations will gain insight into how AI impacts their direct interests, along with increasing their organizational knowledge about evaluation and benchmarking as a practice. That said, as discussed below, leaving these outputs stagnant sets up the broader process for failure; as such, recursive updating of process inputs is key. The data and metrics that result from evaluation studies must be leveraged to inform future AI development.

Updating Inputs. At its heart, the model presented here emphasizes associational politics, deliberative processes, and experimentalism. Each aspect has roots in the literatures of democratic political theory, philosophy, and experimental governance.[124] As such, this model includes the essential function of updating inputs recursively, underscoring the need for constant analysis and iterative learning through reaction to prior operations and deliberation between constituent components of the evaluation cycles. Moreover, this model emphasizes experimentation in that it calls for "ongoing, reciprocal readjustments of ends and means."[125] Actors must reflexively learn and scrutinize their own actions and interests in pursuing the collective goal.[126] This model builds these requirements as new inputs as the process cycles through continuous iteration. Without adapting to new information, actors risk deploying technology to use cases it is not built for, resulting in undesirable consequences for everyday users.

In summary, each component part of the model described above must work in conjunction with the others to achieve the desired outcome of responsive, domain-specific, benchmarking efforts. By including the proper legislative incentive structures and tangible benefits, each actor within the model can be incorporated into a cycle of positive feedback loops that can sustain the model's relevance and success. The following section will identify three recommendations that could serve as enablers to implementing this model in real-world contexts.

Policy Implications and Recommendations

It is worth linking the above discussion to ongoing developments in U.S. AI policy. As addressed in more detail above, recent guidance from the Trump administration has rearticulated much of the work done by the Biden administration in establishing national AI policy. In doing so, the Trump administration, particularly in terms of rhetoric, has preferred a more innovation-forward approach to implementing AI across the U.S. government. This is seen in a range of policy guidance from the administration, including the aforementioned executive order entitled "Removing Barriers to American Leadership in Artificial Intelligence" as well as guidance from the Office of Management and Budget (OMB) in the form of "Accelerating Federal Use of AI through Innovation, Governance, and Public Trust" and "Driving Efficient Acquisition of Artificial Intelligence in Government."[127] While innovation and adoption of AI throughout the U.S. government may be a priority of the Trump administration, the current approach risks failure without reliable and iterative evaluation practices that emphasize input from civil actors. Voices from the bottom-up need to be key drivers of technology implementation.

To be fair, the above-mentioned federal guidance appears to be amenable to many of the items proposed in the associational evaluation model of this report. For example, the OMB's memorandum on "Accelerating Federal Use of AI through Innovation, Governance, and Public Trust" notes the importance of protecting civil rights and liberties and developing responsible AI, and it calls out the critical nature of transparency, governance, and public trust.[128] Moreover, it identifies the need for continuous monitoring of AI systems, particularly in high-impact use cases, along with supporting broader AI literacy among agency employees.[129] Finally, in line with aspects drawn out in the model discussed here, this OMB memo highlights the importance of numerous and diverse stakeholders, including private actors and external experts in AI, along with the need for the independent review of AI systems, particularly in high-impact use cases.[130]

There are a few issues within the OMB guidance worth further assessment, as well as discussion regarding how the model proposed here can support and improve recent policy guidance. For example, independent reviews within the OMB document are stated to be conducted by a reviewer "within the agency" who has not been a part of system development.[131] While this is certainly an important step, within-agency review may not sufficiently isolate the reviewer from political and bureaucratic pressures. As such, incorporating civil society organizations into the process of model assessment and evaluation can add an additional layer of separation between bureaucratic political pressures and AI evaluation, thereby contributing to more objective and transparent analyses of benchmarking results. Additionally, while innovation is key, emphasizing this factor to too great a degree may undermine other components of federal guidance, including developing and implementing AI responsibly, improving public trust, and avoiding harm to U.S. citizens. Depending on the domain, incorporating a range of civil society actors into the evaluation cycle and integrating robust and transparent feedback mechanisms could facilitate a pragmatic balance between technological innovation and building systems that work as intended within specific contexts. Importantly, increasing trust and accountability within AI governance will lead to more robust and sustainable development of AI-associated technologies. Finally, while sensible in general terms, proposed efforts at improving efficiency and reusing elements of AI development

through agency coordination could lead to issues in downstream technology deployment.[132] Unrestrained efforts at improving implementation efficiency and the reuse of technical systems could lead to circumstances in which systems and data designed for one context are applied to a different, non-applicable case, increasing the risk that the system will not perform as desired due to contextual variation. This is why continuous benchmarking on domain-specific tasks, including constant iteration and feedback loops, is key.

This discussion leads to three substantive recommendations (summarized in Table 2 below). **First, legislators at the federal and state levels should pursue legislation that encourages and enables civil society actors to become involved in independent benchmarking efforts, while working with private technology providers and public institutions.** This could include offering funding resources for think tanks, universities, and even local community organizations to become meaningful stakeholders in the process of benchmarking and evaluation. This could also incorporate increasing resources for upscaling broader AI literacy in existing civil society organizations and encouraging these organizations to position themselves as key stakeholders within conversations on AI governance, benchmarking, and evaluation.

Moreover, Congress should begin to explore which congressional committees have the authority to hold routine hearings on model benchmarking and evaluation to ensure that iterative, inclusive, and robust processes remain in play across private and public sector contexts. Currently, a range of committees have at least some form of jurisdiction related to AI policy development, including committees as diverse as the Judiciary Committee in the Senate and the Foreign Affairs Committee in the House, reflecting the wide impact AI will have on questions of governance.[133] Congressional committees with interests in AI should begin the process of holding benchmarking-related hearings to hold AI firms and federal agencies accountable, while also looking to the best interests of the people they are elected to represent. Within such a process, civil society organizations can leverage their traditional advocacy roles to make local interests heard by congressional representatives, as well as offer expert testimony on the evaluation results and technological impact they observe in their local contexts.

Second, legislative bodies at all levels of government should pursue creating meaningful incentives for AI firms and public institutions to coordinate their benchmarking efforts with civic society organizations. This can help to match civil society organizations' domain expertise, local concerns, and user feedback with any deployed AI agent. Incentives could include formal legislation requiring AI companies to operate in specific ways or incentives which hope to shape corporate behavior without specific standards.[134] Legislators could even use state or federal tax law in order to manage AI firms' incentives. Historically, these tools have been used successfully in other contexts, such as influencing corporate hiring practices.[135] Of course, such a strategy would forgo tax revenue from some of the United States' largest companies, requiring a careful legislative strategy that balances desired corporate behavior with necessary tax revenue. Moreover, legislation should encourage open, transparent, and accountable evaluation practices, and could even establish standard practices for benchmark reporting.[136] Similar reporting incentives have been used with regard to corporate disclosures of supply chains and labor abuses, although their effectiveness remains a point of a debate.[137] Private firms have already offered a variety of pledges toward building more robust and safe AI systems,

for example, in the form of creating nonprofit industry groups and voluntary commitments to shared safety practices in model development and deployment.[138] Additionally, legislation could encourage these private organizations to pursue additional commitments aimed at integrating civil society actors—as well as committing to greater degrees of transparency—within the evaluation processes in line with the general model of associational benchmarking proposed here. A key consideration in legislative processes needs to be harmonizing local, state, and federal AI policies into a functioning mosaic of governance. Excessive fragmentation of AI policy could lead to difficulties as organizations attempt to navigate different rules for model evaluation and deployment in different localities. Efforts to disincentivize state-level regulations on AI look to address this issue; however, these efforts challenge core components of the U.S. federalist system. Moreover, lawmakers will need to balance the need for oversight with the desire for further innovation.[139]

Third, and perhaps most critically, the United States' philanthropic foundations can play a fundamental role in coordinating the funding of multiple major cross-disciplinary benchmarking initiatives to enhance and preserve the independence of a modern form of accountable, associational model evaluation. Without independent financial support, there is a risk that evaluation processes will be co-opted by narrow interests, thus turning a potentially productive tool of measuring the technologies' performances into something far more superficial. In fact, recent research has demonstrated that some model leaderboards related to popular AI benchmarks can be gamed by large private companies, which are incentivized to present their products as the best-performing models.[140] Involving independent actors with interests in the public good and not financial profit, and providing them with the resources to meaningfully participate, could help to mitigate such problems and make evaluation practices more robust.

While it is important to recognize that foundations, and their philanthropic contributions, have a more complex history than can be explored here, there is a clear precedent for their inclusion in this process.[141] Researchers studying foundations suggest that philanthropic contributions commonly focus on the creation of "something new," whether that be related to social arrangements, the arts, or science.[142] Moreover, foundations can serve as social entrepreneurs that "respond to needs or problems that are beyond the reach" of current market incentives or government capacity.[143] While the grant-making capabilities of foundations are key, particularly in the specific contexts of supporting research initiatives, these organizations also have the ability to leverage social mechanisms that legitimate new forms of organization and that develop professionals to support that same organizational infrastructure.[144]

Foundations also play a key role in supporting civil society actors, such as NGOs. For example, the MacArthur Foundation and the Mott Foundation both provide grants to various NGOs working in the United States and globally.[145] The grants range from support for organizations working on managing global crises to those addressing climate change. Historically, foundations have enabled key research in the social and hard sciences. The Rockefeller Foundation played a key role in the establishment of biomedical research and molecular biology.[146] Grants by the MacArthur Foundation helped to build up a range of research institutions working on managing risks from nuclear weapons, such as Stanford's Center for International Security and Cooperation and

Harvard's Belfer Center.[147] The Russell Sage Foundation has focused on supporting social science researchers and contributed to the development of "social indicators" to help inform data-driven social policy. The editors of a foundation-supported volume asserted that such indicators are "imperative for the guidance of social policy."[148] As a final example, the Gates Foundation is a critical cog in foundation-based global health and development initiatives.[149] Foundations thus support a wide range of pursuits in the social and hard sciences across a range of domains.

Critical for this discussion is that the broader fields of computer science and artificial intelligence also have roots in foundation support. As a significant example, the Macy Foundation contributed to a series of conferences beginning in the 1940s that set the stage for the development of the field of "cybernetics." While the field is far reaching in scope, impacting disciplines as diverse as neuroscience and various social sciences, cybernetics was highly influential in early research into computing and AI.[150] Its influence largely lies in the introduction of concepts such as feedback mechanisms, control theory, and complex systems. In basic terms, cybernetics focuses on how feedback loops shape behavioral patterns in both human and machine systems. Key attendees of the Macy conferences included Norbert Wiener, generally considered the father of cybernetics; John von Neumann, the creator of the Von Neumann architecture, which remains an influential paradigm in modern computer design; and Warren McCulloch and Walter Pitts, key figures in the development of the initial model of the neural network.[151] Foundation support can be said to have helped create the conditions for today's advances in AI and computing.

Foundation grants in the domain of AI continue to be influential across a diverse spectrum of viewpoints. Philanthropic organizations, such as Open Philanthropy and the Future of Life Institute, provide grants for AI-related research, particularly focusing on managing "existential risks."[152] Moreover, established foundations such as the Ford and MacArthur Foundations have funded AI-related organizations, including the AI Now Institute, which focuses on issues of AI-related surveillance and power consolidation in the hands of technology companies.[153]

The history and current funding profile of U.S. foundations have clear relevance for the CSIS Futures Lab's proposed associational model of AI evaluation and benchmarking. While the foundation approach to funding is not perfect, it does play a critical role in how research organizations and civil society groups gain support for their initiatives. For example, foundation contributions could encourage the emergence of new civil society organizations oriented toward transparent benchmarking practices across a range of domains.[154] This type of foundation support could fill in the gaps where market or government incentives for such funding may be limited. Moreover, foundation-based funding efforts could assist in creating new research fellowship streams that are focused directly on bringing domain-specific experts into benchmarking processes, whether these be individuals from small local organizations focused on community issues or experienced researchers with specific disciplinary skills and knowledge. However, effective relationships will need to result in long-term funding commitments to avoid funding gaps and to reduce the risk that funding organizations rapidly change priorities.

The recommendations outlined above seek to create a policy environment that enables AI firms and civil society groups to work together in a productive fashion to produce more robust and dynamic

benchmarking outputs. Such cooperation will improve the likelihood that AI works for broader segments of society on tasks that require specific expertise and knowledge. Moreover, it suggests that U.S. foundations can play a crucial part in establishing civil society groups that have the talent and resources to participate in this associational model of AI benchmarking and evaluation.

Table 2: Policy Recommendations Summary

Recommendation	Key Actions	Implications and Benefits	Potential Challenges
1. Legislative Support for Civil Society Involvement	▪ Fund think tanks, universities, and community organizations for benchmarking ▪ Increase AI literacy resources for civil society ▪ Hold routine congressional hearings on model evaluation	▪ More independent oversight of AI systems ▪ Enhanced accountability through local stakeholder input ▪ Regular legislative review of benchmarking practices	▪ Coordination across multiple congressional committees ▪ Balancing diverse congressional interests ▪ Ensuring sustained funding commitments
2. Incentivize AI Firm–Civil Society Coordination	▪ Create tax incentives for collaborative benchmarking ▪ Require transparent evaluation-reporting standards ▪ Encourage voluntary industry commitments	▪ Better alignment between AI development and public interest ▪ Improved corporate accountability	▪ Loss of tax revenue from major companies ▪ Risk of regulatory fragmentation ▪ Balancing innovation with oversight requirements
3. Foundation-Led Independent Funding	▪ Support cross-disciplinary benchmarking initiatives ▪ Fund civil society organizations focused on AI evaluation ▪ Provide scholarly fellowships for benchmarking research	▪ Maintains independence from narrow commercial interests ▪ Prevents co-optation of evaluation processes by firms ▪ Builds out professional workforce for AI governance	▪ Possible dependence on philanthropic priorities ▪ Potential gaps in foundation funding ▪ Need for sustained long-term commitments

Conclusion

In the near term, AI agents will not simply function as a tool but will shape (and to a degree already are shaping) how decisions are made and who gets to make them. Whether it be in the context of national security and foreign policy decisions or related to getting access to government benefits, decisions shaped by AI systems are set to influence people's life outcomes. To avoid a scenario in which society relinquishes a detrimental amount of agency to opaque AI systems, thereby challenging democratic principles of governance, a renewed form of associative action is required. This form of association must bring many stakeholders to the table—including policymakers, researchers, civic leaders, and interested members of impacted communities—to generate robust, dynamic, and responsive benchmarking practices. Critically, this process must be transparent and open. As others have argued, "transparency is an essential precondition for public accountability, scientific innovation, and effective governance of digital technologies. Without adequate transparency, stakeholders cannot understand foundation models, who [sic] they affect, and the impact they have on society."[155] Moreover, researchers have suggested that transparency can be fundamental for "reducing the mystique and opaqueness of AI to the general public."[156] While research has shown that transparency is not a direct cause of accountability, it is a necessary condition.[157]

In line with this goal and drawing from the work of Tocqueville, scholarship in deliberative democracy, and experimentalist approaches to governance, this analysis has presented an associational model of benchmarking to assist in visualizing our broader argument. Moreover, this report has offered a variety of recommendations related to government and foundation policy that could assist in enabling this vision to become a reality. Fundamentally, in an open and democratic

society, it is not enough to build powerful and capable models–we must also build the civic institutions and practices capable of questioning them.

About the Authors

Benjamin Jensen is director of the Futures Lab and a senior fellow for the Defense and Security Department at the Center for Strategic and International Studies (CSIS). At CSIS, Dr. Jensen leads research initiatives on applying data science and AI and machine learning to study the changing character of war and statecraft. Under his leadership, Futures Lab has pioneered building AI applications into wargames and innovative scenario exercises. The exercise topics range from major war, competitive strategy, and national mobilization to economic security, energy politics, and national resilience. He is also the Frank E. Petersen Chair for Emerging Technology and a professor of strategic studies at the Marine Corps University School of Advanced Warfighting (MCU). At MCU, he leads a research program on future war and teaches seminars on modern operational art and joint-all domain operations. Dr. Jensen has authored five books, including *Information at War: Military Innovation, Battle Networks, and the Future of Artificial Intelligence* (Georgetown University Press, 2022), *Military Strategy in the 21st Century: People, Connectivity, and Competition* (Cambria, 2018), *Cyber Strategy: The Evolving Character of Power and Coercion* (Oxford University Press, 2018), and *Forging the Sword: Doctrinal Change in the U.S. Army* (Stanford University Press 2016). He also served as senior research director for the U.S. Cyberspace Solarium Commission and is a reserve officer in the U.S. Army, with command experience from platoon to battalion. Dr. Jensen graduated from the University of Wisconsin-Madison and earned his MA and PhD from the American University School of International Service.

Ian Reynolds is the postdoctoral fellow for the Futures Lab in the International Security Program at the Center for Strategic and International Studies. Ian holds a PhD in international relations from American University's School of International Service. His research focuses on the intersection

of technology, science, and international security. Ian's dissertation addressed the history and cultural politics of integrating artificial intelligence into military decision-making processes in the United States. From 2022 to 2023, he was a pre-doctoral fellow at Stanford University's Center for International Security and Cooperation and the Institute for Human-Centered AI. Ian's work has appeared in publications including War on the Rocks, E-International Relations, and the *Bulletin of the Atomic Scientists*.

Endnotes

1 Benjamin Jensen and Matthew Strohmeyer, *Rethinking the Napoleonic Staff: Agentic Warfare and the Future of Military Operations* (Washington, DC: CSIS, July 2025), https://www.csis.org/analysis/rethinking-napoleonic-staff.

2 "Treasury Department Now Using AI to Save Taxpayers Billions," NBC News, October 17, 2024, https://www.nbcnews.com/business/consumer/how-ai-artificial-intelligence-fights-taxpayer-fraud-treasury-dept-rcna175916; U.S. Department of Commerce, "DOC Artificial Intelligence (AI) Use Case Inventory," accessed June 11, 2025, https://www.commerce.gov/data/AI_inventory; "AI for Energy," Energy.gov, accessed June 11, 2025, https://www.energy.gov/cet/articles/ai-energy; Pascaline Gaborit, "A Sociopolitical Approach to Disinformation in AI: Concerns, Responses, and Challenges," *Journal of Political Science and International Relations* 7, no. 4 (2024): 75–88; Elizaveta Kuznetsova et al., "In Generative AI We Trust: Can Chatbots Effectively Verify Political Information?," *Journal of Computational Social Science* 8, no. 15 (2025): 1–31; and Kostas Gemenis, "Artificial Intelligence and Voting Advice Applications," *Frontiers in Political Science* (2024): 1–15.

3 Ian Reynolds and Yasir Atalan, "Calibrating NATO's Vision of AI-Enabled Decision Support," CSIS, *Commentary*, July 8, 2024, https://www.csis.org/analysis/calibrating-natos-vision-ai-enabled-decision-support; and Benjamin Jensen, "Building a Brain of the Army Through Professional Military Education," CSIS, *Commentary*, June 10, 2025, https://www.csis.org/analysis/building-brain-army-through-professional-military-education.

4 Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Hoboken: Pearson, 2021); and Naveen Krishnan, "AI Agents: Evolution, Architecture, and Real-World Applications," arXiv.org, March 16, 2025, https://doi.org/10.48550/arXiv.2503.12687.

5 Alexis de Tocqueville, *Democracy in America and Two Essays on America*, ed. Isaac Kramnick, trans. Gerald Bevan (London: Penguin Classics, 2003), 591.

6 Ibid., 597.

7 Ibid., 592–93.

8 Ibid., 600.

9 Cary Coglianese and David Lehr, "Transparency and Algorithmic Governance," *Administrative Law Review* 71, no. 1 (2019): 1–56.

10 Christian Katzenbach and Lena Ulbricht, "Algorithmic Governance," *Internet Policy Review* 8, no. 4 (November 29, 2019): 1, https://doi.org/10.14763/2019.4.1424.

11 Coglianese and Lehr, "Transparency and Algorithmic Governance," 2.

12 Thomas Carothers and William Barndt, "Civil Society," *Foreign Policy* 117 (Winter 1999–2000): 18, 21, https://doi.org/10.2307/1149558; and Sidney Tarrow, "Making Social Science Work Across Space and Time: A Critical Reflection on Robert Putnam's Making Democracy Work," *American Political Science Review* 90, no. 2 (June 1996): 389, https://doi.org/10.2307/2082892.

13 Larry Diamond, "Rethinking Civil Society: Toward Democratic Consolidation," *Journal of Democracy* 5, no. 3 (1994): 4.

14 Ibid., 4.

15 Robert D. Putnam, Robert Leonardi, and Raffaella Y. Nanetti, *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton: Princeton University Press, 1994), 173–74.

16 Putnam, Leonardi, and Nanetti, 182.

17 Tarrow, "Making Social Science Work Across Space and Time," 390.

18 Margaret E. Keck and Kathryn A. Sikkink, *Activists beyond Borders: Advocacy Networks in International Politics* (Ithaca: Cornell University Press, 2014).

19 Though the impact of such practices is debated. See Yuan Zhou et al., "New Evidence That Naming and Shaming Influences State Human Rights Practices," *Journal of Human Rights* 22, no. 4 (August 8, 2023): 451–68, https://doi.org/10.1080/14754835.2022.2122785; and Emilie M. Hafner-Burton, "Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem," *International Organization* 62, no. 4 (October 2008): 689–716, https://doi.org/10.1017/S0020818308080247.

20 Marcia Grimes, "The Contingencies of Societal Accountability: Examining the Link Between Civil Society and Good Government," *Studies in Comparative International Development* 48, no. 4 (December 2013): 380–402, https://doi.org/10.1007/s12116-012-9126-3.

21 Although, as Fox notes, the common assumption that transparency necessarily leads to accountability is mistaken. See Jonathan Fox, "The Uncertain Relationship between Transparency and Accountability," *Development in Practice* 17, no. 4–5 (August 1, 2007): 663–71, https://doi.org/10.1080/09614520701469955.

22 Manuel Castells, *The Rise of the Network Society*, 2nd ed. (Oxford: Wiley-Blackwell, 2010), xxxi.

23 Lee Rainie and Barry Wellman, *Networked: The New Social Operating System* (Cambridge: MIT Press, 2014), 9.

24 Yanuar Nugroho, "Adopting Technology, Transforming Society: The Internet and the Reshaping of Civil Society Activism in Indonesia," *Transforming Society* 6, no. 2 (2008): 78.

25 Castells, *The Rise of the Network Society*, 3.

26 Michael Chan and Jingjing Yi, "Social Media Use and Political Engagement in Polarized Times. Examining the Contextual Roles of Issue and Affective Polarization in Developed Democracies," *Political Communication* 41, no. 5 (September 2, 2024): 743–62, https://doi.org/10.1080/10584609.2024.2325423; and Tony Roberts and Marjoke Oosterom, "Digital Authoritarianism: A Systematic Literature Review," *Information Technology for Development* November 24, 2024, 1–25, https://doi.org/10.1080/02681102.2024.2425352.

27 Science and technology studies (STS), itself a diverse scholarly field, approaches technology not just as a material or instrumental artifact, but instead as a product of human practices, culture, and values. See Eric Schatzberg, *Technology: Critical History of a Concept* (Chicago: University of Chicago Press, 2018), 8–14. While the material component of technology is not disregarded by such work, STS attempts to broaden mainstream conceptions of technology's relationship with society by illustrating how thin views of technological development do not explain either how technologies come into being or their broader social impacts. See Bruno Latour and Steve Woolgar, *Laboratory Life: The Construction of Scientific Facts* (Princeton: Princeton University Press, 2013); Sheila Jasanoff and Sang-hyun Kim, *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (Chicago: University of Chicago Press, 2015). In defense-related contexts, this perspective has been used to address issues ranging from nuclear weapons to the integration of AI and advanced computation into defense spaces. See, among others, Benjamin M. Jensen, Christopher Whyte, and Scott Cuomo, *Information in War: Military Innovation, Battle Networks, and the Future of Artificial Intelligence* (Washington, D.C.: Georgetown University Press, 2022), 23–24; Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge: MIT Press, 1996); and Ian J. Reynolds, "Speed and War in US Military Thought: Mapping the Conditions for AI-Enabled Decision-Making," *Millennium* (April 1, 2025), https://doi.org/10.1177/03058298251317205.

28 Schatzberg, *Technology*, 10–11.

29 Castells, *The Rise of the Network Society*, 5.

30 Latour and Woolgar, *Laboratory Life*.

31 Langdon Winner, "Do Artifacts Have Politics?," *Daedalus* 109, no. 1 (1980): 121–36.

32 Ibid., 123-24.

33 Hans K. Klein and Daniel Lee Kleinman, "The Social Construction of Technology: Structural Considerations," *Science, Technology, & Human Values* 27, no. 1 (January 2002): 40, https://doi.org/10.1177/016224390202700102; Donna J. Haraway, "A Cybord Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century," in *In Simians, Cyborgs, and Women: The Reinvention Of Nature* (New York: Routledge, 1991), 150; Jasanoff and Kim, *Dreamscapes of Modernity*; Reynolds, "Speed and War in US Military Thought"; and Thomas Christian Bächle and Jascha Bareis, "'Autonomous Weapons' as a Geopolitical Signifier in a National Power Play: Analysing AI Imaginaries in Chinese and US Military Policies," *European Journal of Futures Research* 10, no. 1 (September 2, 2022): 20, https://doi.org/10.1186/s40309-022-00202-w.

34 Winner, "Do Artifacts Have Politics?," 127.

35 Sheri Berman, "Civil Society and the Collapse of the Weimar Republic," *World Politics* 49, no. 3 (1997): 401–29.

36 Dylan Riley, "Civic Associations and Authoritarian Regimes in Interwar Europe: Italy and Spain in Comparative Perspective," *American Sociological Review* 70, no. 2 (April 2005): 288–310, https://doi.org/10.1177/000312240507000205.

37 Putnam, Leonardi, and Nanetti, *Making Democracy Work*, 6, 182. Oddly, Putnam contextualizes and generalizes in the same paragraph in his concluding chapter.

38 John S. Dryzek, "Pragmatism and Democracy: In Search of Deliberative Publics," *The Journal of Speculative Philosophy* 18, no. 1 (2004): 72.

39 Fabrizio Ferraro, Dror Etzion, and Joel Gehman, "Tackling Grand Challenges Pragmatically: Robust Action Revisited," *Organizational Studies* 36, no. 3 (2015): 369, https://doi.org/10.1177/0170840614563742.

40 Christopher K. Ansell, *Pragmatist Democracy: Evolutionary Learning as Public Philosophy* (New York: Oxford University Press, 2011), 5.

41 Ibid., 8.

42 John Dewey, *Reconstruction in Philosophy* (New York: H. Holt, 1920); and Josh Whitford, "Pragmatism and the Untenable Dualism of Means and Ends: Why Rational Choice Theory Does Not Deserve Paradigmatic Privilege," *Theory and Society* 31 (2002): 340.

43 Ferraro, Etzion, and Gehman, "Tackling Grand Challenges Pragmatically," 364.

44 John Dewey, "Creative Democracy: The Task Before Us," in *John Dewey: The Later Works, 1925–1953*, vol. 14 (Carbondale: Southern Illinois University Press, 1976 [1939]), 225.

45 IBM, "What Are LLM Benchmarks?," January 29, 2024, https://www.ibm.com/think/topics/llm-benchmarks; Anka Reuel et al., "BetterBench: Assessing AI Benchmarks, Uncovering Issues, and Establishing Best Practices," arXiv.org, November 20, 2024, https://doi.org/10.48550/arXiv.2411.12990; Inioluwa Deborah Raji et al., "AI and the Everything in the Whole Wide World Benchmark," arXiv.org, November 26, 2021, http://arxiv.org/abs/2111.15366.

46 Reuel et al., "BetterBench."

47 Yubo Wang et al., "MMLU-Pro: A More Robust and Challenging Multi-Task Language Understanding Benchmark," arXiv.org, November 6, 2024, https://doi.org/10.48550/arXiv.2406.01574.

48 Alicia Parrish et al., "BBQ: A Hand-Built Bias Benchmark for Question Answering," arXiv.org, March 15, 2022, https://doi.org/10.48550/arXiv.2110.08193.

49 Benjamin Jensen et al., "Critical Foreign Policy Decisions (CFPD)-Benchmark: Measuring Diplomatic Preferences in Large Language Models," arXiv.org, March 8, 2025, https://doi.org/10.48550/arXiv.2503.06263.

50 Raji et al., "AI and the Everything in the Whole Wide World Benchmark," 5.

51 Reuel et al., "BetterBench," 10; and Shivalika Singh et al., "The Leaderboard Illusion," arXiv.org, April 29, 2025, https://doi.org/10.48550/arXiv.2504.20879.

52 Tocqueville, *Democracy in America*, 600–604.

53 Winner, "Do Artifacts Have Politics?"; and Trevor J. Pinch and Wiebe E. Bijker, "The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other," *Social Studies of Science* 14, no. 3 (August 1, 1984): 425, https://doi.org/10.1177/030631284014003004.

54 The White House, "Executive Order on Maintaining American Leadership in Artificial Intelligence," accessed June 12, 2025, https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

55 Winston Luo, "President Trump Issues Executive Order to Maintain American Leadership in Artificial Intelligence," *Harvard Journal of Law & Technology* (March 6, 2019), https://jolt.law.harvard.edu/digest/president-trump-issues-executive-order-to-maintain-american-leadership-in-artificial-intelligence.

56 Select Committee on Artificial Intelligence of the National Science and Technology Council, "The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update," June 2019, https://trumpwhitehouse.archives.gov/wp-content/uploads/2019/06/National-AI-Research-and-Development-Strategic-Plan-2019-Update-June-2019.pdf.

57 "The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update," 19–21.

58 The White House, "Blueprint for an AI Bill of Rights," accessed June 12, 2025, https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/.

59 The White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," October 30, 2023, https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

60 "1 Year Later, How Has the White House AI Executive Order Delivered on Its Promises?," Brookings (blog), accessed June 12, 2025, https://www.brookings.edu/articles/one-year-later-how-has-the-white-house-ai-executive-order-delivered-on-its-promises/.

61 U.S. Department of Commerce, "U.S. AI Safety Institute Establishes New U.S. Government Taskforce to Collaborate on Research and Testing of AI Models to Manage National Security Capabilities & Risks," press release, November 20, 2024, https://www.commerce.gov/news/press-releases/2024/11/us-ai-safety-institute-establishes-new-us-government-taskforce.

62 The White House, "Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence," October 24, 2024, https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2024/10/24/memorandum-on-advancing-the-united-states-leadership-in-artificial-intelligence-harnessing-artificial-intelligence-to-fulfill-national-security-objectives-and-fostering-the.

63 The White House, "Executive Order on Advancing United States Leadership in Artificial Intelligence Infrastructure," January 14, 2025, https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2025/01/14/executive-order-on-advancing-united-states-leadership-in-artificial-intelligence-infrastructure/.

64 The White House, "Fact Sheet: President Donald J. Trump Takes Action to Enhance America's AI Leadership," January 23, 2025, https://www.whitehouse.gov/fact-sheets/2025/01/fact-sheet-president-donald-j-trump-takes-action-to-enhance-americas-ai-leadership/.

65 The White House, "Removing Barriers to American Leadership in Artificial Intelligence," January 23, 2025, https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

66 Brooke Tanner, "New OMB Memos Signal Continuity in Federal AI Policy," Brookings (blog), 2025, https://www.brookings.edu/articles/new-omb-memos-signal-continuity-in-federal-ai-policy/.

67 Ibid.

68 Ibid.

69 U.S. Department of Commerce, "Statement from U.S. Secretary of Commerce Howard Lutnick on Transforming the U.S. AI Safety Institute into the Pro-Innovation, Pro-Science U.S. Center for AI Standards and Innovation," press release, June 3, 2025, https://www.commerce.gov/news/press-releases/2025/06/statement-us-secretary-commerce-howard-lutnick-transforming-us-ai.

70 "Center for AI Standards and Innovation (CAISI)," NIST, October 26, 2023, https://www.nist.gov/caisi.

71 Zeev Moaz and Bruce Russett, "Normative and Structural Causes of Democratic Peace, 1946–1986," *American Political Science Review* 87, no. 3 (1993): 624–38.

72 Michael Tomz, Jessica L. P. Weeks, and Keren Yarhi-Milo, "Public Opinion and Decisions About Military Force in Democracies," *International Organization* 74, no. 1 (January 2020): 119–43, https://doi.org/10.1017/S0020818319000341; and T. Knecht and M. S. Weatherford, "Public Opinion and Foreign Policy: The Stages of Presidential Decision Making," International Studies Quarterly 50, no. 3 (2006): 705–27.

73 Matthew A. Baum and Philip B. K. Potter, *War and Democratic Constraint: How the Public Influences Foreign Policy* (Princeton: Princeton University Press, 2015).

74 Matthew A. Baum and Philip B. K. Potter, "Media, Public Opinion, and Foreign Policy in the Age of Social Media," *The Journal of Politics* 81, no. 2 (April 2019): 747, https://doi.org/10.1086/702233.

75 Jensen et al., "Critical Foreign Policy Decisions (CFPD)-Benchmark."

76 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2015).

77 Haque M. Shamsul, "The Emerging Challenges to Bureaucratic Accountability: A Critical Perspective," in *Handbook of Bureaucracy* (New York: Routledge, 1994).

78 Tocqueville, *Democracy in America*, 591.

79 Faiz Surani et al., "Cleaning Up Policy Sludge: An AI Statutory Research System," Stanford Institute for Human Centered AI, 2025.

80 Tocqueville, *Democracy in America*, 592.

81 As Tocqueville argues, citizens can meet once, "and forever know how to meet again," Tocqueville, 605.

82 Dewey, "Creative Democracy: The Task Before Us," 226; and Tocqueville, *Democracy in America*, 612.

83 Robert D. Putnam, "Bowling Alone," in *City Reader* (Oxford: Taylor Francis, 2020), 308.

84 Paul M. Barrett, Justin Hendrix, and J. Grant Sims, *Fueling the Fire: How Social Media Intensifies US Political Polarization—And What Can Be Done About It* (New York: NYU Stern Center for Business and Human Rights, 2021), https://bhr.stern.nyu.edu/wp-content/uploads/2024/03/NYUCBHRFuelingTheFire_FINALONLINEREVISEDSep7.pdf.

85 Yen-Chia Hsu et al., "Empowering Local Communities Using Artificial Intelligence," *Patterns 3*, no. 3 (March 11, 2022): 100449, https://doi.org/10.1016/j.patter.2022.100449.

86 Maciej Kuziemski and Gianluca Misuraca, "AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings," *Telecommunications Policy* 44, no. 6 (July 2020): 2, https://doi.org/10.1016/j.telpol.2020.101976.

87 Frontier Model Forum, "FMF Announces First-Of-Its-Kind Information-Sharing Agreement," March 28, 2025, https://www.frontiermodelforum.org/updates/fmf-announces-first-of-its-kind-information-sharing-agreement/; "Voluntary Commitments from Leading Artificial Intelligence Companies on July 21, 2023," *Harvard Law Review* (February 12, 2024), https://harvardlawreview.org/print/vol-137/voluntary-commitments-from-leading-artificial-intelligence-companies-on-july-21-2023/; and "AI Companies Promised to Self-Regulate One Year Ago. What's Changed?," *MIT Technology Review*, accessed April 25, 2025, https://www.technologyreview.com/2024/07/22/1095193/ai-companies-promised-the-white-house-to-self-regulate-one-year-ago-whats-changed/.

88 Kari Briski, "How Scaling Laws Drive Smarter, More Powerful AI," NVIDIA Blog (blog), February 12, 2025, https://blogs.nvidia.com/blog/ai-scaling-laws/. Recent debates about the future of scaling in AI have emerged, see, for example, Arvind Narayanan, "Is AI Progress Slowing Down?," AI Snake Oil, April 15, 2025, https://www.aisnakeoil.com/p/is-ai-progress-slowing-down.

89 Tocqueville, *Democracy in America*, 80.

90 Distributed AI Research Institute (DAIR), accessed April 25, 2025, https://www.dair-institute.org/.

91 Martin Heinrich, "Creating Resources for Every American to Experiment With Artificial Intelligence (CREATE AI) Act," Pub. L. No. S. 2714 (2024), https://www.congress.gov/bill/118th-congress/senate-bill/2714.

92 "Models," Hugging Face, accessed April 25, 2025, https://huggingface.co/models.

93 Hsu et al., "Empowering Local Communities Using Artificial Intelligence," 1.

94 Charles Tyler and Heather K. Gerken, "The Myth of the Laboratories of Democracy," *Columbia Law Review* 122 (2022), https://doi.org/10.2139/ssrn.3902092.

95 James Fallows and Deborah Fallows, *Our Towns: A 100,000-Mile Journey into the Heart of America* (New York: Knopf Doubleday Publishing Group, 2018).

96 Stephen M. Walt, "Alliance Formation and the Balance of World Power," *International Security* 9, no. 4 (1985): 3–43, https://doi.org/10.2307/2538540; Kenneth A. Waltz, *Theory of International Politics* (Boston: Addison-Wesley Publishing Company, 1979).

97 Ken Booth, "Security and Emancipation," *Review of International Studies* 17, no. 4 (1991): 313–26; Amitav Acharya, "Human Security: East versus West," *International Journal* 56, no. 3 (2001): 442, https://doi.org/10.2307/40203577.

98 Advocacy and public pressure do not always work, and are mediated by variables such as broader levels of public awareness, such as in the case of labor rights and consumer behavior. See Morton Winston, "NGO Strategies for Promoting Corporate Social Responsibility," *Ethics & International Affairs* 16, no. 1 (March 2002): 86, https://doi.org/10.1111/j.1747-7093.2002.tb00376.x; and Maria Shao, "Social Pressures Affect Corporate Strategy and Performance," Stanford Graduate School of Business, accessed June 13, 2025, https://www.gsb.stanford.edu/insights/social-pressures-affect-corporate-strategy-performance; Kristen Bell DeTienne and Lee W. Lewis, "The Pragmatic and Ethical Barriers to Corporate Social Responsibility Disclosure: The Nike Case," *Journal of Business Ethics* 60, no. 4 (September 2005): 359–76, https://doi.org/10.1007/s10551-005-0869-x.

99 Urs Gasser and Virgilio A.F. Almeida, "A Layered Model for AI Governance," *IEEE Internet Computing* 21, no. 6 (November 2017): 58–62, https://doi.org/10.1109/MIC.2017.4180835.

100 Amna Batool, Didar Zowghi, and Muneera Bano, "AI Governance: A Systematic Literature Review," *AI and Ethics* 5 (January 14, 2025): 3265–79, at 9, https://doi.org/10.1007/s43681-024-00653-w.

101 Batool, Zowghi, and Bano, 10.

102 Teemu Birkstedt et al., "AI Governance: Themes, Knowledge Gaps and Future Agendas," *Internet Research* 33, no. 7 (January 1, 2023): 133, https://doi.org/10.1108/INTR-01-2022-0042.

103 Birkstedt et al., "AI Governance"; and Emmanouil Papagiannidis et al., "Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes," *Information Systems Frontiers* 25, no. 1 (February 1, 2023): 123–41, https://link.springer.com/article/10.1007/s10796-022-10251-y.

104 Matti Mäntymäki et al., "Defining Organizational AI Governance," *AI and Ethics* 2, no. 4 (November 1, 2022): 603–609, https://doi.org/10.1007/s43681-022-00143-x.

105 Harris Gleckman, *Multistakeholder Governance and Democracy: A Global Challenge* (London: Routledge, 2018), https://doi.org/10.4324/9781315144740; and Birkstedt et al., "AI Governance," 146.

106 "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," National Institute of Standards and Technology, 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf; EU Artificial Intelligence Act, "Up-to-Date Developments and Analyses of the EU AI Act," accessed May 2, 2025, https://artificialintelligenceact.eu/; "Microsoft Responsible AI Standard, V2," Microsoft (blog), 2022), https://msblogs.thesourcemediaassets.com/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf; "Responsible AI: Our 2024 Report and Ongoing Work," Google (blog), February 4, 2025, https://blog.google/technology/ai/responsible-ai-2024-report-ongoing-work/; and Kate Jones, "AI Governance and Human Rights: Resetting the Relationship," Royal Institute of International Affairs, January 10, 2023, https://doi.org/10.55317/9781784135492.

107 Batool, Zowghi, and Bano, "AI Governance," 10.

108 Ibid., 8.

109 NIST, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," 9.

110 Jessica Morley et al., "From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices," *Science and Engineering Ethics* 26, no. 4 (August 1, 2020): 2143, https://doi.org/10.1007/s11948-019-00165-5.

111 Timothy Bresnahan, "General Purpose Technologies," in *Handbook of the Economics of Innovation*, Vol. 2, ed. Bronwyn H. Hall and Nathan Rosenberg (Amsterdam: Elsevier, 2010), 761–91, https://doi.org/10.1016/S0169-7218(10)02002-2.

112 Michael Jacobs, "High Pressure for Low Emissions: How Civil Society Created the Paris Climate Agreement," *Juncture* 22, no. 4 (March 1, 2016): 314–23, https://doi.org/10.1111/j.2050-5876.2016.00881.x; and David Birchall, "The Role of Civil Society and Human Rights Defenders in Corporate Accountability," in *Research Handbook on Human Rights and Business*, ed. Surya Deva and David Birchall (Cheltenham: Edward Elgar Publishing, 2020), 422–25, https://www.elgaronline.com/edcollchap/edcoll/9781786436399/9781786436399.00030.xml.

113 Duri Long and Brian Magerko, "What Is AI Literacy? Competencies and Design Considerations," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu: ACM, 2020), 1–16, https://doi.org/10.1145/3313831.3376727.

114 For a description of application programming interfaces, see IBM, "What Is an API (Application Programming Interface)?," April 9, 2024, https://www.ibm.com/think/topics/api.

115 Mark Jackley, "Using AI in Local Government: 10 Use Cases," Oracle.com, accessed April 30, 2025, https://www.oracle.com/artificial-intelligence/ai-local-government/; Tan Yigitcanlar et al., "Unlocking Artificial Intelligence Adoption in Local Governments: Best Practice Lessons from Real-World Implementations," *Smart Cities* 7, no. 4 (August 2024): 1576–1625, https://doi.org/10.3390/smartcities7040064.

116 Thomas H. Hammond, "Agenda Control, Organizational Structure, and Bureaucratic Politics," *American Journal of Political Science* 30, no. 2 (1986): 379–420.

117 Russel T. Vought, *Accelerating Federal Use of AI through Innovation*, Governance, and Public Trust (Washington, DC: U.S. Office of Management and Budget, 2025).

118 Khalifa Alhosani and Saadat M. Alhashmi, "Opportunities, Challenges, and Benefits of AI Innovation in Government Services: A Review," *Discover Artificial Intelligence* 4 (March 4, 2024): 18, https://doi.org/10.1007/s44163-024-00111-w; and Michael Mayowa Farayola et al., "Ethics and Trustworthiness of

AI for Predicting the Risk of Recidivism: A Systematic Literature Review," *Information* 14, no. 8 (August 2023): 426, https://doi.org/10.3390/info14080426.

119 "Center for AI Standards and Innovation (CAISI)."

120 NIST, "Pre-Deployment Evaluation of OpenAI's o1 Model," update, December 18, 2024, https://www.nist.gov/news-events/news/2024/12/pre-deployment-evaluation-openais-o1-model.

121 U.S. AI Safety Institute Technical Staff, "Technical Blog: Strengthening AI Agent Hijacking Evaluations," NIST (blog), January 17, 2025, https://www.nist.gov/news-events/news/2025/01/technical-blog-strengthening-ai-agent-hijacking-evaluations.

122 Jacob Metcalf et al., "Algorithmic Impact Assessments and Accountability: The Co-Construction of Impacts," in *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (New York: Association for Computing Machinery, 2021), 735–46, https://doi.org/10.1145/3442188.3445935.

123 Klein and Kleinman, "The Social Construction of Technology."

124 Tocqueville, *Democracy in America*; Dewey, *Reconstruction in Philosophy*; John Dewey and Melvin L. Rogers, *The Public and Its Problems: An Essay in Political Inquiry* (Athens: Ohio University Press, 2016), https://www.ohioswallow.com/9780804011662/the-public-and-its-problems/; Ferraro, Etzion, and Gehman, "Tackling Grand Challenges Pragmatically"; and Charles F. Sabel and Jonathan Zeitlin, "Experimentalist Governance," in *The Oxford Handbook of Governanc*e, ed. David Levi-Faur (Oxford: Oxford University Press, 2012), https://doi.org/10.1093/oxfordhb/9780199560530.013.0012.

125 Sabel and Zeitlin, "Experimentalist Governance," 170.

126 Ferraro, Etzion, and Gehman, "Tackling Grand Challenges Pragmatically," 369.

127 "Removing Barriers to American Leadership in Artificial Intelligence"; and Russel T. Vought, *Driving Efficient Acquisition of Artificial Intelligence in Government* (Washington, DC: Office of Management and Budget, 2025).

128 Vought, *Accelerating Federal Use of AI through Innovation*, Governance, and Public Trust, 2.

129 Ibid., 6.

130 Ibid., 11.

131 Ibid., 11, 16.

132 Ibid., 2, 7.

133 Alex Tsalidis, "Staffer's Guide to AI Policy: Congressional Committees and Relevant Legislation," Future of Life Institute (blog), accessed June 17, 2025, https://futureoflife.org/document/guide-to-ai-congressional-committees/.

134 Margaret Ryznar and Karen E Woody, "A Framework on Mandating Versus Incentivizing Corporate Social Responsibility," *Marquette Law Review* 98, no. 4 (2015): 1668–94.

135 Ibid., 1648.

136 Reuel et al., "BetterBench," 10.

137 Jeff Schwartz, "The Conflict Minerals Experiment," *Harvard Business Law Review* 6 (2015), https://doi.org/10.2139/ssrn.2548267; Radu Mares, "Corporate Transparency Laws: A Hollow Victory?," N*etherlands Quarterly of Human Rights* 36, no. 3 (September 1, 2018): 189–213, https://doi.

org/10.1177/0924051918786623; Karen E. Woody, "Can Bad Law Do Good? A Retrospective on Conflict Minerals Regulation," *Maryland Law Review* 78 (2019): 291–322.

138 "Issue Brief: Early Best Practices for Frontier AI Safety Evaluations," Frontier Model Forum, July 31, 2024, https://www.frontiermodelforum.org/updates/early-best-practices-for-frontier-ai-safety-evaluations/; "Voluntary Commitments from Leading Artificial Intelligence Companies on July 21, 2023."

139 Matt Brown, "Senate Republicans Revise Controversial Ban on State AI Regulations," AP News, accessed June 17, 2025, https://apnews.com/article/ai-regulation-state-moratorium-congress-78d24dea621f5c1f8bc947e86667b65d.

140 Singh et al., "The Leaderboard Illusion."

141 David C. Hammack and Helmut K. Anheier, *A Versatile American Institution: The Changing Ideals and Realities of Philanthropic Foundations* (Washington, DC: Brookings Institution Press, 2013).

142 Ibid., 12.

143 Ibid., 13.

144 Rand Quinn, Megan Tompkins-Stange, and Debra Meyerson, "Beyond Grantmaking: Philanthropic Foundations as Agents of Change and Institutional Entrepreneurs," *Nonprofit and Voluntary Sector Quarterly* 43, no. 6 (2013): 950–68.

145 "Civil Society," Mott Foundation (blog), May 16, 2025, https://www.mott.org/work/civil-society/; MacArthur Foundation, "Grant Search," accessed June 17, 2025, https://www.macfound.org/grants/.

146 "Science Spawned by Foundations," CSPO (blog), accessed June 17, 2025, https://cspo.org/areas-of-focus/health-research-policy/science-spawned-by-foundations/.

147 "Reflecting on the Origins and Successes of Our Nuclear Grantmaking," MacArthur Foundation, accessed June 17, 2025, https://www.macfound.org/press/perspectives/reflecting-on-the-origins-and-successes-of-our-nuclear-grantmaking.

148 Kenneth C. Land and Seymour Spilerman, "Introduction," in *Social Indicator Models* (New York: Russel Sage Foundation, 1975), 2.

149 "About," Gates Foundation, accessed June 17, 2025, https://www.gatesfoundation.org/about.

150 In fact, the model proposed here, which leverages non-linear feedback loops, shares many elements discussed by cybernetics.

151 Norbert Wiener, *Cybernetics Or Control and Communication in the Animal and the Machine* (Cambridge: MIT Press, 1961); Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *The Bulletin of Mathematical Biophysics* 5, no. 4 (December 1, 1943): 115–33, https://doi.org/10.1007/BF02478259; M. D. Godfrey and D. F. Hendry, "The Computer as von Neumann Planned It," *IEEE Annals of the History of Computing* 15, no. 1 (1993): 11–21, https://doi.org/10.1109/85.194088.

152 "Potential Risks from Advanced Artificial Intelligence," Open Philanthropy, accessed June 17, 2025, https://www.openphilanthropy.org/focus/potential-risks-advanced-ai/; and "Home," Future of Life Institute, accessed June 17, 2025, https://futureoflife.org/.

153 "About Us," AI Now Institute, accessed June 17, 2025, https://ainowinstitute.org/about.

154 Hammack and Anheier, *A Versatile American Institution*, 13.

155 Rishi Bommasani et al., "The Foundation Model Transparency Index," arXiv.org, October 19, 2023, https://arxiv.org/abs/2310.12941v1.

156 Yogesh K. Dwivedi et al., "Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy," *International Journal of Information Management* 57 (April 2021): 20, https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

157 Fox, "The Uncertain Relationship between Transparency and Accountability," 665.

PHOTO SOURCE: BENJAMIN JENSEN/CREATED USING THE MIDJOURNEY WEB APP

CSIS | CENTER FOR STRATEGIC & INTERNATIONAL STUDIES

1616 Rhode Island Avenue NW
Washington, DC 20036
202 887 0200 | **www.csis.org**