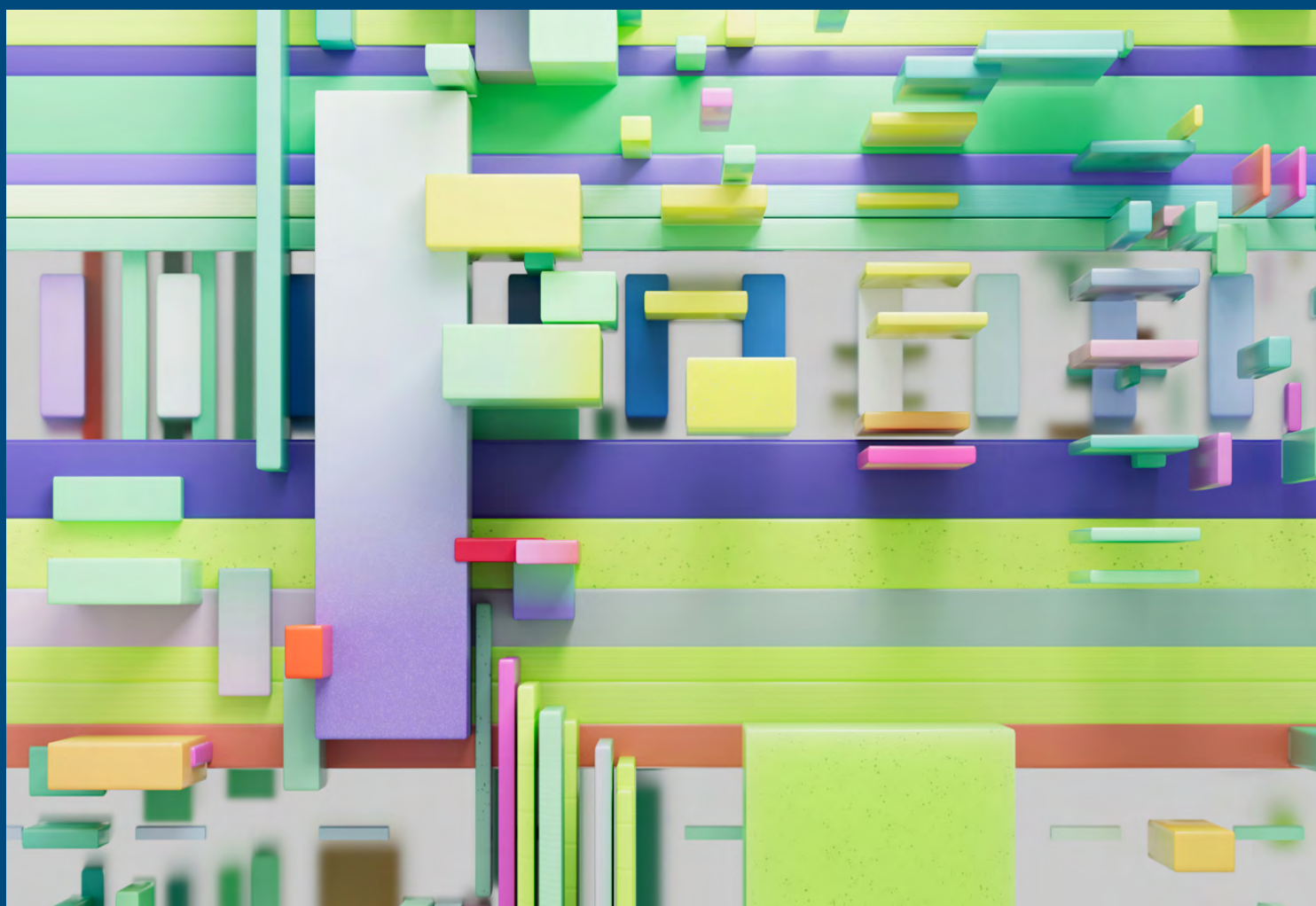


Building the Future of Science

A Policy Playbook for
Infrastructure, Governance, and Trust
in the Age of AI

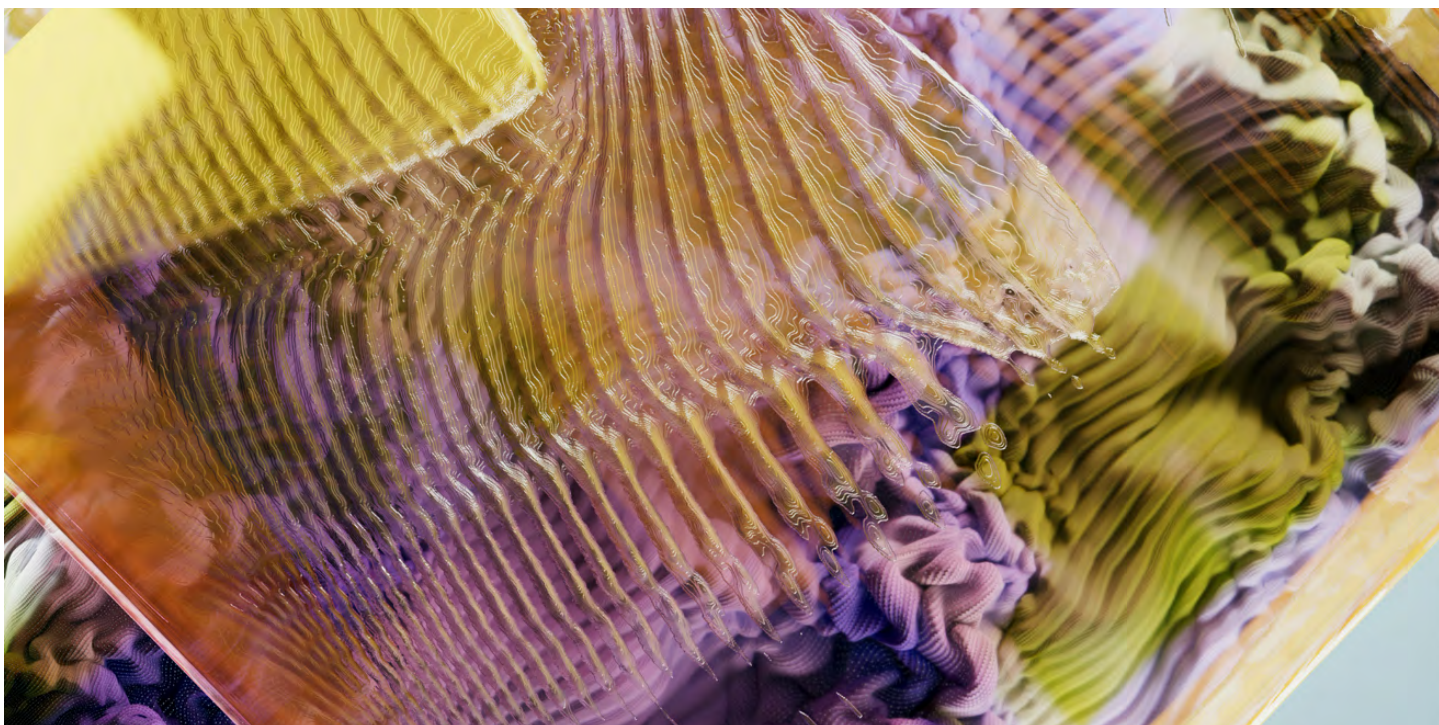


A report by



With gratitude for support from

Google.org



The Aspen Institute is a global nonprofit organization igniting human potential to build understanding and create new possibilities for a better world. Founded in 1949, the Institute drives change through dialogue, leadership, and action to help solve the most important challenges facing the United States and the world. Headquartered in Washington, DC, the Institute has a campus in Aspen, Colorado, and an international network of partners. In 2019, the Aspen Institute launched the Science & Society Program with seed support from the Alfred P. Sloan Foundation and Johnson & Johnson.



The Aspen Institute Science & Society Program strengthens the relationship between scientific progress and public life. Drawing on the Aspen Institute's convening power, cross-sector networks, and public platforms, the program helps people understand science, engage with and shape how science is used, and make decisions grounded in evidence. Through work spanning public understanding, civic engagement, and decision systems, Science & Society connects narrative, institutions, communities, and leaders across sectors to improve how scientific knowledge informs collective choices.

THE SCIENCE & SOCIETY PROGRAM WOULD LIKE TO THANK THE FOLLOWING ENTITIES FOR THEIR SUPPORT:

PROJECT SPONSOR



PROGRAM SUPPORT



Daniel H. Weiss
Rick Stamberger
Nick DeWolf Foundation
Paul Beirne Foundation
Marc E. Elovitz

Photo credits: Cover – Martina Stiftinger / Google DeepMind on Unsplash;
Page 2 and Section Headings – Ariel Lu / Google DeepMind on Unsplash; Back Cover – Nikada / Getty

CONTACT: Please keep in touch with us at aspeninstitute.org/science,
and for questions and comments, please write to
Aaron F. Mertz, Aspen Institute Science & Society Executive Director, at aaron.mertz@aspeninstitute.org

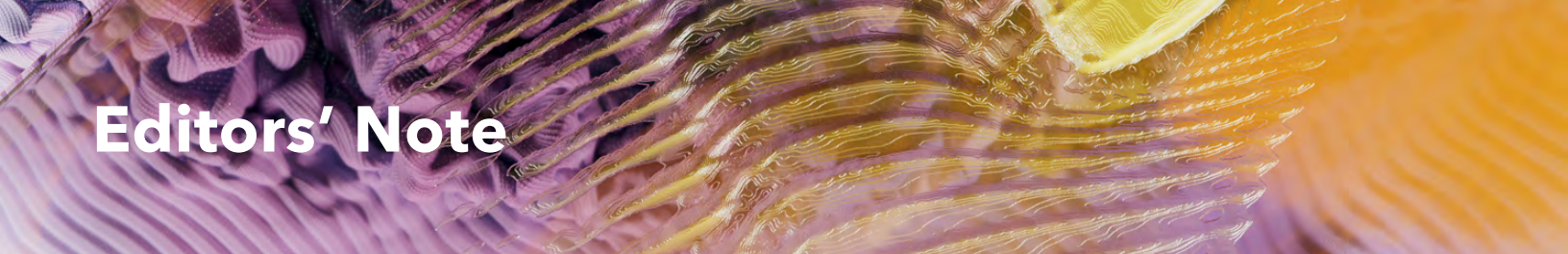
Table of Contents

EDITORS' NOTE..... 4
 Participants 6

**1. STRATEGIES FOR EQUITABLE, SUSTAINABLE, AND
INNOVATIVE INFRASTRUCTURE 9**
 Overview..... 9
 Background 9
 Analysis 9
 Recommendations..... 13

2. DATA GOVERNANCE AND PRIVACY 16
 Overview..... 16
 Background 16
 Analysis 17
 Recommendations..... 20

3. SCIENTIFIC TRUSTWORTHINESS AND PUBLIC CONFIDENCE 23
 Overview..... 23
 Background 23
 Analysis 24
 Recommendations..... 27



Editors' Note

As we highlighted in our September 2026 report, [AI for Discovery: New Frontiers, Challenges, and Opportunities for Scientific Research](#), artificial intelligence is reshaping the practice of discovery. Researchers are using AI to accelerate their work across all parts of the scientific process—including reviewing literature, generating hypotheses, designing experiments, and analyzing data—thereby expanding what scientists can study and how quickly they can study it. These new technologies also raise consequential questions about who has access to the resources required for AI-enabled research, how scientific data should be governed as it moves through increasingly complex computational systems, and how the integrity and trustworthiness of research can be maintained as AI becomes more deeply embedded in the scientific enterprise.

These considerations are closely connected. Access to computing resources shapes who can participate in AI-enabled science. Rules governing data determine what research can be conducted, by whom, and under what conditions. Prevalence of AI throughout the research process affects reproducibility, research integrity, scientific communication, and public confidence.

While the tools available to science will continue to change, the more immediate task is to build the infrastructure, governance systems, and institutional practices that allow technology to advance scientific discovery while protecting the conditions on which a healthy research ecosystem depends.

To chart the path toward such an ecosystem, the Aspen Institute Science & Society Program convened experts from universities, government, technology companies, scientific publishing and journalism, civil society, healthcare, and research and policy organizations. Throughout the summer of 2026, our conversations with these leaders focused on **research infrastructure, privacy and data governance, and scientific trustworthiness**.

The roundtables examined both the opportunities created by AI and the practical constraints emerging as researchers adopt these technologies. Universities must decide how to allocate limited AI resources across researchers and fields. Funders must determine how to evaluate proposals that depend on new models, datasets, and computational methods. Research institutions, journals, and scientific communities must decide what evidence and practices are sufficient to establish the credibility of AI-enabled research. Understanding these institutional decisions—and the tradeoffs, incentives, and standards that shape them—is an important next step in understanding how AI will change not only the practice of science, but the systems that determine what research is pursued and how scientific knowledge is evaluated.

Our participants debated specific questions, including:

- How can public research institutions secure sustained access to computing resources and technical expertise?
- How can governance systems protect sensitive data while enabling legitimate scientific use and reuse?

- How can researchers preserve provenance, reproducibility, and accountability when working with rapidly changing AI systems?
- How can institutions ensure that AI supports scientific expertise without weakening the human capabilities on which future discovery depends?

The Science & Society Program synthesized the discussions into policy recommendations reflecting perspectives raised across the conversations. We then brought the draft recommendations to panels of reviewers (who had not participated in any of the original roundtables) with experience translating policy ideas into practice. These review sessions included civil-society leaders, current and former government officials, legislative staff, researchers and university leaders, and think-tank experts working across state, federal, and international policy. Reviewers were asked to assess whether the recommendations addressed decisions that policymakers and institutions can realistically make, whether responsibility was assigned to the appropriate actors, and where proposals needed greater specificity or attention to implementation. Their feedback informed the final recommendations presented in this playbook.

Across the roundtable discussions and review sessions, our 47 participants brought perspectives from the United States, Canada, Chile, Colombia, the Philippines, South Africa, and the United Kingdom. Models shared during the roundtables and review process, such as the [Latin American Artificial Intelligence Index](#), revealed that each recommendation must be adapted by sector, geography, and audience. In particular, participants highlighted that implementation and downstream implications depend on both cultural and infrastructural realities.

The policy choices made now will help determine whether AI strengthens the scientific research ecosystem or deepens existing disparities and vulnerabilities. This playbook is intended as a practical contribution to decisions already confronting governments, research institutions, funders, technology companies, and scientific organizations.

Sejal Goud – Program Associate, *Aspen Institute Science & Society Program*

Jylana L. Sheats, Ph.D., M.P.H. – Associate Director, *Aspen Institute Science & Society Program*; Clinical Associate Professor, Social, Behavioral, and Population Sciences Department, Celia Scott Weatherhead School of Public Health & Tropical Medicine

Amelia Hanbury – Program Assistant, *Aspen Institute Science & Society Program*

Aaron F. Mertz, Ph.D. – Founder and Executive Director, *Aspen Institute Science & Society Program*

The editors used Google Gemini as a drafting tool to assist in converting detailed outlines, manually prepared by the editors from the roundtable transcripts, into prose. All analysis and final wording were reviewed and substantially revised by the editors. The Aspen Institute Science & Society Program retained full editorial independence and control over the final report.

Participants

We aim to synthesize and share perspectives from the initiative as a whole rather than to attribute any quotations or viewpoints to specific individuals. Participants—with whom we engaged in roundtable discussions, review panels, and individual conversations—are listed below (alphabetically by last name):

- **Susan Ariel Aaronson, Ph.D.** – Research Professor of International Affairs and Director of the Digital Trade and Data Governance Hub, *George Washington University*
- **Michael Akinwumi, Ph.D.** – Chief AI Officer and Head of the Responsible AI Lab, *National Fair Housing Alliance*
- **Gabriela Arriagada-Bruneau, Ph.D.** – Assistant Professor, Institute of Applied Ethics and Institute of Mathematical and Computational Engineering, *Pontificia Universidad Católica de Chile*; Associate Researcher, *National Center for Artificial Intelligence (CENIA)*
- **Solon Barocas, Ph.D.** – Principal Researcher, *Microsoft Research*; Adjunct Associate Professor, *Cornell University*
- **Asmeret Asefaw Berhe, Ph.D.** – Professor of Soil Biogeochemistry and Falasco Chair in Earth Sciences and Geology, *University of California, Merced*; Former Director, *U.S. Department of Energy Office of Science*
- **Elisabeth Bik, Ph.D.** – Independent Scientific Integrity Consultant
- **Lorissa Bounds** – Chief of Staff to Representative Jay Obernolte, *U.S. House of Representatives*
- **Ryan Calo, J.D.** – Professor of Law and Information Science, *University of Washington*
- **Stephanie Russo Carroll, Ph.D., M.P.H.** – Associate Professor of Public Health and Associate Director of the Native Nations Institute, *University of Arizona*
- **Kasia Chmielinski** – Co-Founder and Executive Director, *Data Nutrition Project*
- **Timothy Errington, Ph.D.** – Senior Director of Research, *Center for Open Science*
- **Barbara Evans, Ph.D., LL.M.** – Professor of Law and Stephen C. O’Connell Chair, Levin College of Law and College of Engineering, *University of Florida*
- **Mahmud Farooque, Ph.D.** – Clinical Associate Professor, School for the Future of Innovation in Society, and Associate Director, Consortium for Science, Policy & Outcomes, *Arizona State University*
- **Céline Gounder, M.D.** – Senior Fellow and Editor-at-Large for Public Health, *KFF Health News*; Medical Correspondent, *CBS News*; Clinical Assistant Professor of Medicine and Infectious Diseases, *NYU Grossman School of Medicine*
- **Conor Griffin** – AI Policy Research Lead, *Google DeepMind*

- **Juan David Gutiérrez, Ph.D.** – Associate Professor, School of Government, *Universidad de los Andes*
- **Laura Helmuth, Ph.D.** – Science Journalist; Former Editor-in-Chief, *Scientific American*
- **Raquel Hill, Ph.D.** – Professor and Chair of Computer and Information Sciences, *Spelman College*
- **William G. Hudson** – Head of Business Development, Regulated Industries, *Qualtrics*
- **Ollie Irwin** – Head, Trust & Safety Global Engagement (Europe, the Middle East, and Africa – EMEA) and Lead, Safety Engineering Center, *Google*
- **Jesse Isaacman-Beck, Ph.D.** – Director, Division of AI Policy and Strategy, Office of the National Coordinator for Health Information Technology, *U.S. Department of Health and Human Services*
- **Thorsten Joachims, Ph.D.** – Professor of Computer Science and Information Science, *Cornell University*
- **Veronique Kiermer, Ph.D.** – Chief Scientific Officer, *PLOS (Public Library of Science)*
- **Maggie Koerth** – Senior Science Reporter, *FiveThirtyEight*; Independent Science Journalist
- **Vivian Lee, M.D., Ph.D., M.B.A.** – Executive Fellow, *Harvard Business School*; Senior Lecturer, *Harvard Medical School*; Senior Fellow, *Institute for Healthcare Improvement*; Former President, *Verily Life Sciences*
- **Doc (Dominic) Ligot** – Founder and CEO, *CirroLytix*; Chair of AI Ethics and Safety, *Philippine AI Business Association*
- **Jon R. Lorsch, Ph.D.** – Deputy Director for Extramural Research, Office of the Director, *National Institutes of Health*
- **Harold M. Love Jr., Ph.D.** – State Representative, *Tennessee District 58*; President, *National Black Caucus of State Legislators*
- **Erin Lynch, Ph.D., Ed.D.** – President, *Quality Education for Minorities (QEM) Network*
- **Michelle Lopes Maldonado, J.D.** – Associate Director of AI Policy, *Center for Data Innovation, Information Technology and Innovation Foundation*
- **Margaret Martonosi, Ph.D.** – Hugh Jamieson Professor of Computer Science, *Princeton University*; Former Assistant Director, *National Science Foundation*
- **Hugues de Maupéou** – Privacy Policy Manager, Government Affairs and Public Policy, *Google*
- **Chris Mentzel** – Managing Director of Research and Campus IT, *Stanford University*; Former Program Director, *Gordon and Betty Moore Foundation*
- **Margaret Mitchell, Ph.D.** – Chief Ethics Scientist, *Hugging Face*
- **Mutale Nkonde, M.A.** – Founder and CEO, *AI for the People*; Doctoral Researcher and Harding Distinguished Postgraduate Scholar, *University of Cambridge*

- **Ivan Oransky, M.D.** – Co-Founder, *Retraction Watch*; Editor-in-Chief, *The Transmitter*; Distinguished Writer-in-Residence, *New York University*
- **Sethuraman Panchanathan, Ph.D.** – Former Director, *National Science Foundation*; University Professor of Technology and Innovation and Foundation Chair in Computing and Augmented Intelligence, *Arizona State University*
- **Manish Parashar, Ph.D.** – Director of the Scientific Computing and Imaging Institute and Chair in Computational Science and Engineering, *University of Utah*
- **Andrew Rens, Ph.D., LL.M.** – Senior Research Fellow, *Research ICT Africa*
- **David Rolnick, Ph.D.** – Assistant Professor of Computer Science, *McGill University*; Co-Founder and Chair, *Climate Change AI*
- **Dietram Scheufele, Ph.D.** – Foundation Professor, School of Social and Behavioral Sciences, *Arizona State University*
- **Brian Southwell, Ph.D.** – Distinguished Fellow and Lead Scientist for Public Understanding of Science, *RTI International*
- **Julia Stoyanovich, Ph.D.** – Associate Professor of Computer Science and Engineering and Director of the Center for Responsible AI, *New York University*
- **Holden Thorp, Ph.D.** – Editor-in-Chief, *Science*
- **Hans van Oostrom, Ph.D.** – Director, AI2 Center, *University of Florida*
- **Salomé Viljoen, J.D., M.Sc.** – Associate Professor of Law, *University of Michigan*
- **Jeannette M. Wing, Ph.D.** – Executive Vice President for Research and Professor of Computer Science, *Columbia University*

1. STRATEGIES FOR EQUITABLE, SUSTAINABLE, AND INNOVATIVE INFRASTRUCTURE

Overview

This policy note examines how access to AI infrastructure—including compute, sustained funding, and expert human capacity—is shaping participation across the scientific enterprise. As AI methodologies become central to research, existing resource disparities across institutions, sectors, and geographies risk becoming reinforced. This brief evaluates where access gaps are most acute, analyzes how current models like cloud credits and federal grants enable or constrain participation, and explores considerations for expanding partnerships. The analysis is followed by seven recommendations to build resilient, sustainable, and public-interest research capabilities.

Background

AI is influencing scientific research through two core modalities. In the first workstream, AI is used to assist in traditional tasks such as code generation and text editing. The second, arguably more transformative paradigm involves deep methodological integration, wherein AI architectures are embedded directly into scientific discovery pipelines, enabling methodologies like *in silico* testing. This deeper integration—while highly promising—can create significant operational challenges and requires sustained collaboration among domain scientists, machine learning engineers, and specialized data staff.

Supporting these approaches requires deploying compute infrastructure across two complementary operational scales. Baseline regional and local facilities supply the computing power necessary for daily data processing, privacy preservation, and institutional AI sovereignty. Concurrently, frontier-scale user facilities offer leadership-class, high-performance systems required for foundational modeling, though these centralized public resources currently face oversubscription challenges.

Participants emphasized that failing to maintain public capability across both infrastructural tiers poses risks for innovation and geopolitical standing. With commercial entities activating compute capital at scales that public funding cannot match on a dollar-for-dollar basis, public policy must develop new frameworks that ensure researchers across all institutional types can participate in AI-enabled science.

Analysis

Local Compute Sovereignty, Token Disparities, and Institutional Equity

Sustained, low-cost access to local or regional compute infrastructure—rather than exclusive reliance on temporary commercial cloud credits—is essential to preserving data privacy, safeguarding intellectual property, and maintaining institutional research sovereignty. Participants noted that widespread concerns regarding data security and commercial exploitation currently inhibit many academic researchers from utilizing the full range of AI capabilities available to them. Furthermore, competitive

resource barriers disproportionately disadvantage early-career investigators, who possess immense potential to drive scientific innovation but often lack dedicated institutional support.

While on-premise compute infrastructure offers local control, mandating that every academic institution host its own physical hardware may introduce equity issues. Smaller and under-resourced institutions often lack the capital and operational capacity to maintain full-scale facilities, making localized hosting demands counterproductive. In the U.S., shared hosting frameworks offer a strong path forward. At the sub-national level, states are introducing consortium approaches such as [Empire AI](#) in New York, the [Redtail](#) supercomputer in Utah, and the [Virginia Quantum Hub](#). Federal initiatives include the [National AI Research Resource \(NAIRR\)](#) and the [Department of Energy's Genesis Mission](#). In other countries, expanding national shared facilities is vital for addressing global research inequities, enabling scientists in low- and middle-income nations to conduct advanced research without requiring formal affiliations with wealthy Western institutions or their research agendas.

Concurrently, access to AI agent tokens has emerged as a significant driver of research disparities. One participant illustrated this gap by describing their experience working across two institutions: a private organization with unlimited agent tokens and a highly-resourced university with strict token limits, noting that the operational difference in research capability is “night and day.” International discussions showed that token costs can create even sharper constraints in lower-resource research environments. Participants noted that researchers may be limited to free tools regardless of their technical expertise, while differences in purchasing power and institutional budgets can place paid models and computational resources beyond reach. Access therefore depends not only on whether tools are available, but also on whether researchers can afford to use them consistently.

Technical Enablement and Human Capital Retention

Access to physical compute represents only one requirement for scientific enablement. Without dedicated campus-level technical expertise, existing tools remain underutilized. Although LLMs have reduced basic programming barriers, translating complex domain-specific hypotheses into computational architectures that execute efficiently on high-performance clusters requires specialized machine learning engineers and research data scientists embedded within academic organizations. As one expert explained, “actually translating a scientific problem or any kind of scholarly problem into something that runs efficiently on large compute is something that I think is going to need research data scientists and machine-learning engineers. [We need] people embedded at the research org[anization] that can sit with the researcher with their question and help translate that into something that scales.”

This expertise bottleneck reflects historical precedents, such as the deployment of advanced additive manufacturing in research laboratories, where equipment sat idle until institutions provided dedicated technical staff to guide implementation. In the AI-enabled research ecosystem, this talent challenge is intensified by private-sector recruitment. Public institutions encounter severe retention deficits as commercial technology firms systematically attract specialized academic talent. One participant contrasted the solution with long-standing practice: “You don’t just invest in infrastructure, you don’t just invest in models, you don’t just invest in data—you actually invest in staff and you also dedicate a significant amount of time on workforce development. We [in the federal government] relatively have way, way, way less resources invested in workforce development and retention of technical

staff. That's true in pretty much all public institutions, but it's probably worse in academic institutions."

International discussions also emphasized that effective AI infrastructure rests on prior investments in foundational data systems, technical personnel, and governance. Looking outside traditional scientific research, the Brazilian judiciary provides a clear [case study](#): over more than a decade of digital transformation, it established interoperable information systems and specialized technical teams before and alongside the deployment of advanced AI tools. Though the judicial system still faces ongoing operational challenges, the example illustrates why access to raw computing resources alone cannot substitute for underlying institutional capacity.

Comprehensive and Collaborative Education

Sustainable workforce development rests on a three-part educational model: (1) foundational AI principles, (2) ethical and policy compliance, and (3) domain-specific research applications. Putting this approach into practice means equipping faculty with discipline-specific pedagogical tools and giving students structured, hands-on experience on functional computing infrastructure.

Historical precedents demonstrate the efficacy of centralized user facilities in expanding national technical capacity. During the 1970s, electrical engineers Carver Mead and Lynn Conway developed a standardized pedagogical approach to Very Large Scale Integration (VLSI) microelectronic circuit design. To make this curriculum actionable, the U.S. government established [Metal Oxide Silicon Implementation Service \(MOSIS\)](#), a shared user facility that enabled a substantial number of U.S. undergraduates to design and fabricate a VLSI circuit. Students spent one semester designing a circuit in class, had it fabricated over break, and returned the following semester to test it out. This broad access model built a workforce that subsequently established a world-leading commercial microelectronics industry.

The [New Jersey AI Hub](#) exemplifies a modern state approach to educational alignment. By standardizing AI instructional modules across all state higher education systems, the initiative is reducing curriculum development burdens for community college faculty while ensuring seamless academic credit transfers to four-year research universities. Since a large number of computer science graduates at institutions like Rutgers University transfer from the state's community colleges, curricular coordination prevents wasted time, reduces stress on faculty, and ensures equal educational quality across diverse student demographics. As one participant noted, "anyone in New Jersey who wants to take an AI class is going to take a great AI class, and I think that's one of the things that makes me proud of where the state is trying to go to fill educational gaps."

Sectoral Capital Asymmetries and Uneven Scientific Domain Readiness

A structural mismatch separates the operational capacities of public research institutions from commercial technology firms. As one roundtable participant stated, "Even if you just think about the pace of progress that's happening in industry, there is no realistic way any public institution can match that." At the same time, addressing this gap presents difficult tradeoffs; expanding physical data centers involves human and environmental health burdens, but relying on a small consortium of com-

mercial vendors to host research infrastructure and sell tokens creates unsustainable public dependencies.

Closing the gap will require genuine cross-sector alignment that includes diverse institutions and a [preK to gray educational outlook](#). “It doesn’t matter whether we are at [an Ivy League] or [small state school]. This is probably the most leveling, humbling experience that all of us have. We all are facing [a] similar challenge. The scale of the challenge may be different, but the challenge remains for all of us. The only way this is going to be solved is by having the true partnership of industry and the haves really coming to the table and willing to work with the have-nots in a partnership mindset.”

Simultaneously, AI readiness differs sharply across scientific fields. Highly publicized breakthroughs like [AlphaFold](#) depended on fifty years of curated data from the [Protein Data Bank](#). Most fields outside genomics, materials science, and climate modeling lack machine-readable data infrastructure and remain in an early “seeding phase.” In these domains, organizational capacity must precede hardware deployment. “These are all disciplines that I think, even if they had access to the infrastructure, wouldn’t necessarily know what AI means for their discipline. So supporting that kind of community building and norm setting [and] governance structures, I think that’s also a gap that I think just holds many of the disciplines back,” explained one university leader. They added that funding discipline-specific “thought summits” for areas like veterinary science or microbial design helps faculty build governance and data-sharing frameworks.

Volatile Funding Models and International Partnerships

Federal research funding is frequently characterized by volatile cycles of expansion and contraction. Supporting specialized technical staff, however, requires stable, persistent operational funding. Since human expertise yields compound returns over time compared to depreciating hardware, participants noted that establishing permanent, endowed funding for technical personnel is critical. A related “missing middle” exists in research funding: projects may receive enough support to develop a prototype but lack sustained funding for implementation, maintenance, or continued use.

A related barrier is the continued use of grant structures that require participating countries to finance their own researchers. One former government official noted they “typically had a process where the U.S. provided the money for the U.S. researchers, and the other country had to provide the money for their researchers.... [This] means that it was procedurally very hard to partner with the less wealthy countries that couldn’t bring their own funding to the table.” Policy frameworks should also consider how major technology companies influence international research access and interact with national priorities concerning infrastructure and technological sovereignty.

Finally, institutional partnership frameworks must prioritize formal agreements to prevent tokenization. Historically Black Colleges and Universities (HBCUs), Minority-Serving Institutions (MSIs), community colleges, and [Established Program to Stimulate Competitive Research \(EPSCoR\)](#)-designated states should receive direct, equitable resource allocations rather than passive inclusion on major research grants. International discussions further emphasized that equitable partnerships should build lasting capacity within local institutions. Access to infrastructure, knowledge transfer, training, and fair allocation of intellectual property are important safeguards against arrangements in which local organizations provide data or expertise but remain dependent on international partners. Partnership

agreements should therefore specify how technical capability and control will remain with participating institutions after a project ends.

Recommendations

1. Focus on Agile Policy Approaches

In light of a rapidly moving technological landscape, institutional leaders and standards groups should prioritize the creation of living documents, guidance briefs, and best practices over static directives. The content of these documents, including definitions, should be revisited and updated on a routine basis to ensure relevance and coverage of both common use cases and edge risks.

2. Establish State and Regional Computing Consortiums

Federal funding agencies, state governments, and philanthropic organizations should fund and scale regional computing consortiums to bridge the gap between commercial cloud reliance and isolated university hardware clusters. This model delivers persistent, low-cost compute access to diverse institutions without imposing heavy capital costs on smaller colleges. As roundtable experts highlighted, consortiums effectively link diverse institution types, secure persistent funding, empower local economic development, help coordinate regional public health responses, and in some cases take advantage of renewable energy sources. Current consortiums should expand their focus beyond pure compute to address gaps in storage, data management, and sharing, which also remain cost-prohibitive for much of the public sector.

Consortium design should also account for researcher planning preferences. Although centralized facilities maximize resource efficiency, individual investigators may favor smaller, locally controlled allocations as a way of facilitating project predictability. Centralized and regional consortiums should therefore integrate guaranteed resource scheduling into their operational models to accommodate long-range academic planning.

3. Transition Private Cloud Credits to Persistent National User Facilities

While federal agencies currently operate leadership-class high-performance systems, these facilities remain oversubscribed and unpredictable. Congress and federal funding agencies should improve on national AI infrastructure initiatives, such as NAIRR and the Genesis Mission, by shifting from temporary commercial cloud credit donations toward permanent, publicly owned national user facilities. Participants framed this move as a core and necessary structural transition: “How do we make sure that we’re not just routing public research money into private platforms that will provide cloud credits and access in the short term, but then expire, create dependency, and [in the end leave us] subsidizing to a large extent, the private companies’ work instead of actually creating lasting capability that everybody can access in a national user facility-type of access?”

Following the same logic, philanthropic funders should also provide sustained access to licensed tools and AI agent tokens, with support calibrated to differences in institutional resources and purchasing

power. Access programs should cover recurring use rather than provide only temporary credits that leave researchers unable to continue their work when initial support ends.

4. Embed AI Considerations Within Strategic University Planning

University leadership, academic committees, and governing bodies should distinguish between expanding raw AI deployment and investing in strategic planning, recognizing that “investment in thinking about AI is not the same as investment in more AI.” Uncritical, top-down pushes to adopt AI tools across every potential use case risk wasting resources without improving scientific or educational outcomes. Institutions should evaluate how these technologies function as foundational campus infrastructure—akin to digital networks or physical utilities—and embed them thoughtfully into standard operating procedures where scientifically justified. As one expert emphasized, AI “needs to be something that the university itself invests in, that it not just adds to its fundraising campaign, but that it really just foundationally provides. So, it’s not AI infrastructure, it is a mindset of how essential these tools are becoming in the day-to-day operations of how science is done, and really how all information work is done,” commented one participant.

5. Implement Routine Auditing of Infrastructure Utilization

Funding agencies, national facility administrators, and university research offices should establish systematic data collection protocols to track how AI infrastructure is used. Methodically auditing user demographics, computational workloads, software dependencies, and primary dataset utilization can allow policy decision-makers to identify operational bottlenecks, optimize resource distribution, and direct capital investments toward priority areas. With the rapidly moving pace of AI technologies, audits should follow a continuous loop model featuring frequent reassessments.

6. Right-Size Compute Allocations and Right-Fit Model Architectures

Drawing on utilization audits, funding agencies and compute allocation committees should match infrastructure access to the specific methodological requirements of a scientific problem, rather than treating massive generative models as the default form of AI-enabled research. Many scientific applications are better served by smaller or specialized architectures, including convolutional neural networks (such as [ResNet](#) and [U-Net](#)), tree-based methods (such as [Random Forest](#)), physics-informed models, or time-series models. As one expert explained: “An astronomer who’s looking to process their data may do better with a ResNet than with a trillion parameter model,” because “generative models are not useful in every context.” At the same time, certain complex scientific applications do require substantial compute. Infrastructure policy should therefore support a range of computational needs, giving researchers access to large-scale compute where empirically and scientifically justified, while avoiding incentives that equate larger models with better science.

7. Build Capacity Through Equitable and Secure International Research Partnerships

Structure international AI collaborations so that institutions in partner countries can continue the work long after a project ends. Thoughtfully executed cross-border partnerships can greatly increase data quality and technical capacity across the board, empower researchers in partner nations, and

foster global scientific equity. Achieving these goals requires clear agreements regarding intellectual property sharing, resource control, and joint governance. Restrictive requirements that force participating countries to finance their own researchers often exclude institutions in lower-resource settings and concentrate collaboration among wealthier nations. Funding frameworks should instead support researchers across all participating organizations to ensure sustainable, long-term capabilities.

As governments increasingly apply a securitized approach to global science, international partnerships must balance equitable capacity building with strong guardrails. Governance protocols should be embedded directly into model architecture and workflows, automatically controlling data flows and protecting against security vulnerabilities. Such architectures are particularly critical when deploying agentic AI systems, where autonomous or recursive self-improvement capabilities introduce heightened biosecurity and operational risks.

2. DATA GOVERNANCE AND PRIVACY

Overview

This policy note examines how AI-enabled research is redefining questions of safe data collection and downstream data sharing. As the modern practice of science increasingly relies on large-scale, cross-border, and corporate-controlled datasets, legacy governance frameworks are proving inadequate. This brief analyzes systemic gaps and tradeoffs, concluding with nine recommendations to transition governance from static oversight to dynamic, pipeline-driven stewardship.

Background

From a legal perspective, data governance encompasses any law determining how data is collected, processed, or used. The goals of this governance are two-fold: (1) to control risks by ensuring the security, integrity, and protection of data and systems and (2) to create value by establishing rules and technical standards that facilitate the exchange and combination of information. In an increasingly globalized world, data governance often operates as a [multi-stakeholder and cross-border endeavor](#).

However, current data governance suffers from two fundamental limitations: (1) it tends to focus only on the release of data rather than its collection procedures, use, and reuse and (2) it lacks the inclusion of essential groups—such as Indigenous data stewards and researchers from low- and middle-income countries—and of essential questions—such as which uses of a given dataset are and are not acceptable.

The central challenge facing modern science was captured by one roundtable participant: “The single most important challenge is that our governance reaches the moment of release but not the moment of reuse—and in an AI era, that’s where everything now happens. For decades we built scientific data governance around collection and sharing, [including] consent at the point of collection, ownership of a dataset, accountability for a single study. All of that assumes data sits still. AI makes data fluid—it gets scraped, summarized, recombined, and acted upon by systems we don’t control, in contexts no one consented to and no one owns.”

As the United Nations Statistical Commission [has emphasized](#), keeping pace with this shift will require providing “correct, timely, coherent, human-understandable, and contextually relevant official data and metadata, with a clear indication of their provenance and vintage.”

Analysis

Effectively Anonymizing Sensitive and Interconnected Data

Artificial intelligence is increasingly able to “derive the intimate from the available,” particularly within health contexts. As participants noted, certain biomedical data are intrinsically identifying, while

other data cannot be meaningfully anonymized without reducing their scientific utility. Given these technical limits, a core policy challenge is determining the acceptable bounds of synthetic data and establishing data generation methods that remain genuinely representative.

Consequently, policymakers should re-examine how AI is expanding the category of information needing legal protection. This includes new sources of data like smart glasses and flock cameras. Washington State's health privacy law offers one positive model, as it explicitly covers both health information and [health information that can be derived](#) from machine learning or other AI processes.

Differences across regulatory jurisdictions further complicate this landscape. In the European Union (EU), the [General Data Protection Regulation \(GDPR\)](#) treats data as anonymous only when an individual is no longer identifiable, taking into account the means reasonably likely to be used for identification. The [EU Digital Services Act \(DSA\)](#) operates in parallel with the GDPR, establishing rules about when and how vetted researchers can access certain online information. In the U.S., the [Health Insurance Portability and Accountability Act of 1996 \(HIPAA\)](#) establishes a separate framework for de-identifying protected health information through either the removal of specified identifiers or a qualified expert's determination that the risk of identification is low. Although both the European and American frameworks recognize that identifiability depends on the practical risk of re-identification, they differ in scope, application, and implementation. Some roundtable experts argued that HIPAA regulatory [guidance](#) issued in 2012 has functioned in practice as a permissive approach to enforcement, allowing "vast traffic in weakly de-identified data [to] happen, because it advanced science."

Underlying this issue is a deeper limitation in the prevailing legal philosophy of privacy. As another participant explained, "The problem is that we have been operating since the 1970s under a rights-based model of privacy that viewed privacy as an individual right, [where] people could manage their own privacy by consenting or not consenting. What's apparent is that our privacy is completely interconnected. AI models sort of hit us in the face with the fact that even if we keep our data out, people can still draw a lot of stigmatizing inferences about us without using our data."

Circumventing Traditional Research Ethics Guardrails

In contrast with historical scientific data collection, which operated under defined cohort sizes, explicit research goals, and corresponding regulatory structures, modern commercial mechanisms bypass many of the guardrails applied to academic scientific research.

As one participant detailed, "I feel that there is no such thing anymore as data privacy... We have these governance frameworks to protect the research data and protect the individuals that are in the dataset, [yet] we have this parallel universe where this data is being collected by companies and being monetized, and there are no controls over here. The researcher who wants to understand more about a problem has to go through a rigorous process to get access to data, when the same information is available to companies and they're monetizing that data."

This "parallel universe" extends into active research methodologies. When large language models (LLMs) are deployed to conduct research interviews, discussants explained that "data that is subject to human subjects rules is flowing directly to private companies whose policies may not actually correspond to what you would expect under research ethics." In many cases, researchers themselves may not be fully aware of the extent to which AI capabilities are present within their data collection

tools, hampering disclosure to study participants. Informed consent may also become less meaningful when individuals cannot anticipate how their data will be used or combined, particularly when scientifically controlled data is integrated with corporate data sources to accelerate discovery.

Institutional Gatekeeping and Limited Model Transparency

Corporate control over data may pose significant risks to public-interest research. In some instances, data granted to academic researchers by commercial entities may be incomplete, heavily curated, or structured in ways that favor the organization's interests. Data integrity and scientific reproducibility may also be threatened by reliance on third-party AI models whose training data, curation methods, and update schedules are not disclosed. When researchers lack access to or control over these underlying systems, they may be unable to reproduce findings. Further safeguards are also needed to ensure that, once an AI model has been used to analyze a scientific dataset, the company that owns the model cannot restrict subsequent reuse of the dataset.

Gatekeeping can also occur within public institutions. One participant described a system in which a single university clearinghouse controls access to all K–12 educational data in the state and maintains the sole authority to approve or deny research requests, creating a significant access and transparency bottleneck. At the same time, international participants raised concerns about poor representation in datasets assembled across institutions and national borders. Bias in language data, facial images, medical scans, and biological data may go unrecognized when no institution controls the full dataset.

Tradeoffs Between Corporate and Government Control

Corporate and government control of data bring unique risks. On the one hand, existing regulatory regimes disadvantage public scientific data production while permitting corporate data collection for profit, increasing academic researchers' dependence on corporate data and infrastructure. On the other hand, government control over research infrastructure creates its own vulnerabilities, including prohibitions affecting federally controlled databases used for research on diversity or health equity. National security and export control restrictions may also serve as barriers to academic research freedom.

Power Imbalances in Cross-Border Research

Data governance can be either inclusive or extractive, depending on who controls decision-making and who benefits from the project. Participants identified several elements of a more equitable partnership: centering the intended beneficiaries, establishing clear rules for data use and reuse, addressing the allocation of intellectual-property rights and compensation, and transferring knowledge so that local institutions do not remain dependent on international partners.

From an international perspective, privacy also concerns who controls data after they are shared. As one participant observed, "governance can be extractive or inclusive." The participant described cross-border projects in which a local institution provides data without clear rules governing future reuse, intellectual-property rights, compensation, or acknowledgement. Meaningful privacy protection

therefore requires transparency about how data will be used beyond the original project and a role for local institutions in those decisions.

Increasing Costs of Data Governance

Although AI can drive down raw computation and analysis costs, the cost of maintaining rigorous data controls and excluding malicious actors is escalating. While participants noted the utility of strict regulatory standards such as the [U.S. Department of Justice's Data Security Program](#) and [FedRAMP](#), they also noted that high compliance burdens tax institutional budgets. Clear economic models should be developed to prevent these mounting security compliance costs from pricing smaller and lesser-resourced research institutions out of the AI-enabled scientific ecosystem entirely.

Malicious Uses of Sensitive Data

Data is increasingly conflated with speech, leading to unintended commercial exploitation. In [Vermont](#), medical data originally released to promote accessible generic drugs was co-opted by pharmaceutical firms to identify and reward physicians prescribing name-brand drugs—an outcome the state could have prevented by conditioning access on non-commercial use. Similarly, while initial AI chemical prompting demonstrations focused on benign compounds, the exact same workflows can be leveraged to streamline the creation of novel toxins or chemical weapons.

Overlooking Primary Source Data

Public reliance on unvalidated AI tools can create widespread misalignment with primary scientific data. AI applications frequently pull from secondary or unverified summaries rather than authoritative primary databases, such as the [Integrated Postsecondary Education Data System \(IPEDS\)](#) or the [National Center for Science and Engineering Statistics \(NCSES\)](#). When AI models generate outputs that contradict raw primary data, the quality of the larger literature pool and public trust in research institutions are eroded.

Recommendations

1. Shift to a Collective and Pipeline-Driven Governance Approach

Re-engineer data governance so that policies attach directly to computational pipelines and models rather than static datasets and individual studies. As roundtable experts emphasized, “Our unit of governance is somehow incorrect.... We govern datasets and studies, but AI operates on pipelines and models. This is where governance really should live. The inadequacy isn’t any single rule, it’s that we lack the actual infrastructure, both human but also engineering, to carry governance forward as data moves.”

Participants suggested that governance frameworks should adopt best practices from Indigenous data stewardship approaches, which often directly address group-level harms and establish distributed accountability. Since AI models exploit interconnected data, policymakers must recognize that individual opt-outs are insufficient to protect populations from collective harm. Local institutions contribut-

ing data should also participate in decisions about access, analysis, reuse, and distribution of resulting benefits.

Further, models governing multi-institution and cross-border research should specify the types of authority held by each actor at each stage of the collaboration pipeline.

2. Enhance Provenance and Fitness Metadata

Invest in technical and engineering infrastructure to attach machine-executable metadata to data at the point of collection, ensuring context carries through the entire AI pipeline. Building on the existing [FAIR and CARE principles](#), standard-setting bodies must develop clear metrics to qualify a dataset's "fitness for use" in different research applications and to ensure transparency regarding the original population and purpose of collection. Current efforts include the [Coalition for Content Provenance and Authenticity \(C2PA\)](#). Metadata updates must occur dynamically rather than manually. "You can't govern reuse, enforce conditions of use, propagate a revision or retraction, or give anyone recourse if the context lives in a footnote no AI system reads," noted one participant. However, governance frameworks must also account for field-specific computational constraints; in data-intensive domains like genomics, storing full metadata lineages can prove more cost-intensive than storing the raw data itself, requiring pragmatic standards for compression and retention.

3. Expand Access to Secure Data Enclaves and Trusted Research Environments

Governments, funders, and research institutions should expand access to [secure data enclaves](#) and [Trusted Research Environments \(TREs\)](#), allowing scientists to compute over sensitive datasets without downloading raw files into unmonitored systems. Operating within controlled enclaves enables institutions to meet stringent privacy standards while retaining localized data sovereignty and constraining opportunities for re-identification.

To maximize public benefit, institutions should explicitly specify governance policies dictating if, when, and how models trained within these enclaves are made accessible to commercial entities or the public. Policy design should prevent secure enclaves from being monopolized by large technology firms, as centralized control risks creating further antitrust barriers and disadvantaging citizen scientists. International research institutions should build local technical capacity for these environments to avoid creating long-term dependence on foreign tech companies.

4. Update Notions of 'Adequate Security' for the AI Age

Legislative bodies, regulatory agencies, and research institutions should modernize legal and security frameworks to guard against AI-specific vulnerabilities, moving beyond traditional physical encryption to protect against model manipulation, prompt engineering, and adversarial machine learning. Lawmakers should patch [legal blindspots](#) in legacy frameworks, such as the Computer Fraud and Abuse Act of 1986, to ensure they effectively cover automated computational exploits. Furthermore, statutory definitions of protected health data should expand to cover both direct health information and

inferences derived through machine learning processes, drawing on state-level [precedents](#) like Washington.

Addressing the overarching fragmentation of data privacy calls for comprehensive level-setting across jurisdictions while respecting cultural and infrastructural sensitivities. Corporate actors and research institutions can collaborate with standard-setting bodies to demonstrate how piecemeal, state-by-state legislation complicates compliance and creates operational friction.

On a technical level, because single-layer solutions like secure metadata can be compromised, policy guidelines should recommend an integrated security overlay that combines digital watermarking, data fingerprinting, and dynamic metadata to ensure robust, end-to-end data protection across computational pipelines.

5. Begin with Introspection of Organizational Practices

Research institutions, government bodies, and corporations can drive systems change by beginning with an audit of their own data handling procedures. While broader regulatory environments remain outside any individual organization's direct control, institutions are encouraged to apply maximum security, transparency, and ethical oversight over the datasets and models they currently use and maintain. Examples of stocktaking measures include the U.S. Office of Management and Budget (OMB) [Memorandum M-25-21](#), titled *Accelerating Federal Use of AI through Innovation, Governance, and Public Trust*. The directive calls on all U.S. federal agencies to list their AI use cases, with this information serving the dual purposes of maintaining privacy standards and ensuring that computational resources are right-sized to the needs of each organization. Inventories and audits should follow a regularly scheduled cadence, paying particular attention to tools with AI functionalities that can be enabled and disabled.

6. Train the Next Generation on Ethical Data Collection and Use

Funding agencies and academic leadership should support interdisciplinary educational programs to ensure data governance policy is continuously informed by domain experts who understand the ethical nuances of specific datasets. Curriculum design and institutional incentives should reinforce that scientific research is not merely an exercise in computational efficiency, but a human pursuit driven by the joy of learning and discovery. Governments can actively encourage these educational reforms by embedding ethics, data stewardship, and interdisciplinary requirements directly into existing national training grants.

7. Incentivize Scientist-Led Teaching and Communication Efforts

Scientific institutions should require researchers to lead public engagement and communicate the real-life impacts of their work rather than delegating public communication to unvalidated, AI-generated content or profit-seeking influencers—two key sources of misplaced trust among the public. Direct communication is vital to fulfilling the “trust contract” owed to taxpayers and trial participants. While

human scientists must drive public outreach, generative AI tools can be responsibly used to assist in translating complex papers into accessible, multi-lingual educational summaries.

8. Develop Comprehensive Guidance on Organizational Interoperability

Establish standardized technical protocols to enable multi-institutional computation across combined datasets without raw data sharing. As research increasingly relies on autonomous AI agents interacting across institutional boundaries, platforms must support policy enforcement in confidential compute environments. Standard-setting bodies should expand and build upon open protocols—such as Google’s [Agent2Agent Protocol](#)—to ensure policy constraints are automatically maintained during agent-to-agent data exchanges.

9. Incentivize Data Sharing and Restructure Scientific Currency

Establish regulatory mandates for third-party academic access to proprietary corporate data that impacts public health and societal well-being. Simultaneously, academic leaders must [reform institutional incentive structures](#) such as tenure and grant evaluations to reward data sharing and curation, countering historical gatekeeping practices. The philanthropic sector also has an important role to play in funding ‘datasets for public good,’ allowing researchers to move away from the challenges of relying on curated commercial sources.

Policymakers can [learn from](#) international data sharing models—such as COVID-19 genomic sequencing or the Human Cell Atlas—to advance global cooperation and expand sample sizes. Cross-border agreements should establish clear rules governing future data use, intellectual-property rights, compensation, acknowledgement, benefit-sharing, and knowledge transfer. These protections can help ensure that publicly funded data remain a shared public good instead of becoming a locked commercial or national-security asset.

3. SCIENTIFIC TRUSTWORTHINESS AND PUBLIC CONFIDENCE

Overview

This policy note examines how AI integration is influencing transparency, integrity, and credibility in science, using healthcare and biomedical research as primary analytical lenses given the direct impact of these fields on individual and community well-being. Far from creating entirely new challenges, AI is compounding existing concerns related to peer-review systems, institutional communication, public engagement, and ownership of knowledge. This brief details key structural risks and outlines eight recommendations to promote education, foster genuine participation, and realign incentives across the research enterprise.

Background

Predictive AI technologies have supported clinical practice and biomedical research for decades. [As early as 1999](#), researchers deployed artificial neural networks to predict active pulmonary tuberculosis. Similarly, the field of cardiac electrophysiology has [long relied](#) on machine learning methodologies to assist in screening procedures. Paired with physician judgment, these predictive systems have been fundamental for supporting diagnostic precision and improving patient outcomes.

Over time, however, the role of AI across scientific domains has increasingly shifted from predictive to generative. In clinical environments, AI is now framed as a potential tool for bridging primary care gaps. [Ambient scribes](#) are beginning to assist clinicians with automated notetaking, ideally allowing providers to focus their attention on patient care while empowering non-physician staff. In the U.S., federal agencies—including the Department of Health and Human Services (HHS) and the Centers for Medicare & Medicaid Services (CMS)—are directing [substantial investments](#) toward AI tools designed to close rural healthcare access gaps. While these allocations have the potential to [unlock new and improved care](#) for communities across the country, they have also [drawn scrutiny](#) from critics concerned about the government pushing Big Tech into healthcare delivery.

Within the research environment, AI is expanding both the scale and volume of what is possible while helping scientists “fail faster” when exploring hypotheses. Enthusiasm is particularly high regarding *in silico* screening techniques, protein design, and drug discovery for rare diseases with weaker commercial incentives. Roundtable participants described routinely using synthesis tools like [Elicit AI](#) or using LLMs to write their own program scripts capable of analyzing millions of papers across catalogs like [Scopus](#) and [OpenAlex](#), thereby streamlining systematic reviews and meta-analyses. Other aggregators such as Yale University’s [Population Health Information and Visualization Exchange \(PopHIVE\)](#) pull together disparate, deidentified data streams into AI-readable, real-time dashboards that accelerate basic research and strengthen downstream public health policy decisions. Participants also character-

ized generative writing assistants as a means to reduce language barriers, thereby enabling international research collaborations and high-visibility publications for non-native English speakers.

Analysis

Disclosure Frameworks, Model Transparency, and Accountability Chains

While scientific publishers already mandate disclosure of funding conflicts and research methodologies, current frameworks remain ill-equipped for AI workflows. Editorial policies currently rely on crude distinctions—such as separating minor editing from majority AI-generated text—that fail to capture the nuanced ways computational tools are used throughout the scientific process. Generative AI also introduces sophisticated forms of data and figure manipulation. Historically, falsified data contained human tells such as non-random number preferences. Modern models are capable of producing synthetic data, graphs, and images with convincing realism, leaving human reviewers largely blind to fabrication. Commercial detection software offers little relief, yielding high rates of false positives, while technical solutions like digital watermarking face steep adoption and feasibility barriers.

Verification is further complicated by model opacity and continuous training. While replication challenges long predated generative AI, shifting model weights, proprietary tools, and autonomous agents make independent auditing far more complex. As participants noted, this lack of visibility prompts fundamental questions about the trustworthiness of science across the board. “Without us having any of that transparency, it’s really hard to claim that we produce a replicable and reproducible science over time that will [yield], with the same inputs, the exact same outcomes. In fact, that will be largely impossible. If we steer toward that world, I think trust is not a matter of perception or a matter of restoring something that should be rightfully there, but it actually may be something that shouldn’t be there because our science is no longer very good,” one researcher explained.

At a systemic level, AI obscures accountability chains and strains the social contract between scientific institutions and the public. Current governance frameworks offer little clarity on who bears responsibility for corrective action when errors occur, how AI involvement redefines academic authorship, or what recourse exists when AI-generated syntheses enter the literature without a human name attached.

Institutional Hype and Media Translation Gaps

“The distance between what’s in a scientific paper and what goes in the public discourse is something that nobody really owns,” noted one roundtable expert. This accountability vacuum is partially driven by institutional communications and university press releases, which routinely exaggerate scientific findings when promoting new studies. Experts also anticipated a sharp rise in spurious correlations, [“illusions of understanding,”](#) and [“p-hacking”](#)—manipulating data, running multiple hypothesis variations, or cherry-picking results to force statistical significance. These risks are amplified as [fully](#)

[autonomous AI agents](#) manage research degrees of freedom without human intervention, allowing researchers to selectively report only eye-catching computational pipelines.

While some specialized organizations exist to review news coverage and intervene where statements exceed the published evidence, the translation gap remains a persistent challenge across the media landscape. To combat institutional hype and an influx of low-quality papers, some journalists have begun taking filtering into their own hands, building custom AI agents that combine domain knowledge with cognitive bias filters to screen out tenuous claims. Translation is further complicated when journalists lack the precise technical vocabulary required to describe specific AI methodologies accurately, creating an additional risk of public miscommunication.

“Pollution of the Literature” and Distorted Academic Incentives

Generative AI is lowering the cost of producing manuscripts and contributing to rising submission volumes, in turn straining already-limited editorial and peer-review capacity. A recent [Organization Science analysis](#) found that submissions increased 42% following the release of ChatGPT, with AI-generated writing accounting for nearly all of the increase. These pressures interact with academic incentives that reward publication and citation counts. A separate [Nature study](#) analyzing 41.3 million papers found that scientists engaged in AI-augmented research published 3.02 times as many papers and received 4.84 times as many citations as scientists who did not.

This dynamic leads to a “pollution of the literature,” as seen in [reported trends](#) surrounding datasets like [TriNetX](#) and the [National Health and Nutrition Examination Survey \(NHANES\)](#). While much of this output is not explicitly false, the research questions it answers tend to be trivial. As one participant summarized, AI may simply make researchers more efficient at producing larger quantities of mediocre research, rather than proportionally increasing the number of important discoveries. The long-term risk may intensify as future AI models are trained on this literature, potentially weakening the reliability of scientific information: “We may see declining levels of trust in science, not because of political polarization or anything else, but because we’re actually producing science that is not as reliable as it should be.”

The integrity risk extends beyond manuscript production to peer review itself. Participants raised cases in which reviewers appeared to submit chatbot-generated evaluations that made claims about content not present in the manuscript. Such cases indicate that disclosure and accountability requirements must apply to the full publication process, including reviewers, rather than focusing only on authors.

Sectoral Hardware and Data Asymmetries

Historically, academic researchers possessed the independent lab capacity to replicate, evaluate, and audit industry claims. However, a widening computational and data divide now separates commercial technology firms from academic institutions. Even the best-resourced research universities lag behind industry giants like Google and Meta, hampering independent verification and public-interest comparative research. “In the past, for carcinogenic effects of cigarettes, yes, we could do the same thing that industry could do. For fossil fuels, yes, we could do the same thing in the lab that industry could do

and provide the same data.... I think now we're getting to the point where that asymmetry is another really [big risk] that I would like to put on our mental agenda," one participant highlighted.

Singular Judgment and Misplaced Trust

AI chatbots tend to output straightforward approximations of language in response to complex tasks, often anchoring on a single answer, failing to properly quantify uncertainty, and generally lacking desired nuance in high-stakes scenarios. While [recent research](#) suggests that LLMs are increasingly capable of making emergency-room decisions, roundtable contributors noted that human doctors are better at evaluating differential diagnoses and actively seeking additional information.

These technological limitations can interact with human cognitive biases, particularly the tendency to interpret confidence as evidence of accuracy. "People often read confidence, whether it's from an AI or people, as truth," explained one participant. "One of the fundamental problems in journalism, in science, in medicine, is the lack of humility [but also] false confidence... and being unable to accept or communicate or really work through uncertainty. I'm afraid that at all levels, generative AI is just exacerbating that fundamental problem," they continued. Public trust is further eroded on social media, where unvalidated users deploy AI tools to argue with health experts or generate sophisticated promotional videos for untested 'wellness' products under the guise of scientific expertise.

Public Engagement vs. Scientific Communication Deficits

Building—and in many cases, restoring—trust in science is not merely a communication problem but a structural participation challenge. Without continuous engagement loops, scientific institutions risk losing public confidence before research findings are ever communicated. "The answer is to involve the public in all of the different loops, from agenda setting to actually doing, to finally once the science is done, communicating.... Otherwise, you don't know at what point you would have lost the public, so that even if your answer is clear and communicat[ive] and trustworthy, you may not get back those you have already lost," detailed one leader. Although AI-enabled automation may free up scientists to meet with stakeholders and explain their research to broader audiences, exchanges must be meaningful and inclusive.

Public confidence also varies by discipline and can depend on whether communities receive access to the advancements they help make possible. When AI accelerates research without addressing how its benefits are distributed or how livelihoods will be impacted, existing skepticism toward scientific institutions is likely to become further entrenched.

De-Skilling, Innovation Spirals, and Automation Pressures

Driven by top-down institutional and funding mandates, researchers face growing pressure to automate workflows for speed and cost reduction. This push for rapid automation conflicts directly with methodological precision and diligence in the peer review process. "The last thing I can do is go faster now. I have to go slower because I need more human beings checking everything," commented one contributor. Others noted that a lack of cohesive messaging pushing back against corporate and government AI hype directly fuels public distrust when the tools being promoted fail to deliver in

practice. Automation pressures additionally [disrupt the traditional mentor-apprentice relationship](#) between faculty and junior researchers, generating widespread career anxiety among younger scholars.

Beyond immediate training gaps, shifting human experts into model-annotation roles risks creating an innovation trap. “As you take people’s jobs away and you change the training and people now no longer have that expertise, the knowledge generation—the innovation—is not happening. So, the models get stuck [in] a previous generation of information, and that forefront of innovation doesn’t push forward.... It becomes this spiral, which will then lose public trust over time,” explained one technologist.

Protecting long-term scientific credibility requires mandating hybrid deployment models that keep human expertise at the center of research pipelines.

Hallucination vs. Creativity Trade-offs

Deploying AI agents for hypothesis generation reveals a constant tension between grounding models in existing literature to prevent hallucinations and allowing them to propose novel ideas, such as cross-disciplinary hypotheses in the digital humanities. While some observers have [suggested](#) that AI hallucination rates compare favorably to online misinformation, defining and engineering for scientific creativity remains a hurdle. One expert explained: “The deep problem we have here is that there’s no metric for creativity. You can’t look at a scientific output and ask five scientists how creative [it is] and get five similar answers, right? AI models often do very well if you can optimize them for something, but we don’t know how to optimize for creativity yet.” Although progress is occurring as researchers connect personalized agents to external databases, society must explicitly determine which scientific tasks are appropriate for AI and which core capabilities must remain uniquely human.

Biosecurity Risks and Unstable Governance

Deep polarization surrounding AI biosecurity risks directly undermines public confidence in scientific oversight. While some researchers fully dismiss AI-assisted biological threats, others express severe alarm. This divide erodes trust and creates volatility in the policy discourse: dismissiveness makes leaders appear negligent, while sensationalism fuels public anxiety and reactive policymaking. As [biological and cyber capabilities converge](#), institutions should prioritize clear, transparent, and balanced risk governance.

Recommendations

1. Equip Trustworthy Civic Institutions to Foster Public AI Understanding

Funding agencies and academic institutions should actively leverage trustworthy civic institutions—such as public libraries, museums, and independent educational bodies—to co-design research agendas, co-develop communication strategies to reach different segments of the population, and implement these strategies. Entities that have historically demonstrated poor trustworthiness should direct funding and partnership efforts toward equipping civic institutions with the resources to educate the public about the underlying workings of specific AI technologies and how these tools intersect with

lived experiences. Examples of concrete resources include financial backing and the deployment of AI experts working in different research fields.

2. Utilize Sustained Engagement Approaches to Identify Shortcomings in Trustworthiness

Decisionmakers should involve the public throughout the scientific process, from agenda-setting through downstream implementation and communication of findings. In addition to helping institutions gain insights into questions of public interest, continued engagement allows research teams to identify points in the scientific process where trust falters. Engagement should include transparent discussion of who contributes to research, how the methodology was determined, who benefits from its outcomes, and how resulting tools or discoveries will remain accessible to participating communities.

3. Explicitly Consider Links Between AI Infrastructure and Trustworthiness

Policymakers and research leadership should evaluate how the physical buildout and operational footprints of AI infrastructure, namely data centers, shape public perception and institutional trust. Data center developments [frequently impose localized public health, economic, and environmental burdens](#) on historically marginalized communities, directly driving public skepticism and community pushback against AI-enabled initiatives. Decisionmakers should examine the governance of these facilities, evaluating how different ownership models—including government, academic, or corporate control—influence confidence, equitable resource distribution, and long-term societal buy-in for AI-driven science.

4. Develop Workflows and Tools Around Transparent Research Records

To address lost attribution and declining provenance when AI models train on open scientific literature, standard-setting bodies and publishers should build on [existing disclosure efforts](#) and demand fully auditable, transparent research records, including mandatory code and workflow availability. Technology companies should then create tools with searchable, user-friendly interfaces to make these records accessible, in turn making disclosures enforceable.

5. Apply a Consumer-Protection Approach to AI Hype and IP Compensation

Restoring public trust requires commercial AI firms to demonstrate transparency, eliminate deceptive hype, and reinvest wealth into the public's scientific interest. While some initial lawsuits have proven the possibility of these outcomes, legal frameworks and institutional policies should apply a stronger consumer protection logic to place checks on corporate marketing practices and to support compensation for human-generated intellectual property that has been scraped from the web and commercialized by AI developers. Examples of government bodies with the authority to promulgate such

protection rules include the U.S. Consumer Financial Protection Bureau (CFPB), the U.S. Federal Trade Commission (FTC), and analogous institutions in other jurisdictions.

6. Deploy Precise Vocabulary to Discuss AI Technologies, Capabilities, and Limitations

Policymakers, scientific societies, and journalists should avoid relying on generic “black box” terminology and adopt specific classifications that distinguish between predictive and generative tools. Greater precision enables nuanced discussion of chatbots as outputting approximations of language rather than true language. Clear distinctions should also inform educational initiatives for science journalists, who are among the primary translators of research, and the broader public. Similar intentions are reflected in the newest update to the [Associated Press Stylebook](#), which directs writers to avoid anthropomorphizing AI technologies.

Particularly within the peer review process, editorial boards and research teams should more clearly identify use cases that assist with editing, translation, or preliminary feedback versus use cases that replace substantive analysis, authorship, or peer review. Disclosure policies should make these differences clear without discouraging uses that reduce barriers to scientific participation, such as translation. Such initiatives should build a foundational understanding of scientific methodology, validation processes, international regulatory approaches, and guardrails so citizens know what standards to demand. They should also be adapted to local languages, information channels, and levels of technical familiarity rather than relying on a single public-education model. Public communications from institutions and funders should explicitly set realistic expectations, reiterating that AI is not a scientific panacea.

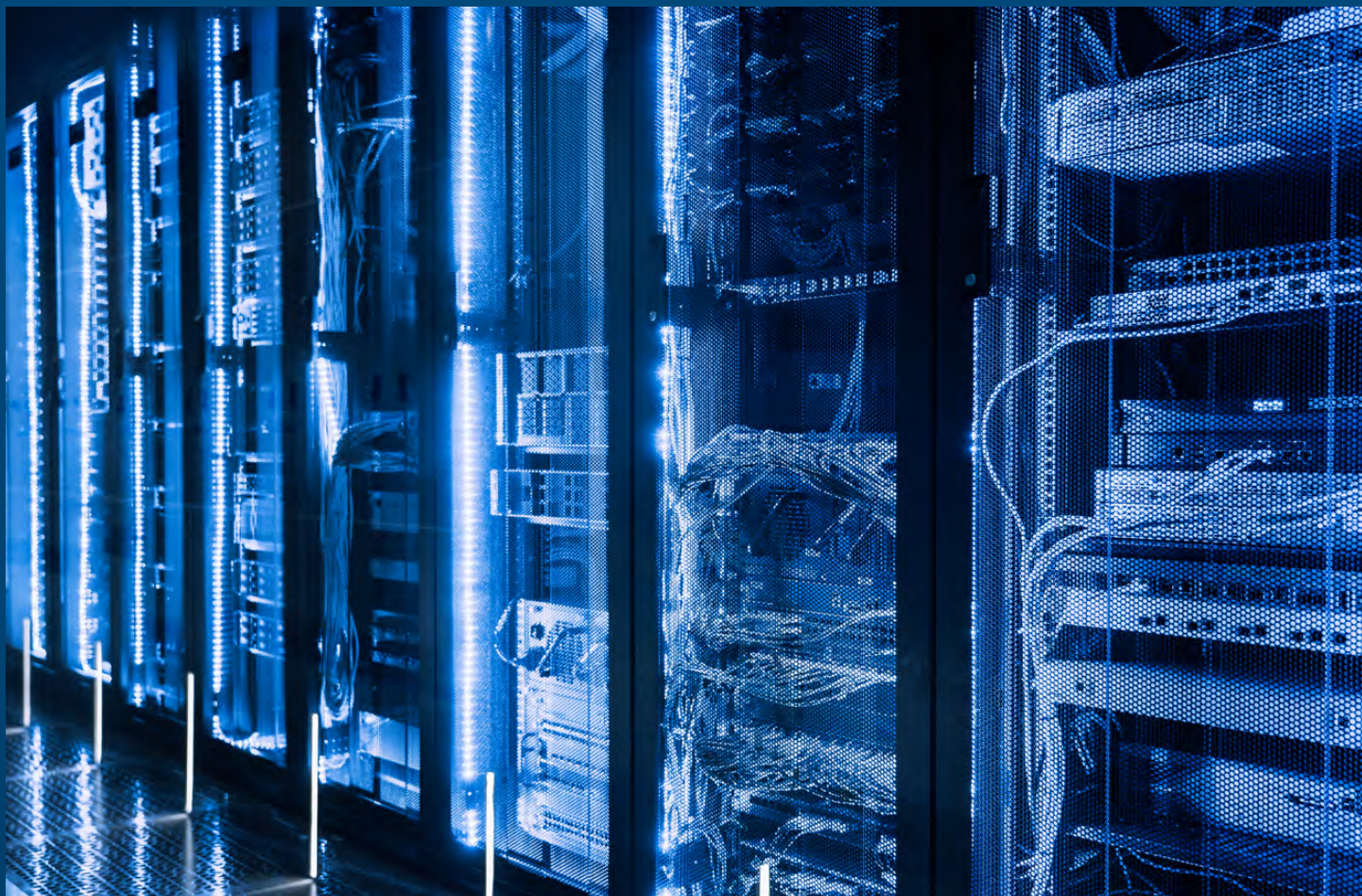
7. Take a Multi-Level Approach to Modernize Academic Incentives and Peer-Review Integrity

Given the interconnectedness of the science research ecosystem, university leadership, journal boards, and funding bodies should incrementally reform tenure, promotion, and grant evaluation criteria to disincentivize submission inflation, citation farming, and low-value research output. Peer-review frameworks should prioritize methodological transparency, data curation, and research question relevance over sheer publication speed. Institutions should support human review processes for AI-generated submissions and potential synthetic-data fabrication. Disclosure requirements should apply to reviewers as well as authors, with journals establishing clear policies governing the use of AI to evaluate manuscripts. Ongoing, academia-led reform efforts include the [Modernizing Academic Appointment and Advancement \(MA3\) Challenge](#) and the [San Francisco Declaration on Research Assessment \(DORA\)](#).

8. Enforce User-Centered, Hybrid Workflows to Safeguard Expertise

Research institutions, healthcare organizations, and technology developers should resist top-down pressures to automate purely for cost reduction. Funding and governance frameworks should mandate user-centered development and continuous human feedback loops, ensuring domain experts actively participate alongside developers throughout the design, deployment, and update processes. In instances where AI tools are being used for public decision making, government procurement mechanisms should explicitly demand measures to keep humans in the loop. Institutions should safeguard

[training pipelines and workforce structures](#) to prevent [cognitive de-skilling](#), protect mentor-student apprenticeships, address youth career anxieties, and ensure human expertise continues to drive scientific innovation. Finally, institutions should elaborate clear human accountability procedures around rogue AI agents.



**SCIENCE
& SOCIETY**



aspen institute