HARVARD Kennedy School
ASH CENTER
for Democratic Governance and Innovation
Allen Lab for Democracy Renovation

GETTING-Plurality Research Network
Allen Lab for Democracy Renovation
Ash Center for Democratic Governance and Innovation
Harvard Kennedy School
124 Mount Auburn Street, Suite 200-North
Cambridge, MA 02138

GETTING-Plurality Research Network Response to the National Science Foundation Request for Information on the Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

The GETTING-Plurality Research Network appreciates the opportunity to provide feedback on the development of the 2025 National AI R&D Strategic Plan to help secure America's position as a leader in artificial intelligence. We have compiled comments below from members of our research community including Sarah Hubbard, Allison Stanger, Shlomit Wagman, Ajeet Singh, and others from the GETTING-Plurality Research Network.

We are supportive of the priorities laid out in the first Trump Administration's National Artificial Intelligence Research and Development Strategic Plan in 2019, as well as the more recent updates to these priorities in 2023. We hope for the continuation of many of these strategic priorities in the Administration's second term, with a renewed focus on addressing the transformative potential of AI, while safeguarding American values and security interests. Below, we offer a few suggestions for federal R&D priorities over the next 3-5 years which we believe will support harnessing opportunities from AI innovations, enhancing U.S. economic and national security, and promoting human flourishing.

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the 2025 National AI R&D Strategic Plan and associated documents without attribution.

---

1. Public AI Infrastructure

Investing in robust public AI infrastructure is a national security imperative for the United States. As outlined in The National Security Case for Public AI, publicly-owned AI tech stack components (e.g. compute resources, datasets) and networking infrastructure would create a more resilient, innovative ecosystem while reducing dependence on a few private firms. It is critical for the U.S. to remain on the cutting edge of AI innovation in our global competition with China, so we can ensure that AI development is oriented towards democracy-, privacy-, and rights-protecting values. This investment will help ensure government independence from market actors with potentially conflicting interests, democratically accountable deployment, and will address other public goods traditionally underserved by corporate actors. We envision this infrastructure as a complement to private sector developments that could improve American democracy, national defense, and benefit the American people in their daily lives.

A foundational step here would be to codify the National AI Research Resource, as outlined in the CREATE Act of 2025.

2. Research into Developing More Ethically-Aware AI Systems

Building off of the previously developed priorities, “Strategy 2: Develop Effective Methods for Human-AI Collaboration” and “Strategy 4: Ensure the Safety and Security of AI Systems”, we would also urge that additional R&D is needed to evaluate AI’s capacity for ethical-moral reasoning and trustworthy decision-making. AI systems are being leveraged to make complex decisions that are often entangled with ethical-moral reasoning, from reshaping global warfare to personal use for therapy and companionship, healthcare, and criminal justice. As AI use expands in these critical domains, emerging evidence reveals issues in the trustworthiness and reliability of current leading models. A recent, notable example of this phenomenon was an update to GPT-4o that OpenAI had to promptly roll back after backlash from users about its sycophantic behavior. Recent studies such as “DarkBench” reveal that leading AI models today contain dark patterns with manipulative behaviors and untruthful communication. Models have also been found to sacrifice truth for sycophancy and even to strategically deceive their users.

In forthcoming research, we present findings from an experiment where we evaluate AI models across dimensions of ethical-moral intelligence, and demonstrate a critical need for more comprehensive ethical evaluation of AI systems. We would recommend additional R&D in this space to ensure the systems that we are building, and relying on, are more ethically aware and dependable.

3. Develop New Evaluation Methods for AI Systems

One of the previously developed priorities, “Strategy 6: Measure and Evaluate AI Technologies through Standards and Benchmarks” emphasizes the need for evaluative techniques for AI. However, we would like to caution against only using technically-oriented benchmarks and would encourage R&D investment into alternative assessment frameworks for the measurement and evaluation of AI systems.

A review from the European Commission Joint Research Center found issues with current benchmarks' weak construct validity, sociocultural context, and industry gaming, among other problems. Despite their existing flaws, policymakers are increasingly integrating these benchmarks into policy development–including the EU AI Act. We would propose exploring other methods for competency-based evaluation such as badging or certifications. Additionally, impact assessments with standards and practices that routinely monitor AI models not just by their technical merits and performance, but also by the real world consequences of their deployment would be valuable. As benchmarks have grown in popularity, and as AI continues to be integrated into critical functions, the stakes for evaluating these systems are higher than ever.

4. Addressing AI-Driven National Security Threats

The convergence of AI capabilities with adversarial intent presents critical research challenges for national security that requires R&D investment beyond private sector capabilities. From autonomous cyberattacks and AI-enhanced disinformation and terrorism, to deepfake-enabled social engineering and synthetic financial fraud, AI is rapidly amplifying traditional threat vectors and introducing new ones. Research priorities should include developing:

- AI tools which are resilient to adversarial manipulation in critical infrastructure contexts, in addition to systems capable of detecting and countering AI-generated attacks in real-time which are transparent and auditable
- Robust authentication methods for distinguishing synthetic information from authentic content
- Advanced techniques for AI misuse detection and clear reporting protocols
- Privacy-preserving techniques that enable threat detection without compromising civil liberties

This research agenda would help position the United States to lead international coalitions and set shared AI security standards while maintaining technological dominance.

5. Post-Section 230 Democratic Digital Infrastructure

The bipartisan move to sunset Section 230 would create a unique opportunity for R&D investment into democratic digital infrastructure. Recent scholarship establishes a crucial distinction between protected human expression and commercial algorithmic amplification, yet current AI systems often conflate these categories, creating constitutional vulnerabilities. Research investment into AI mechanisms that operationalize this distinction would fill a critical gap, such as developing:

- Federated AI governance protocols and technical standards
- AI architectures that distinguish between use-directed content discovery and commercial content amplification
- AI systems which are optimized for civic value rather than attention capture, which could be leveraged in AI-enhanced deliberative democracy platforms, educational systems

which are designed for learning rather than engagement, scientific discovery that prioritizes knowledge advancement over commercial application, and other public information systems

- Technical infrastructure that enables citizen control over personal data and AI interactions through cryptographic protocols that ensure data ownership, interoperable identity systems that prevent platform lock-in, and other privacy-preserving technologies

The convergence of sunsetting Section 230, growing platform manipulation concerns, and advances in AI creates an unprecedented opportunity for American leadership in democratic digital infrastructure. Federal R&D investment in this domain addresses clear market failures while positioning the United States as the global leader in trustworthy AI systems. By investing in AI systems designed around constitutional principles rather than commercial metrics, the United States can demonstrate that democratic values create sustainable competitive advantage in the global technology landscape.

6. Public Goods Opportunities

While there is often a focus on mitigating the risks of AI, we should also seek to ensure that new opportunities are seized. In some cases, there will be new opportunities for R&D where commercialization is not the best vehicle for supporting the development and scaling of novel technologies. As we outline in our paper, A roadmap for governing AI: technology governance and power-sharing, we can look for public goods opportunities aligned with the following dimensions: 1. Individual and community flourishing (consumer protection, user safety, social and mental health, and climate and sustainability); 2. democratic/political stability; and 3. economic empowerment (integration, innovation, and creativity). We recommend the development of national R&D in areas that support these domains such as:

1. *Individual and Community Flourishing:*
    a. personalization of learning and translation of credentials; education and vocational training;
    b. improved access to expert advice and internet literacy;
    c. contextualization engines to help protect against fraud, misinformation, and disinformation.
2. *Democratic/ Political Stability:*
    a. increased opportunities to engage;
    b. translation: cross-jurisdictional possibilities.
3. *Innovation, Creation and Economic Integration*
    a. improved educational and training opportunities;
    b. advances in drug development, cancer research, and other sciences
    c. entrepreneurial opportunities;
    d. potentially new jobs emerging;
    e. "task diversity" - one person can complete many more different kinds of tasks than they could before.

Strategic government investment in these public goods opportunities can complement private sector innovation to benefit the American people, while strengthening American leadership and economic growth.

7. Prepare the American Workforce

AI will continue to play a prominent role in the U.S. economy, and as these technologies continue to improve, we must prepare the American workforce for the implications. A recent report from Pew Research shows that "workers are more worried than hopeful about future AI use in the workplace" and a recent study from Microsoft and Carnegie Mellon demonstrated that the rise of generative AI in knowledge work is negatively impacting workers critical thinking skills and practices. We expect that AI will have a significant destabilizing and transformative impact on workers, their families, and communities. Federal R&D investment should focus on developing evidence-based frameworks for workforce transition and human-AI collaboration, research into training methodologies and opportunities for reskilling, as well as how to equip young people and how to harness innovation for individual and collective good.

Together, these research priorities position the United States to maintain global leadership while ensuring AI development serves democratic values, the American people, and broad American interests over the next 3-5 years.

About the GETTING-Plurality Research Network
Governance of Emerging Technology and Tech Innovations for Next-Gen Governance (GETTING-Plurality) is a multi-disciplinary research network linking philosophers, social scientists, computer scientists, legal scholars, and technologists. We are building a unique collaborative that unites technology and policy initiatives at Harvard University with external impact partners across higher education and the tech industry. More information: https://ash.harvard.edu/programs/getting-plurality/

For any additional information on the comments above, please reach out to Sarah Hubbard at sarah_hubbard@hks.harvard.edu.