

PUBLIC SUBMISSION

| |
|--|
| Received: May 29, 2025 Tracking No. mb9-xk3n-fs6u Comments Due: May 28, 2025 Submission Type: Web |
|--|

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0278
Comment on FR Doc # 2025-07332

Submitter Information

Organization: The Institute for Trustworthy AI in Law and Society

General Comment

See attached file(s)

Attachments

TRAILS RFI Response



Overview

The Institute for Trustworthy AI in Law and Society (TRAILS) submits this response to the 2025 National AI Research and Development (R and D) Strategic Plan Request for Information. TRAILS is a multi-institutional, interdisciplinary research institute dedicated to advancing the science of trustworthy AI through participatory approaches that center the voices of diverse stakeholders. Our work integrates technical, social, legal, behavioral, economic, and ethical perspectives to understand how and when trust is built and maintained in AI systems. Drawing on our members' research and convenings, we offer recommendations aimed at ensuring that federal leadership in AI fosters systems that are not only innovative and secure but also aligned with democratic values and public trust. These priorities support the goals of Executive Order 14179 and the RFI by identifying foundational areas where scientific trust-centered approaches can fill gaps that market forces alone will not address.

Recommendation 1: Focus Research on AI for Information Integrity and Public Trust

We recommend that the federal government prioritize AI R and D that advances information integrity and promotes public trust – including the cognitive foundations of this trust – across sectors critical to public welfare, economic competitiveness, and national security.

As AI systems increasingly generate content across high-stakes domains—such as healthcare, law, public policy, and education—the threat of plausible-sounding but inaccurate outputs poses a growing risk to public trust and informed decision-making. This risk directly intersects with the Administration's priorities of promoting human flourishing and securing national interests. Ensuring that AI-generated information is accurate, complete, verifiable, and interpretable is essential to maintaining U.S. leadership and preventing inaccurate information from eroding social and institutional trust.

The U.S. government must invest in this area of research because the private sector lacks sufficient incentives to do so. Companies tend to prioritize short-term gains such as user

engagement or novelty, rather than foundational work aimed at verifiable accuracy, user understanding, or long-term public benefit. This kind of basic research—critical to aligning AI systems with public values and supporting cognitive resilience—is unlikely to yield immediate commercial returns, making it an essential responsibility of federal leadership.

1. Foundational AI for Accuracy and Interpretability

Support development of AI architectures and training methods that produce accurate outputs, particularly in safety-critical areas. Encourage exploration of smaller foundation models that require less computing power and training data, promoting broader accessibility and reducing reliance on massive centralized infrastructure.

2. Human-Centered AI Design for Cognitive Trust

Invest in research on how users perceive, interpret, and act on AI-generated content. This includes:

- Applying leading scientific paradigms of human cognition and decision-making, such as Fuzzy-Trace Theory¹, to improve users’ ability to detect errors or misleading statements.
- Studying how interface design, prompt structure, and user expertise affect trust and comprehension.

These efforts support the Executive Order’s call to develop AI systems free from engineered agendas and aligned with practical human reasoning.

3. Impacts of AI on Users

Fund research into the cognitive, behavioral, and emotional effects of AI systems on users and communities. Private markets are unlikely to address these impacts, especially over the long term.

4. Accountable and Transparent AI

Investigate recourse mechanisms when AI systems deny access to critical services, and develop tools to promote trust, when warranted, in automated decision-making. These safeguards help ensure AI systems remain consistent with public service goals and bolster institutional legitimacy.

5. AI-Enabled Data and Analysis of Public Institutions

Explore the use of AI for public analysis of institutions, promoting transparency and

¹ Edelson, S. M., Roue, J. E., Singh, A., & Reyna, V. F. (2024). How Decision Making Develops: Adolescents, Irrational Adults, and Should AI be Trusted with the Car Keys?. *Policy Insights from the Behavioral and Brain Sciences*, 11(1), 11-18.

accountability. This reflects a commitment to empowering the public to inform the work of experts and public servants to promote democratic values.

6. Governance to Foster Trust and Innovation

Support research on adaptive governance models that evolve alongside AI technologies, with a focus on sustaining public trust. Policymakers need clearer insights into which governance approaches effectively build confidence without stifling innovation. This includes identifying strategies that promote transparency, responsiveness, and legitimacy, ensuring that regulatory frameworks can keep pace with technological change while enabling responsible progress.

To fulfill the vision laid out in **Executive Order 14179** and shape a National AI R and D Strategic Plan that genuinely enhances American leadership, the federal government should invest in AI research that promotes information integrity and cognitive trust. This aligns with national priorities where market incentives alone are inadequate and supports a long-term foundation for safe, trustworthy, and globally competitive AI development. Advancing these capabilities is vital not only to U.S. innovation but to the resilience of our democratic institutions and public trust in the digital age.

Recommendation 2: Build a National AI Trust Infrastructure to Secure U.S. Leadership

We recommend that the 2025 National AI R and D Strategic Plan prioritize the development of a National AI Trust Infrastructure that is supported by scientifically validated precursors of trust in AI, which is essential for maintaining U.S. leadership and public confidence in AI systems used across critical sectors. As the Executive Order makes clear, retaining global AI dominance requires trust not only in outcomes but also in the underlying technical foundations. The absence of standardized and scalable tools that are known to promote trust threatens both national security and economic competitiveness. Such tools might include, but need not be limited to, standards for model documentation, dataset transparency, and responsible deployment. Without trusted infrastructure, AI innovations may face diminished adoption in safety-critical sectors or be outpaced by foreign systems better prepared for regulatory or reputational challenges.

The federal government is uniquely positioned to lead in this space. Commercial platforms such as GitHub and HuggingFace serve important roles, but they do not provide the vetted, mission-aligned governance needed for national-scale AI applications. This is a classic example of a strategic gap where federal investment can drive long-term public benefit without being constrained by market incentives.

Recommended Actions

1. Validate, Test and Scale Technical Governance Tools

Tools such as the Trustworthy AI Checklist and Data Governance Metric, developed by TRAILS and others, offer pragmatic, actionable steps for ensuring model transparency and responsible deployment. These tools should be tested in federal pilot programs and refined in collaboration with agencies such as the National Institute of Standards and Technology (NIST) and the Office of Science and Technology Policy (OSTP) to ensure scalability and alignment with the national interest.

2. Enable Research Access to Models that Serve as Critical Infrastructure

Facilitate reasonable academic access to data and model internals under controlled, privacy-preserving conditions to support safety and performance research. This promotes innovation and public safety while respecting commercial IP protections—a balance critical to U.S. AI competitiveness.

3. Support Verification and Provenance Tools

Invest in national infrastructure for provenance, (such as watermarking) enabling the public and institutions to distinguish between authentic content and AI-generated material. This is vital for ensuring that the public receives accurate, authentic information in response to queries—supporting EO 14179’s focus on human flourishing.

4. Prioritize Dynamic Evaluation Benchmarks

Fund research and development emphasizing contextualized, evolving benchmarks that reflect deployment conditions. This supports the development of robust, adaptable AI systems better suited for national-scale implementation in key industries central to human flourishing and national competitiveness, such as health, defense, and infrastructure.

5. Fully Implement the National AI Research Resource (NAIRR) pilot

Accelerate and expand NAIRR to provide shared computational infrastructure for vetted, high-impact AI research. This will facilitate a level playing field for academic and public-sector researchers and increase the strategic utility of federal Rand D investment.

6. Establish a National AI Infrastructure Platform

Develop a federally supported platform for sharing tools, models, and datasets intended for public use and public-interest AI development. The platform should be guided by scientifically validated methods and empirical research to ensure reliability and utility.

Key features would include:

- Standardized metadata and documentation protocols to facilitate innovative scientific breakthroughs while maintaining research integrity
- Structured risk disclosures and audit trails to support independent evaluation

- Scientifically valid indicators of robustness, safety, and performance

This platform could be hosted or coordinated through NIST, NSF, NTIA, or other federal research entities, complementing existing research standards without imposing ideologically motivated frameworks.

A National AI Trust Infrastructure directly supports the Administration’s stated policy goals to “promote human flourishing, economic competitiveness, and national security” through AI. It also exemplifies the kind of long-term, foundational investment that falls squarely within the federal government’s unique responsibility, especially in areas where market mechanisms alone are insufficient. Without such infrastructure, the U.S. risks ceding leadership to countries with more integrated public-private AI strategies.

By acting decisively, the federal government can ensure that trust, safety, and reproducibility are not afterthoughts—but competitive advantages in the global AI race.

Recommendation 3: Advance AI for Societal Resilience and Public Sector Application

We recommend that the federal government prioritize AI R and D that strengthens societal resilience and supports public sector transformation. These are mission-critical areas that are unlikely to attract sustained private-sector investment but are vital to the national interest, especially in areas where AI can improve crisis response, healthcare delivery, and public infrastructure.

The Executive Order calls for removing barriers to American AI leadership and emphasizes the promotion of human flourishing, economic competitiveness, and national security. Achieving these goals requires AI systems that function reliably in the real world for all Americans—across decentralized agencies, multiple communities, and evolving crises. These systems must be embedded in public institutions, tailored to the needs of frontline workers, and supported by interdisciplinary research, such as research on decision making under risk and uncertainty.

Such high-impact applications frequently require sustained federal support, particularly when market forces deprioritize long-term or non-commercial domains. The government has a unique role in de-risking research in areas like disaster response, scientific discovery, and public health—where the payoffs are national, but the immediate monetization paths are unclear.

Recommended Federal R and D Priorities

1. AI for Crisis Preparedness and Response

Invest in AI tools for crisis mapping, emergency logistics, and policy simulations—including the use of digital twins to model and rehearse coordinated disaster responses. These applications are critical to public safety and are aligned with EO 14179’s mandate to enhance national security.

2. Mission-Driven AI Co-Designed with Stakeholders

Support real-world deployment of AI systems developed through close collaboration with public institutions. Several examples are shown below.

- *BeFreeBot*,² an AI-powered smoking cessation chatbot developed in collaboration with public health agencies
- Socially assistive robotics, e.g., for in-home care of autistic children

These projects exemplify the kind of high-impact, mission-aligned AI that public-sector R and D can innovate.

3. AI-Human Collaboration in Discovery and Creativity

Fund research into machine-in-the-loop systems that augment human reasoning and creativity in science, design, and the humanities. Such hybrid systems can accelerate U.S. innovation capacity across strategic domains.

4. AI for Edge Environments and Connectivity-Limited Settings

Develop AI systems that function independently of cloud infrastructure, supporting rural, remote, or infrastructure-vulnerable areas. This enhances U.S. resilience and broadens access to AI capabilities across the country.

5. AI to Facilitate Access to Public Services for All Americans

Invest in AI tools that ensure that public services are available to all Americans, including individuals with disabilities, ensuring broader societal participation in the benefits of AI technologies—an outcome squarely aligned with national goals of human flourishing and service equity.

6. Civic and Infrastructure-Focused AI Applications

Expand early-stage research into transportation optimization, energy grid resilience, and local governance tools, with the federal government acting as a catalyst for innovation in

² Abroms, L. C., Yousefi, A., Wysota, C. N., Wu, T. C., & Broniatowski, D. A. (2025). Assessing the Adherence of ChatGPT Chatbots to Public Health Guidelines for Smoking Cessation: Content Analysis. *Journal of Medical Internet Research*, 27, e66896.

domains essential to infrastructure modernization.

Federal investment in public-sector and resilience-oriented AI is a strategic necessity for the next phase of American AI leadership. These domains reflect national priorities and require R and D commitments that extend beyond what commercial interests can reliably support. By acting now, the government can ensure that AI is not only globally competitive, but also domestically transformative—fulfilling the ambitions of Executive Order 14179 and the 2025 National AI R and D Strategic Plan.

Recommendation 4: Fund Research into Participatory AI for Public Confidence and National Performance

The federal government should expand research into participatory methods for AI design, development, deployment, and evaluation that increase public trust, improve model quality, and ensure alignment with real-world users’ needs—especially in applications affecting national services and public institutions.

Executive Order 14179 underscores the importance of maintaining America’s global leadership in AI by developing systems that serve human flourishing and enhance national resilience. In high-impact domains—such as public health, education, transportation, and law enforcement—AI systems must not only perform well, but also operate in a way that is interpretable, accountable, and trusted by the people who rely on them.

Research into practical participatory methods can directly improve technical outcomes by providing earlier identification of failures, biases, or blind spots. Such methods also foster public understanding of how AI systems work, aligning with EO 14179’s focus on ensuring that innovation supports rather than alienates the American public.

Because these participatory approaches are not widely commercialized, they are underexplored and represent a clear area for federal R and D leadership, especially where democratic legitimacy and national-scale deployment are concerned.

Recommended Federal R and D Priorities

1. Fund Applied Research on Participatory Methods

Support research designed to evaluate scalable tools and practices—such as citizen assemblies, deliberative workshops, and real-time user feedback platforms—that might allow affected populations to meaningfully engage with the design, use, and oversight of AI systems. For example, TRAILS’ participatory auditing platforms enable users to test live AI systems and flag concerns, producing data that improve model safety and helps

the public build realistic expectations of system behavior.

2. Study the Conditions for Trust and Effective Engagement

Advance research on when and how public participation contributes to greater trust, technical robustness, or behavior change in deployed systems. This includes managing competing goals among stakeholders and identifying which engagement strategies scale most effectively across domains.

3. Launch Researcher-in-Residence Programs

Fund the placement of AI researchers and technical experts inside federal agencies and NGOs to directly observe operational needs, align system development with real-world use cases, and inform participatory methods from within the institutional context.

4. Incentivize Partnerships Between Academia and Local Stakeholders

Establish funding mechanisms that pair universities and researchers with communities and organizations that can enable the development of AI tools that are grounded in practical experience and domain expertise—especially in health, education, and infrastructure.

5. Promote Operational Transparency in Public Sector AI Use

Support initiatives that evaluate how AI is deployed across government, enabling outside researchers to better align their work with government challenges and increase the strategic value of federal AI investments.

Research into scientifically innovative and participatory AI can improve the performance, accountability, and public value of AI systems that increasingly shape American life without imposing unnecessary compliance requirements or undue regulatory burden. This research supports the core directives of Executive Order 14179 by ensuring AI systems are designed and deployed in ways that are technically robust and publicly trustworthy—strengthening national cohesion, improving mission delivery, and reinforcing U.S. leadership in AI innovation.

Recommendation 5: Establish a Mission-Driven AI Research Laboratory for National Competitiveness and Strategic Depth

We recommend the federal government establish a flagship national research laboratory system focused on long-horizon, mission-driven AI R and D, modeled on successful international examples such as the Max Planck Institutes. This institute would strengthen the foundations of U.S. AI leadership by investing in critical areas underserved by commercial actors—particularly those with national security, infrastructure, and public sector relevance.

Executive Order 14179 emphasizes the need to maintain American leadership in AI through decisive policy action, long-term vision, and investments in foundational research. As generative and autonomous systems evolve rapidly, there is a growing need for a federally supported research laboratory system dedicated to AI challenges with high national value but uncertain near-term commercial returns. These include infrastructure resilience, public-sector transformation, health and safety, national competitiveness, and scientific discovery.

Unlike private laboratories or short-cycle research funding, a standing federal laboratory system would provide strategic continuity, deep technical focus, and talent development pipelines needed to sustain leadership over decades—not just product cycles. Unlike our competitors, the U.S. currently lacks a public-interest AI national research laboratory system with the scale, autonomy, and interdisciplinary mandate to address these long-term needs in a systematic way.

Institute Objectives and Focus Areas

1. Long-Horizon, Foundational AI Research

Advance research in areas such as:

- AI systems relevant to national competitiveness and defense as well as health and safety, such as medical advances, medical decision making, and agentic and embodied systems.
- Next-generation human-AI interaction, including interpretability strategies to promote collaboration and operational trust
- AI for scientific discovery, modeling, and policy simulation
- Public-sector AI systems designed for transparency, robustness, and responsible deployment

2. Infrastructure Stewardship and Standards Development

Host or support development of shared tools, models, datasets, and benchmarks—especially for use by government agencies and public institutions. These efforts would complement existing work by NIST and other federal actors by focusing on non-commercial, mission-critical use cases.

3. Translation and Deployment Support for Government Partners

Facilitate the translation of research into usable tools for federal agencies, state and local governments, and public educators—bridging the gap between academic work and operational needs. This supports EO 14179’s emphasis on practical, national-scale applications.

4. Human Capital Development

Develop programs to train researchers and other scientists and engineers in areas of national need, particularly those related to safety-critical and public-serving AI. The institute could support fellowships, public-sector sabbaticals, and embedded researcher programs to strengthen the talent pipeline for government and academic institutions.

5. Coordination, Not Duplication

Work in partnership with university-based centers such as TRAILS to ensure complementarity and coordination, avoiding duplication of effort and leveraging existing research ecosystems.

Establishing a mission-driven national AI institute would fulfill a strategic gap in the U.S. AI ecosystem and align directly with the vision laid out in Executive Order 14179. It would strengthen U.S. leadership in non-commercial AI applications, support long-term innovation in areas essential to national well-being, and ensure that foundational research continues to serve the public interest and government mission—advancing American prosperity, security, and scientific achievement in the age of AI.

Recommendation 6: Support Research on Governance to Facilitate AI Innovation by Building and Sustaining Trust

Because AI is simultaneously a dual use and general-purpose technology, policymakers in many countries are determined to nurture AI. Most countries have not developed clear strategies to govern AI that build and sustain trust. Although the EU and China have adopted direct regulation of high-risk systems, officials from other nations may fear stifling innovation by overregulating AI or corporate business practices. The absence of such governance strategies threatens US AI leadership, especially if US AI companies lose market share. Research examining which governance strategies promote trust in AI and how this translates to adoption can simultaneously incentivize innovation while sustaining trust.

Conclusion: Advancing American AI Leadership Through Strategic Federal Investment

In alignment with the goals of Executive Order 14179 and the objectives of the 2025 National AI R and D Strategic Plan, the five recommendations presented here outline a cohesive federal strategy to secure long-term U.S. leadership in artificial intelligence. These proposals are rooted in the recognition that commercial incentives alone will not sustain the foundational research, infrastructure development, and public-sector innovation required to ensure AI advances American competitiveness, national security, and human flourishing.

1. **Focus Research on AI for Information Integrity and Public Trust:** The federal government should prioritize AI R and D that advances information integrity and public trust—including the cognitive foundations of that trust—by investing in foundational accuracy, human-centered design, societal impact research, transparent and accountable systems, institutional analysis, and adaptive governance, all of which are essential to counter risks posed by inaccurate AI outputs in high-stakes domains and to ensure that AI systems remain aligned with democratic values, national security, and long-term U.S. leadership in a competitive global landscape.
2. **Build a National AI Trust Infrastructure to Secure U.S. Leadership:** The federal government should establish a National AI Trust Infrastructure grounded in scientifically validated tools and standards to ensure transparency, safety, and accountability in AI systems—an essential investment for maintaining U.S. leadership, enabling responsible innovation, and securing national competitiveness.
3. **Advance AI for Societal Resilience and Public Sector Application:** The federal government should prioritize AI R and D that strengthens societal resilience and transforms public sector capabilities—through sustained, mission-driven investment in areas such as crisis response, healthcare, infrastructure, and public services—because these high-impact applications, essential to national security, human flourishing, and economic competitiveness as outlined in Executive Order 14179, require risk-tolerant, interdisciplinary research and federal support that private markets are unlikely to provide.
4. **Fund Research into Participatory AI for Public Confidence and National Performance:** The federal government should invest in research on practical, scalable participatory methods for AI design, development, deployment, and evaluation—such as user feedback platforms, researcher-in-residence programs, and partnerships with local stakeholders—to enhance model quality, public trust, and alignment with real-world needs in high-impact domains such as health, education, and infrastructure, thereby advancing the goals of Executive Order 14179 to ensure AI systems are interpretable, accountable, and democratically legitimate.
5. **Establish a Mission-Driven AI Research Laboratory for National Competitiveness and Strategic Depth:** The federal government should establish a flagship national AI research laboratory system, modeled on successful international examples, to conduct long-horizon, mission-driven R and D in critical areas underserved by the private sector—such as national security, public infrastructure, and scientific discovery—thereby advancing Executive Order 14179’s vision through sustained foundational research, talent development, public-sector innovation, and coordinated infrastructure stewardship that ensures U.S. leadership in non-commercial, high-impact AI applications.

- 6. Support Research on Governance to Facilitate AI Innovation by Building and Sustaining Trust:** Because AI is both a dual-use and general-purpose technology, the federal government should invest in research on governance strategies that build and sustain public trust without stifling innovation, addressing a global gap in regulatory approaches and ensuring U.S. competitiveness amid international efforts—such as those in the EU and China—to directly regulate high-risk AI systems.

Together, these recommendations reflect the federal government’s unique role in steering AI development toward outcomes that benefit the American people, strengthen national competitiveness, and secure U.S. leadership for generations to come. Each proposal is designed to address domains where the return on investment is societal rather than purely commercial—making them ideally suited for federal prioritization over the next three to five years.

We urge their inclusion in the 2025 National AI R and D Strategic Plan as pillars of a forward-looking, mission-aligned, and impact-driven national AI strategy.