PUBLIC SUBMISSION

Received: May 26, 2025
Tracking No. mb5-vkhg-i2oa
Comments Due: May 28, 2025 **Submission Type:** API

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-0110
Comment on FR Doc # 2025-07332

Submitter Information

Name: Matthew Brooks

General Comment

See attached file(s)

Attachments

Matthew Brooks - Public Comment on the 2025 National AI Research and Development Strategic Plan

Public Comment on the 2025 National AI Research and Development Strategic Plan

Submitted by: Matthew Brooks

Executive Summary

The next 3-5 years represent a critical window for shaping AI's trajectory toward beneficial outcomes. Federal R&D investments should prioritize areas where commercial incentives are misaligned with long-term public good: AI safety research, coordination infrastructure, robust evaluation frameworks, and tools that enhance human decision-making rather than replace it. These investments will determine whether AI amplifies human capability and wisdom or creates systemic risks we're unprepared to handle.

Introduction

As AI capabilities advance at unprecedented rates, we face a fundamental challenge: the most profitable AI applications are not necessarily the most beneficial for society. Market forces drive development of engagement-maximizing algorithms and automation that displaces human judgment, while underinvesting in safety research, coordination tools, and systems that enhance rather than replace human reasoning.

Federal R&D funding must fill these critical gaps. The private sector won't adequately fund research with 5-10 year payoff horizons, work on preventing low-probability catastrophic risks, or infrastructure that benefits competitors. Yet these areas are essential for ensuring AI promotes human flourishing alongside economic competitiveness.

Priority Research Areas

1. AI Safety and Alignment Research

Current Gap: Industry focuses on capabilities over safety, with alignment research receiving <5% of AI R&D investment despite potentially catastrophic consequences of misaligned systems.

Federal Priorities:

- Fundamental research on value alignment and goal specification in AI systems
- Development of interpretability tools to understand AI decision-making processes

- Research on robustness to distribution shift and adversarial inputs
- Investigation of emergent behaviors in large-scale AI systems
- Creation of "AI safety benchmarks" that become industry standards

Why Federal Funding: No immediate commercial return, benefits all players equally, requires long-term thinking beyond quarterly earnings.

2. Coordination and Epistemic Infrastructure

Current Gap: As AI accelerates change, humans struggle to maintain shared understanding of rapidly evolving landscapes. Information overload prevents effective coordination on critical issues.

Federal Priorities:

- AI-powered tools for synthesizing research across disciplines
- Systems for tracking scientific consensus and surfacing disagreements productively
- Platforms for structured transparency that preserve privacy while enabling verification
- Tools for enhancing collective decision-making and reducing information asymmetries
- Infrastructure for testing and certifying AI systems' epistemic virtues (truthfulness, clarity, non-deception)

Why Federal Funding: Classic public goods problem - everyone benefits from better coordination, but no single actor captures enough value to justify investment.

3. Human-AI Collaboration Research

Current Gap: Industry optimizes for automation and replacement rather than augmentation and empowerment of human decision-making.

Federal Priorities:

- Research on AI systems that enhance rather than replace human judgment
- Development of AI advisors that help humans identify and correct cognitive biases
- Tools for preference elicitation that help people understand their own values
- Systems for AI-mediated negotiation and conflict resolution
- Research on maintaining human agency and meaningful choice in AI-integrated systems

Why Federal Funding: Market incentives favor dependency-creating products over empowerment tools.

4. Robust Evaluation and Standards

Current Gap: Current AI evaluation focuses on narrow performance metrics, missing critical safety and societal impact dimensions.

Federal Priorities:

- Development of comprehensive AI evaluation frameworks beyond accuracy metrics
- Research on long-term and systemic effects of AI deployment
- Creation of standardized safety testing protocols
- Investigation of AI failure modes and edge cases
- Establishment of certification processes for high-stakes AI applications

Why Federal Funding: Standards benefit everyone but disadvantage first movers, creating a collective action problem only government can solve.

5. High-Risk, High-Reward Foundational Research

Current Gap: Industry avoids genuinely novel approaches that might obsolete existing investments.

Federal Priorities:

- Alternative AI architectures prioritizing interpretability and controllability
- Research on AI systems with explicit uncertainty quantification
- Investigation of AI approaches inspired by human cognitive development
- Exploration of AI architectures with built-in value learning capabilities
- Development of AI systems designed for graceful degradation and safe failure

Why Federal Funding: Potential to revolutionize the field but too risky for private investment.

Conclusion

The United States has an opportunity to lead not just in AI capabilities but in ensuring AI benefits all of humanity. By investing in safety research, coordination infrastructure, and human-empowering AI tools, we can create competitive advantages that align with our values of human dignity and flourishing.

The next 3-5 years will determine whether AI becomes a force for unprecedented human empowerment or systemic risk. Federal R&D investments in these overlooked areas are essential for navigating this transformative period successfully.