PUBLIC SUBMISSION

Received: May 29, 2025
Tracking No. mb9-jskd-1n7y
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-0194
Comment on FR Doc # 2025-07332

Submitter Information

Organization: Arm, Inc.

General Comment

See attached file(s)

Attachments

Arm_USAIRD_NSF_29May2025

Arm Inc. Response to Request for Information on the 2025 National Artificial Intelligence Research and Development Strategic Plan

Submitted to: Networking and Information Technology Research and Development (NITRD) National Coordination Office, National Science Foundation

Date: May 29, 2025

Submitted by: Vince Jesaitis, Vice President & Head of Global Government Affairs, Arm Inc.

Introduction

Arm appreciates the opportunity to contribute to the development of the 2025 National Artificial Intelligence Research and Development Strategic Plan. As a global leader in semiconductor intellectual property (IP), computing architectures, compute sub-systems and related technology, Arm's products underpin innovation across the entire AI technology stack—from hyperscale data centers to intelligent edge devices. Arm's energy-efficient, high-performance technologies are foundational to U.S. advancement in AI, enabling secure, scalable, and economically viable solutions across diverse industries.

To sustain and expand America's innovative edge in artificial intelligence, targeted federal investments and strategic policy initiatives are crucial. Arm offers recommendations on seven key areas: (1) semiconductor research and development, (2) secure federal AI infrastructure, (3) efficiency in AI compute, (4) edge AI deployment, (5) robust hardware-level AI security, (6) initiatives expanding AI access and research, and (7) advancements in physical AI and robotics.

1. Strengthen Semiconductor Research and Development (R&D)

Semiconductors are foundational to AI innovation, as advances in chip design directly enable greater computational performance, energy efficiency, and scalability of AI systems. Enhanced semiconductor capabilities are essential for supporting increasingly complex AI models and applications, from data centers to edge devices. To maintain America's competitive edge in AI, sustained federal support for semiconductor innovation must remain a priority. The CHIPS and Science Act and programs like the National Semiconductor Technology Center (NSTC) and the National Advanced Packaging Manufacturing Program (NAPMP) provide essential foundations for semiconductor innovation. Arm recommends expanding these efforts:

Expand NSTC and NAPMP programs:

- Prioritize funding for collaborative research in heterogeneous chip architectures, combining CPUs, GPUs, and specialized AI accelerators to optimize performance and energy efficiency.
- Increase investment in advanced packaging technologies, such as chiplets, to enable modular, flexible computing systems.

Enhance federal support for semiconductor startups:

- Leverage programs like the NSF's Regional Innovation Engines and DARPA's Electronics Resurgence Initiative (ERI 2.0) to provide targeted support for emerging semiconductor companies.
- Extend the Department of Defense's Microelectronics Commons program to facilitate prototyping of novel AI-optimized hardware, lowering barriers to entry for startups and small businesses.

2. Develop Secure and Sovereign Federal AI Infrastructure

Sensitive governmental and national security workloads require dedicated, secure AI computing infrastructure. While leveraging commercial cloud capacity is beneficial for scalability and efficiency in many federal AI applications, certain AI workloads involving sensitive or classified data require infrastructure under direct federal oversight. Additionally, in cases where government priorities and commercial financial incentives diverge, such as in long-term research or specific national security missions, dedicated federal investment in sovereign AI infrastructure becomes essential. Arm recommends significant federal investment to build this dual-capability approach, balancing commercial partnerships with sovereign capabilities:

Establish dedicated federal AI compute clusters:

- Expand Department of Energy (DOE) and National Nuclear Security Administration (NNSA) initiatives, including the Advanced Simulation and Computing (ASC) program, to build secure AI clusters for defense, intelligence, and critical infrastructure.
- Utilize the National AI Research Resource (NAIRR) as a model to ensure broad federal access to advanced computing resources, complementing existing commercial infrastructure with dedicated government-managed platforms.

Invest in federal testbeds for classified AI R&D:

- Support platforms like the DOE's AI for Science initiative to develop classified models for national security challenges such as nuclear security, cybersecurity, and defense logistics.
- Facilitate interagency collaboration for AI development through enhanced secure data sharing, standardized AI model frameworks, and integrated classified research environments.

3. Prioritize Efficiency in AI Compute

The operational costs of AI infrastructure—particularly energy consumption—are substantial and growing. In large-scale high-performance computing (HPC) systems commonly used for AI training and inference, energy costs alone can account for approximately 30-50% of the total operational expenditure. Although advances in efficiency techniques can temporarily reduce energy usage per workload, these efficiencies often lead to increased overall demand, driven by more intensive training, broader model deployment, and growing inference requirements.

Moreover, the location of data centers is frequently constrained by the availability of gigawatt-scale grid power and lengthy permitting processes. Additionally, the quality of power supplied by strained grids poses potential safety risks, such as electrical fires. As inference workloads continue to grow, particularly at the edge, decentralized upgrades to local grids will be required, demanding greater attention in infrastructure planning and permitting.

Promote efficient, low-power AI architectures:

- Expand DOE's Energy Efficiency Scaling for Two Decades (EES2) program to fund research into highly efficient AI processors leveraging energy-efficient architectures.
- Invest in R&D for software stacks that dynamically target inference workloads across CPUs, GPUs, and NPUs, ensuring that optimal resources are used, in the most efficient location, at the most efficient time. This work could be performed under the DoE's Advanced Simulation and Computing (ASC) program with expanded authorization and focus to include dynamic targeting of AI inference workloads.

Establish benchmarks and incentives for efficiency:

- Develop comprehensive, standardized energy-per-inference and performance-per-watt metrics for evaluating AI systems through National Institute of Standards and Technology (NIST) initiatives.
- Encourage the adoption of new industry metrics that go beyond Power Utilization Efficiency (PUE), addressing diminishing returns on data center cooling optimization and driving improvements in compute equipment efficiency.
- Incorporate these comprehensive efficiency benchmarks into federal AI procurements, incentivizing adoption of technologies that lower total cost of ownership and operational expenditure.

4. Accelerate Edge AI Deployment

Edge AI brings advanced computational capabilities directly to the point of data generation, enabling rapid analysis and decision-making in real-time without reliance on centralized data centers. This approach significantly enhances responsiveness, security, and reliability, particularly in scenarios where network connectivity is limited or latency-sensitive decisions are critical. Key use cases of edge AI include autonomous vehicles requiring instant processing for safe operation, medical devices offering immediate diagnostics and patient monitoring, and industrial automation systems optimizing manufacturing processes and predictive maintenance. Additionally, edge AI is vital for telecommunications infrastructure, particularly in deploying secure, intelligent 5G and beyond networks. By delivering secure, real-time insights at the edge, these technologies greatly enhance mission capabilities in defense, healthcare, industrial operations, and communication networks. The federal government should enhance funding for R&D in edge AI deployment to capitalize fully on these benefits:

Expand AI-RAN and Open RAN initiatives:

- Fund targeted programs like DARPA's Open, Programmable, Secure 5G (OPS-5G) and NSF's Platforms for Advanced Wireless Research (PAWR) to accelerate integration of AI and Open RAN technologies.
- Promote collaboration between industry, academia, and federal labs to ensure secure, interoperable telecommunications infrastructure.

Support federal testbeds for real-time edge applications:

- Enhance funding for federally sponsored testbeds such as DOE's National Renewable Energy Laboratory (NREL) and DoD's Joint Artificial Intelligence Center (JAIC) for edge AI experimentation and deployment.
- Develop a nationwide network of AI-edge labs to facilitate public-private partnerships, rapid prototyping, and secure deployment of real-time applications in mission-critical contexts.

5. Ensure Robust Hardware-Level AI Security

Hardware-level security is essential for protecting sensitive AI applications due to the increasing sophistication of cyber threats and vulnerabilities inherent in complex digital systems. With AI systems deployed in critical sectors such as defense, healthcare, finance, and infrastructure, ensuring the integrity and security of these systems at the foundational semiconductor level is paramount. The semiconductor layer is the root of trust for the entire computational stack, making it a primary target for potential breaches and cyber-attacks. The federal government possesses unique experience and visibility into such security challenges, gained through managing sensitive and classified data and infrastructure. By leading efforts in hardware-level security, the government can develop standards and practices that significantly enhance not just governmental technology, but also critical infrastructure and commercial technology, fostering broader resilience across all AI systems.

- Expand programs such as DARPA's Automatic Implementation of Secure Silicon (AISS) and NIST's Cybersecurity Framework to develop and deploy secure-by-design features, including secure enclaves and trusted execution environments.
- Support interagency initiatives such as the DoD's Trusted and Assured Microelectronics (T&AM) program to secure and validate semiconductor supply chains.

6. Initiatives Expanding AI Access and Research

NSF-supported initiatives like the "Scalable AI Gateways for Education (SAGE)" project (Award ID 2436842) significantly democratize access to AI, facilitating broader participation in advanced research and innovation. The SAGE project provides scalable, user-friendly gateways allowing nationwide access to cutting-edge computational resources, enabling institutions and researchers without extensive local resources to engage fully in AI research. Expanding AI scalability and access through initiatives such as SAGE can drive substantial economic benefits by fostering innovation and entrepreneurship, particularly in underserved regions. From a national security perspective, broader AI access strengthens resilience by diversifying sources of technological advancement and accelerating breakthroughs essential for defense and intelligence

applications. Additionally, the research community benefits immensely as wider access supports interdisciplinary collaboration, accelerates scientific discovery, and leverages diverse expertise to solve complex societal challenges. Expanding NAIRR could further complement initiatives like SAGE, offering integrated federal resources to researchers nationwide and amplifying these substantial economic, security, and research benefits.

Specific recommendations include:

- Increase funding and resources for the NSF's SAGE initiative to expand the number of institutions and researchers able to leverage these gateways.
- Expand NAIRR's computing infrastructure to further reduce barriers to entry for smaller institutions and enhance scalability of AI research nationwide.
- Establish targeted federal grants, managed by NSF and DOE, to specifically incentivize cross-disciplinary and multi-institutional AI research collaborations.
- Support workforce development programs through collaborations with agencies like NSF and the Department of Education to ensure a broader distribution of AI expertise.
- Foster regional innovation hubs in partnership with NSF's Regional Innovation Engines to accelerate AI-driven economic development in underserved regions.

7. Advancements in Physical AI and Robotics

Physical AI, including humanoid and industrial robotics, can significantly enhance efficiency, productivity, and mission capabilities across government and industry. Specific areas where federal investment would be particularly impactful include autonomous robotic systems for disaster response and hazardous environments, enhancing the safety and effectiveness of emergency and defense operations. Additionally, investment in advanced robotic manipulation and precision control technologies can substantially benefit manufacturing, logistics, and healthcare sectors. The federal government should also focus on developing interoperable robotic platforms that integrate seamlessly with human teams, improving collaboration between robots and personnel in complex, real-time scenarios. Arm recommends federal investment through existing labs like DARPA, NSF, and the Department of Energy's national laboratories, enabling rapid prototyping, rigorous testing, and accelerated deployment of physical AI solutions tailored for critical government and industry applications.

Specific recommendations include:

- Increase investment in DARPA's Robotics Challenge and DARPA's Robotic Servicing of Geosynchronous Satellites (RSGS) program to develop autonomous robotic systems for disaster response, space exploration, and operations in hazardous environments.
- Enhance support for the NSF's National Robotics Initiative (NRI) to accelerate development of collaborative robotic technologies that safely interact with humans in manufacturing, healthcare, and logistical applications.
- Expand funding for robotic R&D within the Department of Energy's National Laboratories, particularly focusing on applications for remote handling and automation in nuclear and hazardous materials management.

- Strengthen Department of Defense (DoD) robotics capabilities through increased investment in the Army Research Laboratory (ARL) and the Naval Research Laboratory (NRL), focusing on robotic platforms for battlefield support, reconnaissance, and logistics.
- Establish interagency initiatives between NSF, DARPA, DOE, and NIST to create standardized frameworks and interoperability protocols for robotic systems, facilitating seamless integration and deployment in critical sectors.

Conclusion

Arm supports a comprehensive approach emphasizing robust semiconductor R&D, secure federal infrastructure, compute efficiency, accelerated edge deployment, hardware security, AI access, and physical AI. Further, despite the wide proliferation of AI, this is still a nascent area with significant advancements and innovation yet to come. Arm would encourage the federal government to continually assess policies and priorities as the technology evolves to ensure policies, funding, incentives, and regulations evolve and adapt to keep pace with AI developments. Arm stands ready to collaborate with stakeholders to ensure continued U.S. prominence in AI.