PUBLIC SUBMISSION

Received: May 29, 2025
Tracking No. mb9-mb28-ouoz **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-0204
Comment on FR Doc # 2025-07332

Submitter Information

Organization: Washington State University

General Comment

See attached file(s)

Attachments

WSU Response AI Research RFI

AI to Advance U.S. Competence in Foundational and Use-inspired AI with Applications in Agriculture, Sciences, and Engineering

Response to Notice of Request for Information (RFI) on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan

Bethany Johns
Assistant Director, Federal Relations, Washington State University

Jacob Dowd
Director, Federal Relations, Washington State University

Ananth Kalyanaraman
Professor, School of EECS, AgAID Institute Director, Washington State University

Jana Doppa
Associate Professor, School of EECS, Washington State University

Kim Christen
Professor, Vice President for Research, Washington State University

Washington State University (WSU) is the land-grant university for the State of Washington and an R1 research institution under the Carnegie classification system with world-renowned programs in the computing areas of artificial intelligence and machine learning, and data sciences and scalable computing. The University also hosts the WA state's flagship research programs in agriculture, animal sciences and veterinary medicine, as well as several areas within engineering such as chemical catalysis, hydrogen and bioenergy, and power grid systems. WSU appreciates the opportunity to provide a response with actionable recommendations to the Office of Science and Technology Policy (OSTP)'s 2025 Request for Information (RFI) on the development of Artificial Intelligence (AI) R&D strategic plan.

I. Opportunity

AI is at the cusp of revolutionizing numerous application domains. Selected examples of key opportunities where AI plays a major role include:

- AI for accelerated scientific discovery at all scales, from molecular to individual to population levels;
- AI for sustained intensification of agricultural production, breeding resilient crops and reducing crop loss, reducing labor- and economic- uncertainties, and increasing farm productivity and farm worker productivity;
- AI for optimized use and smart and efficient stewardship of environmental resources including water, energy, and soil, and for improved air quality;
- AI for securing our power grid, renewable sources, aviation fuels, and bioenergy alternatives;
- AI for accelerated drug discovery, early detection of diseases, preventing and mitigating the threats of infectious diseases, and improving human health and well-being;
- AI for advanced manufacturing, advanced semiconductor design, and advanced materials design; and
- AI for transportation, construction, and advanced structural design.

To fully harness the power of AI, foundational research in AI and machine learning as well as their applications are needed. This two-pronged approach of investing in both foundational AI and use-inspired AI is an essential strategy to ensure that the innovation ecosystem in AI technology is inherently equipped to engage both AI researchers and the users of AI technology as part of AI design. This approach leads to identification of theoretically-sound and practically-grounded AI technology, that is ready for adoption in the real-world as it is developed – closing the gap in the time needed from development to practice and accelerating AI-fueled innovation across science and engineering disciplines.

Public, land-grant universities like WSU are best positioned to both conduct research in the above-mentioned areas as well as to train the next-generation of AI-ready workforce. In fact, the tech industry is and will remain more focused on the use of AI in internet-scale applications (e.g., retail, services). However, universities remain the only place to pursue basic science and engineering research (including those that are high-risk and high-reward) as well as train students who contribute to our next-generation workforce for the industry. Public land-grant universities also have extensions which further serve as the interface between academic research and industrial practice. Therefore, it is critical that this R&D investment prioritizes academic institutions as the home of foundational- and use-inspired AI tool research, development, and deployment.

Academic institutions such as WSU also reach out to rural American communities and provide curriculum and AI technology access and training to the K-12 American youth, including programs such as the Future Farmers of America (FFA), 4-H, and the Mathematics, Engineering, Science Achievement (MESA) programs. This unique combination of basic research, connection to industry, and strongly integrated work-force development programs give American universities (in particular, the public land-grant institutions such as WSU) the advantage of leading the implementation of this AI R&D strategic plan.

II. Recommendations

1) Investment in Foundational AI Research and Development

Foundational AI provides the building blocks of AI that can then be adapted and integrated to create use-inspired solutions. Future research in foundational AI is needed in the following areas over the next 5 years:

- **Foundation models for scientific data** have many challenges including: integrating and reasoning over heterogeneous data; dealing with relatively smaller size datasets; and the ability to characterize their uncertainty and to gather new data through experiments to expand/refine their scientific knowledge bases. There is a great need for such models to create human-AI collaborative systems to generate hypotheses, perform lab experiments, and refine those hypotheses to enable scientific discovery at speed and scale in many domains including drug/vaccine design and materials/catalyst discovery among others. These foundation models can provide the substrate over which to design and implement digital twins and synthetic data frameworks for real-world applications.

- **Machine learning algorithms for adaptive experimental design**: One of the serious impediments to accelerating scientific discovery is the cost of experimental design and evaluation. For example, the number of candidates that need to be screened in drug design, materials design, or in crop breeding experiments (to identify varieties with desirable traits) is often a significant limiting factor that prolongs the discovery timeline to years. Machine learning methods have the potential to reduce this search space of potential candidates for screening and experimentation. It also has the potential to reduce the number of parametric configurations significantly. However, new research is needed to integrate machine learning methods for experimental design, through incorporation of domain knowledge to improve the explainability and reasoning capabilities of those methods. New research is also needed to help make experimental design methods more sample-efficient while maximizing information gain.
- **Theoretically sound uncertainty quantification for structured domains**: To create trustworthy AI systems, we need formal guarantees about their uncertainty. Unlike traditional classification and regression tasks where conformal prediction framework has shown good success, structured output spaces present unique challenges. Some examples include (a) generating a small prediction set from a deep generative model that contains at least one admissible solution (e.g., code and natural language text) where the output space is combinatorial; (b) to mitigate hallucinations from large language models, how can we design an abstention policy to achieve target trade-offs in terms of utility and abstention rate; (c) how to generate localized uncertainty for outputs of large language models such as text and code; and (d) generating prediction intervals for spatio-temporal prediction problems such as predicting streamflow, weather, faults in additive manufacturing process etc.
- **Decision-making from offline/logged datasets**: Reinforcement learning (or RL) has shown great success in many real-world applications to learn decision policies via trial-and-error based on feedback from the environment. However, this approach is not practical in many decision-making scenarios where random exploration is dangerous (e.g., applications such as health, education, managing agricultural farms, and autonomous driving). There are many open challenges in learning decision policies from offline datasets. Selected examples include: (a) small datasets from many related tasks with slightly different dynamics such as logged irrigation decisions from different agricultural fields; or (b) how to compare and select

appropriate strategy from a give set for a specific context such as a given farm and time of the year.

- **AI in open world domains:** Much of today's AI systems are narrow and based on the closed-world assumptions. The key challenge to deal with open world domains is how to deal with novelty (e.g., new objects, novel hazards, and change in system dynamics). We need a fundamentally new set of machine learning and uncertainty quantification methods to build AI systems capable of dealing with open world challenges.

2) Investment in use-inspired AI research, development, and deployment

An increased investment is needed in use-inspired AI, where the research, development, and deployment of AI tools are all inspired by concrete uses. AI has numerous potential applications across various domains. Yet, the type and class of AI needed across these domains varies widely to suit their respective uses:

1) **AI for science** involves accelerating discovery through smarter experimental design, reduction in candidates for testing through virtual screening, active and adaptive learning, automated hypothesis extraction toward causal inference, knowledge and rule extraction from high dimensional and relational data, reasoning frameworks for process modeling, foundation models for scientific exploration and to create simulation abstractions.
2) **AI for engineering** involves optimization, design space exploration and design automation, stress and extreme-condition testing, robotics and automation, transferability across platforms and systems and transferability from simulation to real-world, reproducibility and simulation capabilities through digital twins.
3) **AI for agriculture** involves precision agriculture, trait optimization, robotics, automation and human-AI workflows to reduce labor costs and improve farm productivity, site-specific models for decision support, spatio-temporal modeling, robust training under uncertainty and using limited and noisy data, predictive and forecasting tools, crop response modeling, soil health, water security and water allocation, digital twins for plants, farms, and regions, earth-systems and biophysical modeling of complex systems, accelerated breeding of resilient and resistant varieties, community and economic modeling for adoption, interactive dashboards for decision making, and responsible and safe AI.

4) **AI for health** involves drug design, early disease detection and prevention science through high-throughput screening and imaging, precision medicine, variant detection at molecular, cellular, tissue-specific, and organismal scales, understanding the roles of genetic and environmental variability on diseases and disease evolution, cancer heterogeneity and cancer genomics, antimicrobial resistance and hospital acquired infections, pandemic prevention and epidemic control.
5) **AI for energy and environment** involves smart power grid design and grid resilience, fault tolerance, distribution networks and load distribution, contingency planning and optimization, renewables, grid security, environmental stewardship, air quality, water resource and critical infrastructure management, emergency planning – e.g., wildfires, flood, human influence and behavior modeling.