PUBLIC SUBMISSION

Received: May 29, 2025
Tracking No. mb9-qmkw-c3cl **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-0233
Comment on FR Doc # 2025-07332

Submitter Information

Organization: California Institute of Technology

General Comment

See attached file(s)

Attachments

ostp_input

Response to RFI Docket ID No. NSF-2025-OGC-0001

2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan

Kaushik Bhattacharya on behalf of the California Institute of Technology

I provide this response in my role as the Vice Provost for Research at the California Institute of Technology (Caltech), with input from a number of faculty members at Caltech. Caltech is a not-for-profit university with a century-old tradition of performing research to expand human knowledge and for the benefit of society. Our students and faculty have won 47 Nobel Prizes, and 80 have been honored with the National Medals of Science and Technology. Technologies developed at Caltech in recent decades are at the heart of the development and design of computing platforms and algorithms, digital medicine, and wireless communication.

We are in the midst of a revolutionary change in science and engineering, driven by remarkable advances in artificial intelligence (AI) and data science. The United States of America pioneered AI development, driven by foundational technical breakthroughs originating in US universities, and followed by the creation of influential AI products in industry. However, for all its successes, we are at the beginning of this profound change, and continued leadership in AI requires the development of new algorithms, new computing platforms, and new applications in scientific and technological domains. This requires the leadership of the US government, and a strong alliance between research universities and industry. Universities serve as the ultimate incubator for AI innovation, ideas and talent, and industry excels at translating and scaling up the innovations.

There are a number of areas in which we urgently require US government leadership and investment.

Fundamental AI research: The AI revolution was rooted in US university research, where innovative ideas and cutting-edge findings were pioneered[1]. Academia is also the premier incubator for talent; today's leading AI researchers hold graduate degrees from US institutions. Although current advances in AI are remarkably impactful, we have yet to fully realize AI's potential – or truly understand its inherent constraints. Fundamental research can illuminate these limits, and continue to generate concepts that can overcome them, thereby maintaining US leadership. Achieving this requires support, financial and intellectual, for innovative research at the core of AI as well as cross disciplines.

Infrastructure for AI research: For universities to shape and guide the scientific AI revolution, they need to be equipped with sufficient computing infrastructure. This includes GPU clusters, managed data storage and software engineering support. Immediate action to bolster the computing, software engineering, and automation engineering resources is crucial to remain at the cutting edge of research. There is a significant, and exponentially growing gap, between the

[1] For example, deep neural nets were conceived and developed by Hopfield at Caltech and LeCun at NYU.

computing resources at US universities and industry, and a smaller, but also rapidly growing gap, between the computing resources at US universities and non-US universities. This puts US universities, and ultimately the US AI effort, at risk.

Further, each US university is building its own computing and data infrastructure, and this is wasteful for multiple reasons. First, computing and data infrastructure have significant economies of scale, and consolidating resources reduces unit cost. Second, computing and data infrastructure is energy hungry, and consolidating resources near sources of inexpensive energy also promotes efficiency. The US government can provide leadership in this regard through the development of a network of regional data centers – each on the scale of facilities that the industry is developing – that provide the US research community with adequate computing and data infrastructure to catalyze AI research in the USA. Importantly, such centers have to be built at scale for this effort to be effective.

New computing platforms: Current approaches in AI are loosely inspired by brain processing but have significantly diverged from the implementations in biological systems. For example, current AI systems require massive amounts of energy, making scaling and implementation in autonomous agents impractical, and placing an increasing burden on energy consumption in society. Furthermore, they fail in unpredictable ways and do not generalize gracefully in diverse contexts. In contrast, brains operate with extraordinary energy efficiency, enable continuous lifelong learning in diverse contexts, and exhibit unparalleled resilience and adaptability. Humans can operate with just 2000 calories per day, while compact nervous systems such as those of flies exhibit exquisite sensorimotor processing in complex environments with ultra-low power consumption. Brains achieve this level of performance by employing distinct computational principles that current neuroscience research has begun to reveal. There are significant opportunities to build synergies between neuroscience, computation, and mathematical sciences to lead the development of novel paradigms in computation to achieve unprecedented levels of efficiency and performance.

Another promising emerging technology is quantum computing. Quantum hardware holds promise for solving computationally intractable problems in fields ranging from cryptography to material science, as well as offering the potential for exponential speedups in optimization, machine learning, and simulation tasks, revolutionizing the landscape of computational science and engineering.

Optical approaches present another compelling avenue for developing new computational frameworks. Optical computing capitalizes on the properties of light to perform computations, offering advantages such as high-speed data transmission, low energy consumption, and parallel processing capabilities. Optical hardware, including photonic integrated circuits and optical neural networks, can potentially accelerate tasks such as image recognition and data encryption. Moreover, the inherent parallelism of optical systems enables efficient processing of large-scale data sets and complex algorithms, paving the way for breakthroughs in AI and data science.

Unlocking the potential of AI in multi-disciplinary research: The rapid development of AI, coupled with the exponential growth of computing power and data analytics capabilities, is leading us into an era where the tools and methodologies of scientific inquiry are being redefined. Despite these green shoots, the revolution in science driven by massive-scale data analysis and computation is still in its infancy. Computing and data sciences are still considered specialized activities, used for specific tasks, and have not yet automated the scientific process to the extent they can. However, as computational power and algorithmic advances continue to develop, the impact on science and engineering will continue to grow. AI and data science algorithms can be used directly as tools for scientific discovery. As new algorithms and computational frameworks emerge, the pace of change and the impact on science and engineering will only accelerate.

Unlocking the potential of AI for scientific discovery is a task that is only possible at universities. Universities catalyze multi-disciplinary research by allowing connections to be made between people with domain expertise *both* in core AI and in application areas. This in turn opens the door to fundamental breakthroughs in science that ultimately lead to impactful services, products, and sometimes entire industries.

Towards this goal, new programs should target and foster explicitly inter-disciplinary research agendas. Single- and multi-investigator projects addressing fundamental scientific questions with AI tools should be supported and amplified. Furthermore, a small group of carefully chosen moonshots with measurable scientific payoffs ought to be funded and backed with compute and infrastructure resources at a scale that only government can provide – as was done for e.g., the human genome project.

Beyond funding, infrastructure support must be provided for collecting and maintaining *open* data repositories at scale. Indeed, without data there is no AI. Researchers working in data-rich fields of research should be given incentives and resources to federate, annotate and curate the data they collect making it available to every other researcher in their field on platforms where it is easy to access the data and train models with it. Funding for research should tie funding to making data available in machine readable, standard format.

Training and empowering researchers to use AI: Beyond their role as the driving force behind the development and deployment of AI tools, universities educate the next generation of engineers and scientists. The US role as the world's pre-eminent academic R&D ecosystem not only helps drive innovation, but also creates a virtuous cycle that reinforces the dominance of the US-based research. To strengthen and expand this position – particularly in AI – resources should be provided to rapidly train skilled engineers and scientists who are able to develop both core AI technology and apply these technologies to broader fields in science, as well as to the design of products and services. This includes new courses and curricula aimed at the training of AI scientists, who will form the new generation of AI researchers, the training of AI engineers, who will apply AI to products and services, and the training of scientists in broader application domains to empower them to develop new methods and advance their field. Financial support should be provided for universities to develop and operate world-class courses on cutting edge topics, as should platforms for sharing training materials (slides, homework problem, software, lecture videos) to ensure the widespread dissemination of educational resources.