Society for Industrial and Applied Mathematics

Society for Industrial and Applied Mathematics Response to American Science Acceleration Project Request for Information

The Society for Industrial and Applied Mathematics (SIAM) is an international community of over 13,000 individual members. Almost 500 academic, manufacturing, research and development, service and consulting organizations, government, and military organizations worldwide are institutional members. Our members come from many different STEM disciplines but have a common interest in developing and using mathematical and computational sciences approaches and tools to advance science and engineering. Through publications, research, and community, the mission of SIAM is to build cooperation between mathematics and the worlds of science and technology. Thank you for the opportunity to respond to the American Science Acceleration Project Request for Information.

Federal research investments in AI and related STEM fields, such as applied mathematics and computational science, are critical to advance innovation, discovery, and ensure continued American leadership. AI has the potential to reshape many scientific disciplines as well as other sectors such as health, energy, defense, economy, manufacturing, and education. Fundamentally, the existence and success of AI technology is grounded in many areas of mathematics and computational science. As applied mathematicians, the SIAM community appreciates efforts both to advance foundational artificial intelligence, as well as to build partnerships that enable specific application areas. In this response, we will focus on several key areas for research and development, infrastructure needs, as well as education and workforce development recommendations.

Cross-Cutting

1. How should the United States achieve the goal of accelerating the pace of scientific innovation? What roles should be played by Congress, the administration, industry, civil society, and academia?

Background: The SIAM community has a long-standing record of advancing science and engineering by developing foundational mathematical and computational theories, tools, algorithms, and systems. Advances in AI that are built using rigorous mathematics and computational ideas provide an unprecedented opportunity for disrupting how science and engineering is done.

While recent AI advances have been driven by progress in image recognition and language processing, scientific and engineering domains pose a fundamentally different set of challenges that current AI systems are not well-equipped to address. Scientific data is often sparse, heterogeneous in type and scale, and reflects complex physical processes. Objectives such as interpretability, physical consistency, and causal understanding go far beyond conventional accuracy metrics. *These new challenges require a new class of mathematically sound and scalable AI methods that are capable of learning from limited data, reasoning under uncertainty, providing reliable solutions, and incorporating scientific knowledge.* Meeting these challenges offers a unique opportunity to push AI in transformative new directions—making science not only a key beneficiary of AI but also a driver of its next frontier.

Strategic Opportunity for U.S. Leadership: There is a significant strategic opportunity for the United States to lead in building a next-generation AI ecosystem for science that integrates advanced AI methods with scientific theory, domain expertise, rigorous mathematical foundations, and high-performance computing. This initiative would simultaneously accelerate discovery across critical fields such as materials, health, space, and advanced manufacturing, while also advancing the frontiers of AI. Investing in research in AI methods and tools is essential not only for sustaining U.S. scientific and technological leadership but also for strengthening national security, revitalizing domestic manufacturing, and ensuring long-term economic competitiveness. It will support evidence-based policymaking in areas ranging from manufacturing to public health and will help grow a future-ready workforce by training the next generation of interdisciplinary researchers. These efforts will stimulate high-quality job creation, drive innovation-led economic growth, and enhance the nation's capacity to respond to emerging challenges and opportunities.

Grand challenges to accelerate scientific discovery: One way to accelerate scientific innovation, is to adopt a structured, challenge-driven approach modeled on the strategies that catalyzed rapid advances in fields like natural language processing (NLP) and computer vision (CV), which led us to the AI boom we are experiencing today. We propose a coordinated national initiative to define and make systematic progress toward a set of **Grand Scientific Challenges**—such as dramatically accelerating computational fluid dynamics (CFD) for aerospace, enabling inverse design of novel materials, or constructing digital twins for biological systems. These high-impact goals should be **decomposed into modular, tractable subtasks**, such as physics-constrained approximation techniques, structure-preserving surrogate modeling, solver acceleration, uncertainty quantification, model calibration, or simulation-experiment data fusion. Each subtask should be accompanied by **standardized datasets, open benchmarks, and public leaderboards**, enabling transparent comparison of methods, driving healthy collaboration and competition, and pulling the entire field forward.

Recommendations: To fully realize the transformative potential of AI for science, we offer the following recommendations for federal investment:

1. Invest in the **development of rigorous AI methods and systems that integrate scientific knowledge.** Federal investment should drive, support, and encourage research in knowledge-guided AI, physics-informed learning, and hybrid modeling

frameworks that blend data-driven approaches with domain knowledge, theoretical constraints, symbolic computation, and physical laws. Foundational and applied mathematics have been central to the success of AI systems and will continue to be a key component of the new AI methods for them to be successful. These efforts will not only enhance the reliability of AI systems but also accelerate discovery in data-scarce scientific domains. Targeted federal investment is needed to ensure that the same cycle of innovation that has manifested in massive gains in typical AI domains such as NLP and Computer Vision can also result in similar breakthrough advances in science and engineering.

2. Invest in **advancing foundational mathematical capabilities** that underpin the development, theory, and practice of AI, e.g., optimization, uncertainty quantification, probabilistic reasoning, linear and non-linear algebra, dynamical systems, network science, and neuro-symbolic computing and mathematical programming. Advances in these mathematical areas are critical to ensuring that science and engineering simulation architectures are AI-ready and that AI systems are scalable, explainable, and produce solutions that can be trusted.
3. Invest in a robust ecosystem of **collaboration infrastructure for science**: shared platforms for hosting datasets, managing evaluations, and sharing models. Complementing this, we must expand **scientific compute infrastructure**, including access to advanced simulation platforms, cloud-based model training environments, and low-latency inference tools so that the entire scientific community can participate. Their role is important for the development and deployment of such tools. Equally important is the creation and support of **open development frameworks** for simulation and modeling. This should involve moving away from fragmented, proprietary software environments that currently permeate the field in favor of interoperable, community-driven toolkits, which typically have a faster research and development cycle.
4. Invest in a **diversified portfolio of research efforts across scale and discipline.** Federal investment should include: (a) Small-team, high-risk/high-reward projects to explore novel approaches; (b) Mid-scale centers focused on specific domains or methodological themes; and (c) Large-scale, cross-institutional efforts, including new or expanded AI Institutes, designed to tackle grand challenges and foster deep collaborations across AI, science, and engineering communities. The agencies can also consider grand challenges that can spur a nationwide collaborative effort to drive the scientific and AI fields concurrently.
5. Build **shared infrastructure to enable AI-driven science.** Invest in national-scale data infrastructure, benchmark datasets, simulation libraries, and interoperable tools that allow scientists to train, test, and deploy AI models in ways that are reproducible and collaborative.
6. Invest in **interdisciplinary education and workforce development.** Prepare a new generation of researchers who are fluent in both AI methods and scientific

disciplines. This includes Congressional funding for creating interdisciplinary curricula, research fellowships, and training programs that bring together AI experts, domain scientists, applied mathematicians and engineers.

Summary: AI has already begun to transform the global scientific enterprise. But the true potential lies in developing a new generation of AI systems that are deeply informed by scientific knowledge accumulated over decades and centuries in various scientific disciplines. Strategic federal investment at the intersection of AI and science will ensure U.S. leadership in both domains, drive technological innovation, and provide critical capabilities to address national and global challenges.

The recommendations provided above reframe AI not as an end in itself, but as a powerful accelerator for rapid scientific discovery—one that advances when research is **structured, benchmarked, and collaborative**. With proper alignment of incentives and infrastructure, we can build the scientific analog of the AI research flywheel, using data, computation, and open collaboration to drive sustained progress on the most pressing scientific questions of our time.

Roles that various sector can play:

Federal Sector: (i) Congress should invest in sustained and significant research funding (both applied and foundational) (ii) the executive branch and Congress should implement policies to attract and retain exceptional international talent, (iii) the executive branch and Congress should create incentives and policies that lead to collaborations between academia, industry and the government, (iv) the executive branch should implement policies that will allow international collaborations with strategic partners – this will ensure that technologies are developed and deployed consistently as well as amplify the federal funding dollars.

Industry: Industry can not only play a vital role in embracing new technologies and creating innovative products but can also invest in sustained talent development. The top ten AI companies in the U.S. can make significant internal as well as external investments to support relevant academic programs that will prove beneficial in the long run. Industries and National Laboratories can also explore ways to make data sets they collect accessible to the broader scientific community.

Academia and National Laboratories can play a role by: (i) carrying out foundational mathematical and scientific computing research in topics relevant to AI-based advances in Science and Engineering, (ii) training next generation scientists in a multi-disciplinary setting, and (iii) working closely with the civil society, industry, and government on important meta-problems related to adversarial use, open access, AI ethics and AI acceptance. Academia can additionally develop innovative courses, degree programs, and training materials.

2. What infrastructure needs to be built to make scientists more productive, and for each type of infrastructure you recommend, what should the funding model be for the construction and operation of that infrastructure?

AI research requires large-scale infrastructure that keeps advancements out of reach for many beyond the most highly resourced companies. The federal government can help in the creation of a more robust ecosystem for AI innovation by enabling university, industry, and national lab research efforts through investments in federal AI infrastructure that are available to users through a competitive process to make sure the most promising research gains access. At the National Science Foundation (NSF), the *National AI Research Resource* (NAIRR) pilot is underway following a task force effort initiated in the first Trump Administration. The pilot will end in FY 2026 and NSF should look to build the full NAIRR that will revolutionize AI capabilities. At the Department of Energy (DOE), in addition to the *National Laboratory AI Testbeds*, a facility or facilities should be established to ensure that researchers have access to experimental non-traditional hardware that will inform a holistic codesign cycle between manufacturers and users of transformative computing technologies. In addition, the national laboratory workforce has strong expertise in high performance computing as well as AI methods, and national lab research programs should be supported to bring high performance resources to bear on AI. As the non-traditional computing hardware industry evolves, and the roles of different enterprises change, the government should remain open to establishing new types of collaborations and relationships with both existing and emerging industrial partners to create new devices. These new devices will likely be initially employed as accelerators, so their ability to integrate with existing high-performance computing (HPC) platforms must be part of the design process. Investments are also needed in research and development collaborations between computational scientists and computer vendors to ensure development of future energy efficient computing platforms that meet the needs of the computational science community.

Investments are needed in foundational mathematical and computational models as well as on the research, software tools, and system management tools needed to enable complex workflows that combine highly efficient simulation with machine learning so that AI can better work with existing HPC computing infrastructures. This is primarily a role for DOE along with some investments from NSF and the Department of Defense (DOD). The Aurora Effort at Argonne National Laboratory is an example of efforts underway, but these are still specialized and in the early stages of their development. In addition to large foundation models, there is also the need to produce smaller and more focused foundation models for specific classes of applications. An important topic beyond software and data systems is the development of tools for verification and validation. As the need for foundation models is becoming more prevalent, the SIAM community is poised to develop new scalable algorithms, theorem provers, reasoning and analysis subsystems. For example, SIAM members can improve upon interactive theorem provers, such as Lean, Coq, Isabelle, or equally Automatic Theorem Provers, such as KaymeraX.

Beyond specific AI-focused infrastructure, access to advanced computing resources is essential for many areas of computational science. NSF and DOE play critical roles in providing this infrastructure to the scientific community. NSF's Cloud Bank model, which brings together federated cloud-based and physical advanced computing facilities, is a good example of how federal investment can streamline access to tools that researchers need to innovate. Key components of these systems are management of the resources for optimized use, along with education and support resources for scientists. It is critical to provide national support, software, and data use tools so that not every user has to become an advanced computing expert to implement their science.

Funding Models: Both direct federal funding as well as combinations of public (federal), private, and philanthropic organizations funding is recommended. This has proved successful in recent application domains, such as evidenced by the NSF AI Institutes, as well as the well-established Engineering Research Centers and Science and Technology Centers. National laboratories and universities can also fund these efforts through, for example, their endowments and laboratory-directed funds. These funds are needed for: (i) building and maintaining the hardware, data and software infrastructure, (ii) advanced training of the workforce, and (iii) undertaking foundational research in mathematical and computational science. International partnerships can further leverage U.S. investment – such projects in physics have shown to be very successful. The leading AI companies and National laboratories have collected significant amounts of data that can be useful for advancing the field. Companies also have leadership class AI systems. Access to these data sets as well as state-of-the-art AI tools can be viewed as sources of indirect funding.

5. Grand challenge problems can help provide a concrete direction for how to implement new innovations. What core innovations does America need that can help guide ASAP? If possible, please provide an objective quantifiable metric, such as decreasing the time it takes to get a new drug to market from 10 years to 1 year.

A particular area of promise for AI innovation is at the intersections of AI and health. AI holds tremendous promise to revolutionize biomedical research in the understanding of chronic diseases, diagnostics, and development of treatments. Health agencies, such as the NIH and CDC, lag federal agencies supporting the mathematical and physical sciences in their AI investments and need a focused initiative to take advantage of this promising technology while addressing privacy and other issues. An AI initiative could invest in potential research moonshots to develop AI that will truly work and be trustworthy for healthcare applications. Further research is also needed to address AI use in healthcare to accelerate the development and deployment of reliable AI that provides accurate diagnosis and treatments for all patients. AI is accelerating at such a speed that NIH needs to keep up with the advancements of this emerging field and related technology like digital twins. In the absence of dedicated funding, NIH may fail to adapt to AI technologies that could shape the future of the U.S. healthcare system. OSTP should also encourage participation of relevant Health and Human Service agencies including NIH, Centers for Disease Control (CDC), and the Food and Drug Administration (FDA) in NSF's National Artificial Intelligence Research Institutes program. NSF AI Institutes can be expanded to address these challenges. Recent announcements by the NIH to develop holistic solutions and advance the use of AI in biomedical and healthcare applications are an encouraging first step.

Another promising area is the intersection of AI and the basic science, energy, and national security missions of DOE. Additional research funding is needed to fully realize the potential of AI for these applications. Current and past successes, such as the development of a digital twin for a nuclear power plant to aid in inspections and maintenance, and AI-based screening and design of functional materials for harsh environments, including fusion reactors, are just the tip of the iceberg of what the incredible developments AI can bring to these areas. Additionally, an area to focus on is the joint problem of energy and AI. In one respect, we need to create new AI systems that are more energy efficient (human brain by many accounts is extremely energy efficient), but at the same time prepare for a potential energy wall in terms of the ability to support the continued growth of AI systems. Development of nuclear energy (small modular reactors as well as potential research in fusion) would provide significant new sources of energy. AI methods can be used in the early stages of these technologies for modeling, simulation, and design.

An illustrative example of a grand challenge problem is *to develop an "LLM-like" copilot for scientific discovery in one or more scientific domains.* We have already discussed the broad vision in Question 1, here we tailor the discussion to the overarching task of scientific discovery. Key components in addressing this grand challenge would include:

- Mirroring the components of modern LLMs and intelligent agents;
- Target use cases: grounded in some specific scientific-discovery applications, e.g., materials discovery;
- Knowledge: pretraining on scientific corpora (cf. Galactica from Meta);
- Planning and reasoning: use supervised and reward-driven post-training to mirror the scientific method and provide scientists with a 'copilot' to help in particular with hypothesis generation, experimental design, and data analysis:
- Observation, Question, Hypothesis, Experimentation, Data collection, Analysis
- Tool use: Equip the co-pilot with tools that allow it to execute the scientific method Simulation, Physical experimentation request, Dataset analysis, Scientist query (ask an expert)
- Infrastructure: Discussed in Question 5, but with additional tasks and datasets related to the 'outer loop' of scientific discovery.

6. How can America build the world's most powerful scientific data ecosystem to accelerate American science?

Building a holistic scientific data ecosystem to accelerate American science needs to focus on multiple components concurrently: (i) collection, organization, and management of large data systems that can be accessed at varying levels of openness; (ii) creation of a computing and storage infrastructure for storing and making the data accessible pervasively; (iii) developing new AI assisted technologies for storing and organizing these data sets that provide high end services as described in the literature of digital libraries; (iv) developing a pipeline to train a cadre of scientists who have the expertise in all the above activities.

AI needs data, and data across many agencies requires investment before it can be efficiently used for AI. Investment is needed in research and development to create an integrated suite of data lifecycle methods and tools, informed by the specific needs of federal research communities at the DOE, NSF, National Institutes of Health (NIH), Department of Defense (DOD), and other agencies. Techniques that lead to data harmonization and normalization, such as data cleaning, validation, and error correction, must be developed to overcome the natural heterogeneity of data sources which is in inherent conflict with the need for data fusion. Open research questions remain around analyzing highly distributed data sources, uncertainty quantification, experimental design, enabling data discovery and integration, tracking data provenance, coping with sampling biases and heterogeneity, ensuring data integrity, privacy, security, and sharing, and visualizing massive datasets. Agencies should also consider the creation and use of digital twins, synthetic data and knowledge bases that might be produced by first principle models and enable better training of AI systems in areas where there is limited real data available. The National Institute of Standards and Technology (NIST) also has an important role to play in setting data and AI standards and convening academic, industry, and other stakeholders to share data and AI systems for measurements of reliability, safety, security, and vulnerabilities.

Data ecosystems benefit from stable funding environments that encourage long-term stewardship of resources and incentives, and resources that promote sharing and upkeep. NSF and NIH have both updated policies in recent years that provide for more robust data sharing expectations, but more can be done to scaffold interoperability, usability, and accessibility in a highly distributed system. Mechanisms for privacy and for data creators to have the ability to control the use of data are key issues to address. Data repositories language was included in *CHIPS and Science* to address these issues, but NSF has not fully implemented those provisions due to lack of funding. These provisions could also be updated given the rapid advancement of data and AI technologies since *CHIPS and Science* were passed into law.

9. How can we radically increase the scale, speed, and impact of scientific collaboration across disciplines, institutions, and sectors?

Partnerships

Computational science has always been an interdisciplinary endeavor, drawing on mathematics, computer science, and application domains. With the growth in the importance of data science and AI, the field is becoming even broader. National Laboratories, academia, and industry all have specific strengths to benefit AI research and development. Many of the most pressing scientific and societal challenges can only be solved through the efforts of multidisciplinary, multi-institutional teams. The federal government should create incentives for interdisciplinary education and research collaboration that engage National Laboratories, academia, and industry.

Establishment of additional coordination and engagement between the applied mathematics and computational science community and the wide-ranging federal healthcare and biomedical research will improve the nation's preparedness, health systems, and drug development and approval pathways. For example, NIH and FDA recently collaborated with NSF on a program to catalyze the development of digital twins for biomedical and healthcare technologies to improve clinical trials. The Administration should encourage these kinds of collaborations to bring together researchers from computer science, applied math, and engineering along with those in health operations, behavioral science, regulatory science, and biomedical research to foster new ways to evaluate medical devices, better modeling for disease preparedness, and development of new tools. This model can be adapted to other agencies and application areas to further additional applications such as advanced material design. The NSF *National AI Research Institutes* are another good model for interagency collaboration and involve the partnership of several agencies, including the United States Department of Agriculture (USDA), Department of Homeland Security (DHS), Department of Education (ED), and DOD. More agencies should be encouraged to join this program to harness breakthrough research for addressing their unique mission needs.

Another model that could be extended is the DOE Science Discovery Through Advanced Computing (SciDAC) program that co-funds research projects through mixed lab and university teams composed of scientists or engineers, applied mathematicians, and computer scientists to advance science understanding through the use of high performance computing systems. Projects funded through this program have executed cross-discipline research and have resulted in significant new simulation capabilities and scientific breakthroughs. Having already demonstrated success in combining much of the expertise in disciplines that are needed for successful AI in science, this program can serve as a model for building multi-disciplinary research teams that can advance AI systems for science and engineering innovation.

Education and Workforce

As an organization representing thousands of researchers and educators from hundreds of academic institutions, SIAM is well aware of the AI R&D workforce challenges facing the nation. The potential limitations posed by workforce shortages could significantly constrain our ability to maintain our leadership in AI amid an international environment that is growing increasingly competitive. Federal science agencies have a critical role to play in sustaining the vitality of AI R&D in the U.S. through their support of research and education programs. SIAM recommends that the Action Plan includes support for programs like NSF's *NSF Research Traineeships (NRT), Graduate Research Fellowships (GRF)*, and *CAREER* awards and DOE's *Computational Science Graduate Fellowship* program. These programs are crucial to the training and professional development of the next generation of the mathematical researchers and computational scientists who will underpin U.S. competitiveness in AI in future decades. Agencies should also look for opportunities to pursue partnerships and offer internships that give students exposure to National Laboratories, intramural research, or defense challenges. Exposure to federal laboratories has proven to be an effective way to inspire students at all levels to pursue careers, which is an important avenue alongside industry and academic workforce needs.

Undergraduate curricula are critically important to ensuring American preeminence in AI through education in the mathematical foundations of machine learning. SIAM recommends that the action plan commits to supporting the integration of data science and modeling into undergraduate STEM coursework in order to improve students' familiarity with these subjects as they seek to successfully enter the AI workforce, use AI in their scientific or engineering careers, and shape future innovations.

Compute

What benchmarking improvements do we need to understand the value provided by computing systems, and how should we best measure the strength of U.S. public sector compute against what is available in other nations?

In recent years there has been significant interest in creating new benchmark data sets to assess the performance of new machine learning algorithms, including speed and task capabilities. New journals are being created, new tracks at conferences are now focused on this as well. Much of the work so far, though, has been to create benchmarks for predictive and inference tasks. Not much effort has gone into creating benchmarks for reasoning tasks. Given that many of the reasoning tasks have well known complexity theoretic lower bounds, such reasoning benchmarks will allow the development of new reasoning methods and systems. Some progress has been made in this direction by companies like Epoch in the field of mathematics.

Data

What are the biggest blockers preventing researchers from sharing high-value scientific data today? What technological solutions could allow researchers to analyze sensitive data without compromising privacy?

Th biggest blockers preventing researchers from sharing high-value scientific data today include: (i) regulations, (ii) monetization, (iii) credit, and (iv) ability to control the use of data. We will discuss each of these points in detail. At this point, regulations and policies on how scientific data created at universities, research institutes, and national laboratories can be shared with the broader community are often non-existent. NIH and NSF have begun enforcing data sharing as part of the grants that they give out. The same cannot be said about data produced at the national laboratories. There are often good reasons for it, including national security, continued maintenance in terms of storage, and personnel. For instance, journals and conferences in computer science and other disciplines are increasing requirements to share the data associated with publications. This is an important trend.

The second impediment concerns credit and monetization. They are related issues. Data is the new currency especially in today's world of deep learning and large language models. Good data continues to be hard, and finding good data related to science and engineering applications is even more sparse. It takes increasingly significant amounts of time and money to produce some of these data sets. Machine learning models can pick this data up and provide services that are often quite useful. But the producers of the data neither get credit for their effort nor get any financial benefit, especially given the effort they spend on producing these data sets. This naturally creates a situation wherein individuals and organizations are often uncomfortable sharing their data. Solving this problem is not easy for two reasons. First, how do you attribute how much one data set is used in training and responding to a query and second, how do we assess the quality of such data sets?

The final impediment concerns control, although solutions are being developed as we write this response. In a nutshell, individuals and organizations should have the ability to pull the data back at any point of time under reasonable situations. ML models currently do not allow for this although work is ongoing wherein this will eventually be possible to some degree.

Other related issues have to do with privacy, security, and costs. Some of these points are discussed as a part of the next question as well as in recent related reports

How can we balance data privacy and security with open access to scientific data?

The National Academy reports[1] lay out excellent recommendations for this question. A few additional points worth noting include:

- As current generation models reach a limit in terms of using available data to train models, one needs to consider the use of synthetic data where data generated by AI models can then be used to train AI systems.
- One needs to ensure that blended data sets made by integrating individual data sets that have the required privacy guarantees, continue to have the needed privacy guarantees.
- Open access is an important driving goal but also poses challenges for national security and competitiveness in the era of AI-enabled science and engineering. The US has invested heavily for the last 70 years to create large and important datasets resulting from large equipment (e.g., accelerators, telescopes) as well as surveys for social, behavioral and health data. It is important to consider the use of these data sets to provide a competitive advantage to US companies and academic institutions.

Sincerely,

Carol d. Woodward

Carol Woodward

President

Suzanne L. Weekes

Chief Executive Officer

Alejandro Aceves

Vice President for Science Policy

[1] National Academies of Sciences, Engineering, and Medicine. 2024. *Toward a 21st Century National Data Infrastructure: Managing Privacy and Confidentiality Risks with Blended Data*. Washington, DC: The National Academies Press. https://doi.org/10.17226/27335