RECOMMENDATIONS: Harnessing AI for Scientific Progress

National Artificial Intelligence Advisory Committee (NAIAC)

May 2024

NAIAC MEMBERS

Miriam Vogel, *Chair*
President and CEO of EqualAI

James Manyika, *Vice Chair*
Senior Vice President, Google, President for Research, Technology & Society

Amanda Ballantyne
Director of the AFL-CIO Technology Institute

Jack Clark
Co-founder of Anthropic

David Danks
Professor of Data Science and Philosophy at the University of California, San Diego

Victoria A. Espinel
President and CEO of BSA | The Software Alliance

Paula Goldman
Chief Ethical and Humane Use Officer at Salesforce

Susan Gonzales
Founder and CEO of AIandYou

Janet Haven
Executive Director of Data & Society

Daniel E. Ho
William Benjamin Scott and Luna M. Scott Professor of Law, Professor of Political Science, and Professor of Computer Science (by courtesy), Senior Fellow, Stanford Institute for Human-Centered AI, Stanford University

Ayanna Howard
Dean of Engineering at The Ohio State University

Jon Kleinberg
Professor in the Departments of Computer Science and Information Science at Cornell University

Ramayya Krishnan
W. W. Cooper and Ruth F. Cooper Professor of Management Science and Information Systems, Carnegie Mellon University

Ashley Llorens
Vice President, Distinguished Scientist, and Managing Director at Microsoft Research

Haniyeh Mahmoudian
Global AI Ethicist at DataRobot, Inc.

Christina Montgomery
Chief Privacy & Trust Officer and Vice President at IBM

Liz O'Sullivan
CEO of Vera

Fred Oswald
Professor and Herbert S. Autrey Chair in Social Sciences, Rice University

Trooper Sanders
CEO of Benefits Data Trust

Navrina Singh
Founder and CEO of Credo AI

Swami Sivasubramanian
Vice President for Data and Machine Learning Services at Amazon Web Services

Keith Strier
Vice President for Worldwide AI Initiatives at NVIDIA

Reggie Townsend
Vice President of Data Ethics at SAS Institute

CONTEXT

Rapid advances in artificial intelligence (AI) technology represent real scientific progress — and these developments can go on to accelerate scientific progress in other domains, as well. AI is a powerful tool for discovery and learning, and it has already proven its potential in particle physics, climate science, neuroscience, drug discovery, and elsewhere.[1] AI can also be used to improve educational opportunities, better equipping the next generation of researchers.[2]

While each domain leverages AI in distinct ways, there are broad similarities. AI uncovers insights from vast amounts of data, predicts outcomes, and generates hypotheses, all of which can break bottlenecks and allow researchers to more easily test and iterate in real time.[3]

One of the most remarkable examples of AI's impact is AlphaFold, an AI system developed by DeepMind that achieved unprecedented accuracy in predicting protein folding structures. This breakthrough has significant implications for understanding biological processes and developing new therapies, addressing a challenge that has puzzled scientists for over five decades.[4]

AI's potential extends beyond biology into chemistry through a variety of efforts. For example, IBM's RXN for Chemistry uses transformer models to predict chemical reactions and synthesis pathways, significantly accelerating the process of discovery in synthetic organic chemistry. This tool, alongside IBM's Generative Toolkit for Scientific Discovery (GT4SD), underscores AI's ability to generate novel hypotheses and facilitate rapid experimentation and validation.[5]

In materials science, AI techniques are revolutionizing the way researchers explore new materials. AI-assisted simulations, leveraging neural networks and other machine learning models, have made it possible to predict complex properties of materials, thereby expanding the realm of possibilities for discovering new substances. These AI tools can also help researchers navigate the vast literature on materials, identify relevant data, and optimize simulation workflows, significantly reducing the time and resources required for materials discovery.[6]

---

[1] "Opportunities for AI," NAIAC briefing, September 2023, https://vimeo.com/869532205; "AI and Science," NAIAC briefing, February 2024, https://www.nist.gov/video/national-artificial-intelligence-advisory-committee-naiac-meeting-february-22-2024.
[2] Sal Khan, "Opportunities for AI."
[3] Anima Anandkumar, "AI and Science."
[4] Jeff Dean, "AI and Science."
[5] Kevin Murphy, "AI and Science."
[6] Jeff Dean, "AI and Science."

The integration of AI into scientific discovery is not just about speeding up existing processes, but also about enabling new ways of conceptualizing and approaching scientific questions. By combining AI with high-performance computing and advanced robotics, scientists are able to tackle complex challenges more efficiently and with greater precision. This synergy enhances our understanding of the natural world and accelerates the pace of innovation, promising breakthroughs that were once considered beyond reach.

As AI continues to evolve, its capacity to enable scientific discovery will undoubtedly expand, opening new frontiers in research and helping address some of the most pressing challenges facing humanity. However, successfully harnessing AI for scientific progress can be challenging. There are economic obstacles, like limited computing resources and a dearth of funding. (Building and leveraging AI models can be prohibitively expensive, even for large institutions.)[7] In addition, lack of access to relevant datasets, especially datasets that are inclusive and without bias, can prevent researchers from leveraging AI in certain domains and use of low quality datasets can result in inaccurate and potentially discriminatory behavior by AI models trained on such datasets. A scarcity of interdisciplinary researchers can also inhibit the potential of AI for science. Then there are technical obstacles, like the difficulty of scaling certain AI models and their limited ability to analyze certain large datasets. In their current state, AI systems cannot fully comprehend fundamental concepts such as DNA or gravity.[8]

There are risks in addition to these challenges. While AI systems can find ways to reduce greenhouse gas emissions, their energy requirements can also contribute to them. While AI can help students learn, it can also mislead them or help them cheat. While AI can help diagnose a patient, it can also misdiagnose them.[9] And while AI models can help discover new drugs, they can also be used by bad actors to help create biological weapons.[10]

The U.S. government can play a pivotal role in helping scientists and educators leverage AI in a positive and transformative way.[11]

The above insights and the below recommendations are informed by two recent briefings organized by the NAIAC AI Futures working group: "Opportunities on AI" held in September 2023 and "AI and Science" held in February 2024. Through these

[7] Rachel Mandelbaum, "AI and Science."
[8] Ece Kamar, "AI and Science."
[9] Priya Donti, "Opportunities for AI."
[10] Ece Kamar, "AI and Science."
[11] "Opportunities for AI"; "AI for Science."

briefings we aimed to learn from industry and academia experts on how we can leverage AI in various domains such as climate, healthcare, and science for public benefit, to improve quality of life, and to expedite scientific discoveries and maintain the U.S. leadership in science. In addition, we heard from these experts on how the government can help bolster the effort in responsible use of AI in science and innovation.

RECOMMENDATIONS

Recommendation 1:
Need for sustained funding and investment in AI in science and support for education and training in scientific communities.

AI-enabled scientific discovery requires an investment in education and training of both scholars entering the discipline as well as for scholars who are farther along in their careers. To maintain U.S. leadership in scientific research, it is essential to develop a comprehensive strategy to update training and curricula. This could be funded by funding agencies–including the National Science Foundation (NSF), National Institutes of Health (NIH), Department of Energy (DOE), and Defense Advanced Research Projects Agency (DARPA). Examples of approaches to use include existing mechanisms such as the NSF IGERT (integrative graduate education and research traineeship programs) focused on AI for Science, summer schools and as well as through other novel approaches for post graduate education through NSF AI Institutes focused on Science and research training and senior fellowship programs modeled after the successful NIH NLM programs for training biomedical informatics.

Reports such as "Report of the 2023 Particle Physics Project Prioritization Panel" and other reports from National Academies can shed light on the gaps and required resources that the U.S. government can provide to ensure scientific communities are empowered and have resources available to leverage AI in their research.[12]

Recommendation 2:
Commission an assessment to inform what additional infrastructure investments are needed for leveraging AI for scientific discovery.

[12] An example of incorporation of a multidisciplinary approach in the scientific discovery process: Vertesi, Janet. *Seeing like a Rover : How Robots, Teams, and Images Craft Knowledge of Mars / Janet Vertesi, The University of Chicago Press.* Chicago ; The University of Chicago Press, 2015. Print.

In Year 1, NAIAC recommended adoption of the NAIRR[13]. The recent AI Executive Order greenlit a pilot of the NAIRR. Building on this success, NAIAC recommends that the government develop an improved perspective on science-specific infrastructure commitments in addition to the compute and data infrastructure to that of the NAIRR. The specific infrastructure needs of AI-enabled scientific discovery may have common components as well as specialized components that are specific to particular branches of science (e.g., the biological sciences vs. astrophysics) and which may be distinct from compute and data resources alone. To make these assessments, entities like the National Science Foundation, the White House Office of Science and Technology Policy, the National Academy of Sciences should be exploring and conducting an assessment of the science-specific AI models, infrastructure, data, and compute required to enable the US to remain a leader in AI-enabled Science and to propose a plan to create such infrastructure. Tools such as NSF workshops and convenings such as by the CCC (https://cra.org/ccc/) can be used to provide inputs to the national academies study. The NSF workshop on self-driving labs (see https://nsf-sdl-2023.github.io/) provides a concrete example of convening expertise in both AI and science-specific infrastructure. AI-enabled scientific discovery can significantly impact both public welfare and commercial success.

We anticipate that such an assessment could result in the design of a comprehensive infrastructure and ecosystem that can enable collaboration between AI/computer science researchers and domain experts in science. Research and quantitative measures that enable the scientific community to track progress on AI's contributions to scientific progress and discovery will enable the creation of a system that will continuously learn.

ACKNOWLEDGEMENTS

The NAIAC AI Futures – Preparedness, Opportunities, and Competitiveness working group participated in the preparation of this document. Contributors include:

- Jack Clark
- David Danks
- Ramayya Krishnan
- Ashley Llorens
- Haniyeh Mahmoudian
- Swami Sivasubramanian

A quorum of the membership of NAIAC reviewed and approved this document.

[13] National Artificial Intelligence Advisory Committee Year 1 Report 2023 (ai.gov)

ABOUT NAIAC

The National Artificial Intelligence Advisory Committee (NAIAC) advises the President and the White House National AI Initiative Office (NAIIO) on the intersection of AI and innovation, competition, societal issues, the economy, law, international relations, and other areas that can and will be impacted by AI in the near and long term. Their work guides the U.S. government in leveraging AI in a uniquely American way — one that prioritizes democratic values and civil liberties, while also increasing opportunity.

NAIAC was established in April 2022 by the William M. (Mac) Thornberry National Defense Authorization Act. It first convened in May 2022. It consists of leading experts in AI across a wide range of domains, from industry to academia to civil society. https://www.ai.gov/naiac/

###