NATIONAL ACADEMIES Sciences Engineering Medicine

Policy and Global Affairs (PGA) Division

AI for Scientific Discovery - A Workshop

National Academy of Sciences Building
2101 Constitution Avenue NW
Washington, DC 20418

The goal for this meeting is to explore the future of AI in terms of its role as an autonomous researcher performing scientific discovery. This includes where AI stands, where it needs to go, and which disciplines should invest more in utilizing AI scientists. The workshop will also explore the ethical aspects and potential pitfalls that loom for AI scientists. This workshop is intended to be inclusive to the global community and will engage with international partners.

THURSDAY, OCTOBER 12, 2023

8:30-9:00 am ET — **Coffee**

9:00-9:05 am ET — **Welcome by Workshop Chair**
Bradley Malin, Professor, Vanderbilt University

9:05-9:15 am ET — **Opening Remarks**
Victor Dzau, President, National Academy of Medicine

9:15-10:45 am ET — **I. What is the Goal of an AI Scientist (in Conducting Independent Research)?**

In the future, AI scientists could have sufficient autonomy and knowledge to function as research assistants, carry out independent investigations, and participate in collaborative projects. This would require novel AI research in areas such as harnessing sophisticated scientific knowledge, mastering advanced scientific processes and methods, and proficiency in scientific communication and collaboration. This session will be devoted to envisioning the potential capabilities that will be needed for AI scientists and the foundational research that will be required in AI and other disciplines. The session will also discuss how AI scientists should be designed to fit the science ecosystem and to respect the boundaries desirable in AI systems for science.

Moderator: **Yolanda Gil,** Principal Scientist and Senior Director for Strategic Initiatives in Artificial Intelligence and Data Science, Information Sciences Institute, University of Southern California

Speakers:

- **Hiroaki Kitano,** CEO, Sony AI Inc.
- **Mario Krenn,** Founder, Krenn Research Group
- **Carla Gomes,** Director of the Institute for Computational Sustainability, Cornell University

10:45-11:15 am ET **Break**

11:15 am-12:30 pm ET **II. Global Scientific Discovery in the AI Age**

AI has already made a substantial impact on the research enterprise in disciplines spanning materials research, chemistry, climate science, biology, and cosmology. These advances have involved the development of self-driving laboratories to efficiently perform physical experiments, AI-accelerated analysis of massive datasets, and the fusion of many disparate data streams. This panel will highlight the recent achievements of AI in these domains to identify where AI is currently in the process of scientific discovery. In addition, a focus of this panel will be identifying what technologies the community requires to advance the use of AI in research.

Moderator: **Keith A. Brown,** Associate Professor of Mechanical Engineering, Physics, and Materials Science & Engineering, Boston University

Speakers:

- **Amy McGovern,** Professor, University of Oklahoma
- **Shirley Ho,** Group Leader, Cosmology X Data Science, CCA, Simons Foundation; Professor, NYU
- **Steve Finkbeiner,** Senior Investigator, Gladstone Institutes
- **Lynda Stuart,** Executive Director (CEO) Institute for Protein Design, University of Washington

12:30-1:30 pm ET **Working Lunch – West Court and Great Hall**

1:30-2:30 pm ET **III. Enablers - What are the gaps in AI that would prevent us from achieving independent scientific discovery?**

For AI to have the knowledge, skills, and abilities to perform independent scientific discovery, it should be able to: 1) Determine research gaps from existing literature, 2) Form unique hypotheses (or intend to conduct hypothesis-generating research), 3) Develop and carry out an experimental plan, 4) Interpret the results, and 5) Draw conclusions both about the original research and how new research could follow on from these results. It is well-established that AI is brittle and performs poorly on out-of-sample distributions, and the explainability and interpretability of AI are other critical limitations. This panel will focus on the limitations of AI in conducting independent scientific discovery and what could be done to mitigate these problems.

Moderator: **Missy Cummings,** Professor and Director of Mason Autonomy and Robotics Center (MARC)

Speakers:

- **Gary Marcus,** Professor Emeritus, New York University (Video Presentation)
- **Chitra Sivanandam,** Co-Founder, Rohirrim
- **Subbarao Kambhampati**, Professor, Arizona State University

2:30-3:30 pm ET

IV. Societal Aspects of AI (Barriers and Opportunities - Inclusive of Regulation)

As AI evolves, it will interact with and influence our world in ways that are unexpected, some for the better and some for the worse. Over the past several years, it has become clear that machine learning methods have the potential to embed the biases that arise in society into their resulting models, such as the systemic discrimination and subjugation of certain populations. This has prompted a flurry of investigations into how to assess the fairness of AI technologies as they are developed and deployed. Yet, as AI begins to interact with the physical world, it will induce new types of concerns, such as how materials engineered in an autonomous fashion may influence the wellbeing of humans and the environment. This panel will raise questions about the ethical, legal, and social implications of autonomous AI and the extent to which these can enable and or hinder the development of AI.

Moderator: **Bradley Malin**, Professor, Vanderbilt University

Speakers:

- **Glenn Cohen,** Deputy Dean, Harvard Law School
- **Vukosi Marivate,** ABSA UP Chair of Data Science, University of Pretoria
- **Deborah Johnson,** Professor Emeritus, Engineering and Society, University of Virginia

3:30-3:50 pm ET **Break**

3:50-4:50 pm ET

V. What are the gaps in automated physical experiments and experimental data collection and curations that would prevent us from achieving independent scientific discovery?

Many types of scientific work fundamentally involve physical experiments. In these fields, AI systems that are not able to directly run experiments may be quite limited. Conversely, one of the powers of an AI system is its ability to simultaneously manage more ongoing experiments than a human can. Automation of experimentation has therefore been a frequent topic over the years, most recently under the moniker of “self-driving labs”. However, most automated experimentation systems have significant limitations. In this session, the participants will discuss where we are currently with AI-driven physical experimentation along with the technology and opportunities that are on the horizon.

Moderator: **Patrick Riley,** SVP, AI at Relay Therapeutics

Speakers:

- **Hod Lipson,** Professor of Mechanical Engineering, Columbia University
- **Benji Maruyama,** Autonomous Materials Lead, Materials & Manufacturing Directorate at Air Force Research Laboratory
- **Peter Madrid**, Head of Scientific Discovery, Synfini Inc.

4:50-5:00 pm ET **Day 1 of Workshop Closing Remarks**

5:00-7:00 pm ET **Networking Reception – Great Hall**

FRIDAY, OCTOBER 13, 2023

8:30-9:00 am ET — **Coffee**

9:00-9:15 am ET — **Welcome by Workshop Chair**
Bradley Malin, Professor, Vanderbilt University

9:15-10:30 am ET — **VI. Examples of Grand Challenges: AI Applications Domains. What is impossible currently in supporting the AI researchers/scientist?**

Achieving independent scientific discovery using AI would rely on key milestones that constitute the scientific discovery process. These include ability to: 1) formulate a scientific hypothesis, 2) identify variables that need to be measured, the corresponding data sources (simulations, experiments, literature, etc.), and record their measurements, 3) test the hypothesis reliably using collected data, and 4) refine the hypothesis based on inference. This session will identify fundamental roadblocks and associated grand challenges that can help assess and advance the feasibility of scientific discovery using AI.

Speakers Moderator: **Aarti Singh,** Professor in the Machine Learning Department, Carnegie Mellon University

Speakers:

- **Aisha Walcott Bryant,** Head Google Research Kenya
- **Tapio Schneider,** Professor of Environmental Science and Engineering, California Institute of Technology
- **Anthony Bak,** Head of AI and Machine Learning, Palantir Technologies

10:30-11:00 am ET — **Break**

11:00 am-12:00 pm ET — **VII. Small Steps in the Direction of Grand Challenges: Examples of AI Pilots and (Small) Wins**

Profound challenges lie ahead to develop AI systems capable of scientific discoveries, requiring significant resources and long-term agendas. This session will focus on potential avenues and promising first steps that could move the community forward, disciplines that are better positioned to enable progress, and immediate challenges that can be realistically tackled to demonstrate progress along the way.

Moderator: **Yolanda Gil,** Principal Scientist and Senior Director for Strategic Initiatives in Artificial Intelligence and Data Science, Information Sciences Institute, University of Southern California

Speakers:

- **Manuela Veloso,** Head of JPMorganChase AI Research, Herbert A. Simon University Professor Emerita, Carnegie Mellon University
- **Peter Clark,** Senior Research Director, AI2
- **Steve Chien,** Technical Fellow, Artificial Intelligence, Jet Propulsion Laboratory, California Institute of Technology

12:00-1:00 pm ET — **Working Lunch – West Court and Great Hall**

1:00-2:00 pm ET

VIII. Opportunities and Where to Make Investments in AI

The final panel of the meeting focuses on how extramural funding organizations can help shape this vision. What activities are they interested in and what are the next steps in AI for Scientific Discovery? What are the big movements in AI scientific research that are occurring right now in preparation for the future? What is missing from research that is needed?

Moderator: **Bradley Malin**, Professor, Vanderbilt University

Speakers:

- **Susan Gregurick,** Associate Director for Data Science and Director of the Office of Data Science Strategy (ODSS), NIH
- **Andrey Kanaev,** Program Director, OAC, NSF
- **Patrick Rose,** Innovation Manager, SPRIN-D
- **Mark Greaves,** Executive Director, AI2050, Schmidt Futures

2:00-3:30 pm ET

Closing Discussion

Moderator: **Bradley Malin**, Professor, Vanderbilt University