REPORT TO THE PRESIDENT

Supercharging Research: Harnessing Artificial Intelligence to Meet Global Challenges

Executive Office of the President

President's Council of Advisors on Science and Technology

APRIL 2024

About the President's Council of Advisors on Science and Technology

The President's Council of Advisors on Science and Technology (PCAST) is a federal advisory committee appointed by the President to augment the science and technology advice available to him from inside the White House and from the federal agencies. PCAST comprises of 28 of the Nation's thought leaders, selected for their distinguished service and accomplishments in academia, government, and the private sector. PCAST advises the President on matters involving science, technology, and innovation policy, as well as on matters involving scientific and technological information that is needed to inform policy affecting the economy, worker empowerment, education, energy, the environment, public health, national and homeland security, racial equity, and other topics.

For more information about PCAST see www.whitehouse.gov/pcast.

EXECUTIVE OFFICE OF THE PRESIDENT
PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY
WASHINGTON, D.C. 20502

President Joseph R. Biden, Jr.
The White House
Washington, D.C.

Dear Mr. President,

Your President's Council of Advisors on Science and Technology (PCAST) is excited by the forward-looking approach that your Administration has taken to advance the safe and effective use of artificial intelligence (AI).[1,2,3] As requested in your landmark Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, we are pleased to report here on the *possibilities* that AI can enable when applied to research to address major societal and global challenges.

AI will fundamentally transform the way we do science. Researchers in many fields are already employing AI to identify new solutions to a wide array of long-standing problems. Today, scientists and engineers are using AI to envision, predictively design, and create novel materials and therapeutic drugs. In the near future, AI will enable unprecedented advances in the social sciences, both through new methods of analyzing existing data and the development and analysis of new kinds of anonymized and validated data. Such advances will allow government to better understand how policies affect the American people, and improve those policies to better meet societal needs and challenges. AI will also allow researchers to run millions of computer-based simulated experiments quickly to provide guidance about the most important real-world experiments to run. In industrial laboratories, rich simulations will be able to identify hazards or faults in design so that scientists and engineers can create safer, scalable, and efficient products that American industry and American consumers can depend on. In sum, AI is revolutionizing the research process, enriching scientific models, and accelerating data generation and analysis, with impacts that will be far-ranging.

In addition to its opportunities, we must recognize that AI can create new issues and challenges, such as distilling errors and biases embedded in skewed training data, the enormous—and increasing—amounts of energy required for the computational processes, the possibility that faulty science could be unwittingly generated, and the ease with which nefarious actors could use new powerful AI technologies for malicious purposes. Expert human supervision, building protections into the AI algorithms, and a culture of responsible use that includes appropriate application of regulatory frameworks, as outlined in your Blueprint for an AI Bill of Rights and the AI Risk Management Framework from the National Institute of Standards and Technology, will be essential to mitigating the weaknesses and dangers of AI. Fortunately, reproducibility and validation are core tenets of the

---

[1] The White House. (2022 October). Blueprint for an AI Bill of Rights: Making Automated Systems Work for The American People.
[2] Executive Order 14110, 88 FR 75191. (2023 October). "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."
[3] The White House Office of Management and Budget. (2024 March). Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence. [Memo M-24-10]

scientific method. As such, the scientific community is already engaged in vibrant and pathbreaking research on AI reproducibility, testing, and evaluation so that researchers—and everyone—will eventually be able to use AI tools with the same confidence with which we use calculators.

Ultimately, our vision is that the responsible use of AI will empower scientists and engineers to leverage the opportunities of this transformative technology while navigating and mitigating its weaknesses.

Per your Executive Order, and building on the bold work of your Administration, PCAST offers findings and recommendations for action that will help the United States to harness the full potential of AI to equitably and responsibly supercharge research to meet the world's challenges.

Sincerely,

The President's Council of Advisors on Science and Technology

Table of Contents

The President's Council of Advisors on Science and Technology

Co-Chairs

Frances H. Arnold
Linus Pauling Professor of Chemical Engineering, Bioengineering, and Biochemistry
California Institute of Technology

Maria T. Zuber
Vice President for Research and E. A. Griswold Professor of Geophysics
Massachusetts Institute of Technology

Arati Prabhakar
Director, Office of Science and Technology Policy
Assistant to the President for Science and Technology
The White House

Members

Dan E. Arvizu
Former Chancellor
New Mexico State University System

Dennis Assanis
President
University of Delaware

John Banovetz
Executive Vice President, Chief Technology Officer and Environmental Responsibility
3M Company

Frances Colón
Senior Director, International Climate
Center for American Progress

Lisa A. Cooper
Bloomberg Distinguished Professor of Equity in Health and Healthcare and Director of the Center for Health Equity
Johns Hopkins University

John O. Dabiri
Centennial Professor of Aeronautics and Mechanical Engineering
California Institute of Technology

William Dally
Chief Scientist and Senior Vice President for Research
NVIDIA

Sue Desmond-Hellmann
Former CEO
Bill & Melinda Gates Foundation

Inez Fung
Professor of Atmospheric Science
University of California, Berkeley

Andrea Goldsmith
Dean of the School of Engineering and Applied Science and the Arthur LeGrand Doty Professor of Electrical and Computer Engineering
Princeton University

Laura H. Greene
Chief Scientist, National High Magnetic Field Laboratory
Florida State University, University of Florida, Los Alamos National Laboratory
Marie Krafft Professor of Physics
Florida State University

Paula Hammond
Institute Professor, Vice Provost for Faculty, and member of the Koch Institute for Integrative Cancer Research
Massachusetts Institute of Technology

Eric Horvitz
Chief Scientific Officer
Microsoft

Joe Kiani
Chairman and CEO
Masimo

Jon Levin
Philip H. Knight Professor and Dean of the Graduate School of Business
Stanford University

Steve Pacala
Frederick D. Petrie Professor Emeritus in the Department of Ecology and Evolutionary Biology
Princeton University

Saul Perlmutter
Franklin W. and Karen Weber Dabby Professor of Physics and Director of the Berkeley Institute for Data Science
University of California, Berkeley
Senior Scientist
Lawrence Berkeley National Labs

William Press
Leslie Surginer Professor of Computer Science and Integrative Biology
The University of Texas at Austin

Jennifer Richeson
Philip R. Allen Professor of Psychology and Director of the Social Perception and Communication Lab
Yale University

Vicki Sato
Professor of Management Practice (Retired)
Harvard Business School

Lisa Su
President and CEO
Advanced Micro Devices (AMD)

Kathryn D. Sullivan
Former Astronaut
National Aeronautics and Space Administration
Former Administrator
National Oceanic and Atmospheric Administration

Terence Tao
Professor & the James and Carol Collins Chair in the College of Letters and Sciences
University of California, Los Angeles

Phil Venables
Chief Information Security Officer
Google Cloud

Catherine Woteki
Visiting Distinguished Institute Professor in the Biocomplexity Institute
University of Virginia
Professor of Food Science and Human Nutrition
Iowa State University

PCAST Staff

Lara Campbell
Executive Director

Reba Bandyopadhyay
Deputy Executive Director

Melissa Edwards
Assistant Deputy Executive Director

Bich-Thuy (Twee) Sim
Assistant Director for Transformative
Medicine and Health Innovation

Kimberly Lawrence
Administrative Specialist

Riya Dhar
Intern

Maya Millette
Former Intern

Working Group on Artificial Intelligence

Working Group members participated in the preparation of this report. The full membership of PCAST reviewed and approved the report.

Co-Leads

Terence Tao
Professor & the James and Carol Collins Chair in the College of Letters and Sciences
University of California, Los Angeles

Laura H. Greene
Chief Scientist, National High Magnetic Field Laboratory
Florida State University, University of Florida, Los Alamos National Laboratory
Marie Krafft Professor of Physics
Florida State University

Members

John Banovetz
Executive Vice President, Chief Technology Officer and Environmental Responsibility
3M Company

William Dally
Chief Scientist and Senior Vice President for Research
NVIDIA

Eric Horvitz
Chief Scientific Officer
Microsoft

Jon Levin
Philip H. Knight Professor and Dean of the Graduate School of Business
Stanford University

Saul Perlmutter
Franklin W. and Karen Weber Dabby Professor of Physics and Director of the Berkeley Institute for Data Science
University of California, Berkeley
Senior Scientist
Lawrence Berkeley National Labs

William Press
Leslie Surginer Professor of Computer Science and Integrative Biology
The University of Texas at Austin

Lisa Su
President and CEO
Advanced Micro Devices (AMD)

Phil Venables
Chief Information Security Officer
Google Cloud

Executive Summary

Artificial Intelligence (AI) has the potential to revolutionize our ability to address humanity's most urgent challenges by providing researchers with tools that will accelerate scientific discoveries and technological advances. Generative AI, which can create content based on vast data sets and extensive computation, is poised to be particularly transformative. Examples of generative AI include large language models, image generating models, and generative scientific models. In his comprehensive [Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence](), issued on October 30, 2023, President Biden charged PCAST to report on "the potential role of AI…in research aimed at tackling major societal and global challenges." PCAST is pleased to offer this report in fulfillment of this charge.

With well-designed, equitably shared, and responsibly used infrastructure, AI will enable scientists to address urgent challenges, including improving human health and enhancing weather prediction in a time of climate change. AI can help explore long-standing scientific mysteries that inspire and stretch human creativity, such as uncovering the origin and evolution of the universe. AI will also help researchers address continuing national needs, from accelerated semiconductor chip design to the discovery of new materials to address our energy needs. Furthermore, AI is starting to remove barriers that make scientific research slow and expensive, for instance by providing the means for rapidly identifying the best drug candidates for testing (thus reducing the number of expensive laboratory trials), helping to optimize experimental designs, and uncovering connections in data much more efficiently than can be done by hand or using traditional data science methods. If basic AI resources, validated data, and scientific tools and training are made broadly accessible, AI technologies have the potential to democratize scientific knowledge,[4] bringing interconnected technical concepts to many more people and enabling diverse researchers to bring their expertise and perspectives to societal and global challenges.

Just as with any other new tool or technology, realizing the potentials of AI will require addressing its limitations. These issues include misleading or incorrect results, perpetuation of bias or inequity[5] and sampling errors from patterns embedded in the model-training data, limited access to high quality training data, the challenges of protecting intellectual property and privacy, the significant energy required to train or deploy a model or run the AI algorithms, and the risk that bad or nefarious actors will use readily available AI tools for malicious purposes. Many public and private sector activities addressing these issues are already underway, including government efforts tasked under the October 2023 Executive Order on AI. Reproducibility and validation are key principles underlying scientific integrity and the scientific method and must continue to be held at high value as we develop a culture of responsible AI use and expert human supervision of AI applications.

AI has the potential to transform every scientific discipline and many aspects of the way we conduct science. Scientists are already employing AI to create new functional materials that we presently do not know how to design; these include superconductors and thermoelectric materials which would not only enhance our energy efficiency but also reduce our carbon footprint. In a similar vein, AI models are helping researchers create new designs for manufacturing processes and products, and

---

[4] Dessimoz, C. and Thomas, P. (2024 March). AI and the democratization of knowledge. *Scientific Data.*
[5] Birhane, A. (2022 October). The unseen Black faces of AI algorithms. *Nature.*

develop new drug therapies which in the future could enable individualized treatment of specific cancers and viruses. AI models are also helping engineers design semiconductor chips, producing better designs with less human effort and time. In health care, AI technologies are creating new ways to analyze a broad spectrum of medical data for applications like the early diagnosis of diseases[6] that can lead to timely intervention and the detection of medical errors.[7] PCAST also foresees widely available AI-powered ultra-personalized medicine tailored to a specific individual and disease process that will include details of medical history, genetic information, and signals, such as how healthy and unhealthy cells are behaving.

AI is also transforming science by improving our scientific models. In climate science, AI models are starting to enhance weather prediction, as well as advancing whole-earth models for water management, greenhouse gas monitoring, and predicting the impacts of catastrophes. Scientists have already used AI to successfully predict the structure of proteins; new foundation models will unlock more secrets of cellular biology and power computer simulations of intracellular interactions that can be used to explore new therapies. AI models promise to help us understand the origin of our universe by allowing us to test numerous cosmological hypotheses via rapid simulations. Such AI-enabled modeling may even help scientists discover new laws of physics.

AI will enable unprecedented advances in the social sciences, complementing qualitative methods with new quantitative techniques for analyzing existing data, as well as the development and analysis of newer types of data, e.g., step counts on smartphones, anonymized data drawn with permission from search and browsing, or images posted on social media. AI could supercharge research using vast data sets, such as those that have long been collected and curated by federal statistical agencies—ideally complemented by those held by the private sector—as input for designing effective federal policies. Application of AI to these long-standing and newer social science data sets could facilitate more effective, responsive, and fairer data-driven policymaking and delivery of services.

These few examples of AI-assisted research illustrate that with the responsible use of AI technology, human scientists will be empowered to realize transformational discoveries. Furthermore, PCAST expects that responsible sharing of basic AI resources will help to democratize science and tackle major societal and global challenges.

The use of AI for science and technology research is accelerating rapidly across the globe and therefore demands our commitment to U.S. leadership in the applications of this powerful new tool. Building on the work of the Biden Administration, the United States must act boldly and thoughtfully to maintain our nation's lead in research, in the innovative applications of AI, and in establishing frameworks and norms for the safe and responsible use of AI. In this report, PCAST offers five specific findings and recommendations for action that will help the U.S. to harness the full potential of AI to equitably and responsibly supercharge scientific discovery.

---

[6] Bera, K. et al. (2019 August). Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*.

[7] Nguyen, V. et al. (2023 November). Efficient automated error detection in medical data using deep-learning and label-clustering. *Scientific Reports.*

Summary of Recommendations

Recommendation 1: Expand existing efforts to broadly and equitably share fundamental AI resources.

Extensive support for widely accessible shared models, data sets, benchmarks, and computational resources is essential to ensuring that academic researchers, national and federal laboratories, and smaller companies and non-profit organizations can use AI to create benefits for the nation. In the U.S., the most promising effort in this direction is the National Artificial Intelligence Research Resource (NAIRR), which is currently a pilot project. PCAST recommends that the NAIRR pilot be expeditiously expanded to the scale envisioned by the NAIRR Task Force[8] and fully funded. The full-scale NAIRR, together with industry partnerships and other AI infrastructure efforts at both the federal and state levels, could serve as a stepping stone towards AI infrastructure projects at the national or international level to facilitate high-impact research.

Recommendation 2: Expand secure access to federal data sets for approved critical research needs, with appropriate protections and safeguards.

The benefits of allowing limited, secure access to federal data sets by approved researchers, as well as allowing the release of carefully anonymized versions of such data sets to curated resource centers such as NAIRR, are immense. PCAST strongly encourages expansion of existing pilot programs for secure data access and the development of guidelines for federal database management that incorporate cutting-edge privacy protection technologies as they become available. There is great potential to use modern AI technologies to automate aspects of the curation of such data sets. PCAST encourages the use of AI to improve data curation as a long-term goal of federal data sharing initiatives such as data.gov.

PCAST endorses the efforts of federal agencies to mandate responsible sharing of data sets arising from the research that they fund or conduct.[9] We encourage further enforcement of such mandates, to include sharing of AI models trained on data from federally funded research, in conjunction with sufficient resources to support the required actions.

Recommendation 3: Support both basic and applied research in AI that involves collaborations across academia, industry, national and federal laboratories, and federal agencies as outlined in the vision for the NAIRR developed by the NAIRR Task Force.

The boundaries between federally funded academic research and private sector research are hazy. Many researchers move among affiliations with academic institutions, non-profit organizations, and/or private companies, and a significant share of all AI research and development (R&D) is

[8] National Artificial Intelligence Research Resource Task Force. (2023 January). Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource.
[9] The White House Office of Science and Technology Policy. (2022 August). Ensuring Free, Immediate, and Equitable Access to Federally Funded Research.

currently supported by private companies.[10] To capitalize fully on the potential benefits of AI for science, research that involves a breadth of promising and productive hypotheses and approaches must be supported. This may require that funding agencies broaden their postures regarding how to work with industry and which researchers can be supported in order to facilitate innovative research efforts and collaborations among different sectors. Examples of such collaboration could include creation of high quality curated public scientific data sets from multiple sources or the creation of multimodal foundation models.[11]

Recommendation 4: Adopt principles of responsible, transparent, and trustworthy AI use throughout all stages of the scientific research process.

Managing the risks of inaccurate, biased, harmful, or non-replicable findings from scientific uses of AI should be planned from the initial stages of a research project rather than performed as an afterthought. PCAST recommends that federal funding agencies consider updating their responsible conduct of research guidelines to require plans for responsible AI use from researchers. These plans should include recommended best practices from agency offices and committees that address potential AI-related risks and describe the supervision procedures for use of any automated process.[12] To minimize additional administrative burden on researchers and build a culture of responsibility, after enumerating major risks, agencies should provide model processes for risk mitigation.

Parallel to this, agencies such as the National Science Foundation (NSF) and the National Institute of Standards and Technology (NIST) should continue supporting research in the scientific foundations of responsible and trustworthy AI. This research should include the development of standard benchmarks to measure AI model properties such as accuracy, reproducibility, fairness, resilience, and explainable AI, as well as AI algorithms that monitor themselves for these properties and adjust when the benchmarks are not within defined norms. Another goal of such research should be the development of tools to evaluate biases in data sets and to distinguish synthetic data from real world data.

Recommendation 5: Encourage innovative approaches to integrating AI assistance into scientific workflows.

The scientific enterprise is an excellent "sandbox" in which to practice, study, and assess new paradigms of collaboration between humans and AI assistants. The objective should not be to maximize the amount of automation, but to allow human researchers to achieve high quality science that utilizes AI assistance responsibly.

Funding agencies should recognize the emergence of these new workflows and design flexible procedures, metrics, funding models, and challenge problems that encourage strategic

[10] National Science Foundation. (Accessed 23 April 2024). The State of U.S. Science and Engineering 2022 (Figure 16).
[11] A *foundation model* is an ML model trained (often at great computational expense) on a broad range of data, which can then be fine-tuned relatively cheaply for more specialized applications. For more discussion, see Bommasani, R. et al. (2022 July). On the Opportunities and Risks of Foundation Models. arXiv.
[12] National Institute of Standards and Technology. (2023 November). Artificial Intelligence Safety Institute Consortium. Federal Register.

experimentation with new AI-assisted ways to organize and execute a scientific project. Implementation of these workflows also present opportunities for researchers from a variety of disciplines, such as human factors and industrial and organizational psychology, to advance our knowledge in the area of human-machine teaming.

More broadly, incentive structures across funding agencies, academia, and the academic publishing industry may need to be updated to support a broader range of scientific contributions, such as curating a high quality and broadly usable data set, that might not be given sufficient recognition by traditional metrics of research productivity.

1. Introduction

We stand on the brink of an extraordinary era where innovation driven by artificial intelligence (AI) across the sciences promises to greatly accelerate America's long-term leadership in scientific knowledge and solutions.[13, 14, 15, 16] This AI-powered paradigm shift in scientific tools and methods is timely and critical as humanity confronts daunting challenges in important areas such as energy, climate, healthcare. Beyond solving known challenges, by harnessing AI responsibly, equitably, and effectively for research, scientists can deliver greater resilience to our society, improving our ability to provide a vast array of benefits such as clean water, abundant electricity, and health and wellness to Americans.

Yet the most important payoff of the AI transformation of the sciences will be the ability to realize previously unattainable or even unimagined possibilities with scientific advancements and understandings —developments that have the potential to improve the lives of all Americans. We are already seeing glimmers of possibility for leveraging AI to address difficult health challenges, including cancer,[17] autoimmune diseases,[18] neurodegenerative disorders,[19] and drug-resistant infection.[20] At the same time, researchers are harnessing AI to generate surprising advances in materials science[21] with promise for creating next-generation batteries,[22] superconductors,[23, 24] and computer chips.[25] Beyond advances in core science and engineering disciplines, AI methods promise to provide high fidelity models—"digital twins"[26]—of the world that can help us to cut through uncertainty and complexity to predict, to plan, and to guide policymaking, where scarce data and models currently make it difficult to assess potential pathways forward.

---

13 Hope, T. et al. (2023 August). A Computational Inflection for Scientific Discovery. *Communications of the ACM.*
14 Wang, H. et al. (2023 August). Scientific discovery in the age of artificial intelligence, *Nature.*
15 Mock, M. et al. (2023 September). AI can help to speed up drug discovery — but only if we give it the right data. *Nature*.
16 National Academies of Sciences, Engineering, and Medicine. (2022 May). Automated Research Workflows for Accelerated Discovery: Closing the Knowledge Discovery Loop.
17 Thierry, A. (2023 January). Circulating DNA fragmentomics and cancer screening. *Cell Genomics*.
18 Danieli, M. et al. (2024 February). Machine learning application in autoimmune disease: State of art and future prospects. *Autoimmunity Rev*iews.
19 Cumplido-Mayoral, I. et al. (2023 April). Biological brain age prediction using machine learning on structural neuroimaging data: Multi-cohort validation against biomarkers of Alzheimer's disease and neurodegeneration stratified by sex. *eLife*.
20 Wong, F. et al. (2023 December). Discovery of a structural class of antibiotics with explainable deep learning. *Nature.*
21 See section 3.1 of this report.
22 Chen, C. et al. (2024 January). Accelerating computational materials discovery with artificial intelligence and cloud high-performance computing: from large-scale screening to experimental validation. arXiv.
23 Pogue, E. et al. (2023 October). Closed-loop superconducting materials discovery. *npj Computational Materials*
24 Liu, Y. et al. (2023 December) Materials Expert-Artificial Intelligence for Materials Discovery. arXiv.
25 E.g., Liu, M. et al. (2024 April). ChipNeMo: Domain-Adapted LLMs for Chip Design. arXiv.
26 A *digital twin* is a high-resolution model of a physical system that is continually updated with real-time data from that system. Such twins usually rely on traditional simulation to model the fundamental processes of the system but can additionally use AI models to refine, accelerate, or analyze such simulations. For the current state of such models, see National Academies of Science, Engineering, and Medicine. (2024). Foundational Research Gaps and Future Directions for Digital Twins. *National Academies Press*.

In the October 2023 Executive Order on the Safe, Secure, Trustworthy Development and Use of Artificial Intelligence,[27] PCAST was charged to report on the potential impact of AI on scientific research aimed at tackling major societal and global challenges, and on practices needed to ensure effective and responsible use of these technologies. The effects of AI on such important topics as national security and critical infrastructure, labor markets, authenticity of content such as images or video, intellectual property rights, education, and the criminal justice system will be addressed in other reports and actions mandated by the Executive Order. As such, this report focuses specifically on the role of AI in the sciences, rather than these broader impacts.

An acceleration of science

Broadly speaking, scientific advances have historically proceeded via a combination of three paradigms: empirical studies and experimentation; scientific theory and mathematical analyses; and numerical experiments and modeling. In recent years a fourth paradigm,[28] data-driven discovery, has emerged.

These four paradigms complement and support each other. However, all four scientific modalities experience impediments to progress. Verification of a scientific hypothesis through experimentation, careful observation, or via clinical trial can be slow and expensive. The range of candidate theories to consider can be too vast and complex for human scientists to analyze. Truly innovative new hypotheses might only be discovered by fortuitous chance, or by exceptionally insightful researchers. Numerical models can be inaccurate or require enormous amounts of computational resources. Data sets can be incomplete, biased, heterogeneous, or noisy to analyze using traditional data science methods.

AI tools have obvious applications in data-driven science, but it has also been a long-standing aspiration[29] to use these technologies to remove, or at least reduce, many of the obstacles encountered in the other three paradigms. With the current advances in AI, this dream is on the cusp of becoming a reality: candidate solutions to scientific problems are being rapidly identified, complex simulations are being enriched, and robust new ways of analyzing data are being developed. By combining AI with the other three research modes, the rate of scientific progress will be greatly accelerated, and researchers will be positioned to meet urgent global challenges in a timely manner.[30]

Like most technologies, AI is dual use: AI technology can facilitate both beneficial and harmful applications and can cause unintended negative consequences if deployed irresponsibly or without expert and ethical human supervision. Nevertheless, PCAST sees great potential for advances in AI to accelerate science and technology for the benefit of society and the planet. In this report, we provide a high-level vision for how AI, if used responsibly, can transform the way that science is done, expand the boundaries of human knowledge, and enable researchers to find solutions to some of society's most pressing problems. We will illustrate this potential for seven different areas of science by

---

[27] Executive Order 14110, 88 FR 75191. (2023 October). "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."

[28] Data has of course played an important role in the sciences for centuries; but the capability to collect and process such data in the modern era is unprecedented, see Hey, T. et al. (2009 October). The Fourth Paradigm – Data-Intensive Scientific Discovery. *Microsoft Research*.

[29] Smalheiser, N. and Swanson, D. (1998 November), Using ARROWSMITH: a computer-assisted approach to formulating and assessing scientific hypotheses. *Computer Methods and Programs in Biomedicine*.

[30] Wang, H. et al. (2023 August). Scientific discovery in the age of artificial intelligence, *Nature*.

describing the current state-of-the-art in the field, the ways that AI is beginning to be deployed to overcome barriers to progress, the benefits we think AI can help to achieve in the mid-to-long-term, and examples of challenges that must be overcome. Finally, we provide cross-cutting findings and recommendations about what will be needed to realize our vision for the future of AI-enabled science while mitigating potential risks.

Artificial intelligence and machine learning

AI refers to a wide spectrum of technologies intended to perform or assist with cognitive tasks that were previously only achievable via human intelligence. AI tools do not actually duplicate the mechanisms of human thought, which are themselves still incompletely understood by modern science. Rather, most AI systems currently operate through machine learning (ML), which is an array of techniques that leverages statistical inference "learned" through training of an AI model on large data sets.[31] Such models may then be applied to infer answers to related questions. AI systems do not have a deep conceptual comprehension of the task that they are attempting to solve,[32, 33] and in the absence of external verification, the answers provided by an AI model are not guaranteed to be correct. Nevertheless, these models can perform remarkably well at many complex and imprecisely specified tasks, using their training process to identify patterns and relationships that were previously hidden in the data sets.

Over the last twenty years, numerous ML technologies have matured to the point where they are routinely deployed in consumer and business products. As classic examples, ML powers filters for junk email and is also used to prioritize email. In more recent applications, a commercially available smartphone is capable of recognizing speech commands, identifying its owner through facial recognition,[34] or extracting text from an image and then translating it to a different language.[35] ML technology is also found in cars: sensors, self-piloting features, and other driver assistance systems are being deployed in newer vehicles to significantly improve road safety.[36] Today, many companies routinely rely on ML algorithms to detect fraud, improve their logistics and marketing, deliver targeted advertisements to consumers, or predict customer creditworthiness.

Despite these proven applications, ML and broader AI technologies still suffer from several weaknesses. Their outputs are often arrived at by an opaque process that carries no guarantee of correctness and which may involve usage of data that is either protected by intellectual property rights or contains sensitive private information about individuals. Biases in the training data set, as well as systematic biases in the training process, can lead to problematic algorithmic bias in the

[31] Several AI systems also have an additional "fine-tuning" component to their training, for instance by adjusting the model parameters in response to human responses to its outputs by a process known as "reinforcement learning with human feedback."
[32] Indeed, there is not even a consensus on what it means for an AI to "understand" a concept.
[33] Joksimovic, S. et al. (2023). Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. *Computers and Education: Artificial Intelligence*.
[34] E.g., Idemia Facial Recognition Access Control is now deployed at many airports worldwide to streamline security procedures.
[35] Popular software tools in this category include the Voice Dream Scanner for low sighted persons, or Google Translate.
[36] According to a recent study by the AAA, such systems could prevent as many as 37 million automotive crashes over the next 30 years.

model outputs. To address these issues, the Biden Administration's Blueprint for an AI Bill of Rights[37] set out guiding principles on how to mitigate algorithmic bias and other weaknesses prior to incorporation into real world systems. The NIST AI Risk Management Framework[38] added further specific measurement and managerial approaches to mitigate and reduce adverse outcomes from adoption of AI, and the Office of Management and Budget (OMB) Memo on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence[39] further clarified how federal agencies should utilize AI.

Generative AI

Some of the most prominent and striking recent advances in the field have occurred in the class of machine learning tools known as generative AI, which include popular large language models such as ChatGPT, Gemini, Claude, and LLaMA, image generation models such as Midjourney and DALL-E, and scientific generative models such as AlphaFold, RoseTTAfold, and ChemBERTa. By using large and complex deep neural network[40] models which have been trained on many terabytes[41] of data, and which have often been refined by thousands of hours of human input, these generative AI tools analyze user prompts and produce outputs using models that have billions or even trillions of learned relationships or parameters. Generative algorithms extend inputs into likely sequences or structures, enabling these models to create text, images, and other media in response to prompts that many liken to the responses of a human expert. Reinforcement learning from human feedback can further enable improvement of these generative models.

Generative AI is still a very new and rapidly developing technology, and as such has issues and problems that warrant further research and improvement. For instance, in addition to the general weaknesses of machine learning tools previously mentioned, some generative AI outputs are also prone to "hallucination." Since generative AI models do not exclusively encode factual knowledge, they can confidently make assertions that have no factual basis but are deceptively convincing, or may produce images with artifacts that are inconsistent with physical reality. The challenge with the veracity of AI generations extends beyond the fidelity of training data: the probabilistic nature of generations may lead AI models to create plausible, yet inaccurate generations. In some applications, e.g., working with AI systems to generate out-of-the-box possibilities, such "imaginative" processes can be useful or desired. However, in most scientific applications, the goal is truthful inferences. Strategies and mitigations continue to be developed to reduce instances of erroneous generations and other observed problems with the output of generative AI. Methods are being investigated to determine when scientific AI models generate errors,[42] to provide a well-calibrated confidence level in the output of such models,[43] and to ensure that such outputs are consistent with physical,

[37] The White House. (2022 October). Blueprint for an AI Bill of Rights: Making Automated Systems Work for The American People.
[38] National Institute of Standards and Technology. (2023 January). Artificial Intelligence Risk Management Framework (AI RMF 1.0).
[39] The White House Office of Management and Budget. (2024 March). Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence. [Memo M-24-10].
[40] Deep *neural networks* are a particularly successful class of machine learning algorithms that were inspired by the structure of neurons inside a human brain. Neural networks are composed of interconnected layers each containing a simpler unit or "node" that can conducts part of a computation.
[41] E.g., of data volumes: University of Delaware. Examples of Data Volumes. (Accessed 2024 April 16).
[42] Azaria, A. and Mitchell, T. (2023 October). The Internal State of an LLM Knows When It's Lying. arXiv.
[43] Nori, H. et al. (2023 April). Capabilities of GPT-4 on Medical Challenge Problems. arXiv.

biological, or chemical constraints.[44]

There are further concerns with generative AI technology. The most powerful models currently require large amounts of computational infrastructure and energy[45] to train and fine-tune. Furthermore, most of these models are constructed and controlled by a small number of large technology companies, with the current leaders in the field mostly based in the United States. This situation continues to evolve. For instance, reasonably capable open source and/or open weight[46] generative models have been released, and recent research has demonstrated that small language models can provide high levels of performance—at times comparable to the output of the largest models—with significantly lower resource costs.[47] While AI tools have advanced considerably, the prospect of employing generative AI to fully replace humans in real world processes and workflows is still largely in the realm of the far future, and the future this report envisions is one in which AI tools assist rather than supplant humans.

Despite their imperfect nature, generative AI technologies hold tremendous potential and promise. According to one recent estimate, applications of generative AI could add between $2.6 trillion to $4.4 trillion annually to the global economy.[48] AI technologies present especially transformative opportunities in the realm of scientific research, in which many of the aforementioned weaknesses can be addressed by combining AI models with other scientific tools and methods. There is a growing body of research demonstrating that external scientific validation methods, such as laboratory experiments, clinical trials, numerical simulations, formal verification software, and limiting the AI input to draw from curated scientific literature to draw from curated scientific literature can be used to mitigate the hallucinations of AI models. For example, rather than being trained on large "internet scale" text resources, scientific models can be trained on extensive scientific data sets that are publicly available and better curated, and which come with relatively few intellectual property restrictions.

The scientific community places great value on transparency and reproducibility, and many AI models and methodologies used in research are openly available. The sharing of data and training resources can greatly reduce the considerable total cost in time, computational resources, and energy consumption needed for scientists to build and train these models. Sharing resources also facilitates validation of outputs, verification of results, and building directly on prior work. Basic research can also be a platform for experimental trials of new AI technologies, which could yield valuable insights for how (or whether) to responsibly deploy AI in other contexts.

The advancing capabilities of AI technologies in the sciences does introduce new potential risks, particularly concerning malicious applications. The enhanced capabilities also present novel challenges to and pressures on the scientific community—such as the risk of placing too much trust

---

[44] For a discussion of physics-informed machine learning methods in general, see Karniadakis, G.E. et al. (2021 May). Physics-informed machine learning. *Nature Reviews Physics*.
[45] Additionally, the cooling requirements of the data centers used by AI can lead to significant water usage.
[46] *Open weight* models disclose the weights obtained at the end of the training process, whereas "open source" models disclose the structure of the model and training process, but not necessarily the final weights. This distinction is particularly significant for generative pretrained transformer (GPT) models, for which the training process is extremely computationally intensive.
[47] Li, Y. et al. (2023 September). Textbooks Are All You Need II: Phi-1.5 Technical Report. arXiv.
[48] Chui, M. et al. (2023 June). The economic potential of generative AI: The next productivity frontier. *McKinsey Digital*.

on a convenient AI tool and not checking the results—which could impact established norms and principles of scientific integrity. Nonetheless, if AI tools are deployed in a responsible, human-supervised, and validated manner, the scientific benefits of these technologies can outweigh the risks. Indeed, since our global competitors and collaborators will certainly be developing applications of AI, the best way to mitigate its risks is to lead in the development of norms and best practices. The greater risk would be *not* seizing the opportunity to lead the world on developing and understanding these powerful tools, and applying them to our most pressing global problems.

2. A Vision for the Future of AI-Enabled R&D

Given the experimental and rapidly developing nature of scientific applications of AI technologies (particularly with regards to generative AI), it is difficult to make predictions about how and when the scientific workflow will evolve as AI techniques are incorporated. Nevertheless, we envision that the ideal future of AI-enabled science will require continued attention in three areas:

- *Empowerment* of human scientists;
- *Responsible use* of AI tools; and
- *Sharing* of basic AI resources.

As depicted in the accompanying figure, these three themes mutually reinforce each other. Sharing basic AI infrastructure, like access to time on high performance computing clusters, will enable scientists to work with and understand advanced AI tools, help set common standards for responsible use, and improve equity by providing access to researchers from all institutions rather than just those with the most resources. A culture of responsible use will encourage secure ways to share models and data, as well as promote thoughtful designs of AI-assisted research projects that enhance, rather than degrade, the quality of their scientific output. Finally, a diverse scientific

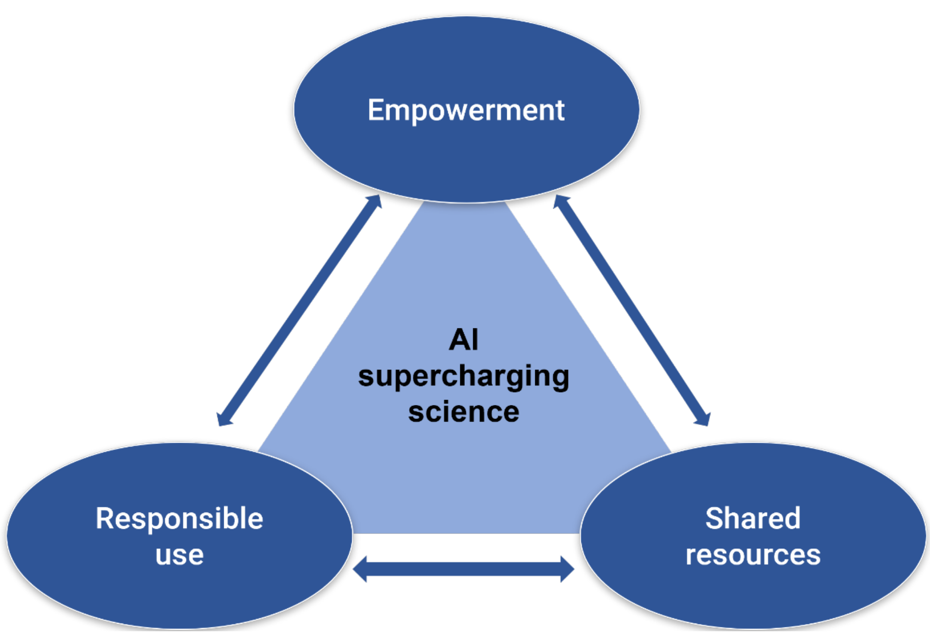

Figure 1. Virtuous cycle of responsible use, shared resources, and empowerment to leverage AI to accelerate scientific research. Multiple dimensions of this interplay are explored in this report.

community that is broadly empowered by AI tools will generate innovative new solutions to pressing challenges, justifying the initial investments in shared resources and in responsible use policies, as well as creating a community of stakeholders to continue to develop and build upon these investments. Our recommendations at the end of this report are designed to support and encourage all three of these themes.

AI methods will help researchers prioritize the most likely solutions

Many of the problems at the frontier of modern science are complex and interdisciplinary. As such it is becoming increasingly difficult for unaided human experts to sift through the large amounts of available data and analysis from all relevant scientific domains or to evaluate hundreds of thousands of candidates (e.g., compounds for medicine or materials for engineering applications) to identify the most promising solutions to a given problem.[49] Scientists are beginning to use machine learning tools, including generative AI, to address this gap. Researchers are now using data-driven models to isolate likely candidates for materials,[50] drugs,[51] and chip designs[52] to test in a laboratory or clinical trial—potentially saving enormous amounts of time and expense by reducing the number of tests with negative outcomes and leveraging the most value out of limited experimental resources. In the future, these tools could also suggest possible explanations for empirically observed phenomena, or uncover connections or analogies between two areas of science that would otherwise have gone unnoticed. Hypothetically, AI could even help scientists to discover new laws of nature, which could then be validated by more traditional theoretical calculations and laboratory experiments. It is important to be clear that we do not envision that AI-driven or AI-assisted reasoning will replace the uniquely human capabilities and genius to make inspired connections and conclusions. Rather, we expect that traditional forms of research will continue to play an essential role in the broader scientific enterprise for the foreseeable future.

By handling routine tasks, AI will allow scientists to focus on core research

Some of the most immediate productivity gains offered by AI will come not from directly assisting with the most difficult scientific research challenges, but rather from the more mundane support AI can offer with secondary tasks that can take up a large portion of the working time of a scientist, such as developing computer code, assisting with papers and reports,[53] performing literature reviews, and acquiring expertise in adjacent scientific fields. Already, general purpose large language models are

---

[49] Bloom, N. et al. (2020 April). Are Ideas Getting Harder to Find?. *American Economic Review*.
[50] Zeni, C. et al. (2024 January). MatterGen: a generative model for inorganic materials design. arXiv; See also the A-Lab and the Materials Project, both operated by Berkeley Lab.
[51] Mock, M. et al. (2023 September). AI can help to speed up drug discovery—but only if we give it the right data. *Nature*.
[52] E.g., Liu, M. et al. (2024 April). ChipNeMo: Domain-Adapted LLMs for Chip Design. arXiv.
[53] Currently, the output from existing large language models is not of sufficiently reliable quality to be acceptable for direct use in writing scientific documents, though it can already serve to create useful first drafts or experimental variants of such texts. Nevertheless, we expect more professional quality AI writing assistants to be integrated into many text editing platforms in the near future, and to eventually gain substantial adoption within the scientific community, in parallel with an updating of professional writing standards and guidelines that takes into account the capabilities and limitations of such AI assistants, for instance through the development of benchmark datasets for scientific writing, similar to recent benchmarks such as "LegalBench" in the practice of law. Furthermore, human authors should be held accountable for any errors in AI-generated writing output, and be expected to maintain full editorial control of content.

used routinely.[54, 55, 56, 57, 58] Nevertheless, we expect hands-on, non-AI-assisted research experiences to remain an essential component of the training of junior scientists for the foreseeable future.

Rote laboratory processes will be automated and improved

Many routine laboratory processes are ideal for labor saving with AI, allowing humans to spend their time and energy on the things we do best and find most interesting, such as design, analysis, and collaboration. A number of AI technologies, including object recognition, reinforcement learning,[59] and generative AI, are beginning to be used to control robotic systems, allowing them to process complex instructions in unpredictable environments and adapt to sensory feedback.[60] This flexible problem-solving capability will soon allow AI-powered robots to be used in the laboratory to perform a wide range of experiments, or synthesize a large number of materials, without the need to carefully reprogram the robot with each new task. Virtually every aspect of the laboratory workflow, from experimental design to data collection to data interpretation, could be partially or fully automated through AI assistance,[61] although we view expert human supervision of such automated laboratories to be essential and highly desirable for decades to come.

Previously intractable simulations will become possible

AI algorithms are successfully being used to greatly accelerate and enhance the computationally expensive computer models used to simulate complex systems, such as Earth's climate, the quantum chemistry of materials, and the intricate dynamics of proteins and cellular structures, reducing the need to always return to time-consuming modeling of these systems from first principles. For example, AI models are being used to provide more efficient approximations for solving the Schrödinger equation for chemical compounds, the solutions of which help to define compound

[54] Ray, P. (2023 April). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*.
[55] Huang, J. and Tan, M. (2023 April). The role of ChatGPT in scientific communication: writing better scientific review articles. *American Journal of Cancer Research*.
[56] Boiko, D. et al. (2023 December). Autonomous chemical research with large language models. *Nature*.
[57] It is however possible that a version of the "Jevons paradox" (or the "myth of the paperless office") will take hold, namely that the increased efficiency of writing afforded by AI tools is converted (in the near term, at least) into a greater amount of scientific writing being produced, rather than a reduction in time spent on writing tasks.
[58] Sellen, A. and Horvitz, E. (2023 November). The Rise of the AI Co-Pilot: Lessons for Design from Aviation and Beyond. arXiv.
[59] *Reinforcement learning* is a method of training algorithms to make desired actions by requiring the process to maximize a given function or result.
[60] Zhou, C. et al. (2021 November). A review of motion planning algorithms for intelligent robots. *Journal of Intelligent Manufacturing*.
[61] National Academies of Sciences, Engineering, and Medicine. (2022 May). Automated Research Workflows for Accelerated Discovery: Closing the Knowledge Discovery Loop.

stability and other properties.[62] Multiple efforts are now underway[63, 64, 65, 66, 67] to develop AI-powered "foundation models"[68] and "digital twins" for many applications. When complete, entire communities of researchers will be able to build on these powerful and broad platforms to rapidly create more customized models for a wide variety of scientific and engineering purposes. "Lightweight" versions of large models will also be developed—these will be small enough to run and fine-tune on individual computers, while still retaining much of the capability of the original model.[69] This will reduce the cost and environmental impact of deploying AI in scientific applications, and should enable wider, more equitable access to the models and resulting innovations. As previously noted, care should be taken to avoid reliance purely on AI models. Whenever possible, AI simulations should be validated and benchmarked against traditional methods of verification such as numerical simulations, theoretical calculation, and agreement with experimental data.

Shared models and data will reduce duplication of effort, democratize research, and reduce the total cost of using AI

For AI tools to be used effectively, high quality data sets must first be collected, made accessible, and put into a usable format. Next, models need to be trained on that curated data, taking care to limit algorithmic bias while also protecting the privacy of any humans whose personal information might be part of the data set.[70, 71] Repeating these steps for each individual research project is inefficient, wasting considerable time and resources. For example, instead of the multiple efforts currently underway to build large-scale foundation models for biology, it is feasible that these efforts could be combined or connected to focus resources and talent on model scale and quality. Through central resources such as the envisioned National Artificial Intelligence Research Resource (NAIRR),[72] researchers will gain access to standardized models and curated data sets, and share best practices

---

[62] Radu, A. and Duque, C. (2022 February). Neural network approaches for solving Schrödinger equation in arbitrary quantum wells. *Scientific Reports*.

[63] E.g., Mukkavilli, S. et al. (2023 September). AI Foundation Models for Weather and Climate: Applications, Design, and Implementation. arXiv.

[64] E.g., Houben, M. (2020 November). Digital Twins, the future in plant phenotyping – TechnoHouse by Rijk Zwaan. *Phenospex*.

[65] E.g., Costin, A. et al. (2023 November). Digital Twin Framework for Bridge Structural Health Monitoring Utilizing Existing Technologies: New Paradigm for Enhanced Management, Operation, and Maintenance. *Transportation Research Record: Journal of the Transportation Research Board*.

[66] E.g., van Willegen, B. et al. (2022 September). A review study of fetal circulatory models to develop a digital twin of a fetus in a perinatal life support system. *Frontiers in Pediatrics*.

[67] E.g., Geddes, L. (2023 November). How digital twins may enable personalised health treatment. *The Guardian*.

[68] A *foundation model* is an ML model trained (often at great computational expense) on a broad range of data, which can then be fine-tuned relatively cheaply for more specialized applications. For more discussion, see Bommasani, R. et al. (2022 July). On the Opportunities and Risks of Foundation Models. arXiv.

[69] Javaheripi, M. and Bubeck, S. (2023 December). Phi-2: The surprising power of small language models. *Microsoft Research Blog*.

[70] Dilmaghani, S. et al. (2019 December). Privacy and Security of Big Data in AI Systems: A Research and Standards Perspective. 2019 IEEE International Conference on Big Data, *IEEE Xplore.*

[71] Metcalf, J. and Crawford, K. (2016 June). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*.

[72] The National Artificial Intelligence Research Resource (NAIRR) Pilot. (Accessed 2024 April 9).

that will advance the fundamental science of AI.[73] A central resource would also facilitate the installation, fine-tuning, and operation of the aforementioned foundation models by researchers in a broad variety of domains without requiring significant levels of specialized AI expertise.

Once the right foundational infrastructure[74] is in place, AI-assisted research will become possible not only for highly resourced companies and research groups, but also smaller institutions and private sector organizations, or even members of the general public,[75] creating more equitable opportunities for discovering and developing innovative ways to utilize AI tools. However, for models and data that could have potentially harmful applications or which require privacy protections, some restrictions on access will be needed. The U.S. government is taking steps to facilitate safe and trustworthy AI. For instance, the recently released OMB memo outlines a set of minimum risk practices that federal agencies will be mandated to follow to mitigate risks of AI in rights- and safety-impacting contexts.[76] The NAIRR pilot is providing access to data, training, and compute resources to researchers who are conducting research on trustworthy machine learning. In addition, the recently formed U.S. Artificial Intelligence Safety Institute and Artificial Intelligence Safety Institute Consortium (AISIC) will bring together experts from across government, academia, and civil science to collaborate on AI safety research and develop resources.

Multimodal foundation models will bring together multiple forms of data and create new synergies among branches of science

Large-scale language models have been referred to as "foundation models" because they provide a basis or "foundation" for efficient refinement that transforms general AI models into AI systems for specialized tasks and domain-specific applications. Such refinement is usually done via learning mechanisms that use specialized, domain-specific data. The process of refining foundation models into high-performance models in specific domains is a type of "transfer learning."[77] In machine learning research, this term has been used for decades to refer to methods involving the adaptation of an ML model trained on one domain, e.g., inorganic chemistry, to harness its "pretraining" to improve the model's ability to learn in other domains, such as predicting the function of proteins and cells.

---

[73] National Artificial Intelligence Research Resource Task Force. (2023 January). Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource.

[74] Such infrastructure includes not only hardware and software resources, but also benchmarks, regulatory principles, and guidelines on the responsible usage of AI (such as the IRB Considerations on the Use of Artificial Intelligence in Human Subjects Research); and, most importantly, the human capital of experts in the use and deployment of AI in both public and private sectors.

[75] For certain "dual use" scientific applications, such as gain of function research for viruses, broadening access of AI-assisted scientific advances to the general public is not necessarily desirable, and will require some additional regulation and oversight. However, we see many areas of science where it would be beneficial to have increased participation and engagement with the public, and to not have the most powerful AI tools limited to only a small number of well-resourced groups, or to researchers based in other countries.

[76] The White House Office of Management and Budget. (2024 March). Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence.. [Memo M-24-10].

[77] E.g., Zhuang, F. et al. (2021 January). A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE, *IEEE Xplore*.

In recent work, researchers have been exploring the construction of multiscale[78] and multimodal foundation models that can take advantage of joint representations[79] learning as well as harnessing transfer learning. The aim of multimodal learning is to combine a diverse array of data sets, including data of different types, of different scales, and even from different fields of science.[80, 81] Beyond learning jointly from datasets containing multiple types of data, efficient methods are being developed for joining models that have been trained independently with data drawn from different scientific areas or foci. Work in this realm includes the use of adaptors: lightweight connector models that are trained to link two or more pre-existing models together. While multimodal models and associated capabilities can be constructed in different ways, we anticipate that developing AI tools that span multiple disciplinary realms and multiple spatial and temporal resolutions[82] will provide scientists with powerful emergent capabilities to describe or simulate complex systems. These capabilities may greatly exceed what can be accomplished by domain-specific models alone, and will open up rich new opportunities for interdisciplinary thinking and collaboration.

AI will help researchers do more with data

Through the demonstrated ability of AI to infer context and make use of complex, noisy, non-quantitative real-world data, such as natural language text, AI algorithms show great potential for automatically organizing, combining, and "cleaning"[83] the extremely large and heterogeneous data sets that underpin modern data-driven science, as well as identifying anomalies and uncovering important correlations and patterns within that data. AI tools are also providing new ways to achieve "superresolution" to enhance the quality of individual images.[84] In addition, in many fields AI tools are already being used routinely to help generate "synthetic" data sets that—when used responsibly—can greatly enhance the quality and predictive power of empirically generated data, protect the privacy of sensitive information in such data sets, reduce the risk of algorithmic bias, and help extrapolate from the underlying data to draw conclusions in broader domains.[85, 86, 87, 88] However, such synthetic data will need to be carefully and permanently labeled as distinct from data

---

[78] *Multiscale modeling* is a modeling strategy that uses multiple models at different scales simultaneously to describe a system.

[79] *Joint representations* are a means for machine learning models to learn from multiple types of data or features, such as images combined with text, in a unified manner.

[80] Fei, N. et al. (2022 June). Towards artificial general intelligence via a multimodal foundation model. *Nature Communications*.

[81] Li, Z. et al. (2024 February). MLIP: Enhancing Medical Visual Representation with Divergence Encoder and Knowledge-guided Contrastive Learning is a recent example of a multimodal model that learns from both the image and text of annotated medical images.

[82] Poli, M. et al. (2023 March). Ideal Abstractions for Decision-Focused Learning. arXiv.

[83] *Data cleaning* refers to the process of removing or repairing portions of a data set that are duplicated, incomplete, inaccurate, or unrelated, in order to improve the quality of that data set for analysis or training.

[84] E.g., Chen, H. et al. (2022 March). Real-world single image super-resolution: A brief review. *Information Fusion*.

[85] Laboratory for Information and Decision Systems. (2020 October). The real promise of synthetic data. *MIT News.*

[86] Savage, N. (2023 April). Synthetic data could be better than real data. *Nature Outlook: Robotics and Artificial Intelligence.*

[87] Jordon, J. et al. (2022 May). Synthetic Data—what, why and how? arXiv.

[88] Listgarten J. (2024 January). The perpetual motion machine of AI-generated data and the distraction of ChatGPT as a 'scientist'. *Nature Biotechnology*.

collected from real world observations, sensing, surveys, and experiments, to avoid contributing to the longer term problem of "data pollution."[89]

New forms of collaboration will emerge

A modern scientific research project is typically led by a small number of senior scientists directing a larger group of postgraduate researchers and students to perform more specialized subtasks. AI tools will automate, or at least assist with, many of these subtasks, and through their facility with natural language, greatly facilitate the way that researchers from different scientific backgrounds, levels of expertise, or primary languages can communicate with each other. Because of this, new paradigms for collaboration will emerge, such as AI-augmented experts, human-advised AI systems, hybrid AI techniques combining complementary AI technologies,[90] or large crowdsourced, decentralized, and/or highly interdisciplinary projects in which individual contributions are validated and collated through a combination of AI tools and more rigorous assessment methods.[91]

Beyond facilitating communications between scientists, AI assistance could allow lay individuals to provide input to cutting-edge research projects. For instance, the public could have opportunities to contribute directly and meaningfully to research in novel ways via specialized chatbots that could engage on demand in a scientifically accurate, accessible, and genuinely two-way fashion in which public comments fold back into hypotheses or formulation of outputs.[92] Even in a future where AI assistance becomes commonplace, we envision a continuing role for traditional forms of scientific research and engagement. Traditional research methods and approaches are an essential and complementary approach to AI-assisted science, providing qualitative context, vital proof through experimentation, crucial methods for training scientists so they understand the mistakes that can come from tools like AI, and, perhaps most importantly, supplying the intellectual diversity of the scientific enterprise that refines scientific thought.

Responsible AI practices will be integrated into research workflows

The scientific method inherently incorporates self-correction mechanisms through independent replication and review. In particular, the design of experiments has continuously evolved as science has matured, with standard procedures introduced over time to reduce bias, minimize harm to human or animal subjects, improve replicability, and avoid waste. A similar evolution will take place with the use of AI tools in research—but it will require that academic institutions, scientific professional organizations, and funding agencies commit to building a culture of responsible and ethical AI practice,[93] and help to make powerful models available to researchers to develop, test, and evaluate these practices. Scientific, academic, and government leaders must encourage best practices, such as scientists ensuring appropriate citations of AI model usage in their work and results, and sharing with their communities the details about the specific version of model used; these citations will be critical for facilitating replication and verification of results. PCAST also

---

[89] Ben-Shahar, O. (2019 September). Data Pollution. *Journal of Legal Analysis*.
[90] Marcus, G. (2020 February). The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. arXiv.
[91] For some emerging examples of such large-scale collaborations in the mathematical sciences, see Artificial Intelligence to Assist Mathematical Reasoning: Proceedings of a Workshop, National Academies of Sciences, Engineering, and Medicine. (2023 June)..
[92] E.g., PCAST. (2023 August). Advancing Public Engagement with the Sciences.
[93] E.g., The Hastings Center. (2024 April). AI Code of Conduct Draft is Released: Submit Your Comments.

anticipates advances in the development of attribution tools and affordances[94] that can provide scientists with an understanding of the sources in the training data, which is fundamental to the generation of output, so that prior research and results can be cited appropriately.

Through advances in the foundational computer science of AI, metrics are being developed to determine the quality and optimize the choice of AI training data sets, along with standardized methods to compensate for biases and omissions in such sets. PCAST expects that new AI architectures will be developed[95] that can offer comparable or superior performance to the largest existing models but at a fraction of their energy consumption and environmental impact. Sophisticated privacy protection methods[96, 97, 98] are beginning to be deployed to protect the privacy of sensitive personal data, such as medical information, that could otherwise leak into and out of an AI model. Hallucinations and other inaccuracies in AI models can be compensated for by refining methods that can "ground" AI outputs with real world sources, and to attach well-calibrated confidence levels to those outputs; we can also harness more rigorous validation methods, such as formal verification.[99] Models, data sets, and weights should be responsibly open sourced when feasible, to assist with replication and transparency,[100] with a preference for "explainable AI" models that can draw explicit links between their conclusions and their input data.[101] Best practices for maintaining human supervision of automated laboratory processes should also be established.

Through increased dialogue between the physical sciences, social sciences, and the humanities, we expect to achieve a greater understanding of the potential unintended consequences of deploying AI tools in research, and to develop standard processes to assess, prevent, minimize, and mitigate such consequences. Conversely, while there are significant concerns that the scientific literature is at risk of experiencing an influx of low-quality articles that are partially or even fully AI-generated, we also expect AI tools to be useful in upholding standards of academic conduct, for instance by detecting image manipulation, plagiarism, and missing citations.

Once the necessary AI infrastructure is in place, new scientific "moonshots" will become possible

When foundational resources such as the NAIRR are in place to provide access to computational power, secure data sharing services, open source (and open weight) AI models, and other key infrastructure, it will become possible to plan ambitious, complex, and large-scale "moonshot"

---

[94] The uses or purposes that a thing can have; the qualities or properties of an object that define its possible uses or make clear how it can or should be used.
[95] One potential future architecture is that of neuromorphic computing; see, e.g., Schuman et al. (2022 January). Opportunities for neuromorphic computing algorithms and applications. *Nature Computational Science*.
[96] E.g., Papernot, N. and Thakurta, A. (2021 December). How to deploy machine learning with differential privacy. NIST Cybersecurity Insights.
[97] Savage, N. (2023 April). Synthetic data could be better than real data. *Nature Outlook: Robotics and artificial intelligence.*
[98] Kaissis, G. et al. (2020 June). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*.
[99] Urban, C. and Miné, A. (2021 April). A Review of Formal Methods applied to Machine Learning. arXiv.
[100] National Academies of Sciences, Engineering, and Medicine. (2018 July). Open Science by Design: Realizing a Vision for 21st Century Research.
[101] E.g., see Xu, F. et al. (2019 September). Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. *Natural Language Processing and Chinese Computing*.

scientific projects involving multiple public and private partners. Examples of such projects could include a foundation model to simulate the complexities of the human cell, allowing for *in silico* (rather than *in vitro* or *in vivo*) study of diseases and experimental treatments; a detailed whole Earth model that uses both conventional and AI models to describe components of the Earth system while also being continually updated with highly multimodal real-time data; or a large collaborative effort to discover practical room-temperature superconductors through systematic collection, processing, and AI-assisted analysis of existing data and literature, together with automated laboratory synthesis and testing of viable candidates. While there are existing projects along these lines already being undertaken by individual research groups or organizations, the larger scale collaborations that could become possible with an infrastructure of shared AI resources would benefit substantially from economies of scale and simultaneously reduce duplication of effort. In the next section, we highlight some of these exciting opportunities for advances in discovery that will be made possible using AI.

3. Key Opportunities for AI to Supercharge Discovery and Address Global and Societal Challenges

The applications of AI tools in science are extremely diverse. In order to illustrate the potential for AI tools to address global and societal challenges, we describe seven representative examples of areas of scientific inquiry to serve as concrete case studies. Each of these fields has faced substantial barriers to progress, which, if overcome, could lead to discoveries that improve people's lives, mitigate global risks, or inspire us as humans. For each of these areas, we characterize the current state of the field, identify barriers to progress, provide examples of how AI is currently being leveraged, and describe the future we envision to be possible, along with potential risks that must be considered and resources that would be necessary for progress to be made. This list of examples is not intended to be comprehensive; there are many other fields of science beyond those mentioned here that are also likely to be transformed by AI. However, these vignettes illustrate the cross-cutting themes, broad range of possibilities, and critical needs that AI offers across the sciences.

3.1. A Phase Change for Materials Discovery

The use of generative AI for the discovery of new materials[102] is already proving to be crucial for driving economic expansion, improving health outcomes, and creating new materials critical for national defense. Historically, eras of major improvement in the human condition were powered by advances in materials science: bronze, iron, concrete, steel. Today, we live in an age of silicon, hydrocarbons, and nitrates. The near future may be an era of nanomaterials,[103] biopolymers,[104] and quantum materials.[105] Novel materials will be the basis for climate-friendly energy technologies, including improved batteries and energy storage, carbon capture, and hydrogen production. New biologics will improve health outcomes and open new pathways for care. Finally, AI-assisted R&D will open possibilities that previously only existed in the realm of imagination, such as room temperature superconductors or large-scale quantum computer architectures. Our future will be built on new materials; they will revolutionize our society.

---

[102] Snyder, A. (2023 December). An AI boost for developing new materials. *Axios*.
[103] PCAST (2023 August). The Seventh Assessment of the National Nanotechnology Initiative.
[104] PCAST (2022 December). Biomanufacturing to Advance the Bioeconomy.
[105] Stanev, V. et al. (2021 October). Artificial intelligence for search and discovery of quantum materials. *Communications Materials.*

Scientists have already had success in leveraging deep learning models for materials discovery.[106, 107] For example, interdisciplinary teams of researchers at private companies have used AI to develop designs for millions of novel materials, where nearly a half a million of those predicted were candidates likely to be stable enough for possible growth in the lab.[108] PCAST expects future advances in AI-assisted materials discovery to be capable of narrowing similarly large feature spaces[109] of candidate stable materials to isolate those that are most likely to achieve specified target properties.[110] AI has also been used to improve on existing materials; for example, AI tools can help scientists optimize material composition to reduce or eliminate potentially environmentally hazardous materials while maintaining performance. In addition, AI can incorporate processing parameters to optimize not only composition but also methods of material manufacturing that can increase efficiency, reduce waste, and lead to new reaction and processing pathways that are more sustainable.

NSF has made a $72.5 million investment to drive the design, discovery and development of advanced materials needed to address major societal challenges.[111] The Designing Materials to Revolutionize and Engineer our Future (DMREF) program will fund 37 new four-year projects that work across many directorates of the NSF in science, engineering, computation, including partnerships with the private sector; each employing deep learning and AI.[112] The Department of Energy has funded the Quantum Science Center at $115M over the next five years, which among other things supports AI-assisted discovery and design of quantum materials.[113]

AI methods could be particularly valuable in the search for new superconductors, materials which can conduct electricity so efficiently that there is little to no loss of energy. Superconductors are essential components for MRI machines, particle accelerators, certain experimental quantum computing technologies, and (in limited places) in our power grid for lossless energy transmission; but they currently have several undesirable features. First, all known practical superconductors must be cooled to cryogenic temperatures (-298 degrees Fahrenheit or less) using liquid helium, an impractical procedure involving an expensive and limited resource. Second, in contrast to conventional conductors such as copper, existing superconductors are not malleable and will lose their superconducting properties with damage. Third, they are very expensive to use, both in terms of the cost of their precursor materials and the effort required to engineer them into wires. Better superconductors—those that can work at more easily achievable temperatures, are easier to engineer into applications, and are cheaper—would be transformative. They would democratize access to MRI, lower energy costs through reduction of resistive losses in our power grid, and enable further electrification of our economy. These materials would also have applications in our

[106] Chen, C. et al. (2024 January). Accelerating computational materials discovery with artificial intelligence and cloud high-performance computing: from large-scale screening to experimental validation. arXiv.
[107] Zeni, C. et al. (2023 December). MatterGen: a generative model for inorganic materials design. *Microsoft*.
[108] Merchant, A. et al. (2023 November). Scaling deep learning for materials discovery. *Nature*.
[109] *Feature spaces* are collections of qualities or dimensions that are used to characterize data.
[110] E.g., see Zhang, Y. and Kim, E. (2017 May). Quantum Loop Topography for Machine Learning. *Physical Review Letters* for an example of applying machine learning techniques to identify materials exhibiting a topological quantum phase transition.
[111] National Science Foundation. (2023 September). NSF invests $72.5M to design revolutionary materials.
[112] National Science Foundation. Designing Materials to Revolutionize an Engineer our Future (DMREF). (Accessed 2024 April).
[113] Oak Ridge National Laboratory. (2020 August). ORNL, partners receive $115 million to establish Quantum Science Center.

transportation sector, for example enabling magnetically levitated trains that can travel with minimal friction for a smoother ride and greater efficiency—truly making science fiction come to life.

Another class of materials that scientists do not know how to design are thermoelectrics, which can convert heat, even waste heat from power transmission or engines, into energy, with many applications such as cooling and electronics.

Researchers have never been able to predictively design a superconductor or thermoelectric material, because these quantum materials require a unique composition of matter.[114] Previous efforts have relied on combinatorial chemical methods—empirical experimentation involving creating and screening vast numbers of material combinations—with limited success. Essentially, all discoveries of these critically important materials have been serendipitous, made by experimental trial and error. The sheer number of variables involved, as well as the need to keep such materials affordable, makes these materials discovery problems overwhelming, and near-impossible to solve by conventional methods.

There are three areas where AI tools will be a game-changer for materials science. First, the predictive abilities of AI modeling are enabling a new approach to materials discovery by connecting and utilizing the vast quantities of data available on existing materials, their processing conditions, and their properties. From this dataset, patterns across the chemistry, physics, and engineering of materials can be determined and combined in unique ways to provide researchers with insights and new approaches.[115] Second, AI models can predict performance (for instance, predicting the coherence time of a quantum bit, the efficiency of a thermoelectric material, or the critical temperature of a superconductor), thus reducing wasteful experimentation and testing of non-viable candidate materials.[116] Third, by combining process information with material composition, practical boundaries can be placed on the material design, accelerating scale up and commercial introduction of the new materials.

In addition to investigating "hard" materials like superconductors and thermoelectrics, AI is poised to revolutionize the development of "soft" materials like polymers and fluids. Materials discovery for soft materials requires the same vast datasets and predictive capabilities as hard materials as well as the complex structure-function relationships found in material science. Unfortunately, to date, the application of AI to polymer discovery and processing remains an emerging field with vast untapped potential—most likely the next frontier for materials science.

Looking further into the future, AI tools could lead to new or improved materials such as cold atoms,[117] topological insulators,[118] or superconductor-based qubits[119] that serve as building blocks for quantum computers, which would be able to perform certain large-scale computations that would require impractical levels of computation or energy if performed with traditional supercomputers.[120]

---

[114] Keimer, B. and Moore, J. (2017 October). The physics of quantum materials. *Nature Physics*.
[115] Liu, Y. et al. (2023 December). Materials Expert-Artificial Intelligence for Materials Discovery. arXiv.
[116] E.g., Biamonte, J. et al. (2017 September). Quantum machine learning. *Nature*.
[117] E.g., Castelvecchi, D. (2023 June). IBM quantum computer passes calculation milestone. *Nature News.*
[118] E.g., Breunig, O. and Ando, Y. (2021 December). Opportunities in topological insulator devices. *Nature Reviews Physics*.
[119] E.g., Ballon, A. (2024 March). Quantum computing with superconducting qubits. *PennyLane.*
[120] McKinsey & Company. The Rise of Quantum Computing. (Accessed 2024 April 9).

While in recent years major progress has been made to realize small-scale quantum computing devices, these systems are unstable, susceptible to noise, and difficult to scale to the size and complexity required to compute solutions to useful problems. Nevertheless, the potential applications of quantum computing have drawn substantial international interest and funding commitments. AI assistance could be a key factor in determining the rate of progress in this emerging technology.[121]

Overall, AI tools offer the opportunity, which we are starting to realize today, to usher in a new era of materials discovery and in doing so, accelerate economic growth and a new future.

3.2 AI for Designing Advanced Semiconductors

The modern electronic devices that underpin the global economy and our national security run on integrated circuits etched onto small pieces of semiconductor material, commonly known as "chips". As these chips become more capable, they also become much more complex, with current state-of-the-art chips now containing tens of billions of components. Today, only the largest corporations can afford to fabricate the most advanced chips, because of the significant engineering resources and infrastructure required to design chips.

Revitalizing the U.S. semiconductor ecosystem has been a significant priority for this Administration.[122] Broad usage of AI in the design of future chips can significantly increase the quality and reduce the time and number of engineers required to design the most advanced chips. For example, an aspirational goal for the U.S. semiconductor industry would be to create platforms, methodologies, and tools to enable chips to be built using a tenth of the person-hours that are required today, which would vastly lower barriers of entry to the semiconductor market, encourage innovation with a larger and more diverse set of participants, and continue to broaden our lead in semiconductor design in the world.

Several AI assistants for chip designers now exist.[123, 124] These types of tools can allow a junior designer to ask questions that previously would have consumed the time of senior designers. Some chip design AI assistants can also summarize bug reports and other design documents, or generate scripts for other design automation tools to run, all from simple English-language prompts. These AI tools do not replace designers, but instead empower designers to be considerably more productive, helping to mitigate the overall shortage of trained chip designers.

Despite circuit design being a mature field, there are AI tools available and under development that promise to provide surprising design improvements.[125, 126] Some of these tools are able to generate circuits that are faster or smaller than the best circuits designed using conventional methods. One

[121] Hart, B. et al. (2023 August). Is China a Leader in Quantum Technologies? *China Power*.
[122] PCAST (2022 September). Revitalizing the U.S. Semiconductor Ecosystem.
[123] Verma, P. (2024 March). DSO.ai – A Distributed System to Optimize Physical Design Flows. ISPD'24 Proceedings. ISPD'24: Proceedings of the 2024 International Symposium on Physical Design.
[124] Liu, M. et al. (2023 October), ChipNeMo: Domain-Adapted LLMs for Chip Design. arXiv.
[125] Roy, R. et al. (2021 December). PrefixRL: Optimization of Parallel Prefix Circuits using Deep Reinforcement Learning. 2021 58th ACM/IEEE Design Automation Conference (DAC).
[126] Budak, A. et al. (2022 February). Reinforcement learning for electronic design automation: Case studies and perspectives. 2022 27th Asia and South Pacific Design Automation Conference (ASP-DAC), *IEEE Xplore*.

mechanism for improving the performance of circuit design AI assistants is a technique known as reinforcement learning. As the AI tool explores a "state-space" of possible circuits receiving reinforcement via positive "reward" and negative "punishment" values for generating good or bad circuits, respectively. It changes its approach in response to these rewards, ultimately learning which tactics lead to those circuits with desirable features.

For each new semiconductor dimension, a library of thousands of "cells"—the standard design building blocks of a modern chip—must be redesigned to meet the constraints of the process. For many manufacturers, this task can require about 80 person-months of effort. Using a combination of AI methods, including generative AI for clustering and reinforcement learning to fix design-rule errors, a recently developed tool[127] is able to automate the generation of a new library and reduce the effort by a factor of over a thousand. In a similar manner, a machine-learning based "floorplanning" tool (used to determine optimal location, shape, and size of components on a chip) uses reinforcement learning to reduce design time and improve the quality of layout for placing these standard cells on a chip.[128] In addition, field-programmable gate arrays enable fast iterations on the latest AI-driven placement and routing[129] techniques demonstrating over 3x relative improvements in efficiency.[130]

As a chip design is created, it is subject to a number of analyses to verify whether the design meets its specifications and the constraints of the manufacturing process. AI has been applied to speed up a number of these analyses as well. For example, predicting timing before detailed routing is performed,[131] or predicting undesirable "parasitic" features of a circuit.[132] Using AI tools such as these, a designer can iterate through many circuit ideas quickly. In the past, for instance, a layout for the circuit had to be generated to get accurate understanding of parasitic features—often adding days of manual effort to each iteration of the design cycle. Now the entire design iteration loop can be completed in a few minutes to obtain a circuit that meets the desired specifications.

As we move forward, we expect large language models (LLMs) to evolve into design assistants that answer questions, critique and verify designs, and carry out routine design tasks. We also expect AI techniques to significantly elevate the productivity of designers, potentially by a factor of ten or more. Designers will work at the algorithmic and system level; AI assistants will then work out the details at lower levels of design. AI synthesis and analysis tools will greatly shorten the design cycle, allowing a design to be carried from a high-level description to verified layout in a few hours, compared to the weeks it takes today.

---

[127] Ho, C. et al. (2023 March). NVCell2: Routability-Driven Standard Cell Layout Advanced Nodes with Lattice Graph Routability Model. ISPD '23: Proceedings of the 2023 International Symposium on Physical Design
[128] Mirhoseini, A. et al. (2021 June). A graph placement methodology for fast chip design. *Nature*.
[129] *Routing* is the process of optimal path selection in any network between interconnected nodes.
[130] Bustany, I. et al. (2023 October). The 2023 MLCAD FPGA Macro Placement Benchmark Design Suite and Contest Results. 2023 ACM/IEEE 5th Workshop on Machine Learning for CAD (MLCAD), *IEEE Xplore.*
[131] Chhabria, V. et al. (2023 October). A Machine Learning Approach to Improving Timing Consistency between Global Route and Detailed Route. arXiv.
[132] Ren, H. et al. (2020 July). ParaGraph: layout parasitics and device parameter prediction using graph neural networks. DAC '20: Proceedings of the 57th ACM/EDAC/IEEE Design Automation Conference.

PCAST expects that by integrating emerging techniques like these into the chip making process, the U.S. will maintain its position as the leader in semiconductor design while relieving a critical workforce shortage in this area.

3.3. Understanding and Addressing Climate Change and Extreme Weather

Each year in the United States, climate disasters such as hurricanes, wildfires, and floods cause hundreds of deaths and tens of billions of dollars of damage, and severely impact the physical and mental well-being of those individuals and communities forced to relocate or rebuild. While rare, the most extreme of these weather events are responsible for the majority of these costs. For instance, in 2023, the U.S. experienced damages of $92.9 billion from 28 separate weather and climate disasters, including a drought and heat wave in the South and Midwest in the spring and fall ($14.5 billion) and severe weather in the south and east in early March ($6.0 billion).[133] Many of these extreme events are becoming more frequent and are predicted to become even more so with climate change, with millions of Americans projected to be at risk of being displaced from their homes by such events by the end of the century.[134, 135, 136]

In order for individuals, communities, and governments to prepare for these events and mitigate their impact, it has become increasingly important to obtain accurate and detailed models for both weather and climate, in order to predict the trajectory of extreme weather events as they occur in real time, as well as projections of future climate risk over the longer term. Ideally, we would like to have a "crystal ball" that would show us what strength of wildfire, flood, heatwave, or hurricane one could expect to affect any given location in the next ten, twenty, or fifty years.

Until recently, the best "crystal balls" available were supercomputers running huge weather and climate models that are not using AI. These older models cover the Earth in a virtual grid with millions of grid points, each one simulated for millions of time steps. But these "crystal balls" are slow, blurry, and uncertain. For instance, the Global Forecast System of the U.S. National Weather Service runs four times a day in order to produce weather forecasts up to 16 days in advance, but with a spatial resolution of about 13 km, and with input data assimilated in a six hour cycle.[137] For longer term climate modeling, one state-of-the-art example is the Energy Exascale Earth System Model supported by the Department of Energy, which when fully operational could produce one possible climate outcome at a 3 km resolution for the next ten years after some weeks of supercomputer time.[138] Although this is an impressive improvement, supercomputer models that do not use AI only produce a few possible outcomes among many, allowing only a limited exploration of uncertainties, and do not pinpoint precisely where in a given grid square the extreme events are most likely to occur.

---

133 Smith, A. (2024 January). 2023: A historic year of U.S. billion-dollar weather and climate disasters. Climate.gov.
134 U.S. Global Change Research Program. (2023 November). Fifth National Climate Assessment.
135 Emmanuel, K. (2017 March). Will Global Warming Make Hurricane Forecasting More Difficult? *Bulletin of the American Meteorological Society*.
136 Bhatia, K. et al. (2019 February). Recent increases in tropical cyclone intensification rates. *Nature Communications*.
137 National Centers for Environmental Information. Global Forecast System (GFS). (Accessed 2024 April 11).
138 Singer, N. (2023 April). Cloud-resolving climate model meets world's fastest supercomputer. *Sandia LabNews*.

The ability of AI to directly learn from data will enhance these "crystal balls," allowing them to become much faster and sharper. An AI can be trained on data about tornadoes in Oklahoma, and it will learn how to model tornadoes in Ohio. If a scientist trains an AI on local climate and weather data, it can be used to upgrade a low-resolution climate model into a high-resolution model (for example, at 1 km resolution or less) that takes the local geography into account, a procedure known as "downscaling."[139] Once the AI model is trained, this process is blindingly fast—thousands of times faster than traditional simulations, and thus also cheaper to run in terms of computational costs and energy usage. AI-based downscaling of coarser climate simulations has the potential to generate hundreds or thousands of possible future outcomes of processes that cannot be resolved by the coarser models, such as tropical cyclones and storm surges, allowing a quantitative assessment of climate and weather hazards.

In the future, any citizen (or a local government) will be able to enter their ZIP code into a government climate portal[140] and obtain multiple detailed scenarios of expected climate changes and potential weather disasters in their area in the next ten, twenty, or fifty years. These data can then be combined with data on the built environment to estimate the economic impact of these events. This information will be invaluable for designing appropriate building codes, emergency response preparations, insurance policies, and city planning. With regards to climate change mitigation efforts, these fast AI climate model emulators will also enable scientists to efficiently compare the outcome of many different mitigation strategies, allowing researchers and policymakers to perform an informed cost-benefit analysis to achieve the greatest climate impact with the least cost and disruption.

The speed of AI weather simulation models also makes them very suitable for real time prediction of extreme weather events, such as wildfires or hurricanes, though at the current state of the art, these models still must rely on more traditional forecasting models to provide initial data assimilation.[141] For instance, a physics machine learning model, when provided with initial conditions from a traditional model, is able to provide global week-long forecasts of weather variables at a 31 km resolution in a matter of seconds, with an accuracy comparable to base models, and is particularly suited for forecasting such extreme events as hurricanes and atmospheric rivers.[142] The European Centre for Medium-Range Weather Forecasts (ECMWF) is beginning to incorporate such AI enhancements to their integrated forecasting system on an experimental basis.[143] If provided with sufficient satellite data on historical wildfires, PCAST also expects machine learning models to be able to predict the trajectory of wildfires in real time with high accuracy, saving the lives of both residents and firefighters.[144] These models will not replace traditional supercomputer simulations, or the real-time acquisition of weather and climate data; but they will greatly enhance and complement these basic components of climate science, with the slower but more reliable traditional simulations being

---

[139] See PCAST (2023 April). Extreme Weather Risk in a Changing Climate: Enhancing prediction and protecting abilities.
[140] Climate Mapping for Resilience & Adaptation. (Accessed 2024 April 11).
[141] European Centre for Medium-Range Weather Forecasts. (2023 September). How AI models are transforming weather forecasting: a showcase of data-driven systems.
[142] Kurth, T. et al. (2023 June). FourCastNet: Accelerating Global High-Resolution Weather Forecasting Using Adaptive Fourier Neural Operators. PASC '23: Proceedings of the Platform for Advances Scientific Computing Conference.
[143] European Centre for Medium-Range Weather Forecasts. Charts. (Accessed 2024 April 11).
[144] PCAST (2023 February). Modernizing Wildland Firefighting to Protect Our Firefighters.

used to initialize, validate and benchmark the faster, but sometimes inaccurate, AI models. While perfect accuracy in weather prediction and climate modeling may not be physically attainable even with the most sophisticated models, being caught unprepared and off guard by extreme weather events will become a thing of the past.

3.4 Revealing the Fundamental Physics of the Universe

Perhaps the most fundamental questions that every human being might wonder is, “Where did this all come from? How did the universe start? What are the rules that make it look like it does and led to us being here today?” Such questions can unite us in awe of the beauty of the world around us, and in the excitement of each advance in our human understanding of its workings. We come together to watch the intricate, darkly luminous images from the Hubble or Webb space telescopes, and, even in our lifetimes there have been major advances in our understanding of an expanding “Big Bang” universe (it now appears to be speeding up in its expansion!) The cosmologists and particle physicists that explore these questions are some of the earliest adopters—and developers—of AI, so an epoch of advanced AI is an epoch of exciting discoveries in fundamental physics and cosmology.

Not only do scientists now have new opportunities to make discoveries using AI, but we know there are discoveries waiting to be made. Our strikingly detailed and surprisingly well-tested picture of the universe has revealed some remarkable puzzles: What is the "dark matter" holding galaxies together? Or the "dark energy" that is accelerating the expansion of the distances between all galaxies? What is the significance of the recently observed galaxies that appear old so soon after they could have formed? Such fundamental questions about our universe are our primary route to deep insight about how it functions. Each time we understand more of these fundamentals, we can delight and inspire a next generation of learners. Crucially for our practical goals and needs, it is these new fundamental understandings that have also allowed us to leapfrog over technological gaps to open up new ways to shape our world that allow us to thrive. For example, it is hard to imagine a more abstract, impractical-sounding fundamental theory than General Relativity; yet it underpins GPS, solving location and navigation problems in a way we could not have expected—with economic benefits reckoned in the hundreds of billions of dollars.

What does AI make possible for this challenge? Physicists and cosmologists are using AI throughout their research—it has quickly become a feature of most steps of the experimental and observational efforts, their design, implementation, and analysis. Some of the uses of AI build on the current approach to comparing and testing theories against data by computationally simulating what the data would look like if different theories were correct. These simulations can be some of the most challenging supercomputer tasks, as they calculate each step of the behaviors of each of myriad particles, stars, or galaxies. One key use of AI is to learn the larger patterns hidden in such simulations, so that scientists can shortcut these supercomputer tasks, making it possible to see in less than a minute an excellent approximation to a month of supercomputer work.[145, 146, 147] Now

[145] Dai, B. and Seljak, U. (2021 April). Learning effective physical laws for generating cosmological hydrodynamics with Lagrangian deep learning. *Proceedings of the National Academy of Sciences*.
[146] de Oliveira, L. et al. (2017 September). Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis. *Computing and Software for Big Science*.
[147] Mustafa, M. et al. (2019 May). CosmoGAN: creating high-fidelity weak lensing convergence maps using Generative Adversarial Networks. *Computational Astrophysics Cosmology.*

researchers can scan through millions of possible theories, each with different starting pictures of our universe, to see which one would lead to the data we are in fact observing with a telescope.[148]

What major results could we see in this AI-accelerated domain? By the late 2030s, we will be analyzing a decade of data from the Nancy Grace Roman Telescope using such AI approaches. AI-assisted analysis of these data might make it possible to find, for instance, surprising evidence that our universe will not end in the cold death of exponential expansion (as many cosmologists currently assert), but instead is actually cyclically reinitiated with repeated new Big Bangs.

Other uses of AI in physics take advantage of its ability to find patterns in extremely complex datasets, where the number of variables is more than we can easily track. Then researchers can look for surprises: events that are extremely rare indicators of new discoveries that break the usual rules, so they stand out amidst these AI-identified regularities. Particle physicists have held competitions to find the best approaches to searching for these "anomalies" that can point us to new physics discoveries.[149, 150] The winners have been AI-based and have themselves also pushed the boundaries of AI techniques.[151]

These AI approaches might allow us to find some extremely rare, unanticipated particle in the next generation of CERN and Fermilab accelerator experiments (perhaps even by the 2040s) that would help build the long-sought Theory of Everything[152] that would combine gravity with the other forces.

Fundamental physics and cosmology are built on statistical analyses of data to test theory, so they require a deep understanding of the probabilities in the interpretation of data. This requirement is driving the mathematical development of AI that can handle probabilistic rigor. Rather than just provide a most likely answer ("that is a photograph of a cat"), which may be right or wrong, the goal is to develop AI systems that can provide a range of possible answers and the probability that each one is correct ("69% odds that that is a cat, 22% that it is an aardvark, 8% that is a blimp, and 1% that it is a refrigerator"). For a measurement of a key number, it would provide a range of possible values that are, say, 68% likely, 95% likely, or 99.9% likely. Assessing uncertainties is crucial for fundamental physics, and probabilistically rigorous AI would be a game changer for many other fields of science as well, in addition to being invaluable for applications beyond science. Ideally, the AI tools we envision would therefore also know that its "hallucinations" are very unlikely to be correct. In a similar vein, physicists and AI researchers have begun exploring the development of AI that can offer

---

148 Cranmer, K. et al. (2020 May). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*.
149 Kasieczka, G. et al. (2021 December). The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics. *Reports on Progress in Physics*.
150 Dai, B. and Seljak, U. (2024 February), "Multiscale Flow for robust and optimal cosmological analysis. *Proceedings of the National Academy of Sciences*. This same approach is being pursued in many other areas of science, as mentioned above with respect to the simulations of novel materials and novel proteins for drug discovery. The difference here is that in the previous examples we are simulating known processes, whereas in cosmology and physics we can be exploring unknown processes.
151 E.g., Stein, G. et al. (2020 December). Unsupervised in-distribution anomaly detection of new physics through conditional density estimation. arXiv.
152 Pultarova, T. (2022 December). The Theory of Everything: Searching for the universal rules of physics

results that are constrained by known physical principles.[153, 154]

There are several opportunities related to fundamental physics and cosmology that arise because of the capabilities of even currently available generative AI models to translate between the jargon of different fields. Perhaps most obviously, the excitement and richness of our understanding of the universe will become increasingly accessible to ever-larger numbers of people, since it is possible to ask questions about physicists' current understanding and get an answer at an appropriate level to take the questioner to the next step. Clearly, this is a valuable route into this field for the public. But even for the experts in a subdomain of physics or cosmology, these capabilities allow access to other neighboring (or more distant) fields, offering the possibility of greatly accelerating the work that ties them together. Such cross-disciplinary efforts are often the most fertile paths to scientific advances.

Anticipating what such AI-enabled interdisciplinary research would discover is, of course, a challenge to the imagination. Perhaps in 20 years, researchers might see such exciting results as the establishment of a parallel between an on-chip quantum-computer device and a black hole, opening up a novel bench-top test of general relativity—and a powerful new chronometric technology[155]. What scientists will really discover will undoubtedly be very different—but perhaps even more surprising.

3.5 Studying Human Behavior, Organizations, and Institutions

Over the last two decades, large and detailed datasets have allowed social scientists to better understand issues such as economic mobility, demographic disparities, market power, and the labor market effects of offshoring and automation.[156] The findings have captured public attention and influenced policy debates. A public meeting of PCAST in September 2023 showcased this social science "data revolution."

AI has the potential to unleash a new wave of discovery and advances in the social sciences. Already, social scientists are embracing many of the tools of machine learning.[157] This includes natural language processing for text analysis, data-driven model selection, and the integration of machine learning prediction methods with decision protocols.[158, 159, 160] Practical examples include the design

[153] Shanahan, P. et al. (2022 September). Computational Frontier Report of the Particle Physics Community Planning Exercise. arXiv
[154] Karniadakis, G.E. et al. (2021 May). Physics-informed machine learning. *Nature Reviews Physics*.
[155] *Chronometric technologies* are methods and tools that aid in measuring the history and future of the universe according to cosmology.
[156] E.g., see "Manifesto Research on Political Representation" dataset.
[157] Athey, S. (2019 May). The Impact of Machine Learning on Economics. *National Bureau of Economic Research*, provides an overview from the perspective of economics, including many observations that apply more broadly across the social sciences.
[158] E.g., Rho, E. et al. (2023 May). Escalated police stops of Black men are linguistically and psychologically distinct in their earliest moments. Proceedings of the National Academy of Sciences.
[159] E.g., Gentzkow, M. et al. (2016 July). Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech. *Stanford Institute for Economic Policy Research*.
[160] E.g., Lerner, J. et al. (2023 March). Financial Innovation in the 21st Century: Evidence from U.S. Patents. *Social Science Research Network*.

of experimental platforms on internet sites,[161, 162] or high-dimensional regression analysis for credit scoring or the pricing of financial securities.

A booming area of study brings machine learning together with causal inference—the methods that social scientists use to infer cause and effect.[163] Machine learning in social science promises to provide more detailed, and often individualized, estimates on how medical treatments, education programs, and other interventions or public policies affect different populations. Already, these new models are being adopted in practice, and in some cases refined—for example, by internet platforms to highlight information or marketing offers personalized to the individual level. The debate about the power and effects of these algorithms, especially on social media platforms, highlights some of the complexities raised in applying AI tools in ways that interact with human and social behavior.

Generative AI methods may allow us to reimagine what qualifies as "data" in studying human behavior, organizations, and institutions. As an example, social scientists appreciate the importance of many aspects of people's well-being—such as their identities, the pride they feel in their work, the way organizations and communities function—that are not always possible or practical to measure at scale through traditional survey methods. However, these beliefs, attitudes, and interactions are reflected in what people say, write, and do, representing a whole new category of data that researchers are only beginning to explore. The emerging AI models can organize and make sense of exactly these types of unstructured data—like the words people use to describe their city or community in a social media post—at large scale. The opportunity to apply AI to unstructured language data creates an enormous opportunity to incorporate qualitative data and research methods in these studies and relate them to more conventional measures of well-being such as income, productivity, educational attainment, or longevity.

AI models also have significant potential to generate better information on which to base government policy.[164] Data on key variables such as inflation, productivity, employment, and GDP are backward-looking and have well-known flaws. Many of these measures are survey-based: however, surveys take time, and obtaining a high response rate is challenging. At the same time, a vast amount of relevant data exists in real-time and goes unused because it is unstructured, incomplete, or inconsistent over time. While AI models cannot magically correct for incorrect or biased data, they may dramatically improve our ability to make predictions from unstructured and incomplete data, providing better real-time data for input to policy decisions for all levels of government. Further, these same arguments apply much more broadly beyond federal policy, to issues encountered and decisions made by a wide range of other large and complex organizations.

Another promising use of AI models is to improve the delivery of federal programs. Many federal programs involve highly manual processes on the part of both federal employees and citizens, for instance filling out forms or being on the receiving end of large quantities of paperwork. There are many opportunities to use AI methods to deliver government services faster, more effectively, with

---

[161] E.g., Mastodon Instance Comparison Experiments.
[162] E.g., Social Science Prediction Platform
[163] Leist, A. et al. (2022 October). Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. *Science Advances*.
[164] Coglianese, C. and Lehr, D. (2019). Transparency and algorithmic governance. *Administrative Law Review*.

improved accessibility options, and in a more personalized manner.[165] Recent work has highlighted one concern with automating different processes, namely that algorithms may be biased in ways that reflect their underlying data; for example, a computer vision algorithm trained primarily on one population may be less accurate in recognizing another.[166] However, social science researchers also are exploring ways in which AI-assisted decision-making might reduce biases and yield fairer outcomes, for instance in parole decisions.[167] One recent study revealed how the algorithms governing IRS audit selection could be modified to reduce the disproportionate auditing of Black taxpayers.[168] This emerging field of research takes an algorithmic approach to thinking about the implementation of policies that must be applied consistently millions of times.

Scientists are still in the early days of applying AI in social science research, but it is already possible to see that we are entering a tremendously exciting period. PCAST is currently studying the steps the federal government might take to support future advances in the social sciences. These may include: (1) promoting secure access to federally collected data, including administrative data, that is the foundation for much social science research; (2) supporting larger-scale "team" social science that is rapidly becoming the norm to analyze large, complex datasets; and (3) encouraging federal agencies to collaborate with social scientists to make policy design and implementation more evidence-based, more efficient, and more sensitive to the subtleties of human behavior.

3.6 Advancing Fundamental Understandings in the Life Sciences

Advances in AI show exciting promise for supercharging scientific advances in the life sciences, enabling breakthroughs in our understanding of biology with deep implications for healthcare, agriculture, energy, and materials. The harnessing of machine learning in the biosciences is already leading to unprecedented advances. These recent advances are only glimmers of the possibilities ahead—and the breakthroughs that can be expected on the horizon. PCAST believes that AI-powered tools, analyses, and results will fundamentally shift how we explore and understand the very building blocks of life and the composition of those building blocks into living systems with broad applications from agriculture to medicine.

We expect that AI-driven technologies will come to play a role in the daily life of bioscientists. Capabilities include harnessing AI methods as computational microscopes, providing scientists with new capabilities to see and understand the structure and function of proteins and other molecules underpinning the functioning of biological systems, as simulators of the activity of living systems, and as generators and design tools, for proposing and refining new candidate molecules.

Unraveling the mysteries of cellular functioning

Deciphering the intricate workings of cells—the fundamental units of life—has been a centuries-long quest. Biologists have been challenged by the sheer complexity and interconnectedness of cellular

---

[165] Engstrom, D. and Ho, D. (2020 July). Algorithmic Accountability in the Administrative State. *Yale Journal on Regulation*.
[166] Buolamwini, J. and Gebru, T. (2018) Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*.
[167] Ludwig, J. and Mallainathan, S. (2021). Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System. *The Journal of Economic Perspectives*.
[168] Elzayn, H. et al. (2023 January). Measuring and Mitigating Racial Disparities in Tax Audits. Stanford Institute for Economic Policy Research.

mechanisms. AI is beginning to provide scientists with powerful tools to help with unraveling the mysteries of viruses, cells, and organisms.

Starting at the foundations of biology, AI has already provided new lenses on proteins, the intricate molecular machines that drive a vast array of cellular processes. For decades, identifying the structure of proteins has been a costly, tedious task. Several AI-powered protein folding prediction systems,[169, 170] for instance, harness machine learning to predict the structure of millions of proteins. These systems learn from data about known proteins and structures, as well as from foundational chemical knowledge such as physical constraints on distances among atoms. The tools for efficiently identifying structure are supercharging further bioscience advances. Moving beyond structure, researchers have more recently harnessed the power of AI to decipher the function of proteins, including how proteins interact with one another, providing insights into the molecular mechanisms that drive cellular signaling, metabolism, and gene regulation.[171] AI tools are also being used to design proteins, given goals of interacting in specific ways with receptors and other targets.[172] AI-powered protein design is already showing successes with creating vaccines and new kinds of therapeutics. Some of the design methods employ notions of "diffusion modeling" and inpainting and outpainting first demonstrated in AI-powered image generation systems.[173]

AI tools are now pushing further towards the horizon, enabling bioscientists to build and study rich representations of thousands of interactions that define networks of activity that govern cellular behavior. In one approach, researchers have used an AI methodology known as graph neural networks to directly learn the mesh of interactions among proteins and medications—and take the transformative next step to identify links to physiologic responses.[174] By unraveling the intricate mechanisms that govern details of cellular functioning, researchers can gain a deeper understanding of the molecular basis of diseases, the operation of therapeutics, medication side effects, and other interventions, paving the way for the development of more effective and targeted therapies. Additionally, the ability to model and simulate larger cellular systems "*in silico*" opens up new avenues for building insights about health and disease.

Building and leveraging bioscience foundation models

A promising approach to building rich *in silico* experimentation tools for the biosciences is to construct multimodal, multiscale bioscience foundation models aimed at whole-cell modeling. AI methods allow scientists to construct multi-modal biological representations or "embeddings" of diverse forms of data, including protein sequence and structures, DNA, RNA expression data, phenotypic data such as clinical observations, imaging data, and data drawn from electronic health records. These rich bioscience models can provide insights at multiple scales, from molecular interactions to cellular processes to the health or illness of whole organisms. Building cellular simulations may seem far out on the frontier of possibility, but we are seeing initial strides in this

---

[169] Jumper, J. et al. (2021 July). Highly accurate protein structure prediction with AlphaFold. *Nature*.
[170] Baek, M. et al. (2021 August). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*.
[171] Humphreys, I. et al. (2021 November). Computed structures of core eukaryotic protein complexes. *Science*.
[172] Krishna, R. et al. (2024 March). Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science*.
[173] Wu, K. et al. (2024 February). Protein structure generation via folding diffusion. *Nature Communications*.
[174] Zitnik, M. et al. (2018 June). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*.

direction, such as the development of the foundation model Evo that integrates large datasets, combining DNA, RNA, and protein data to elucidate interactions that orchestrate the overall functioning of cells.[175] This multimodal and multiscale model can provide both prediction of outcomes and generation of molecules and behaviors, spanning a range in scale from atoms to physiologies.

Bioscience foundation models are poised to become commonplace tools, enabling researchers to run virtual experiments via perturbing aspects of the system and noting the effect of medications and other interventions. We expect that these AI systems will enable explorational efforts within the computational realm, such as studies of genetic engineering to inactivate or remove one or more specific genes, drug treatments, or environmental stressors. These new tools will enable scientists to more quickly probe the foundations of health and disease, for instance by building models of cancers and exploring how cellular interactions and networks that underlie cancers might be disrupted or "cured" in simulation. These computer-based studies will guide the identification or development of medications by enabling virtual screening and optimization of potential therapeutic compounds before embarking on costly and time-consuming experimental validations. The systems will be able to provide scientists with guidance on ideal experiments to run, guided by expected value of the in-world studies to reduce key uncertainties on the way to key results and objectives.

New powers and responsibilities

We anticipate significant advancements in the field of life sciences through the use of AI-driven methodologies. These innovative techniques are poised to aid life scientists in unraveling countless biological mysteries, paving the way for unprecedented capabilities in modulating cellular machinery. Nonetheless, it is imperative to exercise diligence and assume responsibility to safeguard against potential misuse. Particularly, as we delve into realms such as AI-powered protein design, measures must be in place to prevent the creation of biological threats, whether such creation is inadvertent or driven by malicious goals. Establishing and adhering to best practices encompassing screening, monitoring, and prompt response mechanisms will be important in mitigating inadvertent or malicious uses of AI, which could result in the generation of new toxins or enhancements of the virulence of infectious diseases.

3.7 Breakthrough Applications of AI in the Life Sciences

We expect significant breakthroughs in multiple areas of the life sciences. We foresee AI tools enabling key developments in agriculture, energy, sustainability, and healthcare.

Advances in AI and agriculture

New AI-based tools allow understanding, modulating, and engineering of processes that will have great implications for agriculture, from the molecular level to the ecosystem level. We foresee AI techniques providing farmers with new approaches to growing and protecting crops and building resilience in the face of climate change and depletion of resources.[176] Some promising examples include, efforts to apply genomic and proteomic techniques from plant model systems to food

[175] Nguyen, E. et al. (2024 February). Sequence modeling and design from molecular to genome scale with Evo. bioRxiv. Cold Spring Harbor Laboratory.
[176] Usigbe, M.J. et al. (2023 August). Enhancing resilience in agricultural production systems with AI-based technologies. *Environment, Development, and Sustainability*.

crops,[177] imaging to identify plant susceptibility to pests more efficiently,[178] increasing the capacity of plants to "fix" nitrogen directly to reduce the need for chemical fertilizers,[179] and improving the combination of irrigation and fertigation to support better overall soil fertility.[180] Another example is harnessing AI methods to speed up the typically protracted process of seed research and development, spanning over a decade or more.[181, 182] We expect to see AI-based methods bringing the cycle of developing seeds with heightened resilience and adaptability down to just a few years or less.

In the pursuit of sustainability, one pivotal area ripe for innovation pertains to the reduction of methane emissions from livestock, particularly cattle.[183] Within the United States, cattle methane emissions alone contribute to approximately 45% of total agricultural methane emissions. This alarming statistic underscores the importance of pursuing the opportunity to develop novel solutions to mitigate such emissions. Generative AI methodologies offer promising avenues for devising innovative strategies aimed at optimizing animal diets and refining policies pertaining to manure storage and handling. By harnessing the power of AI-driven insights, there exists a tangible opportunity to develop tailored interventions appropriate to each farm that can significantly curtail methane emissions from livestock, thereby advancing sustainability goals in the agricultural sector.

On another front, the integration of AI-powered discovery pipelines stands poised to revolutionize traditional agricultural practices by offering alternative pathways beyond reliance on chemical-based approaches. These pipelines facilitate the exploration and development of novel biological alternatives that hold potential for enhancing both sustainability and productivity within agricultural systems. By leveraging AI-driven insights, researchers can unlock innovative solutions that not only mitigate environmental impact but also bolster agricultural resilience. However, the ability of AI tools to help develop approaches that may not fit with the values of communities or consumers, as has been the case with genetically modified organisms in Europe, points again to the need to maintain strong public engagement[184] when using AI tools.

Advances in AI for medical diagnosis and prediction

AI methods will become known as powering a revolution in healthcare over the next several decades. Moving beyond illness, AI methods can be harnessed to promote health, vibrancy, and longevity. To start, there is great promise in pressing today's well-understood AI solutions into service. However,

---

[177] Van den Broeck, L. et al. (2023 August). Functional annotation of proteins for signaling network inference on non-model species. *Nature Communications*.
[178] Złotkowska, E. et al. (2024 April). Automated imaging coupled with AI-powered analysis accelerates the assessment of plant resistance to *Tetranychus urticae*. *Scientific Reports*.
[179] Li, M. et al. (2023 May). Nano-enabled strategies to enhance biological nitrogen fixation. *Nature Nanotechnology*.
[180] Lou, S. et al. (2022 December). The formulation of irrigation and nitrogen application strategies under multi-dimensional soil fertility targets based on preference neural network. *Scientific Reports.*
[181] Rai, K. (2022 August). Integrating speed breeding with artificial intelligence for developing climate-smart crops. *Molecular Biology Reports*.
[182] Dadlani, M. (2023 February). Emerging Trends and Promising Technologies. *Seed Science and Technology*.
[183] Altshuler, Y. et al. (2023 November). From Microbes to Methane: AI-Based Predictive Modeling of Feed Additive Efficacy in Dairy Cows. arXiv.
[184] This point is made more thoroughly in PCAST's Letter to the President on Advancing Public Engagement with the Sciences. PCAST (2023 August). Advancing Public Engagement with the Sciences.

impressive breakthroughs will be achieved with leveraging advances with applications of AI in biosciences to discover and design novel therapeutic agents, using AI technologies in the early detection and response to the first signs of illness, and in "ultra-personalizing" healthcare.

Strides have been made over decades with using conventional machine learning and reasoning in systems that can provide recommendations to clinicians with expert-level assistance with diagnosis and prediction. As examples, systems have been developed for predicting the onset of sepsis,[185] risk of acquiring hospital-associated infections,[186, 187] detecting avoidable errors in patient care,[188] predicting unexpected patient deterioration,[189] and risk of readmission following discharge from the hospital.[190, 191] Thus far, the adoption of these AI systems has been modest. Nevertheless, future integration of these AI technologies into practice has significant potential for increasing the quality of care while reducing healthcare costs. We believe that such conventional machine learning technologies[192] should be pursued with energy, with a focus on addressing challenges with integration with workflows and adoption observed to date with these systems being pressed into real world service, for instance by ensuring full compliance with patient privacy regulations. At the same time, more recently developed deep learning and uses of foundation models in generative AI also hold great promise for reducing the drudgery of administrative tasks such as the reporting, writing and summarizing of information. However, PCAST see the biggest wins coming with the leveraging of advances in the biosciences as described in the prior section. In the next subsections, we describe some of the exciting possibilities ahead.

AI in the discovery and design of novel therapeutic agents

PCAST expects that AI-powered molecular discovery and design will deliver a torrent of innovation in the biosciences that promises to revolutionize *how* we develop therapeutic agents. Generative AI is emerging as a powerful tool for both discovering and designing novel molecular structures that can interact with specific cellular targets, offering immense potential for new ways to disrupt the functioning of infectious bacteria and viruses, to block disease pathways, such as those active in cancers and auto-immune diseases, or to boost immune responses.

On the molecular discovery side, by integrating AI into the drug discovery pipeline, researchers can rapidly screen millions of potential compounds, prioritizing those with the most promising

---

[185] Henry, K. et al. (2015 August). A targeted real-time early warning score (TREWScore) for septic shock. *Science Translation Medicine*.
[186] Wiens, J. et al. (2014 July). Learning Data-Driven Patient Risk Stratification Models for Clostridium difficile. *Open Forum Infectious Diseases*.
[187] Oh, J. et al. (2018 April). A Generalizable, Data-Driven Approach to Predict Daily Risk of Clostridium difficile Infection at Two Large Academic Health Centers. *Infection Control Hospital Epidemiology*.
[188] Hauskrecht, M. et al. (2016 December). Outlier-based detection of unusual patient-management actions: An ICU study. *Journal of Biomedical Informatics*.
[189] Lee, D. et al. (2020 July). Predicting severe clinical events by learning about life-saving actions and outcomes using distant supervision. *Journal of Biomedical Informatics*.
[190] Bayati, M. et al. (2014 October). Data-driven decisions for reducing readmissions for heart failure: general methodology and case study. *PLoS One*.
[191] Ross, J. et al. (2008 July). Statistical models and patient predictors of readmission for heart failure: a systematic review. *Archives of Internal Medicine*.
[192] By *conventional machine learning*, we refer to the rich array of supervised machine learning methods that have been explored for decades, involving the training of diagnostic and predictive systems on corpora of examples of cases, representing positive and negative cases.

properties and reducing the time and cost associated with experimental validation. For example, a powerful new antibiotic compound was recently identified through an AI-driven computational screening process.[193] By leveraging AI algorithms to navigate the vast chemical space of potential molecular structures, researchers were able to identify a novel compound with potent antibacterial activity against a wide range of pathogens, including those resistant to existing antibiotics. This type of discovery-based method shows great promise for providing new insights about molecular structures and functions, for example leading to whole new classes of antibiotics.[194]

Beyond discovery, one of the most promising uses of generative AI models in biology is to design novel molecular structures tailored to specific therapeutic targets, such as custom-tailoring molecules to block the binding regions of viruses, stopping their ability to interact with human tissues—and thus shutting down viral infections. Several classes of AI algorithms are now available to generate new molecular structures to achieve specific target structures and functions, and with desired characteristics, such as binding affinity, specificity, and pharmacokinetic properties.[195] These generative models are the basis for rising fields of AI-powered drug and protein design.

Beyond helping scientists to come up with new designs, AI is also proving invaluable in the field of drug repurposing. Leveraging machine learning techniques and expanding knowledge of human cell interaction networks, researchers are identifying new uses for existing, approved drugs, by unraveling their interactions with diverse cellular networks and biological pathways. AI models which integrate diverse data sources, such as gene expression profiles, protein-protein interactions, and drug-target associations, have enabled the identification of numerous drug repurposing opportunities, offering a cost-effective and accelerated path to expanding our therapeutic arsenal.[196]

A coming revolution in disease detection and early intervention

PCAST expects AI methods to revolutionize our ability to provide insights into special risks and propensities for disease as well as early detection and intervention to halt the onset of illness or to begin to respond and heal disease at the earliest stages. Beyond helping to boost the accuracy and timeliness of detection from image-based diagnostics in fields like radiology, AI algorithms will play an increasingly pivotal role in interpreting complex molecular signals about the presence of proteins, DNA, and RNA, enabling earlier and more accurate disease detection. By analyzing large datasets encompassing genomic, transcriptomic, proteomic, and metabolomic data, AI algorithms can uncover subtle patterns and associations that may have gone unnoticed using traditional analytical methods.

Exciting work in AI is already enabling physicians to detect minute traces of molecular signatures that are indicative of early-stage cancers.[197, 198] This breakthrough represents a significant leap forward in our ability to diagnose cancers that were previously challenging to detect in their initial

---

193 Stokes, J. et al. (2020 February). A Deep Learning Approach to Antibiotic Discovery. *Cell*.
194 Wong, F. et al. (2023 December). Discovery of a structural class of antibiotics with explainable deep learning. *Nature*.
195 E.g., Han, R. et al. (2023 September). Revolutionizing Medicinal Chemistry: The Application of Artificial Intelligence (AI) in Early Drug Discovery. *Pharmaceuticals*.
196 E.g., Pun, F. et al. (2023 September). AI-powered therapeutic target discovery. *Trends in Pharmacological Sciences*.
197 Zviran, A. et al. (2020 June). Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nature Medicine*.
198 Thierry, A. (2023 January). Circulating DNA fragmentomics and cancer screening. *Cell Genomics*.

stages—when therapeutic interventions are most effective. The implications of the trajectory of this AI technology are profound: early detection holds the potential to significantly improve cancer survival rates by enabling earlier intervention and more tailored treatment strategies. Beyond cancer detection, AI-driven analysis of genetic data promises to revolutionize the diagnosis and management of other complex diseases. For instance, researchers are leveraging AI algorithms to interpret the vast amounts of genomic data, enabling the identification of rare genetic variants and their associations with a propensity for disease and developing early interventions via developing and making available therapeutics.[199, 200, 201]

Toward the ultra-personalization of medicine

AI methods will enable the ultra-personalization of healthcare, when precision and fine-grained monitoring and adaptation is needed. Directions include leveraging proteomic and genomic information, in addition to data about the patterns of expression of RNA and proteins across different cell types, for personalized risk assessments and tailored treatment plans. By analyzing the intricate interplay between an individual's genetic makeup, lifestyle factors, and environmental exposures, AI algorithms have the potential to provide patients and physicians with insights into disease susceptibility, disease progression, and potential therapeutic responses.[202] The integration of electronic health records and phenotypic data into multimodal embeddings has paved the way for personalized medicine, enabling the development of predictive models that can stratify patients based on their molecular profiles and optimize treatment strategies accordingly. As an example, AI tools for studying variants of human genomes can boost our ability to understand genetic variations and their impact on cellular functions. By accurately identifying and interpreting genetic variants from sequencing data, molecular underpinnings of complex disease can be identified, providing valuable insights for the development of personalized therapies and targeted interventions.

One of the most promising applications of ultra-personalized, adaptive therapy is in cancer. Cancer is a complex disease characterized by a myriad of genetic and molecular aberrations that can vary significantly from one patient to another. Traditional "one-size-fits-all" approaches to cancer treatment have often fallen short, as they fail to account for this inherent diversity. Also, cancers continue to evolve, and can become resistant to therapies—requiring an updating of approach as well as multi-point attacks on the underlying molecular networks to cut through the resilience that many cancers show. AI algorithms can integrate multi-omics[203] data sets with clinical and environmental factors, enabling the construction of personalized disease models that capture the intricate network of cellular processes driving an individual's cancer progression. These models can then be used to simulate and predict the likely response to various therapeutic interventions, paving the way for tailored treatment strategies optimized for each patient's unique molecular profile. We see great opportunities ahead to leverage AI-driven analysis of genomic, transcriptomic, and proteomic data,

---

[199] Frazer, J. et al. (2021 October). Disease variant prediction with deep generative models of evolutionary data. *Nature*.
[200] Qi, H. et al. (2021 January). MVP predicts the pathogenicity of missense variants by deep learning. *Nature Communications*.
[201] Quazi, S. (2022 June). Artificial intelligence and machine learning in precision and genomic medicine. *Medical Oncology*.
[202] The Economist. (2024 March). The AI doctor will see you now... eventually.
[203] Esteban-Gil, A. et al. (2019 October). ColPortal, an integrative multiomic platform for analysing epigenetic interactions in colorectal cancer. *Sci Data*.

to gain unprecedented insights into the unique molecular signatures of each patient's cancer and to respond with precision therapeutic targeting.
Beyond cancer, the integration of AI and cellular biology holds immense promise for personalized care across a wide range of diseases, including cardiovascular disorders, neurological conditions, and autoimmune diseases. By modeling the intricate cellular networks and signaling pathways involved in these diseases, AI systems can identify potential therapeutic targets and predict the efficacy of various interventions based on an individual's unique molecular and environmental factors.

Transitioning from healing to maintaining health and vitality, PCAST envisions a future of personalized preventive strategies, grounded in AI-driven insights into cellular processes and individual genetic predispositions. Tailored lifestyle recommendations, dietary interventions, and targeted therapies can be implemented proactively to mitigate disease risk and promote overall well-being, ushering in a paradigm shift towards truly preventive and proactive healthcare. In this future, ultra-personalized healthcare becomes a reality that is broadly available to the public, propelled by AI advancements in cellular biology and innovative models linking molecular biology to health and wellness, and trained on medical data sets that encompass a representative range of demographics. This vision is rooted in our deepening understanding of both the similarities and differences that define us as individuals. AI systems are poised to harness data that captures our shared molecular and cellular structures as well as our unique characteristics—from the molecules forming our proteins, DNA, and RNA, to the networks driving our cellular activities, and extending to our overall physiologies. Elevating this pursuit to a higher level of health and vitality, we anticipate AI's role in ultra-personalizing medicine to transcend molecular and physiological aspects. Future AI systems could help us to encompass information about our environments, activities, and personal preferences, including our aspirations and hopes about living happy, productive, and healthy lives.

4. Findings & Recommendations

AI technologies, responsibly deployed, can be harnessed to radically accelerate scientific progress and revolutionize its practice across all subfields of science, helping researchers to identify novel, effective options for addressing today's global challenges. There is already substantial interest, investment, and momentum across the private sector, public sector, and academia directed towards achieving potentially significant advances in science through the application of AI. However, the resources required to use AI tools are highly concentrated in a small number of private sector entities. The speed with which AI has been adopted by individuals, industry, and institutions has sometimes outpaced the implementation of responsible practices and safeguards. Further, the existing incentive structures can be insufficient to attract talent to work on important foundational AI research with no clear commercializable purpose in the near term, hindering the progress of those aspects of AI research which are primarily in the public, rather than commercial, interest.

In this section, we present our findings on these issues and recommendations for advancing responsible and equitable deployment and application of AI for accelerating scientific discovery. Our findings build upon previous reports, particularly the 2021 final report of the National Security Commission on Artificial Intelligence (NSCAI)[204] and the 2023 NAIRR task force report.[205] Some of

[204] National Security Commission on Artificial Intelligence. (2021). The Final Report.
[205] National Artificial Intelligence Research Resource Task Force. (2023 January). Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem An Implementation Plan for a National Artificial Intelligence Research Resource.

the recommendations from those prior reports have already been partially or fully implemented, including through the blueprint for the AI Bill of Rights (2022),[206] the CHIPS and Science Act of 2022,[207] the NIST AI Risk Management framework (2023),[208] an expansion of the number of NSF AI institutes in 2021 and 2023, the extensive Executive Order on the Development and Use of Artificial Intelligence (2023),[209] the OMB Memo on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence,[210] the recommendations of the National AI Advisory Committee,[211] and the NAIRR pilot[212, 213] (launched in January 2024). Despite this extensive and rapid array of actions over the last three years, more needs to be done if we are to realize the full potential of AI for addressing urgent societal and global challenges.

Finding 1: Important research is being hindered by lack of access to advanced models.

> "*Progress at the current frontiers of AI is often tied to access to large amounts of computational power and data. Such access today is too often limited to those in well-resourced organizations. This large and growing resource divide has the potential to limit and adversely skew our AI research ecosystem... A widely accessible AI research cyberinfrastructure that brings together computational resources, data, testbeds, algorithms, software, services, networks, and expertise would help to democratize the AI research and development landscape in the United States for the benefit of all.*"
> - NAIRR Task Force report (2023)

Scientific research is a collaborative effort. In order for the findings of one group of researchers to be replicated, verified, and built upon by others, it is important that the data and models used are made broadly available to the research community. At present, the largest and most powerful AI models are often proprietary and available only to select groups of researchers, making it difficult to independently validate or adapt research using these models. If this situation persists, we are at risk of creating a new "digital divide" in AI,[214] as well as a potential "AI monoculture" in which AI applications are dependent on a very small number of providers or architectures. Research and the scientific method thrive when diverse hypotheses and methods approach the same problem, so an "AI monoculture" would impede scientific progress by missing potentially better answers or leaving entire avenues unexplored.

Already, the cost of access to cutting-edge AI models and the extreme disparities in compensation between the leading technology companies and academic or government institutions is leading to a

---

206 The White House. (2022 October). Blueprint for an AI Bill of Rights: Making Automated Systems Work for The American People.

207 H.R.4346. (2021 July 1). Chips and Science Act.

208 National Institute of Standards and Technology. (2023 January). Artificial Intelligence Risk Management Framework (AI RMF 1.0).

209 Executive Order 14110, 88 FR 75191. (2023 November 1). "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."

210 The White House Office of Management and Budget. (2024 March 28). Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence. [Memo M-24-10].

211 As of April 2024, the National AI Advisory Committee had 11 relevant reports.

212 National Science Foundation. National Artificial Intelligence Research Resource Pilot. (Accessed 2024 April 23).

213 The National Artificial Intelligence Research Resource (NAIRR) Pilot. (Accessed 2024 April 23).

214 Klinger, J. et al. (2020 November). A Narrowing of AI Research? Social Science Research Network.

loss of valuable human capital in these latter sectors.[215] Furthermore, to learn how to deploy these powerful AI models responsibly, and to detect and counter the impacts of irresponsible uses of these tools, it is a matter of crucial public interest to have a broad range of studies that probe the various weaknesses and biases of AI systems and quantify the effectiveness of different ways to mitigate defects as accurately as possible. As the performance of AI models depends significantly on their size, it is essential that *all* researchers in the foundations of AI have access to both small- and large-scale models. The public and private sector have different incentives and goals in their use of AI—yet they can valuably work together to create a culture of responsible AI use.

Box 1. The National AI Research Resource (NAIRR)

The National AI Research Resource (NAIRR) is envisioned as an infrastructure to help democratize access to resources necessary for conducting responsible AI research and development, including computational resources, data sets, and a research environment with user support for both open and secure projects. The idea of a NAIRR as a key element of the U.S. AI innovation ecosystem has been advanced by the U.S. National Security Commission on AI [a], the United States Congress [b], a NAIRR Task Force [c], and the Executive Office of the President [d].

The NAIRR task force, established as part of the National AI Initiative Act[e], released a report with an implementation plan for establishing a NAIRR to spur innovation, increase the diversity of talent, improve research capacity, and advance trustworthy AI. The NAIRR Task Force estimated that Federal investment on the order of $500 million per year over six years, along with in-kind and other contributions, will be needed to establish the NAIRR [c].

A NAIRR pilot was launched on January 24, 2024 [e]; however, as of the time of this writing, no dedicated funds have been appropriated for establishing and sustaining the NAIRR.

[a] National Security Commission on Artificial Intelligence. (2021). The Final Report.
[b] S.2714. (2023 July). CREATE AI Act of 2023.
[c] National Artificial Intelligence Research Resource Task Force. (2023 January). Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource.
[d] Executive Order 14110, 88 FR 75191. (2023 November). "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."
[e] National Science Foundation. National Artificial Intelligence Research Resource Pilot. (Accessed 2024 April 11); Department of Energy. (2024 January). DOE Advancing Safe and Secure AI Research Infrastructure Through the National Artificial Intelligence Research Resource Pilot; The National Artificial Intelligence Research Resource (NAIRR) Pilot. (Accessed 2024 April 11).

Recommendation 1: Expand existing efforts to broadly and equitably share basic AI resources.

To support these scientific needs in an open and cost-effective fashion, encourage experimentation and innovation, and democratize the performance of and benefits from AI research, it is essential to have extensive support for shared AI models, data sets, benchmarks, and computational resources

[215] Nix, N. et al. (2024 March). Silicon Valley is pricing academics out of AI research. Washington Post.

that benefits not only the largest private AI companies, but also academic researchers, national and federal laboratories, smaller companies, and non-profit organizations. In the U.S., the most promising effort in this direction is the National Artificial Intelligence Research Resource (NAIRR), which is a program designed to achieve all of these goals. NAIRR is currently operating as a small-scale two-year pilot project (see Box 1). However, at its current size, the NAIRR pilot does not offer access to computational resources or models at the scale of leading efforts in the private sector (most of whose models are closed-source and proprietary), or of comparable efforts in other countries (see Box 2). Further expansion of these shared resources, and in particular full and timely implementation of the NAIRR, will be an important first step to providing a cost-efficient means for a broad base of researchers to fully benefit from the most advanced AI models available while avoiding duplication of effort; it will also allow the US to maintain its current technological lead in AI capabilities.

PCAST sees a particularly promising use case for NAIRR in enabling the development of smaller-scale versions of AI models. Further expansion of these shared resources, and in particular[216] of the models that are less resource intensive, yet still powerful enough for many research applications, is needed. As discussed in the subsequent recommendations, a fully functional and well-resourced NAIRR (together with other AI infrastructure efforts at both the federal and state level) could serve in the future as a stepping stone towards even more ambitious AI infrastructure projects at the national or international level. A full-scale NAIRR and subsequent AI infrastructure could be vital for projects such as the construction of the next generation of multimodal foundation models or the development of further large scale data processing capabilities (including cloud-based computing as well as computing in physical proximity to large data sources) for academic research purposes.

Box 2. Global AI Investments

Between 2016 and 2021, at least 30 nations publicly announced national AI strategies and initiatives, including explicit investment into research and development for AI [a]. For example, India, [b] Singapore, [c] and Finland[d] each announced nationwide plans to elevate their nation as a leader in advanced AI research and development through investments in both public and private resources. China announced new legislation and national and regional bodies to incentivize AI research and development [e].

Several nations' AI investments—including in the UK, Australia, Germany, Turkey, South Korea, New Zealand, South Africa, and France—correspond to hundreds of millions to billions of dollars. For example, the UK launch of an Artificial Intelligence Research Resource aims to provide researchers with access to compute capacity for AI R&D at a cost of 300 million pounds [f]. Australia's National AI Centre [g] included 44 million AUD for four Artificial Intelligence and Digital Capability Centres—that aggregate compute infrastructure, data, and existing expert collaborations—as part of its 124 million AUD AI Action Plan for development and adoption of responsible AI [h]. The German government promised over 5 billion Euros for AI and related efforts over 5 years [i].

[a] See the annual "The AI Index Report" issued by the Stanford University that measures global trends in artificial intelligence and their Global AI Vibrancy Tool.
[b] INDIAai. (Accessed 2024 April 11).

[216] Javaheripi, M. and Bubeck, S. (2023 December). Phi-2: The surprising power of small language models. Microsoft Research Blog.

[c] AI Singapore. (Accessed 2024 April 11).
[d] Business Finland. Artificial Intelligence From Finland. (Accessed 2024 April 11).
[e] Alper, A. (2023 February). U.S. investors have plowed billions into China's AI sector, report shows. Reuters.
[f] UK Research and Innovation. (2023 November). £300 million to launch first phase of new AI Research Resource.
[g] Commonwealth Scientific and Industrial Research Organization. National Artificial Intelligence Centre. (Accessed 2024 April 11).
[h] The Hon Melissa Price MP: Archived content. (2022 March). $44 million to build AI and digital capability centres; Australian Government: Department of Industry, Science and Resources. (2021 June 18). Australia's Artificial Intelligence Action Plan.
[i] Federal Ministry of Education and Research. (2023 November). BMBF action plan "Artificial Intelligence."

Finding 2: Cutting-edge research requires access to high quality data.

> "*A major distinguishing factor of the sciences (specifically, biology, chemistry and physics), as compared to AI fields such as natural language processing and computer vision, is the relative lack of publicly available data suitable for these domains. There are simply vastly fewer data [sets] existing in the sciences, and these are often siloed by academics and companies. Acquiring appropriate scientific data for AI typically requires not only highly trained humans, but also high-end facilities with expensive equipment, making for an overall costly and slow endeavor compared to humans simply going about their day by adding to the vast trove of images, text, audio and video on the World Wide Web.*"
> - Jennifer Listgarten, "The perpetual motion machine of AI-generated data and the distraction of ChatGPT as a 'scientist'," *Nature Biotechnology* (2024)

The future will be increasingly data-driven. Machine learning technologies, in particular, rely crucially on large, high-quality data sets for their training. For instance, successful protein folding models have been trained on existing databases of hundreds of thousands of protein structures, such as the Protein Data Bank and UniProt.[217] AI climate models rely on a diverse set of existing data, including historical records, supercomputer simulations, satellite data, and combinations of all three using reanalysis methods.[218] In the social sciences, secure access to federal administrative data has allowed researchers to document the geographic variation in healthcare costs, the effects of neighborhoods on economic mobility, and the extent to which children from low-income families have a smaller chance to become inventors and innovators, among many other findings.[219, 220, 221]

The federal government possesses many additional data sets that would be invaluable for many scientific applications, as well as for informing policy decisions on important challenges facing our

[217] Jumper, J. et al. (2021 July). Highly accurate protein structure prediction with AlphaFold. *Nature*.
[218] Gibson, P. et al. (2021 August). Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts. *Communications Earth & Environment*.
[219] Finkelstein, A. et al. (2016 November). Sources of Geographic Variation in Health Care: Evidence from Patient Migration. *The Quarterly Journal of Economics*.
[220] Chetty, R. and Hendren, N. (2018 August). The Impacts of Neighborhoods on Economic Mobility I: County-Level Estimates. *The Quarterly Journal of Economics.*
[221] Bell, A. et al. (2019 May). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *The Quarterly Journal of Economics*.

nation. A substantial amount of useful raw data is already available at the data.gov portal, but—as our past PCAST reports have repeatedly discovered—much more value from these data could still be unlocked. For example, efforts to use ML to predict and model wildfires would be significantly boosted by access to declassified satellite data from the Department of Defense and other federal sources.[222] Similarly, multiple agencies hold records on the physical and economic impact of extreme weather events, which can be used as input to ML approaches to catastrophe modeling.[223] Addressing another significant issue that impacts many Americans, access by researchers to anonymized healthcare data will be invaluable in assessing and improving patient safety measures.[224] Data created by the private sector also has the potential to improve public welfare. For example, as noted in the quote at the beginning of this section, data on chemical and physical properties of compounds and materials is more commonly held by the private sector.

Many federal data sets contain sensitive personal information and so cannot be released openly; however, it is still possible to perform scientific research using these data while protecting the privacy of individuals, e.g., by limiting access to aggregated and anonymized information, by introducing controlled amounts of random noise, or by working with synthetic data sets derived from the underlying data. Pilot federal programs to allow such secured access are now underway, such as the National Secure Data Service Demonstration Project[225] and the Federal Statistical Research Data Centers project.[226]

Recommendation 2: Expand secure access to federal data sets for approved critical research needs, with appropriate protections and safeguards.

PCAST recognizes that the issue of secure access to federal data is both sensitive and technically complex, and that protection of personal information—and in some cases, national security concerns—must always be a primary consideration. Nevertheless, the benefits of allowing limited access to such data by approved researchers, as well as the release of carefully anonymized versions of such data sets to curated resource centers such as a future NAIRR or National Secure Data Service, are immense, and so we strongly encourage expansion of existing pilot programs for secured data access. PCAST also encourages further development of existing guidelines[227] on federal database management that incorporate cutting-edge developments in privacy protection technologies such as differential privacy,[228] homomorphic encryption,[229] federated learning,[230] and use of synthetic data.[231] Such technologies are not "silver bullets" for attaining absolute privacy protection, which is

---

[222] PCAST (2023 February). Modernizing Wildland Firefighting to Protect Our Firefighters.
[223] PCAST (2023 April). Extreme Weather Risk in a Changing Climate: Enhancing prediction and protecting communities.
[224] PCAST (2023 September). A Transformational Effort on Patient Safety.
[225] National Center for Science and Engineering Statistics. The National Secure Data Service Demonstration Project. (Accessed 2024 April 11).
[226] United States Census Bureau. (2023 July). Federal Statistical Research Data Centers.
[227] Subcommittee on Open Science of the National Science and Technology Council. (2022 May). Desirable Characteristics of Data Repositories for Federally Funded Research.
[228] Papernot, N. and Thakurta, A. (2021 December). How to deploy machine learning with differential privacy. NIST Cybersecurity Insights.
[229] Torkzadehmahani, R. et al. (2022 June). Privacy-Preserving Artificial Intelligence Techniques in Biomedicine. *Methods of Information in Medicine*.
[230] Kaissis, G. et al. (2020 June). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*.
[231] Savage, N. (2023). Synthetic data could be better than real data. *Nature Outlook.*

an infeasible goal, but they can significantly reduce the risks of leakage of private information.

While raw data sets, especially when labeled with relevant "metadata," are already of significant research value for both traditional data science and for recent AI-based data analytics, more carefully curated and maintained databases, geared specifically towards research use,[232, 233] will have an even more transformative impact in multiple scientific fields. At present, high-quality data curation is an expensive and difficult task, requiring a large amount of expert human attention; however, PCAST sees great potential in using modern AI technologies to partially automate the curation process. In the future, we expect it to become feasible to use AI tools to upgrade many existing federal data sets into such a curated form, which we would recommend adopting as a long-term goal of federal data sharing initiatives such as data.gov.

There are further scientific data sets of public value beyond those directly held by the federal government which are worth collecting and sharing in a manner that respects privacy and intellectual property rights. PCAST endorses the efforts of federal agencies such as the National Institutes of Health (NIH)[234] and NSF[235] to mandate responsible sharing of data sets arising from the research that it funds or conducts. We encourage further development and enforcement of such mandates, in conjunction with sufficient supporting resources to meet them. We should also consider ways to encourage sharing of the AI models trained on that data when appropriate. In the longer term, we envisage the launch of more ambitious projects involving multiple public and private partners to create secure and high-quality scientific databases as part of the national AI and data infrastructure. We endorse initiatives such as the NSF FAIROS Research Coordination Networks[236] to develop practices and methods to facilitate the creation of such databases.

Finding 3: AI provides a unique resource for collaborations across academia, industry, and all branches of the federal government.

> *"The collaborative nature of the pilot, bringing together academia, industry, nonprofit and government sectors, is intended to promote cross-sector partnerships. Industry collaboration can lead to the development of commercially viable AI applications and solutions, fostering economic growth by creating new markets and revenue streams."*
> - NAIRR report (2024)

U.S. industry invests tens of billions of dollars[237] in fundamental and applied AI research. NSF recently announced[238] a $140 million investment to establish seven new National AI Research Institutes that

[232] For a discussion of some desirable properties of data sets for scientific research, see Wilkinson et al. (2016 March). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, Vol. 3, No. 160018.
[233] For extremely large (petabyte scale) data sets, bandwidth limitations in downloading the raw data are also an issue; it can be desirable in such cases to be able to perform analysis and model training close to the data source.
[234] National Institutes of Health. Final NIH Policy for Data Management and Sharing. (Accessed 2024 April 11).
[235] National Science Foundation. NSF Public Access Initiative. (Accessed 2024 April 11).
[236] National Science Foundation. Findable Accessible Interoperable Reusable Open Science Research Coordination Networks (FAIROS RCN). (Accessed 2024 April 11).
[237] Our World in Data. Annual private investment in artificial intelligence. (Accessed 2024 April 11).
[238] National Science Foundation. (2023 May). NSF Announces 7 new National Artificial Intelligence Research Institutes.

involve every directorate within NSF, including the newly established Technology, Innovation, and Partnerships (TIP) Directorate. The NAIRR pilot program already contains both a "NAIRR Open" area providing open access to democratize the research, as well as a "NAIRR Secure" area co-led by NIH and DOE to support AI research requiring privacy and security.

The AI infrastructure provided by these resources, as well as the fundamentally cross-disciplinary nature of AI itself, will enable highly interdisciplinary research, bringing together national and federal laboratories and user facilities, private industry, and academics from the basic and applied sciences in novel, and potentially transformative, forms of collaboration. An example of such a collaboration is the NSF Materials Innovation Platforms,[239] part of the Materials Genome Initiative.[240] These platforms are developing a data-sharing infrastructure while employing AI tools as part of community building, with the vision of bringing in other agencies and more industrial partners.

Recommendation 3: Support both basic and applied research in AI that involves collaborations across academia, industry, national and federal laboratories, and federal agencies as outlined in the vision for the NAIRR developed by the NAIRR Task Force.

The boundaries between federally funded academic research and private sector research are hazy, with many researchers moving among affiliations with academic institutions, non-profit organizations, and/or private companies, and with a significant share of research R&D currently supported by private companies. To fully capitalize on the potential research benefits of AI, we need to support research that involves a breadth of promising and productive hypotheses and approaches. This may require that funding agencies broaden their postures regarding how to work with industry and which researchers can be supported in order to facilitate innovative research efforts and collaborations among different sectors. While the private AI sector can bring significant and highly beneficial computational resources, expertise, and funding to a collaboration, care will need to be taken to ensure that the market incentives of the industry partners align with the public and scientific goals of the project, along with clarifying intellectual property rights and plans for licensing at the outset. Examples of cross-sector collaboration could include creation of high-quality curated public scientific data sets from multiple sources, the creation of multimodal foundation models, or development of next-generation technologies such as new quantum computer qubit architectures.

As the rapidly evolving field of AI grows, federal agencies and the private sector will need to explore and re-evaluate how the emerging national AI infrastructure, such as the proposed NAIRR platform, may need to evolve in order to maintain U.S. leadership competitiveness and innovation.

---

[239] National Science Foundation. Materials Innovation Platforms (MIP). (Accessed 2024 April 11).
[240] Materials Genome Initiative. About the Materials Genome Initiative. (Accessed 2024 April 11).

Finding 4: Without proper benchmark metrics, validation procedures, and responsible practices, AI systems can give unreliable outputs whose quality is difficult to evaluate, and which could be harmful for a scientific field and its applications.

> "*Just as the cost of addressing problems after the fact is often much higher than the cost of addressing them during the initial design of a system, so too in research it is much easier to address oversights, unanticipated consequences, and unexpected outcomes if more thought is given to these issues at the outset... If they do not engage the relevant stakeholders, computing researchers run the risk of building systems that may work for themselves (or their friends and colleagues) but may not perform equally well for other populations including ones not originally targeted by the developer*."
> - "Fostering Responsible Computing Research," National Academies (2022)

Scientific research is an interconnected ecosystem—the findings, models, and data produced by one research group will impact the work of other researchers in both basic and applied fields. If used irresponsibly, AI tools could negatively impact this ecosystem with findings that are subject to algorithmic bias, cannot be replicated, lack quantifications on their uncertainty, are overfitted to training data at the expense of accuracy against real world data, or cannot be validated against more transparent techniques, such as numerical simulation and laboratory experimentation. Furthermore, the privacy of personal data, or the intellectual property rights of creators of that data, must be protected and respected. In the longer term, the release of improperly labeled synthetic data could lead to "data pollution"[241] that would negatively impact further scientific work. These risks are particularly high if fully automated AI tools are deployed without expert supervision.

The academic community must develop and refine standards for publishing research and data that rely on AI-derived inputs. Institutions of higher education will need to modernize their training of junior scientists to allow them to fully take advantage of the new AI tools, to be able to independently verify and benchmark the output of such tools, and to use them ethically and responsibly. The increasing pace of scientific advances will also require higher levels of two-way public engagement[242] by researchers and government science agencies. Particularly as AI begins to be used to impact policies of government and the private sector, the priorities and values of the public that will be impacted must be part of developing the inputs and even the questions. Some dual use research applications, such as gain-of-function research for biological pathogens, will require higher levels of monitoring, regulation, and supervision.

Recommendation 4: Adopt principles of responsible, transparent, and trustworthy AI use throughout all stages of the scientific research process.

Agencies such as NSF and NIST should continue supporting research[243] in the scientific foundations of responsible and trustworthy AI, including the development of standard benchmarks to measure

[241] Ben-Shahar, O. (2019 September), Data Pollution. *Journal of Legal Analysis*.
[242] E.g., PCAST (2023 August). Advancing Public Engagement with the Sciences.
[243] One current mechanism for doing this is through the NSF-funded National Research AI Institutes, which are already funding some research regarding trustworthiness in AI as well as NSF's Safe Learning-Enabled Systems program and Responsible Design, Development, and Deployment of Emerging Technologies program.

AI model properties such as accuracy, reproducibility, fairness,[244] resilience, and explainability, and to develop tools to evaluate biases in data sets and to distinguish synthetic data from real world data. Agencies should solicit input from researchers in those fields, as well as from relevant disciplines in the social and behavioral sciences, to help craft responsible AI use policies, which will need to be continually updated as our understanding of the theory and practice of these technologies mature. More generally, PCAST strongly encourages continued dialogue on broader topics of AI governance, use, and impact between developers and users of AI technologies, policymakers, and experts in the humanities, law, and social sciences. Broad dialogue and diverse input is especially important as there is increasing consolidation in industry around governance tools, with a small set of technology and consulting organizations controlling much of the market.[245]

In parallel, managing the risks of inaccurate, biased, harmful or non-replicable findings from scientific uses of AI should not be a mere afterthought to a scientific research project, but needs to be planned from the initial stages of a project. PCAST recommends that federal funding agencies consider updating their responsible conduct of research guidelines to require responsible AI use plans from researchers using such technologies. These plans should include an assessment of potential AI-related risks such as algorithmic bias, disclosure of sensitive information, lack of transparency or reproducibility in the results, or potential harmful applications, and describe the measures taken to mitigate these risks, in particular describing the supervision procedures of any automated process. Such requirements could be modeled on (or combined with) existing data management plan requirements,[246] and be modeled after the Blueprint for an AI Bill of Rights as well as the NIST AI Risk Management Framework. To minimize additional administrative burden on researchers and build a culture of responsibility, agencies can help enumerate major risks and provide potential processes for risk mitigation. PCAST also recommends updating data management plan requirements to require disclosure and documentation of any AI tools that will generate or process the data used in proposed research projects.

Finding 5: Optimal performance requires both AI and human expertise.

> "*The government needs AI systems that augment and complement human understanding and decision-making so that the complementary strengths of humans and AI can be leveraged as an optimal team. Achieving this remains a challenge.*"
> - NSCAI report (2021)

Science fiction often portrays artificial intelligence as fully autonomous entities, operating independently without human oversight, with individual AIs serving as a complete replacement for individual humans in complex tasks. Such a level of automation is not infrequently stated as a long-term goal by supporters of AI technologies, but replacing human creativity is not the outcome PCAST

[244] E.g., Ferryman, K. (2020 December). Addressing health disparities in the Food and Drug Administration's artificial intelligence and machine learning regulatory framework. *Journal of the American Medical Informatics Association*.
[245] Fortune Business Insights. (2024 April). AI Governance Market Size, Share & Industry Analysis, By Deployment (Cloud and On-premise), By Enterprise Type (Large Enterprises and Small & Medium Enterprises (SMEs)), By End-user (Healthcare, Retail, IT and Telecom, BFSI, Automotive, Media and Entertainment, Manufacturing, and Others), and Regional Forecast, 2024-2032. [Market Research Report].
[246] E.g., National Science Foundation. Preparing Your Data Management Plan. (Accessed 2024 April 11).

envisions in this report. PCAST contends that the most effective and valuable deployment of AI to help science and research solve pressing problems will be in the form of assisting tools. In addition, at their current level of technological maturity, generative AI resources are far too unreliable to enable fully automated continuous operation independent of human oversight or intervention, particularly in scientific domains where accuracy, explainability, and replicability are essential.

Furthermore, the strengths of human intelligence and artificial intelligence are largely complementary. AI can tirelessly locate patterns from enormous data sets and perform repetitive tasks, while humans can extrapolate and draw conclusions from far smaller amounts of data, perform systematic and strategic reasoning, and coordinate actions with both other humans and AI assistants.

Current AI tools are still quite weak at demonstrating true creativity, analysis, or high-level strategic thinking. Even in a future world where AI assistance is commonplace, and additional AI technologies beyond the current state of the art ML models (such as large language models) become available, we expect that scientific research will (and should) continue to be directed by human scientists. Scientists will employ the uniquely human abilities to draw high-level conclusions from relatively small amounts of data to complement the ability of AI to generate suggestions and connections from large data sets, and to automate the more routine and tedious aspects of research. In short, PCAST expects that the paradigm of AI-assisted science will often be superior in the near and medium term to both unassisted science and fully automated science.

Recommendation 5: Encourage innovative approaches to integrating AI assistance into scientific workflows.

The scientific enterprise is an excellent "sandbox" in which to practice, study, and assess new paradigms of collaboration between humans and AI assistants. Examples of such paradigms could include scientific advisory AI agents[247] that would serve as sophisticated natural language interfaces for operating complex software or laboratory equipment; coupling generative AI algorithms with expert human feedback and formal verification methods (or with other types of AI systems) to improve accuracy of generative AI; or the use of AI tools to enable interdisciplinary, decentralized, or crowdsourced research projects at a scale that would not otherwise be feasible. The objective should not be to maximize the amount of automation, but rather to take advantage of complementarity and allow human researchers to do high quality science that utilizes AI assistance responsibly.

Funding agencies should recognize the emergence of new workflows, and design flexible procedures, metrics, funding models,[248] and challenge problems[249] that encourage (but do not mandate) strategic experimentation with new AI-assisted ways to organize a scientific project. More broadly, incentive structures (not only in funding agencies, but also in academia and the scientific publishing industry) may need to be updated to be able to support different types of scientific contributions, such as training an AI system with expert human feedback, developing new AI software tools that are specifically tailored for scientific applications,[250] or curating a high-quality and broadly usable data set, that might not be given due recognition by more traditional metrics of research productivity.

[247] E.g., Boiko, D. et al. (2023 December). Autonomous chemical research with large language models. *Nature*.
[248] E.g., National Science Foundation. Artificial Intelligence, Formal Methods, and Mathematical Reasoning. (Accessed 2024 April 25).
[249] Donoho, D. (2023 October). Data Science at the Singularity. arXiv.
[250] Examples include better tools for learning from unlabeled data that are difficult to "tokenize" (e.g., due to high dimensionality of the data), or for performing uncertainty quantification.

These new metrics would complement traditional metrics; the combination of these metrics would provide fuller understanding of, and credit for, the various types of intellectual contributions that are integral to the advancement of science.

5. Conclusion

Today, our world has an urgent need for scientific advances to address a wide variety of global and societal challenges. This report has provided "snapshots" of such advances that may be realized with the assistance of AI tools: new materials to enable the transition to an energy-efficient, low-carbon economy; accurate climate models to forecast the impact of climate change and extreme weather events; more effective and data-driven delivery of essential public services; and much more. When supervised properly and responsibly by human experts, trained on high quality data, and verified using reliable scientific techniques, artificial intelligence tools can become engines of innovation that can supercharge the ability of scientists and policymakers to address such challenges. To achieve these goals, advanced models, data sets, and benchmarks need to be broadly available to the scientific community, responsible AI practices need to be interwoven into scientific workflows, and sustained investments need to be made in both the foundations of AI and in its various scientific applications.

The resources and effort needed to build such an AI infrastructure and culture for science are significant, but will yield outsize rewards: a broad and democratized ecosystem for research and development that is both open and secure, in which scientific ideas can be rapidly translated into experiments, prototypes, and successful solutions to pressing problems. AI tools can complement traditional research practices, relying on AI-assisted, but human-directed, paradigms of collaboration and research that still conform to the highest scientific standards of validation, replicability, and objectivity.

Science is an arena in which artificial intelligence technologies can make enormous positive contributions, if we are vigilant about rigorously identifying and managing its weaknesses and risks. While AI will be the means of this transformation, it will ultimately be we humans—both scientists and the general public—who will be empowered, through an increased ability to collaborate to address important challenges, and to more deeply understand the complexities of the world around us. We should embrace this opportunity enthusiastically—while being fully aware of the potential risks if these tools are not deployed responsibly—and make thoughtful investments to turn this vision of addressing global challenges with AI-supercharged, but still very much human-oriented, science into reality.

Appendix A: Glossary

- **Affordances:** The uses or purposes that a thing can have; the qualities or properties of an object that define its possible uses or make clear how it can or should be used.
- **Artificial intelligence (AI):** Refers to computer systems capable of performing complex tasks that historically only a human could do, such as reasoning, making decisions, or solving problems.
- **Chips:** Referring here primarily to application specific integrated circuits (ASIC). A specialized computer processing chip that improves speed and efficiency for specific tasks, such as neural network calculations.
- **Cleaning (of data):** Refers to the process of removing or repairing portions of a data set that are duplicated, incomplete, inaccurate, or unrelated, in order to improve the quality of that data set for analysis or training.
- **Conventional machine learning:** A rich array of supervised machine learning methods that have been explored for decades, involving the training of diagnostic and predictive systems on corpora of examples of cases, representing positive and negative cases.
- **Data Pollution:** This term is ambiguous and its meaning often depends on context. Most relevant to this report are: (a) inclusion of data generated by LLMs into the data flows that are used as or categorized as authentic user-entered data,[251] and (b) data that does not thoroughly and completely represent the subject that an AI effort is addressing. However, there are many other ways of considering data pollution that are also valid.[252, 253]
- **Data sets:** A collection of related data; however, what constitutes a dataset is not clearly demarcated. For instance, one could consider all the data associated with a research project, regardless of the type of data, to be a single dataset.[254]
- **Deep neural network models:** A particularly successful class of machine learning algorithms that were inspired by the structure of neurons inside a human brain. Neural networks are composed of interconnected layers each containing a simpler unit or "node" that can conducts part of a computation.
- **Digital twin:** A high-resolution model of a physical system that is continually updated with real-time data from that system. Such twins usually rely on traditional simulation to model the fundamental processes of the system but can additionally use AI models to refine, accelerate, or analyze such simulations.
- **Field programmable gate arrays:** A type of integrated circuit (IC) that enables the development of custom logic for rapid prototyping and final system design. FPGAs are different than other custom or semi-custom ICs due to their inherent flexibility that allows it to be programmed and re-programmed via software download to adapt to the evolving needs of the larger system in which it is designed into. FPGAs are ideally suited for today's fastest growing applications, like edge computing, artificial intelligence (AI), system security, 5G, factory automation, and robotics.[255]

---

[251] Kniaz, R. (2023 May). The Incoming Tidal Wave Of Data Pollution In AI. Forbes.
[252] Ben-Shahar, O. (2019 September). Data Pollution. *Journal of Legal Analysis*.
[253] Hasselbalch, G. (2022 July). Conclusion to White Paper: A Data Pollution Movement. Data ethics.
[254] National Library of Medicine. Dataset. (Accessed 2024 April 25).
[255] Lattice Semiconductor. What is an FPGA? (Accessed 2024 April 25).

- **Foundation models:** A ML model trained (often at great computational expense) on a broad range of data, which can then be fine-tuned relatively cheaply for more specialized applications.
- **Generative AI:** A kind of artificial intelligence capable of generating new content such as code, images, music, text, simulations, 3D objects, videos, and so on. It is considered an important part of AI research and development, as it has the potential to revolutionize many industries, including entertainment, art, and design.
- **Hallucination (by AI):** Generated content that is nonsensical or unfaithful to the provided source content [; ...] there are two main types of hallucinations, namely intrinsic hallucination and extrinsic hallucination. [An intrinsic hallucination is a] generated output that contradicts the source content; [an extrinsic hallucination is a] generated output that cannot be verified from the source content (i.e., output can neither be supported nor contradicted by the source).
- **Image generation models:** AI image generators are computer programs that use deep learning algorithms to produce digital images from scratch (usually text) or modify existing ones (usually images). These generators can create highly realistic and complex images, including landscapes, faces, objects, and more. They have practical applications in various fields, such as art, design, advertising, and gaming.[256]
- **Joint representations:** A means for machine learning models to learn from multiple types of data or features, such as images combined with text, in a unified manner.
- **Large Language Models (LLMs):** A class of language models that use deep-learning algorithms and are trained on extremely large textual datasets that can be multiple terabytes in size. LLMs can be classed into two types: generative or discriminatory. Generative LLMs are models that output text, such as the answer to a question or even writing an essay on a specific topic. They are typically unsupervised or semi-supervised learning models that predict what the response is for a given task. Discriminatory LLMs are supervised learning models that usually focus on classifying text, such as determining whether a text was made by a human or AI.
- **Machine learning (ML):** An array of techniques that leverages statistical inference "learned" through training of an AI model on large data sets.
- **Multimodal models**: A multimodal model is a ML model that expands on generative AI and is capable of processing information from different modalities, including images, videos, and text.[257]
- **Multiscale modeling:** A modeling strategy that uses multiple models at different scales simultaneously to describe a system**.**
- **Open source:** Models that disclose the structure of the model and training process, but not necessarily the final weights.
- **Open weight:** Models that disclose the weights obtained at the end of the training process.
- **Personalized medicine:** A form of medicine that uses information about a person's own genes or proteins to prevent, diagnose, or treat disease.[258]

[256] NYU Libraries. (2024 April). Machines and Society.
[257] Google Cloud. Multimodal AI. (Accessed 2024 April 25).
[258] National Cancer Institute. Personalized medicine. (Accessed 2024 April 25).

- **Quantum computers:** The general term for a device (whether theoretical or practically realized) that carries out quantum computation. A quantum computer may be analog or gate-based, universal or not, and noisy or fault tolerant".[259]
- **Reinforcement learning:** A method of training algorithms to make desired actions by requiring the process to maximize a given function or result.
- **Responsible conduct of research guidelines:** The shared values, including "rules of the road" for conduct of research with scientific integrity, whether when planning research, conducting research, or reporting and reviewing research.[260]
- **Routing**: The process of optimal path selection in any network between interconnected nodes.
- **Superconductors:** Materials which can conduct electricity so efficiently that there is little to no loss of energy. Superconductivity currently requires low cryogenic temperatures.
- **Synthetic data (generation):** A process in which seed data are used to create artificial data that have some of the statistical characteristics of the seed data.[261]
- **Systematic biases:** Result from procedures and practices of particular institutions that operate in ways which result in certain social groups being advantaged or favored and others being disadvantaged or devalued. This need not be the result of any conscious prejudice or discrimination but rather of the majority following existing rules or norms.
- **Thermoelectric materials**: Solid-state semiconductors that transform heat into electric power or produce cold from an applied voltage.[262]

Additional Glossary Resources

- IEEE Ethically Aligned Designed Glossary
- NIST's Computer Science Glossary
- NIST's Trustworthy AI Glossary

[259] National Academies of Sciences, Engineering, and Medicine. (2019). Quantum Computing: Progress and Prospects. *The National Academies Press*.
[260] Steneck, N. (2007 August). Introduction to the Responsible Conduct of Research. The Office of Research Integrity.
[261] National Institute of Standards and Technology. Synthetic data generation. (Accessed 2024 April 25).
[262] Shevelkov, A. (2010). Thermoelectric materials: an introduction. *Dalton Transactions*.

Appendix B: External Experts Consulted

PCAST sought input from a focused group of additional experts and stakeholders. PCAST expresses its gratitude to those listed here who shared their expertise. They did not review drafts of the report, and their willingness to engage with PCAST on specific points does not imply endorsement of the views expressed herein. Responsibility for the opinions, findings, and recommendations in this report and for any errors of fact or interpretation rests solely with PCAST.

Jatin Aythora
Director, Research & Development
BBC

Rian Bahran
Assistant Director, Nuclear Technology & Strategy
Office of Science & Technology Policy
The White House

Ben Buchanan
White House Special Advisor for AI
National Security Council
The White House

Robert Chesney
John Jeffers Research Chair in Law
University of Texas at Austin

David Danks
Member
National AI Advisory Committee

Hany Farid
Professor
Berkeley School of Information

Paula Goldman
Member
National AI Advisory Committee

Andrew Jenks
Principal Program Manager
Microsoft

Michael Jordan
Professor, Department of Electrical Engineering, Computer Science & Statistics
UC Berkeley

Eun-Ah Kim
Professor of Physics
Cornell University

Ramayya Krishnan
Dean, Heinz College of Information Systems and Public Policy
Carnegie Mellon University

Haniyeh Mahmoudian
Global AI Ethicist
DataRobot

Nik Marda
Chief of Staff, Tech Division
Office of Science & Technology Policy

Liz O'Sullivan
Member
National AI Advisory Committee

Manish Parashar
Director, Scientific Computing & Imaging
University of Utah

Andy Parsons
Senior Director, Content Authenticity Initiative
Adobe

Wade Shen
Director of National AI Initiative Office
Office of Science & Technology Policy
The White House

Justin Vail
Special Assistant to the President, Democracy & Civic Participation
The White House

Acknowledgments

The members of PCAST wish to thank the staff in the Office of Science and Technology Policy (OSTP) for their various contributions throughout the preparation of this report. We thank the Institute for Defense Analyses (IDA) Science and Technology Policy Institute (STPI) team, especially Emily Grumbling, Sara Jordan, Matthew Ishimaru, and Kate Ross for assistance with research and analyses. We also thank our many colleagues who provided insights in addition to the formal consults listed in Appendix B, including:

Anima Anandkumar
Tanya Klowden
Tapio Schneider
Ivan K Schuller
Vivian Zapf