National Artificial Intelligence Research and Development Strategic Plan 2023 Update

A Report by the

SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE

of the

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

May 2023

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976 to provide the President and others within the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, health, foreign relations, the environment, and the technological recovery and use of resources, among other topics. OSTP leads interagency science and technology policy coordination efforts, assists the Office of Management and Budget with an annual review and analysis of federal research and development in budgets, and serves as a source of scientific and technological analysis and judgment for the President with respect to major policies, plans, and programs of the federal government. More information is available at https://www.whitehouse.gov/ostp.

About the National Science and Technology Council

The National Science and Technology Council (NSTC) is the principal means by which the Executive Branch coordinates science and technology policy across the diverse entities that make up the federal research and development enterprise. A primary objective of the NSTC is to ensure that science and technology policy decisions and programs are consistent with the President's stated goals. The NSTC prepares research and development strategies that are coordinated across federal agencies aimed at accomplishing multiple national goals. The work of the NSTC is organized under committees that oversee subcommittees and working groups focused on different aspects of science and technology. More information is available at https://www.whitehouse.gov/ostp/nstc.

About the Select Committee on Artificial Intelligence

The Select Committee on Artificial Intelligence advises and assists the NSTC to improve the overall effectiveness and productivity of federal efforts related to artificial intelligence (AI) to ensure continued U.S. leadership in this field. It addresses national and international policy matters that cut across agency boundaries, and it provides formal mechanisms for interagency policy coordination and development for federal AI activities. It also advises the Executive Office of the President on interagency AI priorities; works to create balanced and comprehensive AI R&D programs and partnerships; leverages federal data and computational resources across department and agency missions; and supports a national technical AI workforce. The National Artificial Intelligence Initiative Office provides technical and administrative support for the Select Committee on AI.

About the Subcommittee on Machine Learning and Artificial Intelligence

The Machine Learning and Artificial Intelligence (MLAI) Subcommittee (MLAI-SC) monitors the state of the art in machine learning (ML) and AI within the federal government, in the private sector, and internationally to watch for the arrival of important technology milestones in the development of AI, to coordinate the use of and foster the sharing of knowledge and best practices about ML and AI by the federal government, and to consult in the development of federal MLAI R&D priorities. The MLAI-SC reports to the NSTC Committee on Technology and the Select Committee on AI.

About the Subcommittee on Networking & Information Technology Research & Development

The Networking and Information Technology Research and Development (NITRD) Program has been the Nation's primary source of federally funded work on pioneering information technologies (IT) in computing, networking, and software since it was first established as the High-Performance Computing and Communications Program following passage of the High-Performance Computing Act of 1991. The NITRD Subcommittee of the NSTC guides the multiagency NITRD Program in its work to provide the R&D foundations for ensuring continued U.S. technological leadership and for meeting the Nation's needs for advanced IT. The National Coordination Office (NCO) supports the NITRD Subcommittee and its Interagency Working Groups (IWGs) (https://www.nitrd.gov/about/).

About the NITRD Artificial Intelligence R&D Interagency Working Group

The AI R&D Interagency Working Group (IWG) coordinates federal AI R&D and supports activities tasked by both the NSTC Select Committee on AI and the Subcommittee on Machine Learning and Artificial Intelligence. This vital work promotes U.S. leadership and global competitiveness in AI R&D and its applications. The AI R&D IWG reports investments to the AI R&D Program Component Area.

About This Document

This document includes relevant text from the 2016 and 2019 national AI R&D strategic plans, along with updates prepared in 2023 based on Administration and interagency evaluation of the *National AI R&D Strategic Plan: 2019 Update* as well as community responses to a Request for Information on updating the Plan. The 2019 strategies were broadly determined to be valid going forward. The 2023 update adds a new Strategy 9, which establishes a principled and coordinated approach to international collaboration in AI research.

Copyright

Note: Any mention in the text of commercial, non-profit, academic partners, or their products, or references is for information only; it does not imply endorsement or recommendation by any U.S. government agency.

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

Chair

Arati Prabhakar, Director, Office of Science and Technology Policy (OSTP), Assistant to the President for Science and Technology

Acting Executive Director

Kei Koizumi, Principal Deputy Director for Policy, OSTP

SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE

Chair

Arati Prabhakar, Director, OSTP, Assistant to the President for Science and Technology

Rotating Co-Chairs

Laurie Locascio, Undersecretary of Commerce for Standards and Technology, Department of Commerce

Sethuraman Panchanathan, Director, National Science Foundation (NSF)

Geraldine Richmond, Under Secretary for Science and Innovation, Department of Energy (DOE)

SUBCOMMITTEE ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Co-Chairs

Stephen Binkley, National Nuclear Security Administration, DOE

Erwin Gianchandani, Assistant Director for Technology, Innovation and Partnerships, NSF

Tess deBlanc-Knowles, Senior Policy Advisor, OSTP

Elham Tabassi, Associate Director for Emerging Technology, Information Technology Laboratory, National Institute of Standards and Technology (NIST)

Executive Secretary

Faisal D'Souza, NITRD National Coordination Office (NCO)

SUBCOMMITTEE ON NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT (NITRD)

Co-Chair

Margaret Martonosi, Assistant Director for Computer and Information Science and Engineering, NSF

Co-Chair

Kathleen (Kamie) Roberts, NITRD NCO

Executive Secretary

Nekeia Butler, NITRD NCO

ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT INTERAGENCY WORKING GROUP

Co-Chairs

Steven L. Lee, Office of Advanced Scientific Computing Research, Office of Science, DOE

Michael L. Littman, Directorate for Computer and Information Science and Engineering, Division of Information and Intelligent Systems, NSF

Craig I. Schlenoff, Intelligent Systems Division, Engineering Laboratory, NIST

Technical Coordinator

Faisal D'Souza, NITRD NCO

Writing Team

Gil Alterovitz, VA
Faisal D'Souza, NITRD NCO
Allison Dennis, NIH
Kyle Fox, NIJ
Craig Greenberg, NIST
William Harrison, NIST
Robert Hickernell, NIST
Jonnie Bradley, DOE/AITO
David Kuehn, DOT
Steven Lee, DOE/SC
Michael Littman, NSF
Zhiyong Lu, NIH
Jillian Mammino, State
Aaron Mannes, DHS
Nikunj Oza, NASA
Richard Paladino, MDA
Pavel Piliptchak, NIST
Craig I. Schlenoff, NIST
Adria Schwarber, State
Ram D. Sriram, NIST
Ann Stapleton, USDA
Elham Tabassi, NIST
Mary Theofanos, NIST
Steven Thomson, USDA
Jay Vietas, NIOSH
James Warren, NIST
Megan Zimmerman, NIST

Table of Contents

Executive Summary

Artificial intelligence (AI)[1] is one of the most powerful technologies of our time. In order to seize the opportunities that AI presents, the Nation must first work to manage its risks. The federal government plays a critical role in this effort, including through smart investments in research and development (R&D) that promote responsible innovation and advance solutions to the challenges that other sectors will not address on their own. This includes R&D to leverage AI to tackle large societal challenges and develop new approaches to mitigate AI risks. The federal government must place people and communities at the center by investing in responsible R&D that serves the public good, protects people's rights and safety, and advances democratic values. This update to the *National AI R&D Strategic Plan* is a roadmap for driving progress toward that goal.

This plan defines the major research challenges in AI to coordinate and focus federal R&D investments. It will ensure continued U.S. leadership in the development and use of trustworthy AI systems, prepare the current and future U.S. workforce for the integration of AI systems across all sectors, and coordinate ongoing AI activities across all federal agencies.[2]

This plan, which follows national AI R&D strategic plans issued in 2016 and 2019, reaffirms eight strategies and adds a ninth to underscore a principled and coordinated approach to international collaboration in AI research:

Strategy 1: **Make long-term investments in fundamental and responsible AI research**. Prioritize investments in the next generation of AI to drive responsible innovation that will serve the public good and enable the United States to remain a world leader in AI. This includes advancing foundational AI capabilities such as perception, representation, learning, and reasoning, as well as focused efforts to make AI easier to use and more reliable and to measure and manage risks associated with generative AI.

Strategy 2: **Develop effective methods for human-AI collaboration.** Increase understanding of how to create AI systems that effectively complement and augment human capabilities. Open research areas include the attributes and requirements of successful human-AI teams; methods to measure the efficiency, effectiveness, and performance of AI-teaming applications; and mitigating the risk of human misuse of AI-enabled applications that lead to harmful outcomes.

Strategy 3: **Understand and address the ethical, legal, and societal implications of AI.** Develop approaches to understand and mitigate the ethical, legal, and social risks posed by AI to ensure that AI systems reflect our Nation's values and promote equity. This includes interdisciplinary research to protect and support values through technical processes and design, as well as to advance areas such as AI explainability and privacy-preserving design and analysis. Efforts to develop metrics and frameworks for verifiable accountability, fairness, privacy, and bias are also essential.

Strategy 4: **Ensure the safety and security of AI systems**. Advance knowledge of how to design AI systems that are trustworthy, reliable, dependable, and safe. This includes research to advance the ability to test, validate, and verify the functionality and accuracy of AI systems, and secure AI systems from cybersecurity and data vulnerabilities.

Strategy 5: **Develop shared public datasets and environments for AI training and testing**. Develop and enable access to high-quality datasets and environments, as well as to testing and training resources. A broader, more diverse community engaging with the best data and tools for conducting AI research increases the potential for more innovative and equitable results.

Strategy 6: **Measure and evaluate AI systems through standards and benchmarks**. Develop a broad spectrum of evaluative techniques for AI, including technical standards and benchmarks, informed by the Administration's Blueprint for an AI Bill of Rights and AI Risk Management Framework (RMF).

Strategy 7: **Better understand the national AI R&D workforce needs**. Improve opportunities for R&D workforce development to strategically foster an AI-ready workforce in America. This includes R&D to improve understanding of the limits and possibilities of AI and AI-related work, and the education and fluency needed to effectively interact with AI systems.

Strategy 8: **Expand public-private partnerships to accelerate advances in AI.** Promote opportunities for sustained investment in responsible AI R&D and for transitioning advances into practical capabilities, in collaboration with academia, industry, international partners, and other non-federal entities.

Strategy 9: **Establish a principled and coordinated approach to international collaboration in AI research.** Prioritize international collaborations in AI R&D to address global challenges, such as environmental sustainability, healthcare, and manufacturing. Strategic international partnerships will help support responsible progress in AI R&D and the development and implementation of international guidelines and standards for AI.

The federal government plays a critical role in ensuring that technologies like AI are developed responsibly, and to serve the American people. Federal investments over many decades have facilitated many key discoveries in AI innovations that power industry and society today, and federally funded research has sustained progress in AI throughout the field's evolution. Federal investments in basic and applied research[3] have driven breakthroughs enabled by emerging technologies like AI across the board, including in climate, agriculture, energy, public health, and healthcare. Strategic federal investments in responsible AI R&D will advance a comprehensive approach to AI-related risks and opportunities in support of the public good.

Introduction to the *National AI R&D Strategic Plan: 2023 Update*

Advances in generating, collecting, processing, and storing data have enabled innovation in AI, allowing this technology to become ubiquitous in modern life and touch nearly every facet of daily activities, directly or indirectly. Besides the AI-enabled applications in smartphones and personal computers, applications of AI have streamlined logistics, accelerated scientific discovery, enabled more efficient design and manufacturing, and aided in detecting financial fraud. However, realizing AI's potential social and economic benefits and aligning it with American values requires considerable research investments, pursued in accordance with the principles of scientific integrity.

In February 2022, the Office of Science and Technology Policy (OSTP) issued a Request for Information (RFI)[4] requesting input from all interested parties on the development of this plan. Over 60 responses were submitted by researchers, research organizations, professional societies, civil society organizations, and individuals; these responses are available online.[5]

Many of the RFI responses reaffirmed the analysis, organization, and approach originally outlined in the 2016 and 2019 strategic plans. It is noteworthy that a majority of the RFI responses referred to aspects of ethical, legal, and societal implications of AI (Strategy 3) or safety and security of AI systems (Strategy 4). These responses underscore a heightened priority across academia, industry, and the public for developing and deploying AI systems that are safe, transparent, and improve equity, and that do not violate privacy. Responses to the RFI also emphasized the importance of supporting AI R&D that will develop systems capable of helping to address some of the foremost challenges and opportunities before the Nation today, including advancing personalized medicine; improving cybersecurity; addressing inequities; bringing efficiencies to manufacturing, transportation, and other critical sectors of the economy; ensuring environmental sustainability; and enabling the scientific discovery and innovation that will power the next generation of technological breakthroughs.

AI as a National Priority

The Biden-Harris Administration is committed to advancing responsible AI systems that are ethical, trustworthy, and safe, and serve the public good. The fiscal year (FY) 2023 President's Budget Request included substantial and specific funding requests for AI R&D, as part of a broad expansion of federally funded R&D to advance key technologies and address societal challenges.[6] The CHIPS and Science Act of 2022[7] and Consolidated Appropriations Act, 2023[8] reflect Administration and Congressional support for an expansion of federally funded R&D, including AI R&D.[9]

The memorandum on Multi-Agency Research and Development Priorities for the FY 2024 Budget[10] issued jointly by the Office of Management and Budget and OSTP likewise calls for agencies to prioritize R&D funding toward advancing national security and technological competitiveness, including trustworthy AI, among other critical and emerging technologies of national interest. This plan pursues the advancement of fundamental and translational AI research to make AI trustworthy, equitable, and both rights- and privacy-preserving.

The National AI Initiative Act (NAIIA) of 2020 established the National AI Initiative Office (NAIIO) to coordinate key AI activities across the federal government. This office, based in the White House OSTP, is the central point of contact for technical and programmatic information exchange on activities related to the National AI Initiative across the federal government, academia, industry, nonprofit organizations, professional societies, civil society, and state, local, and tribal governments. In addition, the NAIIO helps advance progress on the priorities outlined in this plan and implements a comprehensive approach to AI-related risks and opportunities in support of the public good.

While R&D activities and outputs inform governance and regulatory approaches, this plan leaves discussions of regulation or governance to other federal documents, such as the *Blueprint for an AI Bill of Rights* and the *AI Risk Management Framework*. In addition, issues related to scientific integrity and public access, while directly relevant to AI R&D, are largely left to other federal government documents as well.

Strategy 1: Make Long-Term Investments in Fundamental and Responsible AI Research

The United States has maintained its leadership in AI in large part because of continued and consistent investment in long-term, fundamental AI research. For example, many of today's AI-enabled products and services have their roots in federally funded fundamental research dating back decades. This trend has continued since the release of the 2019 Strategic Plan, with a notable increase in AI R&D funded by the federal government. For example, the AI Initiative and companion efforts funded by the Department of Energy (DOE) Office of Science have enabled groundbreaking discoveries in areas from fusion energy to SARS-CoV-2 understanding. The Nation must continue to foster long-term, fundamental, and responsible research in AI to allow for further discoveries and innovations with long-term benefits.

Investments in fundamental AI R&D span the spectrum from foundational to use-inspired research. For example, foundational investment in AI R&D drives forward learning, reasoning, planning, knowledge representation, computer vision, and beyond, with potential for scale-up and adoption in practice. Use-inspired AI research, meanwhile, contributes to advances in AI while also advancing areas such as agriculture, healthcare, manufacturing, economics, critical infrastructure, and sustainability, with the goal of engaging and improving all of society while respecting individual freedoms.

Of particular importance is the investment in the development of assurance and trust in AI systems, as reflected in Strategy 3 and Strategy 4. Research in these areas is essential for using AI in all fields, but it is particularly important in systems that involve safety or applications in which AI decisions affect individuals, groups, communities, and the environment. Most AI R&D thus far has focused on the advancement of AI for individual tasks. Additional work is needed to solve increasingly difficult science and technology challenges covering multiple domains and applications, moving toward the vision of general-purpose AI. AI R&D increasingly attempts to consider how various areas of AI work can fit together into an integrated system. As a result, this strategy includes priorities that continue to advance AI for individual tasks but also aim toward the vision of general-purpose AI systems. The priorities involve using the significant amount of available data for machine learning (ML) and knowledge discovery, improving the abilities of AI to perceive and act, and developing scalable, general-purpose systems to work in real and virtual environments.

Finally, developing a theoretical understanding of the capabilities and limitations of AI systems can inform what R&D should be done and is critical for enabling safe use of AI. For example, a better understanding of how deep networks construct effective representations could lead to new network designs that can reason about uncertainty more directly and without requiring as much data to train.

This strategy is divided along ten lines of effort: Advancing Data-Focused Methodologies for Knowledge Discovery; Fostering Federated ML Approaches; Understanding Theoretical Capabilities and Limitations of AI; Pursuing Research on Scalable General-Purpose AI Systems; Developing AI Systems and Simulations Across Real and Virtual Environments; Enhancing the Perceptual Capabilities of AI Systems; Developing More Capable and Reliable Robots; Advancing Hardware for Improved AI; Creating AI for Improved Hardware; and Embracing Sustainable AI and Computing Systems.

Advancing Data-Focused Methodologies for Knowledge Discovery

As discussed in the *Federal Big Data Research and Development Strategic Plan* from 2016,[11] new tools and technologies are needed to achieve intelligent data understanding and knowledge discovery. For example, progress on the development of more advanced AI systems will help identify useful information hidden in big data. Many open research questions revolve around the creation and use of data, including its veracity

and appropriateness for AI system training and its role in creating interpretable, reproducible algorithms. While much research has dealt with veracity through data quality assurance methods to perform data cleaning and knowledge discovery, further study is needed to improve the efficiency of data cleaning and labeling techniques, to create methods for discovering inconsistencies and anomalies in the data, to address privacy considerations, and to develop approaches for incorporating human feedback. Researchers also need to explore new methods to enable data and associated metadata to be mined simultaneously. Another major issue is the lack of adequate and representative data in many domains, such as healthcare. Techniques need to be developed to deal with the generation and curation of redacted data to facilitate ML for domains with sensitive data. These and other data concerns are addressed in Strategy 5.

Many AI applications are interdisciplinary in nature and involve heterogeneous data. Further investigation of multimodal ML is needed to enable knowledge discovery from a wide range of heterogeneous data types (e.g., discrete, continuous, text, spatial, temporal, spatiotemporal, graphs).

In addition to data, one of the fundamental challenges in current AI systems is the lack of a standard infrastructure to encode knowledge AI systems must process and interpret significant amounts of data to approximate human-like responses. Hence, it is important to have different kinds of data (e.g., causal, temporal, heuristic) encoded in a form that is open and accessible. As an example, an Open Knowledge Network[12] is one concept for making this knowledge accessible,[13] but there is a need for considerable research, including developing domain-specific knowledge repositories in standardized formats.[14]

Fostering Federated ML Approaches

New federated approaches to ML will be important in an increasingly interconnected world and amid growing concerns around data privacy and security.[15] Federated learning allows multiple computers or devices to collaborate in building a shared global ML model based on the data that is locally stored on each device. The overall process is a back-and-forth iteration that involves each device training a local model on its own data and then sharing only the model updates (not the data) to improve the global model. The global model is distributed back to the devices for further local training until the global model reaches a specified level of accuracy. Federated learning can improve the accuracy and fairness of such global ML models by including locally-protected data from a diverse and more representative range of users, devices, and other sources that may have data-sharing restrictions due to competitive, regulatory, or privacy concerns. The ability to process confidential information is critical to industries such as healthcare, finance, and telecommunications. Federated learning is one among a range of approaches for privacy-preserving data sharing and analytics.[16] Major research challenges arise in dealing with the heterogeneous characteristics of devices (memory capacity, computing power, network connectivity) and data (skewed data samples, different modalities such as images, video, text). Improved efficiency in ML model communication and updating from multiple devices into a shared global model, as well as better data protection and security approaches, are areas for continuing research focus.[17]

Understanding Theoretical Capabilities and Limitations of AI

While the goal for many AI algorithms is to address open challenges with general-purpose systems, there is not yet a good understanding of the theoretical capabilities and limitations for AI, nor of the extent to which such solutions are even possible with AI algorithms. Theoretical work is needed to better understand how some AI techniques, especially generative AI, work and their emerging properties. Building this understanding of what advanced systems can and cannot do is important for enabling safe and responsible use of AI. While different disciplines (including mathematics, control sciences, and computer science) are studying this issue, the field currently lacks unified theoretical models or

frameworks to understand AI system performance. Additional research is needed on computational solvability, which is an understanding of the classes of problems that AI algorithms are theoretically capable of solving, and likewise, those that they are not capable of solving. This understanding must be developed in the context of existing hardware, to see how the hardware affects the performance of these algorithms. Understanding which problems are theoretically unsolvable can lead researchers to develop approximate solutions to these problems, or even open new lines of research on new hardware for future AI systems.

Pursuing Research on Scalable General-Purpose AI Systems

A development toward scalable general-purpose AI is the emergence of so-called foundation models that are trained on large amounts of unlabeled data, usually using self-supervised learning, and can be adapted to many application domains such as law, healthcare, and science. Innovations continue to advance the frontiers of what foundation models can do on language and image tasks. Familiar examples of large pre-trained language models include BERT (Bidirectional Encoder Representations from Transformers), GPT-4 (Generative Pre-trained Transformer), and other AI systems with skills that might begin to resemble intelligence within certain domains. Additional R&D is necessary to minimize unwanted fabrications and harmful biases in generative AI. These models are prone to "hallucinate" and to recapitulate biases derived from unfiltered data from the internet used to train them. Further research is needed to enhance the validity and reliability as well as security and resilience of these large models, especially in response to adversarial attacks. Further research is also needed to develop techniques for explaining and interpreting model outputs. Additional work is needed to address privacy concerns related to training models on such large corpuses of data. Finally, appropriate safeguards will need to be conceptualized and designed into these systems.

Developing AI Systems and Simulations Across Real and Virtual Environments

An emerging trend in modeling and simulation is the development of "digital twins." A digital twin is a virtual representation or model that serves as the real-time digital counterpart of a physical object or process. Real-world applications include predictive maintenance of aircraft engines, urban planning and the management of smart cities, and additive manufacturing. A key requirement is that the physical system is instrumented so that the collected data is interactively shared with the digital or computational model of itself. The digital-twin approach enables smart automation of physical systems across real and virtual environments. Challenges specific to various applications, such as data completeness, quality, latency, and privacy, and the varying accuracies with which different phenomena can be modeled, are likely to lead to additional challenges for digital twins.[18]

Enhancing the Perceptual Capabilities of AI Systems

Perception is an intelligent system's window into the world. Perception begins with sensor data, which come in diverse modalities and forms, such as the status of the system itself or information about the environment. Sensor data are processed and fused, often along with *a priori* knowledge and models, to extract information relevant to the AI system's task, such as geometric features, brightness, velocity or vibration. Integrated data from perception forms situational awareness to provide AI systems with the comprehensive knowledge and a model of the state of the world necessary to plan and execute tasks effectively and safely. AI systems would greatly benefit from advancements in hardware and algorithms to enable more robust and reliable perception. Sensors must be able to capture data at long distances with high fidelity, often in real time. Systems for perception need to be able to integrate data from a variety of sensors and other sources, including edge devices and cloud systems, to determine what the AI system is currently perceiving and to allow the prediction of future states. Detection, classification,

identification, and recognition of objects remains challenging, especially under cluttered and dynamic conditions, and privacy considerations add additional complexity to designing systems for real-world applications. In addition, the perception of humans, including the states of their attention and emotion, must be greatly improved by using an appropriate combination of sensors and algorithms so that AI systems can work more effectively with people,[19] and as discussed in Strategy 2. Methods and techniques for calculating and propagating uncertainty throughout the perception process are needed to quantify the confidence levels that AI systems have in their situational awareness and to improve overall accuracy.

Developing More Capable and Reliable Robots

Robotics continues to harness most fields of AI, with special emphasis on perception, physical manipulation, and navigation. Significant advances in robotic technologies over the last decade are leading to potential impacts applications including manufacturing, logistics, medicine, healthcare, defense and national security, agriculture, and consumer products. One noteworthy development involves the introduction of AI-controlled robots into the research environment, yielding "autonomous laboratories" that can enable closed-loop synthesis characterization and testing systems capable of designing new drugs, chemicals, advanced electronic materials, and countless other materials far faster and with greater variety and precision than previously possible. Introducing autonomy into manufacturing can further accelerate the efficiency of product design coupled to product performance, while in biological systems, it can drive evolution of organisms to act as living sensors of specified environmental signals. While robots were historically deployed in static industrial R&D environments, recent advances involve close collaborations between robots and humans. Robotic technologies are now showing promise in their ability to complement, augment, enhance, or emulate human physical capabilities or human intelligence. However, scientists and engineers need to make these robotic systems more capable, reliable, easy-to-use, and safe.

Researchers need to improve robot perception to better extract information from a variety of sensors to provide robots with real-time situational awareness to inform decision-making. Progress is needed in cognition and reasoning to allow robots to better understand and interact with the physical world. An improved ability to adapt and learn, building abstract representations of low-level physical tasks, will allow robots to generalize their skills, self-assess their current performance, and learn a repertoire of physical movements from human teachers. Mobility and manipulation, especially when dealing with heavy objects, are areas for further investigation so that robots can move across rugged and uncertain terrain and handle a variety of objects dexterously. Robots need to learn to team together in a seamless fashion and collaborate with humans in a way that is trustworthy and predictable. Robotic systems must safely and cooperatively interact with humans and other actors in complex built and natural environments. Research is also needed to deal with adversarial systems, or systems that operate in disguise to collect data or interfere with legitimate operations. In general, robotic systems require research advances that will make them more capable and reliable, easier to use, and safer.

Advancing Hardware for Improved AI

While AI research is often outwardly associated with advances in software, the performance of AI systems has been heavily dependent on the hardware on which they run. The current renaissance in deep learning and generative AI is directly tied to progress in graphics processing unit (GPU)-based[20] and accelerator-based hardware technology and the associated improved memory, input/output, clock speeds, parallelism, and energy efficiency.

Developing hardware optimized for AI algorithms will enable even higher levels of performance than those of GPUs. One example is "neuromorphic" processors that are inspired by the organization of the brain and, in some cases, optimized for the operation of neural networks.[21]

Hardware advances can also improve the performance of AI methods that are highly data intensive. Advances in storage technology would also benefit the deployment of AI systems. Continued research is also needed to allow ML algorithms to efficiently learn from high-velocity data, including distributed ML algorithms that simultaneously learn from multiple data pipelines. More advanced ML-based feedback methods will allow AI systems to intelligently sample or prioritize data from large-scale simulations, experimental instruments, and distributed sensor systems (e.g., smart buildings and the Internet of Things). Such methods may require advances in input hardware, including dynamic input or output decision-making, in which choices are made in real time to store data based on importance or significance, rather than simply storing data at fixed frequencies.

Creating AI for Improved Hardware

Just as improved hardware can lead to more capable AI systems, AI systems can also improve the performance and resource (e.g., energy) usage of hardware.[22] This reciprocity will lead to further advances in hardware performance, since physical limits on computing require novel approaches to hardware designs.[23] One example is where AI is being used to predict high-performance computing (HPC) performance and resource usage and to make online optimization decisions that increase efficiency; more advanced AI techniques could further enhance system performance. AI can also be used to create self-reconfigurable HPC systems that can manage system faults when they occur, without human intervention.[24]

Improved AI algorithms can increase the performance of multicore systems by reducing data movements between processors and memory. In practice, the configurations of processes in HPC systems are never the same, and different applications are executed concurrently, with the state of each different software application evolving independently over time. AI algorithms need to be designed to operate online and at scale for HPC systems. HPC systems are governed by physical and mathematical laws, which both determine and constrain their performance, and AI algorithms that incorporate these laws into their design will be able to more efficiently optimize AI hardware design in a virtuous loop, leading to even more powerful AI implementations.

CHIPS and Science Act of 2022[25]

Passed by Congress and signed into law by President Biden to boost U.S. semiconductor production and the Nation's research and innovation enterprise,[26] *the bipartisan and bicameral CHIPS and Science Act of 2022 provides a generational opportunity to advance U.S. leadership in semiconductor design and manufacturing, as well as in accelerating breakthroughs in emerging technologies such as AI. AI is an essential element of microelectronics research, development, and manufacturing, and is intimately tied to the Nation's ability to sustain leadership in this industry. Advances in AI R&D will power future microelectronics devices and systems. Of particular interest is the timely opportunity to develop an overall co-design framework for all stages of microelectronics design, development, fabrication, and application.*[27] *Co-design involves multidisciplinary collaboration that accounts for the interdependencies among materials design, device physics, computer architectures, memory, interconnects, and the software for developing next-generation computing and networking systems. At the same time, a strong microelectronics innovation ecosystem resulting in next-generation computing and network systems will lay the foundation for the breakthroughs and innovations that will ensure that the United States remains in the vanguard of competitiveness across a wide range of fields and sectors, including AI R&D.*

Embracing Sustainable AI and Computing Systems

The rising computational cost of developing and operating state-of-the-art AI systems warrants significant attention. The proliferation of data-intensive AI is expected to dramatically increase computational demands and the associated environmental impacts. There is an urgent need to design resource-aware AI algorithms, systems, and applications that consider broader notions of sustainability beyond simply energy consumption. Sustainable AI also depends on research in environmental sustainability within and across all layers of the computing stack and the data management and use lifecycle. This requires a shift in research toward embracing design for sustainability that treats sustainability impacts as first-order metrics and on equal standing with performance, reliability, usability, and operational energy efficiency.

Strategy 2: Develop Effective Methods for Human-AI Collaboration

Effective methods for human-AI collaboration have become an increasingly important priority as AI becomes more prevalent throughout society. Fully autonomous systems that involve little or no human interaction will continue to be crucial for applications in industry (e.g., automated factories, control of energy systems), hazardous domains (e.g., deep space, radioactive environments), and other areas. However, other applications, ranging from disaster recovery to scientific discovery, are most effectively addressed by a combination of humans and AI systems working together in a way that leverages their respective strengths and mitigates risk. Indeed, the promise of future AI applications requires fully understanding human-AI teaming and collaboration.

This strategy recognizes the growing importance of sociotechnical and human factors and addresses the need for multidisciplinary research in enabling effective human-AI collaboration. It is divided along five lines of effort: Developing the Science of Human-AI Teaming; Seeking Improved Models and Metrics of Performance; Cultivating Trust in Human-AI Interactions; Pursuing Greater Understanding of Human-AI Systems; and Developing New Paradigms for AI Interactions and Collaborations.

Developing the Science of Human-AI Teaming

Teaming is a complex relationship requiring a deep understanding of human decision-making processes and their interactions. Human-human teaming is supported by a substantial body of knowledge, models, and methods for enhancing team performance. The relevance of this body of work for enabling more effective human-AI teams is unclear.[28] Research is needed to understand the human side of human-machine interactions. Studies are needed to gain an understanding of the attributes and requirements of successful human-machine teams for efficient and effective task performance. These studies will involve understanding the additional capabilities that a machine needs in order to become an effective teammate for the relevant tasks and environments and includes the modeling of human interactions. The first *National AI R&D Strategic Plan* defined three functional roles for AI systems in teaming contexts:[29]

- ***AI performs functions alongside the human***: AI systems perform peripheral tasks that support the human decision-maker. For example, AI can assist humans with working memory, short- or long-term memory retrieval, and prediction tasks.
- ***AI performs functions when the human encounters high cognitive overload***: AI systems perform complex monitoring functions (such as ground proximity warning systems in aircraft), decision-making, and automated medical diagnoses when humans need assistance.
- ***AI performs functions in lieu of a human***: AI systems perform tasks for which humans have very limited capabilities, such as for complex mathematical operations, control guidance for dynamic systems in contested operational environments, aspects of control for automated systems in harmful or toxic environments, and in situations to which a system should respond very rapidly (e.g., in nuclear reactor control rooms).

Fully understanding human-AI teaming requires moving beyond these three functional roles, or today's models of humans as operators, and on to the idea of teammate relationships. To become true teammates, machines will need to be flexible and adaptive to the states of their human counterparts, as well as to the environment—to intelligently anticipate their human teammates' capabilities and intentions, and to generalize specific learning experiences to entirely new situations.[30] Each of these capabilities represents a research challenge. Other open questions that impact human-AI teaming include team composition, management of situational awareness, and interaction paradigms that govern the

amount of control given to the AI system, when that control is granted, and how that control is distributed and transitioned.[31]

Seeking Improved Models and Metrics of Performance

A traditional approach for building effective human-AI teams is to consider the capabilities of the humans and AI systems separately, and then to investigate how the team can be brought together in an optimal fashion. Qualitative and descriptive models of human-AI performance will need to develop into predictive computational models that can assess the relative value of teaming compositions, processes, interface mechanisms, and other characteristics. Human-AI team collaborations are difficult to model well. Ensuring that the team's collective abilities are significantly better is a grand multidisciplinary challenge across such areas as psychology, decision sciences, economics, and human factors engineering, among others. The challenges are compounded when accounting for unexpected events and the issues of situational awareness, trust, and the potential for human and AI biases. The collaboration types of human-AI teaming models will also differ among human-assisted AI decision-making, AI-assisted human decision-making,[32] pure AI decision-making, and AI-assisted machine decision-making. Significant amounts of research are required on the theories, models, data, and computational tools needed for measuring, modeling, simulating, analyzing, and understanding the effectiveness of human-AI teams.

Cultivating Trust in Human-AI Interactions

The opaque nature of the programming and decision processes within AI systems is a potential barrier to the trust needed for effective human-AI teaming. One key challenge for humans is an expectation that mechanical and automated systems will behave in a deterministic way. Given similar conditions and inputs, the system should respond in the same way as before. However, AI systems may behave in non-deterministic, or unpredictable, ways in response to imperfect, noisy, and complex real-world information or even simply because they are stochastic by design. Furthermore, continuous learning systems will evolve over time. Another challenge is related to the accuracy of AI systems and appropriately calibrating understanding of system outputs that could be incorrect. Trust is recognized as a key factor associated with the use of AI systems.[33] Research is needed on how to establish and maintain appropriately calibrated trust among teammates in uncertain conditions and environments.[34]

Pursuing Greater Understanding of Human-AI Systems

Greater trust in and overall success of human-AI teaming will stem from the lessons learned from failures that can be replicated and studied to determine what went wrong. "Recorders" are important in all AI applications, and diagnosing failures in human-AI teams is a particularly acute need. As the science of teaming evolves, the need for testbeds and methodologies to measure the effectiveness of human-AI teaming in settings that replicate the complexity of the operational environments also becomes critically important. Pursuing research in virtual environments and developing testing methodologies that measure human teaming components and the user experience are important next steps for the deployment of successful systems that provide assurance.[35]

Developing New Paradigms for AI Interactions and Collaborations

Usability and human-centered design research demonstrate that interaction mechanisms, designs, and strategies highly influence user performance. Similar research is required to understand the usability and impact of interaction design in human-AI teaming. Specifically, research is needed to understand the influence of interaction design on decision-making, skill retention, training requirements, job satisfaction, and overall human-AI team performance and resilience. Research should also include the development of new paradigms for human-AI interaction to facilitate collaboration, decision-making actions, human

oversight, accountability, and control. A particular challenge is conveying enough information to the user while avoiding cognitive overload. Other interaction challenges include enabling the user and the machine to understand when to pass control back and forth, and how to maintain user engagement for proper situational awareness. Early research has shown that relying on a "human in the loop" is not a universally effective method for catching errors or ensuring sound decision-making, even though these human-in-the-loop applications may give the impression of a more robust or fair system. Finally, research into human-AI interactions and paradigms requires controlled experiments with end users. There is currently little research on the application of usability, human factors, and human-centered design to the development of AI-teaming applications.[36] Open research areas include understanding user needs and user requirements; the role of context in AI-teaming application use; the use of task analysis and iterative design methods; and ways to measure efficiency, effectiveness, and performance of AI-teaming applications. A research focus that includes end users, including the public where appropriate, provides a lens for studying how best to address existing structural inequalities in human-AI collaboration, promote the development of tools for safe and effective human-AI collaboration, and effectively train the human in human-AI collaborative situations.

Strategy 3: Understand and Address the Ethical, Legal, and Societal Implications of AI

AI technologies hold significant opportunity, but they also pose risks that can negatively impact individuals, groups, organizations, communities, society, the environment, and the planet. Like risks for other types of technology, AI risks can emerge in a variety of ways and can be characterized as long- or short-term, high or low-probability, systemic or localized, and high- or low-impact.[37] Without proper controls, AI systems can amplify, perpetuate, or exacerbate inequitable or undesirable outcomes for individuals and communities.

Since the *National AI R&D Strategic Plan: 2019 Update*, investment in AI and public awareness of the technology have increased. This has been accompanied by an increasing focus on the ethical, legal, and societal implications of responsible AI. According to the 2022 AI Index Report, publications addressing AI fairness and transparency have quintupled over the past decade.[38]

As a step toward addressing concerns related to the use of AI in society, the White House issued *A Blueprint for an AI Bill of Rights* that lays out five core protections to which everyone in America should be entitled when interacting with AI and automated systems: Safe and Effective Systems; Algorithmic Discrimination Protections; Data Privacy; Notice and Explanation; and Human Alternatives, Consideration, and Fallback.[39] In January 2023, NIST published a framework to better manage risks to individuals, organizations, and society associated with AI.[40] These complementary frameworks provide useful guidance to researchers as well as important avenues for further research.

This strategy identifies R&D priorities that can help to instantiate these principles—viewing them as design objectives, system properties, or requirements. Centering these principles in the development process is key to ensuring that AI broadly benefits the American people. The interdisciplinary field of values in design develops methods and approaches to build support and protection for rights and values into sociotechnical systems, or systems that integrate social and technical aspects. The study of ethical, legal, and societal aspects of AI is critical because decisions about the use and design of AI can require trade-offs between competing values, such as equity, fairness, privacy, and autonomy. These issues are challenging, even outside the realm of AI. But AI systems bring these concerns to the fore because they often do not attempt to model the decision-making processes, including the ethical and legal constraints on them, of humans or organizations, but rather analyze the results of such decision-making processes to develop their own heuristics for decision making.

The extent of work in developing AI principles and guidelines highlights growing concerns about the ethical, legal, and societal implications of AI. Ensuring that AI can be developed and used in accord with these principles will require an expansive R&D program. This strategy divides this R&D program along four lines of effort: Making Investments in Fundamental Research to Advance Core Values Through Sociotechnical Systems Design and on the Ethical, Legal, and Societal Implications of AI; Understanding and Mitigating Social and Ethical Risks of AI; Using AI to Address Ethical, Legal, and Societal Issues; and Understanding the Broader Impacts of AI.

Making Investments in Fundamental Research to Advance Core Values Through Sociotechnical Systems Design and on the Ethical, Legal, and Societal Implications of AI

There are several areas in which fundamental research is needed to advance our ability to design values-aligned AI systems and to understand the ethical, legal, and societal implications of AI. The use of design, in addition to policy, to protect security, accessibility, privacy, and accountability is an active area of research and practice. It moves beyond the retroactive analysis of impacts, developing the tools and

methodologies to reason about how best to protect values through mixed technical and policy choices. Research that supports values-aligned design approaches that consider multiple values, rather than one at a time, are essential to support the development of safe, equitable, and accountable AI systems. Technical work on issues such as explainability and interpretability are important to this work, as is technical work on privacy, harmful bias mitigation, and accountable design. For example, with many types of AI, such as deep learning models, explainability, and effective auditing of the model are difficult technical problems. Resolving the technical problems is only part of the challenge. Ensuring that users can make sense of system behavior in context (i.e., interpretability) is also essential. This is a sociotechnical problem that requires understanding the context in which the model will be operating, the needs and capabilities of the people who require the explanation, and the most effective methods of communicating the explanation. Research into communications and psychology finds that individuals generally overestimate how well they understand others' perspectives and how well their communications are understood.[41] Given this reality, interpretability will require fundamental research into communications.

There is also a need for technical research to develop metrics and frameworks for accountability, fairness, privacy, and bias. This includes research into language models and other generative AI systems to mitigate the production of harmful and biased outputs.[42] This must be accompanied by basic social science research into AI governance, which will include understanding how to engage stakeholders most effectively on AI issues throughout the AI life cycle, establishing legitimacy for AI development and implementation decisions, and performing intersectional research into how different people and communities understand, interact with, and are impacted by technology.

This work must be accompanied by research examining the potential implications of AI and developing evaluation and mitigation strategies. This research is needed to inform policy and governance approaches.

Understanding and Mitigating Social and Ethical Risks of AI

There is an immediate need for research to identify effective AI governance structures that can mitigate risks, build systems and implement AI worthy of public trust, and foster appropriately calibrated public trust in it through effective engagement. One possibility is to study and adapt approaches from other fields, such as medicine, that have robust governance and regulatory ecosystems. For example, institutional review boards for AI research to consider AI R&D's potential harm could be explored. Such an Ethics, Scientific Integrity, and Society Review Board could help steer the research community away from research questions that pose risks of downstream harm without any clear benefits, and could learn from past engagement with nuanced questions of harm and value tradeoffs. Similarly, the random control trials, validation, and ongoing monitoring used for drugs and medical devices may provide models for AI governance more generally. However, the governance of AI will vary depending on the context of use and approaches to validating efficacy and safety vary across sectors in relation to risk. The need for robust governance and oversight structures appropriate to domains of use, which is relevant to all fields of scientific endeavor, is particularly acute in AI as the pathways from ideas to impacts have become especially short.

Social science research exploring the introduction of AI systems into organizations, professions, and fields is necessary to develop a richer understanding of how AI alters the production of knowledge, shapes understandings of professional responsibility, shifts accountability across institutional actors, and shapes the relationships between organizations and the populations they serve.

Stakeholder engagement can be advanced by studying how to adapt deliberative civic engagement processes to AI governance and develop new methods to elicit stakeholder feedback. These social science and regulatory tools can empower communities to weigh in on AI's public- and private-sector uses, legal

and ethical issues, and societal implications. Broadened participation can also promote diversity and equity in shaping data collection, storage, and management practices; developing regulatory oversight and guidance; and creating equitable policy solutions.[43]

Finally, R&D can determine how best to teach and communicate about AI governance structures and sociotechnical approaches for various audiences, be they researchers, research subjects, technologists, policymakers, other stakeholders, or the public. As previously stated, a fundamental truth in the field of communications[44] is that people overestimate how well they are understood by others. There are ethics and scientific integrity requirements in some technology curricula, but it is important to systematically identify and promulgate the most effective ways to integrate these concepts into the learning process to ensure that people have the tools to engage with these issues effectively and consider their actions and decisions in broader contexts. There are other urgent technology issues that require ongoing R&D as well. For example, the use of personal data in AI systems raises privacy concerns, highlighting the importance of privacy-enhancing technologies such as homomorphic encryption, differential privacy, and secure multiparty computation to mitigate these concerns. There is also a need for tools to identify and mitigate harmful bias across datasets, particularly in new training data. Overall, mechanisms to develop, assess, and maintain AI systems that mitigate risk and maximize benefit are keenly needed.

Using AI to Address Ethical, Legal, and Societal Issues

AI system development, when approached in a manner that mitigates bias and harm and is done in accordance with the civil rights, civil liberties, and interests of those affected by the system, can help address complex societal challenges. Properly developed, AI can help provide data-driven inputs as society tries to address issues in domains that advance equity, climate change adaptation and mitigation, employment, and healthcare, especially for those traditionally underserved. AI often exacerbates bias, but ongoing research has shown that it can also be used to identify and mitigate harmful bias in current practice.[45] Different AI tools need to be developed and adapted to face the challenges in different domains: the AI capabilities needed to optimize healthcare will differ from those needed to address environmental sustainability. There is also reason for caution in these endeavors, as technology solutionism, where technological solutions are advocated for challenges for which they may be inappropriate or ineffective, has been problematic in a number of scenarios.

A few general capabilities are needed for AI to better be able to help address broader societal issues. First, as noted above, AI can be used to counter harmful bias. Understanding how AI can reduce inequities stemming from systemic, structural, and individual bias is an important area of research. This would enable a range of analyses of the use of AI in managing harmful bias.[46] Existing research has shown that some well-known mathematical definitions of bias[47] make inherently conflicting recommendations, so an ongoing challenge is developing sociotechnical mechanisms to resolve conflicts in the decision-making pipeline. Indeed, research in this area must be sociotechnical, focusing on real world implementations, in particular institutional and regulatory contexts, and account for the policies, professional and organizational obligations that structure interactions and reliance between humans and AI results.

Second, research is needed to ensure that use of AI capabilities advance equity rather than exacerbating inequity. For example, if only wealthy hospitals can take advantage of AI systems, the benefits of these technologies will not be equitably distributed. Research to make beneficial AI accessible in historically underrepresented communities will help ensure that those in greatest need of these capabilities can use them. This research will include making AI capabilities affordable and ensuring that AI is understood and can be integrated into existing systems.

Many historically underrepresented communities may not be represented in datasets typically used to train AI systems, nor included in development processes. This limits the ability of these communities to benefit from the AI systems. While noteworthy efforts are being made to connect with a broader set of communities, additional research is needed to identify these types of gaps and address them more fully.

Finally, there is an international dimension to these challenges. For truly global concerns (e.g., pandemics), international approaches are needed, as discussed in Strategy 9. In addition to the concerns about access and serving the underserved, AI that can be adapted to societies with different legal, ethical, and political commitments while respecting human rights and democratic values is essential.

Understanding the Broader Impacts of AI

AI promises to bring vast changes to society. While many of those changes will be positive, there are likely to be negative consequences, and these impacts are also likely to be unequally distributed. R&D in the ethical, legal, and societal implications of AI is needed to understand, anticipate, and mitigate harm as well as understanding the distribution of likely benefits. Large-scale research into sociotechnical feedback loops, using the tools of systems engineering and complexity theory, is needed to understand how AI interacts with society. This includes the systematic study of the tradeoffs in societal benefits and risks of using, using in different permutations, or not using AI in each context.

One specific area that requires this approach is the future of work.[48] There has been some attention to the future of work, the potential for AI to displace workers, and the need to retrain workers for a rapidly changing economy.[49] There is also a need to understand what AI does to workplaces and how it impacts work safety and overall well-being.[50] This is especially needed with the growing popularity and abilities of generative AI systems. Similar inquiry is needed across social institutions, such as research into how AI will change how patients experience the healthcare system and how students are educated.

Finally, R&D is needed to identify means to counter malicious uses of AI, for example the generation of deep fakes and manipulation of social media. Here, too, there may be technical responses, but sociotechnical study is needed as well. The Information Integrity Research and Development Interagency Working Group (IWG) recently published recommendations, which will in turn require innovative approaches to implement in future AI systems.[51]

Strategy 4: Ensure the Safety and Security of AI Systems

While AI systems offer promise in providing performance improvements in several different applications, their increased complexity, rapidly evolving technology base, and significant data needs can lead to increased risks derived from their deployment. The result is an emerging emphasis on the safety and security of AI systems, which requires an inherently interdisciplinary approach. For the purposes of this strategy, to appropriately discuss the needs surrounding these risks, the terms "safety" and "security" will carry the definitions laid out in the "Assessing and Improving AI Trustworthiness: Current Contexts and Concerns" workshop report,[52] which defines safety as mitigating against a system producing new harm, and security as monitoring a system's integrity. This usage is consistent with the NIST AI RMF.

Critical areas for research focus include the development of testing methods that can scale with the increasing demands of modern AI systems and complex systems-of-systems, and improved methods for ensuring the security of AI systems against input data manipulation, model inversion, and other forms of adversarial attack. Ultimately, a combination of additional investment in standards, systems, and research is needed to calibrate trust in the performance of deployed AI systems.

Key to this strategy is addressing the fundamental question of what level of testing is sufficient to ensure the safety and security of non-deterministic and/or not fully explainable systems before their deployment. The process of securing and making safe AI, as discussed in the first *National AI R&D Strategic Plan* and the 2019 update, must be incorporated in all stages of the AI system life cycle, from the initial design and data/model building to verification and validation, deployment, operation, and monitoring. "Safety by Design" must therefore be an important part of the AI R&D portfolio, particularly as models are increasingly used by non-technical users and incorporated across a broad range of platforms and applications. Adopting AI systems that are unsafe or insecure will potentially lead to harm, and uncertainty about the safety and security of these systems will stymie AI adoption (as discussed in Strategy 3).

Standards setting (discussed in depth in Strategy 6) is critical in the effort to develop safe and secure AI; it requires research into how effective and meaningful standards can be developed and adapted to the broad array of applications in which AI will be used.

This strategy divides the safety and security R&D program along two lines of effort: Building Safe AI and Securing AI.

Building Safe AI

As AI becomes commonplace and its applications proliferate, the need for a national approach to research on AI and safety becomes increasingly urgent. This research includes developing methods for creating, evaluating, deploying, and monitoring AI that are focused on safety.

With datasets and models growing larger and more complex, there is an urgent need for solutions that can scale with these larger systems. Additionally, there is a need for a national innovation ecosystem[53] that can democratize the tools for accessing AI models at this scale, making analysis of such large models accessible to the broader community and beyond the groups that are capable of investing in the infrastructure to develop and deploy them. This approach would enable a larger field of researchers to address safety and security concerns relating to these larger models, including those related to bias, accuracy, and functionality.

More research is needed to develop safe human-machine interactions. Exploration of new formal methods could characterize boundaries of behavior and bring much-needed rigor to safety-critical AI algorithms and applications. These techniques include novel programming languages and compilers to

develop more robust AI, formal verification techniques for AI systems that could provide assurances of safety, and neurosymbolic programming that could bridge the areas of deep learning and program synthesis. Addressing AI systems-of-systems, in which the AI system is only one component of a larger system, or a large system composed of many AI and classical subsystems, is one of the most pressing challenges in testing systems at scale. Methods and approaches need to be developed to independently verify subsystems within the context of their operating framework and to evaluate the performance of the overall construct to ensure that the ensemble will operate safely, and that security of the overall system is not harmed by subsystem interactions.[54] New testbeds and prototyping facilities could enable this area of research.

Long-term risks remain, including the existential risk associated with the development of artificial general intelligence through self-modifying AI or other means. Other long-term risks are related to the possible deep entangling of AI systems into all parts of daily life and systems, which may cause large-scale societal or environmental issues that are currently difficult or impossible to predict; or specification gaming, whereby an AI system gradually learns to achieve numerical requirements but somehow avoids accomplishing the desired task. These risks are difficult to quantify currently and need additional research.

Securing AI

The national need for secure AI is growing as software and systems are growing more complex but also increasing our collective vulnerability to cybersecurity threats.[55] This is echoed in both the desire for additional training among practitioners within government agencies, and in the recognition of AI security as an independent field of study adjacent to cybersecurity and AI. Such a field is needed to address the many open questions still surrounding the multifaceted issues of AI security, such as the need for appropriate metrics for goal alignment, protection against adversarial attack, scalable methods, and trade-offs between interpretability and accuracy.[56]

Adversarial AI includes “data poisoning,” in which AI training or input data are manipulated, and other forms of adversarial attacks against AI, such as targeting systems linked to the AI or manipulating objects in the physical world. Changes to audio or visual data that cannot be perceived by humans can change how an AI system processes data. This is particularly salient for ML systems. Some of these risks can be identified through red teaming, where trusted partners act as adversaries in a simulated compromise attempt, and other risks can be mitigated through mathematical modeling. Research is needed to better enable both approaches.

An additional threat to AI systems is the existing AI-development supply chain. As only a few tools are currently used for AI system development and deployment, there is a risk that these tools could present a vehicle for systems to be compromised. Efforts should be made to protect these tools from manipulation, and to develop a more robust toolset to protect the AI-development supply chain.[57]

Research into improved methods for ensuring the security of AI systems is critical, including work on improving the capability of systems against input data manipulation, model inversion, and other forms of adversarial attack. The many-against-one nature of an AI system’s vulnerability needs to be addressed, as AI systems open more pathways for disruption that most systems. Only one needs to be successful, while the AI system must protect against all.[58]

Strategy 5: Develop Shared Public Datasets and Environments for AI Training and Testing

Progress in AI is increasingly linked to data and computation. The availability of well-purposed (i.e., legally, and ethically collected and managed) data for AI training and testing enables research applications, scientific discovery, and operational efficiencies. A well-designed cyberinfrastructure can aid data federation, support metadata, track provenance, and enable reproducibility. Similarly, access to advanced computing, including HPC, edge computing, cloud resources, traditional desktop computing, and emerging computing paradigms, drives AI innovation. At the same time, the challenges for researchers to access at-scale data and computing resources continue to pose significant obstacles for the field. For example, many AI researchers are departing academia for industry settings where such resources are more readily available. Similarly, with resources concentrated in large technology companies and well-resourced universities, the divide between those with access and those without has the potential to adversely skew AI research. Researchers who lack access to rigorous data and computation will simply not be competitive.

With the goal of democratizing access to at-scale AI data and compute resources, the NAIRR Task Force has published a roadmap and implementation plan for a national research cyberinfrastructure that would connect researchers to data, computation, testbeds, and associated training.[59] The work of the NAIRR Task Force builds on other efforts to enhance access to these diverse resources. For example, since their introduction in 2016, the Findable, Accessible, Interoperable, and Reusable (FAIR) Guiding Principles[60] for data have seen tremendous acceptance by the scientific research community. In 2019, the OPEN Government Data Act[61] mandated that the federal government, through collaboration and coordination, provide open data, engage in evidence-building activities, enhance statistical efficiency, uphold confidential information protection, and, where data is about humans, respect privacy.

This strategy is divided into four lines of effort: Developing and Making Accessible Datasets to Meet the Needs of a Diverse Spectrum of AI Applications; Developing Shared Large-Scale and Specialized Advanced Computing and Hardware Resources; Making Testing Resources Responsive to Commercial and Public Interests; and Developing Open-Source Software Libraries and Toolkits.

Developing and Making Accessible Datasets to Meet the Needs of a Diverse Spectrum of AI Applications

Sustaining access to well-purposed training and testing datasets is crucial for ensuring scientifically reliable, reproducible, ethical, and equitable results. While there is value in simplified and synthetic datasets for algorithm research, development, and testing, other datasets must be sufficiently representative to effectively tackle challenging, real-world problems. Dataset documentation must include data provenance and references to previous work with the data. These will facilitate the ability of researchers to compare multiple datasets generated by the same system or process and clearly describe changes in the system that yield any differences in the data. The technical and sociotechnical infrastructure necessary to support reproducible research has been recognized as an important challenge—and is essential to AI systems as well. The current infrastructure and the level of documentation and curation of datasets are mostly inadequate and vary significantly by research area.

Many machine learning applications need their training data to be integrated, cleaned, and refined in order to be usable. Specific, detailed user and system requirements, methods by which the data were collected, and any factors (e.g., sensor noise) that lead to noise or other artifacts in the data will drive how the data can be made "ready" for use in AI applications (Strategy 6). As with computational infrastructure and testbeds, data infrastructure needs to be designed to meet the specific demands of AI applications. The infrastructure should be developed with community input and continually re-evaluated and updated as technology advances and the research problems and endeavors evolve. Many government datasets are already available to researchers and students on various websites and platforms (e.g., substantial National Aeronautics and Space Administration [NASA] Earth Science datasets are available through NASA's Distributed Active Archive Centers[62]), though not all are well known, easy to find, or easy to use.[63, 64] For example, the different processes by which government data and government-funded data are publicly available or licensed for external use can be confusing and time consuming to navigate.

Easing access to federal government data,[65] when appropriate, can increase the use of existing resources for developing and studying AI.[66] This includes potential benefits from creating agreements, templates, or processes for data access that can be shared across agencies so that researchers and students no longer face an array of different requirements to gain access to different datasets. Government data are often suitable for inclusion in standardized training datasets and benchmarks within the AI research community. When appropriate, agencies may identify opportunities to share data across agency boundaries or contribute agency data to standardized resources that have the potential for widespread usage by researchers and practitioners. Likewise, the development of better methods for moving data and computing closer together is important, as these methods decrease the costs and time associated with analysis. For example, the National Institutes of Health (NIH) S&T Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative[67] enables access to rich datasets and advanced computational infrastructure, tools, and services. This can involve investment in federated and distributed learning approaches and supporting infrastructure (a potential crosscut with Strategy 8) and support of federation of data archives that increasingly make use of cloud services. This includes improving data access and facilitating use of ML and other data analytics methods.

As needed, because of privacy or security, the federal government can provide secure platforms for access to government data, as the National Geospatial-Intelligence Agency and the Department of Defense's Chief Data and Artificial Intelligence Office have done, and as is envisioned for the recently authorized National Secure Data Service demonstration.[68] Data providers also may be able to make otherwise confidential data available through removal of some information not critical to analytics, as is done with NASA's commercial aviation data.[69] Such data access also requires having Identity Access Management in place.

Alongside easing access to data is making data more discoverable and usable. There is a research need for exploring rational methods for linking related datasets, such as the development of open knowledge[70] graphs for data discovery and compilation across sources and users.

In addition, increasing capabilities for and public access to synthetic data generation can be helpful when corresponding real data cannot be made available because of privacy concerns or because they are available only in small quantities. Other means by which to increase the amount of useful data, such as crowdsourcing and active learning to increase the number of labels, should also be considered. Thoughtful investment may be necessary to supplement existing datasets through the careful collection of more representative data (for example, the *All of Us* Research Program at NIH, which has focused participant enrollment to ensure a large percentage includes individuals who have been traditionally under-represented in biomedical research[71]). These resources can be made available more easily and in greater

quantity through the use of public-private partnerships. This requires investment in additional mechanisms to facilitate collaborations with the private sector through funding and novel mechanisms that allow academia, non-government organizations (NGOs), and other researchers to use private-sector and public-sector resources. Additionally, government or government-funded data should account for historically underrepresented communities and groups being underrepresented in data. For example, support is needed to generate natural language processing tools for underrepresented languages and those that do not have a written form.

Datasets also need to document context (e.g., labeling process and sample bias) to be useful, in part by encouraging the capture of metadata that can be computationally queried and assessed. Depending on the problem, additional effort may be needed to capture sociological and contextual information and to enhance secure and privacy-preserving data linkages between informative data assets. Computer science and data experts in government may need to engage social scientists and other relevant experts in this process.

An important part of the government provision of data for AI is ensuring their use in a manner that reflects American values (a crosscut with Strategy 3). For example, it is critical to advance both technological and governance methods that preserve privacy, including protecting against revelations of personally identifiable information when publicly available government data are combined with other data. There also is a need for research on effective data governance that allows releasing data under secure platforms that control access to or removal of content.

Beyond ensuring that AI does not result in harm from the release of personally identifiable information, there is also a public need to demonstrate how datasets can help overcome inequities. This could include creating curated datasets for analysis of past inequities, such as digitizing “redline” maps originally developed by the Home Owners Loan Corporation in the 1930s. This analysis can be used to avoid replicating disparities, and can help increase access to safe and sanitary housing combined with flood maps.

Developing Shared Large-Scale and Specialized Advanced Computing and Hardware Resources

Innovation in AI is dependent not just on data, but also on access to advanced computing. Large universities, federal laboratories, and private-sector firms often have access to such capability, which can take the form of HPC, cloud, hybrid, and/or emerging systems. But many researchers and students at smaller institutions of higher education, minority-serving institutions, community colleges, secondary schools, and startups and small businesses may have less access or fewer resources to purchase the computing needed to conduct AI R&D.

To lower barriers to entry into AI R&D, enhanced access to advanced computational resources is necessary, particularly for the variety of new users who otherwise would face financial, logistical, or capacity challenges to engaging in the AI research ecosystem. Expanded access should be provided by leveraging existing resources in all sectors, augmenting the capacity of federally provided resources as appropriate, creating new research computing infrastructure to serve the AI R&D community, and providing financial support where needed.

To this end, the NAIRR Task Force has put forward a roadmap and implementation plan leading to a mix of computational resources (i.e., on-premises and commercial cloud, dedicated, and shared resources) with a range of central processing unit (CPU) and graphics processing unit (GPU) options with multiple accelerators per node, high-speed networking, and sufficient memory capacity.

Making Testing Resources Responsive to Commercial and Public Interests

The growing complexity of AI systems has created a need for equally robust AI testing resources. In many cases, these resources are developed alongside the technology itself by private industry or the research community at large. However, this approach to AI testing leaves certain concerns unaddressed. First, novel AI research often experiences limited testing because of difficulties with replication.[72] Second, AI systems developed by private industry often do not have mechanisms for public qualitative evaluation and testing.[73] Finally, for research institutions or private industry, certain areas of testing, especially surrounding large-scale AI models, are not economical to pursue in isolation, and these areas are left underexplored as a result.

Federal AI testing resources, primarily in the form of testbeds and testing frameworks, can address the limitations of existing testing paradigms. For example, the NIST Facial Recognition Vendor Test (FRVT) helps provide insight into the accuracy of otherwise private facial recognition algorithms,[74] and the Guaranteeing AI Robustness against Deception program at the Defense Advanced Research Projects Agency supports novel testing mechanisms in ML security by means of a virtual testbed, toolbox, and benchmarking dataset.[75] Similar approaches could be employed for other common AI applications, including voice-assistant software and recommender systems.

Expanding the scope of federal testing resources is critical to the healthy adoption of emerging AI systems. As agencies develop new testbeds, both foundational AI and application-specific AI should be considered. In addition, new testing efforts may also inform (or conversely, be contingent on) emerging AI standards and benchmarks. Awareness and coordination between these efforts is likely to improve the efficacy of both. Finally, because of the rapid rate of AI R&D, test framework designers should pay close attention to changing trends in software, hardware, and research focus to plan for the longevity of their work.

Developing Open-Source Software Libraries and Toolkits

Another area for government investment involves providing access to and support for open AI software libraries. Access to and continued support for libraries and toolkits can accelerate R&D, from conducting fundamental research through facilitating technology translation, as the same libraries may be used for a wide range of services, including commercial ones. The growth in open software libraries and toolkits has enabled a corresponding growth in AI applications and skills. Researchers and students across sectors use open-source tool sets. Government agencies also develop open software libraries and toolsets specific to mission needs in which industry lacks market incentives to develop the products for the government or other sectors. Many agencies and agency-funded researchers make code available through GitHub or other commonly used commercial platforms that provide resources for researchers and students. Also, prior to commercial interest, the federal government may need to incentivize continued development, maintenance, and curation of software and tools to prevent them from becoming outdated. As an example, NSF's Pathways to Enable Open-Source Ecosystems program aims to harness the power of open-source development for the creation of new technology solutions to problems of national and societal importance.

Strategy 6: Measure and Evaluate AI Systems through Standards and Benchmarks

Standards, benchmarks, testbeds, and their adoption by the AI community are essential for guiding and promoting R&D on AI systems, and the recognition of this role continues to rise in the United States and globally. Both the 2019 Executive Order on Maintaining American Leadership in Artificial Intelligence[76] and the NAIIA[77] explicitly call out the importance of standards. In addition, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) have convened a joint technical subcommittee on AI (ISO/IEC Joint Technical Committee 1,[78] Subcommittee 42 on Artificial Intelligence[79]) to develop standards and associated considerations for AI systems. The surge in AI-related standards activities has outpaced the launch of new AI-focused benchmarks and evaluations, particularly as related to the trustworthiness of AI systems. Considerations of fairness and bias in benchmark datasets have become increasingly important. Much more plentiful are benchmarks that test the application-level performance of AI algorithms (e.g., false-positive or false-negative rates for classification algorithms) and benchmarks that quantify the compute-level performance of AI software and hardware systems. These efforts need to be expanded to sociotechnical evaluations of AI systems and assessing the broader impact of AI technologies.

Assessing, promoting, and providing assurances on all aspects of AI trustworthiness require measuring and evaluating AI technology performance through benchmarks and standards. Beyond being safe, secure, reliable, resilient, interpretable, and transparent, trustworthy AI must preserve privacy while detecting and avoiding inappropriate bias. Claims of trustworthiness and compliance must also be verifiable and certifiable. As AI systems evolve, so will the need to develop new metrics and testing requirements for validation of these essential characteristics.

The evaluation, standards, and benchmarks of AI systems need to acknowledge underrepresentation of certain communities, and approaches are required to solve this issue theoretically and practically. There is a need to proactively identify underrepresented communities, including Indigenous groups, and to include diverse stakeholders and domain experts from government, academia, the private sector, and civil society, including representatives from differently sized organizations and lower- and middle-income socioeconomic groups and countries, to ensure fairness and prevent bias in the development of standards and benchmarks.

It is necessary to develop standard ways to measure, test, and report the potential societal effects of commonly used datasets such as ImageNet.[80] A user-friendly acknowledgment of privacy considerations regarding an organization's collection, use, and sharing of personal information as well as a label of ethical assurance could help promote trustworthy AI systems. A potential example of a voluntary program that effectively fosters trust based on standards-based assessments is the Leadership in Energy and Environmental Design program for building certifications.[81]

This strategy divides the areas in which additional progress must be made along five lines: Developing a Broad Spectrum of AI Standards; Establishing AI Technology Benchmarks; Increasing the Availability of AI Testbeds; Engaging the AI Community in Standards and Benchmarks; and Developing Standards for Auditing and Monitoring of AI Systems.

Developing a Broad Spectrum of AI Standards

The development of standards must be hastened to keep pace with the rapidly evolving capabilities and expanding domains of AI applications. Standards provide requirements, specifications, guidelines, or characteristics that can be used consistently to ensure that AI systems meet critical objectives for

functionality and interoperability, and that they perform reliably and safely throughout their operational lifecycle. There is a need to achieve consensus-based provision of precise definitions of technical terms and consistent terminology (e.g., AI, autonomy, transparency, explainability, and interpretability) within the domain of safety and security.

Adoption of standards brings credibility to technological advancements and facilitates an expanded interoperable marketplace. One example of an AI-relevant standard that has been developed—by the Institute of Electrical and Electronics Engineers—is P1872-2015 (Standard Ontologies for Robotics and Automation). This standard provides a systematic way of representing knowledge and a common set of terms and definitions. These allow for unambiguous knowledge transfer among humans, robots, and other artificial systems. Another example of an AI-relevant standard is ISO/IEC 22989:2022 (Information technology—Artificial intelligence—Artificial intelligence concepts and terminology), developed within ISO/IEC Joint Technical Committee 1 SC 42,[82] which defines terminology and concepts related to AI. Additional work in AI standards development is needed across all subdomains of AI. For example, there should be support for an Open Knowledge Network standard to address the limitations of current deep learning systems.[83,84]

As part of the NAIIA, Congress directed NIST to develop an AI risk management framework, a voluntary tool that organizations can use to evaluate, assess, and manage risks that may result from the use of AI. The Framework leverages standards and best practices that organizations can use to achieve stated outcomes. Further support is needed for research initiatives that tackle questions related to understanding and operationalizing the risks and harms of applications of AI systems so that risk ratings, certifications, and insurance become feasible for AI systems.

One of the key challenges to developing standards in new areas is achieving a sufficient shared understanding of the underlying issues necessary for the standards to serve useful purposes. Additional efforts are needed to inform and create standards that support the following:

- ***Software engineering*** to manage system complexity, sustainment, and security, and to monitor and control emergent behaviors.
- ***Functionality and trustworthiness*** to assess an AI system's validity and reliability, safety, security and resilience, privacy, interpretability, and bias as well as the tradeoff among the mentioned trustworthiness characteristics.
- ***Metrics*** to quantify factors impacting performance and compliance with standards.
- ***Safety*** to evaluate risk management and hazard analysis of systems, human-computer interactions, control systems, and regulatory compliance.
- ***Usability*** to ensure that interfaces and controls are effective, efficient, and intuitive.
- ***Interoperability*** to define interchangeable components, data, and transaction models via standard and compatible interfaces.
- ***Security*** to address the confidentiality, integrity, and availability of information, and cybersecurity.
- ***Privacy*** to control for the protection of information while being processed, when in transit, or while being stored or used.
- ***Fairness and interpretability*** to ensure that AI systems' harmful bias is managed and able to help humans understand their operation and outputs.

- ***Flexibility*** to avoid a rigid lock-in that may lead to workarounds, lack of compliance, and other harmful spillover effects.
- ***International collaborations*** to that can support responsible AI development and thoughtful policy both domestically and abroad.
- ***Traceability*** to provide a record of events (their implementation, testing, and completion), and to curate data.
- ***Domains*** to define use-inspired standard lexicons and corresponding frameworks.

For example, consider the domains of healthcare and manufacturing. In the United States, nearly $4 trillion is spent on healthcare each year, and healthcare data today are fragmented, often incomplete, and difficult to access. This limits AI capabilities in healthcare. A significant improvement to AI algorithms—for safety, reliability, and trust—can be obtained through improvements to data access, standards for metadata that capture important social characteristics, and a balance that achieves privacy for the individual and enables ethical, legal, and societal validation. Manufacturing is a major contributor to the U.S. economy, and research is needed on data standards for AI in manufacturing.[85] With the passage of the CHIPS and Science Act of 2022,[86] there will be an expanded role for AI in semiconductor design and manufacturing, where standards will aid in further innovation.

Impact assessments can expose preventable harm, encourage consultation with affected communities, and standardize the information available for further research about which AI systems are used in which contexts and for what purposes. The development of methodological standards for these assessments is especially critical for ensuring that impact assessments are done in the public interest, and for preventing the proliferation of assessments that manipulate or obscure harmful impacts of applications of AI systems.

Finally, the real-world performance and energy efficiency of AI models remain poorly quantified. One recent study found that the carbon footprint of a large language model nearly doubled when equipment manufacturing and idle consumption during training were taken into account.[87] Development and adoption of standards, including the documentation of hardware and training details, may allow better management of the nuances of the environmental performance of AI, which in turn, informs its responsible use.

Establishing AI Technology Benchmarks

Benchmarks, comprising tests and evaluations, provide quantitative, qualitative, or mixed method measures for developing standards and assessing compliance to standards. Benchmarks drive innovation by promoting advancements aimed at addressing strategically selected scenarios; they additionally provide objective data to track the evolution of AI science and technologies. To effectively evaluate AI systems, relevant and effective testing methodologies and metrics must be developed and standardized. Standard testing methods will prescribe protocols and procedures for assessing, comparing, and managing the functionality and trustworthiness of AI systems. Standard metrics are needed to define measures to characterize AI systems, including, but not limited to accuracy, complexity, trust and competency, risk and uncertainty, explainability and interpretability, unintended bias, comparison to human performance, and economic impact. It is important to note that benchmarks are driven by data. Research needs to be done on how to construct benchmarks that test for more than accuracy under assumptions that data are independent and identically distributed. Strategy 5 discusses the importance of datasets for training and testing.

Frequently, AI performance is evaluated using only a handful of typical metrics (e.g., accuracy, precision). While these metrics are useful for development, they do not provide end-to-end contextual information. For instance, for AI systems developed to improve maintenance, metrics associated with repair times and overall system availability will be more informative than the accuracy of a maintenance action prediction. Hence, testing should also use metrics that are operationally relevant to the use context for an AI system. Further, datasets used should be dynamic in the sense that they should be enhanced by new data and connected to domain problems with human committees and evaluators, not just provide metrics numbers.

While prior efforts provide a strong foundation for driving AI benchmarking forward, they are limited by being domain-specific. Additional standards, testbeds, and benchmarks are needed across a broader range of domains to ensure that AI solutions are broadly applicable and widely adopted. The federal government should validate and collate evaluations created by independent researchers to create a catalog of approved tests for deployed models and those in development. It is useful to emphasize characterizing performance across use conditions; an AI system can be deployed with constraints that limit it from working in conditions where its performance is degraded, or it is more likely to do harm.

Increasing the Availability of AI Testbeds

As noted in one recent report: "Testbeds are essential so that researchers can use actual operational data to model and run experiments on real-world system[s] … and scenarios in good test environments."[88] While some AI testbeds exist,[89] adequate testbeds are needed across all areas of AI. As an example, although the federal government has massive amounts of unique and mission-sensitive data, many of these data cannot be distributed to the extramural research community. Appropriate programs should be established for academic and industrial researchers to conduct research within secured and curated testbed environments established by federal agencies. AI models and experimental methods can be shared and validated by researchers if they have access to these test environments, affording AI scientists, engineers, and students unique research opportunities not otherwise available. It is necessary to create standardized testing frameworks and benchmarks that allow for effective evaluation of AI systems to ensure that they are performing appropriately for a given use case in a way that is fair, safe, secure, and reliable, as well as to develop new tools for test, evaluation, validation, verification, and monitoring—and to assure the reliability of AI systems over their full domain of use and life cycle. A NAIRR, as outlined by the NAIRR Task Force, would support this goal.

Engaging the AI Community in Standards and Benchmarks

Government leadership and coordination are needed to support standardization and encourage its widespread use in government, academia, and industry. The AI community—comprising government, academia, industry, and civil society, including end users—must be energized to participate in developing standards and benchmark programs. As each government agency engages the community in different ways based on its role and mission, community interactions can be leveraged through coordination to strengthen their impact. This coordination is needed to collectively gather user-driven requirements, anticipate developer-driven standards, marshal the expertise of the AI R&D community, and promote educational opportunities. User-driven requirements shape the objectives and design of challenge problems and enable technology evaluation. Community benchmarks allow R&D to define progress, close gaps, and drive innovative solutions for specific problems. These benchmarks must include methods for defining and assigning ground truth. The creation of benchmark simulation and analysis tools will also accelerate AI developments. The results of these benchmarks will help match the right technology to the

user's need, forming objective criteria for standards compliance, qualified product lists, and potential source selection.

Industry and academia are the primary sources for emerging AI systems. Promoting and coordinating R&D subject matter expert participation in standards and benchmarking activities are critical. As solutions emerge, opportunities abound for anticipating developer- and user-driven standards through sharing common visions for technical architectures, developing reference implementations of emerging standards to show feasibility, and conducting precompetitive testing to ensure high-quality and interoperable solutions, and to develop best practices for technology applications.

AI practitioners carry critical domain expertise on testbeds for AI, and their expectations can play a major role in developing AI systems. As a result, it is crucial to broaden AI education to a variety of industries and encourage the AI community to further engage in standards development for evaluating AI systems. Furthermore, it is even more crucial to bridge the gap between practitioners' expectations and AI researchers to achieve a harmonious development cycle between AI technology developers and users. It is also important to collaborate with industry consortia and affected communities.

Developing and adopting standards, as well as participating in benchmark activities, comes with a cost, and R&D organizations engage in these activities when they see significant benefit. Updating acquisition processes across agencies to include specific requirements for AI standards in requests for proposals will encourage communities to further engage in standards development and adoption. Community-based benchmarks such as the Text Retrieval Conference[90] and FRVT[91] also lower barriers and strengthen incentives by providing types of training and testing data otherwise inaccessible, fostering healthy competition between technology developers to drive best-of-breed algorithms, and enabling objective and comparative performance metrics for relevant source selections. There is also a need for improved testing methodologies and resources that would allow agencies to directly evaluate cloud-hosted AI capabilities.

Developing Standards for Auditing and Monitoring of AI Systems

AI systems will need to be properly audited and regularly monitored to identify and mitigate risks, both technical (e.g., accuracy, reliability, and robustness) and sociotechnical (e.g., bias and privacy). There are many unresolved research questions about how to effectively audit and monitor AI systems, and the scalability of auditing is emerging as a significant practical challenge. As AI systems proliferate and find their way into more realms of human activity, it is imperative to develop scalable auditing techniques, create new types of qualitative analysis tools, train enough people to carry them out, receive feedback from humans in the loop, and build institutional capacity in government and industry to undertake, oversee, and respond to audits.

Strategy 7: Better Understand the National AI R&D Workforce Needs

Rapid advancements in AI continually impact the workforce by creating a growing demand for qualified computer and information science professionals and for new skills in the broader workforce now or soon using AI systems daily.

Within the United States, computer and information science positions are projected to grow by 22 percent between 2020 and 2030.[92] Private industry is expected to lead this demand with its sustained financial support and access to advanced computing facilities and datasets.[93] The resulting economic growth could be large: AI research is expected to contribute as much as $11.5 trillion in cumulative growth across G20 countries alone over the same period.[94]

Fortunately, interest in AI study and careers remains high. However, U.S. academic institutions are struggling to keep pace with the explosive growth in student interest and enrollment in AI.[95] Furthermore, while booming enrollments are common at the undergraduate level in AI-related fields such as computer science, doctoral enrollment trends show steady decreases in U.S. citizen and permanent resident enrollments. This has impacts on the AI workforce, particularly in government positions such as those requiring security clearances. Overall, these trends put an onus on government to better understand workforce needs and take steps to develop and support AI talent, with the goal of creating a sustainable AI workforce for government, academia, and industry. Moreover, the trends in computing and information science need to be complemented by those in other disciplinary areas that also contribute to AI discovery and innovation, such as the social and behavioral sciences, economics, and systems engineering.

This strategy is divided into ten lines of effort: Describing and Evaluating the AI Workforce; Developing Strategies for AI Instructional Material at All Levels; Supporting AI Higher Education Staff; Training/Retraining the Workforce; Exploring the Impact of Diverse and Multidisciplinary Expertise; Identifying and Attracting the World's Best Talent; Developing Regional AI Expertise; Investigating Options to Strengthen the Federal AI Workforce; Incorporating Ethical, Legal, and Societal Implications into AI Education and Training; and Communicating Federal Workforce Priorities to External Stakeholders.

Describing and Evaluating the AI Workforce

The *National AI R&D Strategic Plan: 2019 Update* described some elements of the AI workforce, marking it interdisciplinary, dynamic, and data-centric, and called for "additional studies on the current and future national workforce needs for AI R&D."[96] Much work remains to adequately and accurately define who makes up the "AI workforce"—including their demographics—and what those persons need to know and do.[97] Moreover, given the dynamic nature of the AI field, this analysis must be redone periodically to keep pace with changes in AI and the workforce.

Data on the current AI workforce, including its participants, their roles and tasks, and the knowledge and skills required to perform these tasks, is critical to properly understanding the workforce's abilities, gaps, and needs. With extensive, properly prepared, and well-ordered data, the United States can gain reliable clarity on the status quo of the AI workforce. Clarifying the understanding and priorities for a strong AI workforce in the United States will help focus efforts and investments across sectors. Further, illuminating the demographic disparities and gaps in the AI workforce will provide policymakers and human resource professionals with information necessary to address these disparities and increase equity and diversity. Facilitating this work could provide incentives for employers in various sectors to improve their data collection methods, consolidate existing workforce datasets, and support the creation of a modernized labor database. Research necessary to facilitate and reinforce this effort should focus on developing

proper data, knowledge and skills, and workforce curation and analysis techniques, including the enterprise and architectural needs of a modernized workforce.

The CHIPS and Science Act of 2022 takes a step in this direction by authorizing NSF to generate a study of U.S. universities that conduct high-impact AI research to better understand what factors enable AI progress. In particular, the report should contain information about university computing power, dataset availability, specialized curricula, faculty and graduate students, sources of federal and non-federal research funding, and industry and other partnerships, with the intention of implementing successful practices across the academic ecosystem. Such a study could help ensure that AI workforce needs such as adequate resources and institutional support are well-understood and integrated with complementary workforce needs such as beneficial training and skills.

Developing Strategies for AI Instructional Material at All Levels

The United States would benefit from making AI research accessible to a wide range of Americans. Moreover, exposing students at all levels, starting at the primary and secondary levels, to AI and data science prepares them for successful integration into a world that is rapidly adopting AI.

High-quality, domain-specific, and appropriately challenging lessons are needed for introducing students to critical thinking skills that will help them understand and evaluate AI systems. The research required to properly identify and curate the right content for a given area and level of study requires considerable effort. Further research is needed to sort out the best pedagogy and media through which to convey this content, as well as to identify and curate best practices for training instructors. It is important to facilitate the engagement of other public- and private-sector entities in this research and ensure demographic and cultural equity in that engagement.

Additionally, it is important that any AI materials, training programs, or systems are accessible, equitably promulgated, and broadly representative, especially given current inequalities among students' and educators' access to resources.

Supporting AI Higher Education Staff

At the most advanced levels, some AI researchers in university positions (e.g., tenured or tenure-track faculty) are moving toward industry R&D. Workforce efforts should also study opportunities to ensure a sufficient university workforce to educate future generations of the AI workforce in two-year and four-year colleges and universities, spanning associate's, bachelor's, master's, and doctoral degree programs. These efforts could include joint appointments enabling faculty to engage across sectors.

Training/Retraining the Workforce

Similarly, there are opportunities to upskill individuals who will be using AI systems in their current lines of work. To do so, the federal government must prioritize developing programs and systems that support the identification and recruitment of AI talent and the assessment, training, and validation of AI skills and knowledge. These programs and systems should leverage AI to maximize their relevance and impact. They should instill standardization, interoperability, and democratization. Once developed, these programs and systems will continue to foster the development of AI-competent workforce and support personnel displaced by AI deployment.

Pursuant to this, research partnerships among government, academia, and industry must be cultivated. These partnerships should prioritize creating on-demand courses that benefit from the best pedagogy and oversight available to a diverse workforce. These courses must be equitably and accessibly available to all.

Additionally, rapid and well-informed development of grand challenges for worker training and retraining programs and systems should be explored. Grand challenges are an exceptional joint research, development, and acquisition method that allows the government to leverage its partnerships, technologies, and other assets to tackle hard problems such as workforce development.

Exploring the Impact of Diverse and Multidisciplinary Expertise

Safe and equitable AI development and deployment requires a broad understanding of the people and places affected by AI as much as deep technical knowledge of the AI itself. Multidisciplinary education across diverse fields can be beneficial for ensuring fair and equitable access to information and opportunity, democratization of new and emerging technologies, and the development of a diverse a marketplace of ideas around technology use and development. Moreover, AI must be developed and managed from a holistic perspective that integrates knowledge from various disciplines and backgrounds to foster an interdisciplinary and transdisciplinary approach that considers the needs of all Americans. As such, hiring for teams that make and/or use AI should emphasize diversity from academic, professional, and experiential perspectives.[98]

To facilitate this approach, federal researchers should leverage their unique position and perspective to spearhead research into the roles and impacts of different areas of study on the realities and future of AI. As a result, researchers will understand how to engage diverse perspectives and align their efforts and resources with national needs and priorities, as well as across all sectors.

These actions should be taken in addition to other efforts to increase the diversity of communities, identities, races, ethnicities, backgrounds, abilities, cultures, and beliefs involved in AI R&D. The federal research community should prioritize research on the best way to increase demographic and cultural representation in the federal AI workforce.

Identifying and Attracting the World's Best Talent

The United States is home to an abundance of talent in many areas but has historically relied on foreign-born talent to bolster its technology workforce—especially in R&D in emerging technologies. Half of the current AI experts in U.S. academia and industry were born outside of the United States.[99,100,101] Federal resources can support university, industry, and civil society efforts to host visiting students and scholars with pathways to U.S. citizenship.

Fostering international partnerships with foreign governments and universities in support of Strategy 9 also serves to address this strategy.

Developing Regional AI Expertise

The size and diversity of the United States makes it useful to synthesize inputs and expertise from various parts of the country. Leveraging different geographical regions can facilitate equitable and broad dispersion of AI training and the economic opportunities, while also accessing a diversity of represented perspectives for contribution and feedback. In addition, by coordinating geographically concentrated resources such as data and computing infrastructure, a highly skilled local workforce, and complementary industry presence (e.g., cybersecurity, data science), regional synergies could foster local participation in the AI-enabled economy, facilitate high-quality AI training, and accelerate AI research progress at the national level.

In complementary fashion, federal efforts should be directed toward fostering regional efforts that enable access to the AI economy in historically underserved areas, including in rural areas and on tribal lands.

Such efforts will ensure that opportunities are provided to a broad array of Americans, allowing AI research efforts to draw upon diverse perspectives that may be underrepresented in current efforts.

Investigating Options to Strengthen the Federal AI Workforce

The federal government should fund and execute research efforts to determine the feasibility of different options for strengthening the federal AI workforce. Federal efforts could accelerate and leverage the growing number of AI K-12 education and workforce development programs to build partnerships among government, academia, and industry, helping to recruit and train early-career private-sector professionals and traditional students to engage with federal agencies in the areas of digital transformation, data management, analytics, and AI. Such partnerships could potentially also include rotations and/or work in local, state, and federal government organizations, accelerating and supporting the deployment of AI across the public sector. Efforts to strengthen the federal AI workforce should include a focus on training federal AI professionals so they are able to design systems that support the rights and safety of the public and mitigate the residual risks to them.

Along these lines, the AI Training Act directs the federal government to develop and provide an AI training program for a substantial portion of the federal AI workforce. The CHIPS and Science Act of 2022 authorizes NSF to study and establish a federal AI scholarship-for-service program to recruit and train the next generation of AI professionals across the federal government. It also clarifies that individuals studying AI-related topics are eligible for the existing NSF CyberCorps: Scholarships for Service program, enabling the program to begin specifically recruiting individuals with an interest in applying AI skills to federal projects in the future.

Incorporating Ethical, Legal, and Societal Implications into AI Education and Training

The ethical, legal, and societal implications of AI have become increasingly salient in recent years and will continue to be so. As such, it is vital for those who develop, use, and oversee AI systems to be conversant in these topics and committed to upholding the associated values. Experts are needed who are conversant in these issues and in data science and AI systems, and who can help educate the workforce and inform education and upskilling curricula. Also needed are policy, law, and governance experts who are conversant in the ethical, legal, societal, and technological aspects of AI topics.

Unfortunately, current academic programs that create qualified experts in any one of these three areas are challenged to offer education in the others. To address this challenge, the federal government should support undergraduate and graduate programs, as well as postdoctoral opportunities that designed to build interdisciplinary competencies, and support research into and dissemination of education materials on ethical, legal, and social aspects of AI for integration in AI education and training programs.

Communicating Federal Workforce Priorities to External Stakeholders

Educating private-sector institutions, higher-education institutions, and the public about the federal government's workforce needs and priorities and how to support fulfilling them is a critical step along the path to intersectoral alignment and optimization. Workforce description, recruitment, and development must be fair, transparent, and accountable, and that expectation should be conveyed consistently to all stakeholders in all lines of effort. Federal agencies can carry out these communications through posting of success stories in the media, outreach to small and minority-owned businesses, representation in talks and booths at industry trade shows, participation in scientific conferences that span the spectrum of disciplines surrounding AI, and program funding announcements. Other opportunities include education and workforce programs intertwined with research, as in the National AI Research Institutes and extant collaborations among university faculty and students, industry representatives, and the federal government.

Strategy 8: Expand Public–Private Partnerships to Accelerate Advances in AI

American leadership in science and engineering research and innovation is rooted in the U.S. government-university-industry R&D ecosystem. As the American Academy of Arts and Sciences has written, "America's standing as an innovation leader" relies on "establishing a more robust national Government-University-Industry research partnership."[102] Since the release of the first *National AI R&D Strategic Plan*, multiple administrations have amplified "the increasing importance of effective partnerships between the federal government and academia, industry, other non-federal entities, and international allies to generate technological breakthroughs in AI and to rapidly transition those breakthroughs into capabilities."[103]

Over the last several decades, fundamental research in information technology conducted at universities with federal funding, as well as in industry, has led to new multibillion-dollar sectors of the Nation's economy. Concurrent advances across government, academia, and industry have been mutually reinforcing and have led to an innovative, vibrant AI sector. The growing importance of public-private partnerships was reflected in the addition of Strategy 8 in 2019, and has become more apparent since then, as highlighted here. The three forward-looking themes of this strategy are: Achieving More from Public-Private Partnership Synergies; Expanding Partnerships to More Diverse Stakeholders; and Improving, Enlarging, and Creating Mechanisms for R&D Partnerships.

Achieving More from Public-Private Partnership Synergies

The private sector often views AI as a high-potential new tool for business and operational interests, whereas public funding in AI research has focused on longer-term impacts and societal good. These complementary perspectives can and should be further integrated into an overall whole.

By leveraging resources, including facilities, datasets, and expertise, the strategists and participants in public-private partnerships will more rapidly advance science and engineering innovations. For example, sharing AI artifacts, models, data, and results serves to reduce resource use and redundancies. Similarly, government-university-industry R&D partnerships bring pressing, real-world challenges faced by industry to university researchers, enabling use-inspired research, and leveraging industry expertise to accelerate the translation of open and published research results into viable products and services in the marketplace for economic growth. Public-private partnerships are especially well served when they build on joint engagements among federal agencies that enable collaboration and better return on investment in areas where agencies' missions intersect.

Continued support for cross-government efforts[104] such as the National AI Research Institutes[105] is key to long-term R&D partnership progress. These coordinated investments advance responsible foundational and use-inspired AI research in collaborations that benefit from a range of direct and indirect partnerships among governments, academia, industry, non-profits, communities of practice, and civil society. Researchers trained in these environments are well-prepared to expand on such approaches in years to come.

Expansion and extension of multiple types of programs that provide opportunities for researchers from government, academia, and industry to spend time working in another sector would additionally enable federal funding agencies, academia, and the private sector to work more effectively with one another. The unique perspectives and capabilities of each sector enable mutual benefit. Industry's commercialization and scale-up of AI systems is assisted by universities' early-stage R&D and federal laboratories' focused materials, device, and measurement research, and specialized computing resources.

Expanding Partnerships to More Diverse Stakeholders

Expanding partnerships between the public and private sectors to include civil society organizations serves to involve those organizations' unique perspectives in the discussion of future developments regarding the implications of AI research, development, and use. Furthermore, development of R&D approaches that focus on accountability, equity, and respect for democratic values and human rights is critical in additional considerations of AI design, development, and deployment. Equitable access to partnerships, ethical guidelines in charters, early experience with developing technologies by a wider stakeholder community, and diverse insight into the strengths and weaknesses of participant approaches yield a more robust AI infrastructure and ecosystem. Also recommended is a more concerted effort to produce international collaborations with like-minded governments, multinational corporations, and the civil society organizations of other nations, which has the potential to accelerate advances in AI for global benefit, as detailed in Strategy 9.

Translation to practice that emphasizes ethics, safety, and public good is also of high importance. Involvement of civil society and its representative organizations is critical for discussion of equitable access and use, and of trustworthiness issues. Companies of all sizes publish guidelines and focus on reducing their risks in AI product development.[106] Small nonprofit organizations are major contributors to societal "AI for Good" efforts, often with substantial volunteer programs that leverage the growing pool of AI talent in the United States. Efforts to increase capacity for advisory services across sectors were recommended by the National Academies[107] to help build partnerships for public good.

Collaborations between public-private partnerships and civil society organizations are particularly critical in striving for equitable access to and use of AI, and in addressing concerns about societal implications to the global ecosphere (e.g., climate change, energy security, agricultural challenges, and healthcare). Governments and international bodies play a key role in setting standards for just and responsible use.[108, 109] An open-access AI collaboration ecosystem that includes large and small corporations, advanced computing capabilities and other resources only available in government agencies, and a diversity of organizations having varied perspectives, expertise, and capabilities can lead to a more ethical use of AI. These diverse collaborations lead to innovations and support new models such as partnerships between minority-serving institutions and National AI Research Institutes.[110]

Partnerships can also support the inherently interdisciplinary nature of AI R&D, which requires convergence between computer and information science, cognitive science and psychology, economics and game theory, the physical sciences, engineering and control theory, medicine, ethics, linguistics, mathematics and statistics, and philosophy. Bringing together this wide diversity of disciplines poses a significant research and logistical challenge (for example, in a common taxonomy), but the ultimate outcomes drive the development and evaluation of future AI systems that are fair, transparent, accountable, safe, and secure.

Improving, Enlarging, and Creating Mechanisms for R&D Partnerships

R&D is a team effort, often conducted by diverse groups operating in multiple institutions. Public-private partnerships require institutional arrangements to facilitate the pooling of resources for efficient return on investment of time and funding, faster outcomes, and positive impacts, and avoiding duplication of efforts. An array of potential configurations and mechanisms for public-private partnerships has been developed over the past few decades for a variety of AI applications.[111] Expanding the reach of existing mechanisms, improving their functioning and outputs for a more diverse set of participants and application spaces, and creating new forms of public-private partnerships are significant and valuable endeavors. Examples include the following:

- ***Individual project-based collaborations***. In these partnerships, government agencies pool resources and/or expertise with industry, NGOs, foundations, and academics to address a critical issue, such as safety and trustworthiness. This is a flexible and rapid approach, but often challenging to sustain and expand.
- ***Joint programs to advance open, precompetitive, fundamental research***. Government has traditionally played a critical role in supporting foundational research through grants and contracts (primarily at universities), for which there is no short-term commercial application, but instead advances the field as a whole.[112] Given the massive needs for expanded fundamental and use-inspired research, innovative methods to bring private-sector resources to these ends are critical but often challenging, given the short project timescales that profit-driven companies typically operate on. One example that addresses this challenge is the NSF Industry-University Cooperative Research Centers program,[113] which provides an NSF-supported institutional framework for industry to support precompetitive research at universities. In general, non-federal partners contributing research resources can receive intellectual property rights as governed by the Bayh-Dole Act.[114]
- ***Collaborations to deploy and enhance research infrastructure***. Large-scale AI research will require significant research infrastructure, including compute and storage resources. Joint projects between the government and private-sector partners can achieve economies of scale that enable access to necessary resources for all engaged parties. The NAIRR[115] is one example of a concept that could transform the national AI research ecosystem by providing researchers with access to computational, data, and training resources. Provision of such resources equitably to a large segment of stakeholders is critical to maximizing the impact of such collaborations.
- ***Collaborations to enhance workforce development, including broadening participation***. As discussed in Strategy 7, there is a tremendous demand for workers with AI skills. Every sector is competing for these valued workers. While there are many programs to encourage students to enter science, technology, engineering, and mathematics (STEM) fields, public-private partnerships should explore opportunities to pool resources to broaden the overall pipeline of AI R&D skills. New types of partnerships for curriculum development and new approaches to developing and implementing curricular standards for programs could be especially impactful by building broader capacity for AI education and training.
- ***Federal prize competitions***. Organizing competitions to address difficult research challenges has significant advantages for supporting R&D. In this form of partnership, the risks are introduced by the participant, not the government. Prize competitions represent only a tiny fraction of federal R&D spending, but they have proved effective at addressing a host of complex scientific and technical challenges. One difficulty has been getting from research to usable product. Research on how best to maximize impact should be enlarged. For example, competitions that are embedded in a broader structure of public-private partnerships might better enable the transition of the competition winners to deployment.[116]
- ***Data and model sharing***. Creating partnerships with the goal of sharing data and testbeds at scale could make a big difference in the breadth of availability of cutting-edge ML models. There are challenges, however, because trained models are a potential source of income and competitive advantage for the organizations that train them, and partnerships that require the release of these models to the public or other private organizations would likely cause these organizations to withdraw from such an arrangement. Innovation in standards and processes for equitable and responsible data sharing is urgently needed.

In each case, leveraging each partner's strengths for the benefit of all is vitally important to achieving the greatest impact.

Strategy 9: Establish a Principled and Coordinated Approach to International Collaboration in AI Research

The 2019 Organization for Economic Cooperation and Development (OECD) Recommendation on Artificial Intelligence included investing in AI R&D as the first recommendation for national policies and international cooperation.[117] While the United States leads the world in annual R&D spending, competitors seek to outpace these investments. The National Science Board's *U.S. State of Science & Engineering (S&E) 2022* report[118] found that no single nation leads in all aspects of science and engineering in today's world. In AI, the annual number of publications in the field has doubled between 2010 and 2020, and research production has become increasingly geographically dispersed.[119] Ensuring that the United States remains a central hub within the AI R&D ecosystem requires ongoing participation in international programs, infrastructures, datasets, and secure data-sharing mechanisms; continued access to global talent; sustained productive international cooperation; working with existing international structures that may already regulate the data, infrastructure, and talent that the AI R&D ecosystem needs; and effective public-private partnerships. International partnerships play a key role in facilitating efforts in all these areas.

In recognition of the importance of AI to economies across the globe, the U.S. government is working to address the pressing need for better access, sharing, management, standards, and common frameworks for data and computational resources, in addition to building out the design, development, verification, validation, and use of trustworthy AI. To support this, and future AI research, development, and deployment, the AI R&D community can facilitate opportunities for international research and exchange of ideas and expertise in line with Strategy 3, including the mutual cultivation of AI international standards and cross-border frameworks that promote responsible and trustworthy AI.

This strategy is divided into four lines of effort: Cultivating a Global Culture of Developing and Using Trustworthy AI; Supporting Development of Global AI Systems, Standards, and Frameworks; Facilitating International Exchange of Ideas and Expertise; and Encouraging AI Development for Global Benefit.

Cultivating a Global Culture of Developing and Using Trustworthy AI

Groundbreaking scientific research is an inherently collaborative and international activity. Given this, global partnerships for the development and deployment of AI capabilities are integral to advancing the state of the art in AI while ensuring that the full scale of its benefits is realized in a secure, equitable, and ethical way. Around the world, "trustworthy AI" is understood as AI with attributes that conform to various ethical, legal, and societal standards. For the United States, these attributes are lawful and respectful of our Nation's values; purposeful and performance-driven; accurate, reliable, and effective; safe, secure, and resilient; understandable; responsible and traceable; regularly monitored; transparent; accountable; and advancing equity.[120]

Federal research and partnership efforts can benefit from international collaboration with likeminded nations to discover and promulgate methods to support AI R&D and innovation that build public trust and confidence and realize shared values and social priorities such as equity, fairness, accountability, transparency, reliability, security, and safety. These collaborations come in many forms and through many mechanisms; examples include MOU10 (2022) with Australia's Commonwealth Scientific and Industrial Research Organization,[121] which has initiated a jointly funded research program that includes equitable and trustworthy AI; and an administrative arrangement between the United States and European Commission to further research on AI in application areas including extreme weather and climate

forecasting, emergency response management, health and medicine improvements, electric grid optimization, and agriculture optimization.[122]

U.S. leadership in multilateral fora such as the OECD and the Group of Seven (G7) has resulted in the 2019 OECD Recommendation on AI and the launch of the Global Partnership on AI.[123] This work has paved the way for promoting research that aligns with U.S. interests and values, including safe and ethical use of AI and building a global community of practice. The United States should continue to engage and lead in these international organizations and fora to signal an interest in R&D cooperation and to send a clear message about shared interests in supporting AI R&D, innovation, and cooperation that builds public trust and confidence and respects applicable international law, individual privacy, and human rights.

Additionally, U.S. agencies should evaluate the risks of pursuing AI R&D collaboration with partners in countries that might not share democratic values or respect for human rights. When identifying opportunities for dialogue on shared AI concerns and priorities, careful consideration should be given to the benefits and risks of discussions with adversaries and competitors. In partnership with countries that share its core values, the United States should develop strategies to combat nefarious uses of AI, such as political oppression and coercion, criminal activities, violations of applicable international law, or social manipulation. Alignment of activities with the aims stated in Strategy 3 is vital.

Not only does international engagement foster research collaborations, but it also provides opportunities to directly engage international stakeholders to amplify the impact of R&D ties and showcase U.S. leadership. One can look, for example, to recent engagements with the United Kingdom and India.[124]

U.S. agencies can also consider R&D engagement with nations that currently lack robust AI R&D ecosystems to build research capacity and strengthen ties.

Supporting Development of Global AI Systems, Standards, and Frameworks

International cooperative research is needed to inform the development of shared and best available metrics, test methodologies, quality and security standards, development practices, and standardized tools for the design, development, and effective use of trustworthy AI systems. Of particular value are methods for secure data-sharing and methods for applying AI to areas of importance such as public health and sustainability. Also valuable are systems and environments that provide nations' domestic enterprises with access to the expertise and infrastructure garnered from increased international collaboration and investments. All of this is ultimately a prerequisite for achieving optimum scale and collaboration with international partners, and critical for bringing about an ecosystem around AI R&D designed from the beginning around principles such as those in the 2020 trustworthy AI executive order.[125]

Also in need of consideration are effective mechanisms for public-private partnerships and international arrangements, as discussed in Strategy 8. This work is especially complex and intersectional, but small-scale and similarly focused examples could help to guide agencies in pursuit of this research. One such example is the Declaration of the United States and the United Kingdom on Cooperation in AI R&D[126] to advance a shared vision of AI and to work toward a mutually supportive AI R&D ecosystem. Another is the recent commitment of the Quad (the United States, India, Australia, and Japan) to establishing various technical standards contact groups,[127] including a group for advanced communications and AI focusing on standards-development activities as well as foundational pre-standardization research.[128] Other fruitful avenues include investigating and optimizing the potential of joint solicitations for AI R&D with international partners, and of joint international AI research and computing infrastructures.

Along the way, it is critical that international cooperative research also focuses on data management, governance, and sharing. One key area of consideration is how to share data, especially if it is sensitive

data, in a safe and secure way among countries that have different information security standards and capabilities. Another is research into allowing interoperability among nations' systems while protecting data and data ownership so that data is treated in a safe and consistent way, leading to the development of trusted and durable mechanisms for cross-border data transfers for AI R&D collaboration. A third consideration could be how best to ensure a culture of transparency and disclosure that aligns with the principles of research integrity, both domestically and with allies and partners. Overall, it is key that U.S. agencies develop and establish appropriately rigorous standards, policies, and procedures for data sharing, data privacy, and the protection of intellectual property to safeguard data, privacy, and national security.

Facilitating International Exchange of Ideas and Expertise

Leading experts and innovators in emerging technologies are spread out over multiple countries and continents. Ensuring that ideas can flow among them and across locations is necessary for a shared global future of effective and trustworthy AI. Agency-to-agency collaborations and broader bilateral and multilateral cooperative arrangements provide an opportunity for the United States to address gaps by leveraging AI research expertise around the world. Such collaboration could be realized through existing programs, such as the Embassy Science Fellows Program,[129] U.S. Science Envoy Program,[130] Fulbright Program,[131] International Visitor Leadership Program,[132] and TechCamps,[133] through AI-centric tracks.

Agencies should consider how undergraduate and graduate AI R&D internships, international fellowships, and exchange initiatives can help build the U.S. STEM workforce. These international collaborations can expose researchers to diverse ideas, attract and retain top AI R&D talent, and foster long-term partnerships among U.S. AI researchers. Current programs, such as the U.S. Intergovernmental Personnel Act (IPA),[134] illustrate what potential partnerships would look like. IPA facilitates temporary exchanges among federal agencies and other organizations, including state, local, and tribal governments, colleges, and universities. Developing similar programs for short-term international exchanges could foster R&D activities and outcomes in the international context.

Additionally, grand challenges are effective and efficient mechanisms for governments to leverage partnerships, technologies, and other assets for the purposes of cooperative research, development, and acquisition that could be more widely used. Grand challenges have provided a leveling platform for multilateral approaches to international collaboration and have enabled highly directed and innovative means of finding solutions to complex societal and industrial challenges of interest to the United States as well as to global partners, such as those related to health and natural disasters as well as food security. Among the strengths of grand challenges is their ability to garner a highly varied set of participants across a diversity of sectors, including academic, industrial, and individual technology enthusiasts, and from all manner of origins and backgrounds. This strength is amplified and realized in an international context. The current U.S.-led series of Grand Challenges on Democracy-Affirming Technologies[135] are an example, having already demonstrated success via a U.S.-United Kingdom collaboration on a prize challenge for accelerating the development and adoption of privacy-enhancing technologies.

Encouraging AI Development for Global Benefit

Certain uses of AI run counter to the values and well-being of the United States, especially when AI is utilized for the purposes of political oppression, coercion, criminal activities, violations of international law, and social manipulation. To combat this threat, additional research is needed into the ways in which nefarious usage of AI may be countered. This research presents further opportunities to engage with the international community and leverage bilateral and multilateral partnerships with allies and partners to restrict competitors and adversarial nations from gaining access to or acquiring advanced AI tools and

associated technologies critical to U.S. national security and other interests. Mutually beneficial alliances and partnerships around AI provide the United States with a durable means of addressing global AI challenges, deterring aggressive behavior, assuring allies and partners, and supporting stability.

Though not created by AI, other existential threats to peace and security could also be countered via AI innovations. For example, as described previously, there is opportunity for co-investment with values-aligned countries in novel AI techniques to solve long-term global challenges such as those related to health, natural disasters, pollution, food production, and sustainability. In addition, investigation of methods of public outreach and engagement with the broader stakeholder community is important to spread awareness regarding capabilities and limitations of AI.

As global interest in and use of AI continues to grow, so does the importance of international cooperation in research and coordination in the field. The United States is already positioned as a leader in AI research and innovation. This existing leadership may be leveraged to realize the aims of safe and secure use of trustworthy AI; standardized effective AI infrastructure, including robust and equitable data-sharing practices; international cooperation and coordination of AI research; and development of AI for global benefit.

Evaluating Federal Agencies' Implementation of the NAIIA and Strategic Plan

It is important to evaluate federal agencies' efforts in support of the NAIIA and the nine strategies described in this Strategic Plan. The proposed metrics that follow, consistent with directives in the legislation (NAIIA, Division E, Section 5103(d)(2)), align with the strategies laid out in this document. These metrics serve as a strong basis to quantify progress by federal agencies in addressing the key challenges laid out in this Strategic Plan and will be included in the future federal AI R&D progress reports, which are updated by federal agencies every three years.

- Level of investment in AI R&D
- Level of investment in AI education and workforce development
  - Numbers of scholarships, fellowships, and traineeships awarded
- Number of multiagency programs supporting AI R&D and education and workforce development
  - Number of multiagency programs with non-federal partners
- Level of investment in a NAIRR
- Number and diversity of active users of a NAIRR
- Number of distinct datasets made available through a NAIRR
- Number of federally supported AI testbeds listed in the NITRD AI R&D Testbed Inventory[136]

The metrics listed represent an initial set and may be revised over time. Data supporting an associated evaluation will come from various sources, including the annual NITRD Supplement to the President's Budget and the federal government's AI Research Program Repository.[137]

List of Abbreviations and Acronyms

Acronym	Definition
AI	Artificial Intelligence
DHS	Department of Homeland Security
DOE	Department of Energy
DOE/AITO	Department of Energy, Artificial Intelligence & Technology Office
DOE/SC	Department of Energy, Office of Science
DOT	Department of Transportation
FRVT	Facial Recognition Vendor Test
FY	Fiscal Year
GPU	Graphics Processing Unit
HPC	High-Performance Computing
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
IWG	Interagency Working Group
MDA	Missile Defense Agency
ML	Machine Learning
MLAI	Machine Learning and Artificial Intelligence
MLAI-SC	Machine Learning and Artificial Intelligence Subcommittee
NAII	National Artificial Intelligence Initiative
NAIIA	National Artificial Intelligence Initiative Act Of 2020
NAIIO	National Artificial Intelligence Initiative Office
NAIRR	National Artificial Intelligence Research Resource
NASA	National Aeronautics and Space Administration
NCO	National Coordination Office
NGO	Non-Governmental Organization
NIH	National Institutes of Health
NIJ	National Institute of Justice (Department of Justice)
NIOSH	National Institute for Occupational Safety and Health
NIST	National Institute of Standards and Technology
NITRD	Networking and Information Technology Research and Development
NSF	National Science Foundation
NSTC	National Science and Technology Council
OECD	Organization for Economic Cooperation and Development

Acronym	Definition
OSTP	Office of Science and Technology Policy
R&D	Research and Development
RFI	Request For Information
RMF	Risk Management Framework
State	Department of State
STEM	Science, Technology, Engineering, and Mathematics
USDA	U.S. Department of Agriculture
VA	Department of Veterans Affairs

Endnotes

[1] There are multiple definitions of AI and AI systems used by the federal government, including from in the National Defense Authorization Act for Fiscal Year 2019 [Public Law 115-232, sec. 238(g)], the National Defense Authorization Act for Fiscal Year 2020 [Public Law 116-617, sec. 5002(3)], and the NIST AI Risk Management Framework, as well as a more encompassing view of automated systems articulated in the Blueprint for An AI Bill of Rights. The R&D priorities defined in this document are applicable and important to the full breadth of technologies covered by these definitions.

[2] NAIIO. (n.d.). National Artificial Intelligence Initiative. *NAIIO.* https://www.ai.gov

[3] The federal government supports 41 percent of basic research funding in the United States. Burke, A., Okrent, A. & Hale, K. (2022, January 18). *The State of U.S. Science and Engineering 2022.* The National Science Board. https://ncses.nsf.gov/pubs/nsb20221. Throughout this document, *basic research* includes both pure basic research and use-inspired basic research—the so-called Pasteur's Quadrant defined by Donald Stokes in his 1997 book of the same name—referring to basic research that has been used for society in mind. For example, the fundamental NIH investments in IT are often called use-inspired basic research.

[4] OSTP. (2022). Request for Information to the Update of the National Artificial Intelligence Research and Development Strategic Plan. *Federal Register*. https://www.federalregister.gov/documents/2022/02/02/2022-02161/request-for-information-to-the-update-of-the-national-artificial-intelligence-research-and

[5] NAIIO. (2022). Public input to the update of the National artificial intelligence R&D strategic plan. *NAIIO*. https://www.ai.gov/87-fr-5876-responses/

[6] The White House (2022). The Biden-Harris Administration FY 2023 budget makes historic investments in science and technology. OSTP. https://www.whitehouse.gov/ostp/news-updates/2022/04/05/the-biden-harris-administration-fy-2023-budget-makes-historic-investments-in-science-and-technology/

[7] CHIPS and Science Act of 2022, Pub. L. No. 117-167, 136 Stat. 1370, Divs. A and B (2022). https://www.congress.gov/117/plaws/publ167/PLAW-117publ167.pdf

[8] Consolidated Appropriations Act, 2023, Pub. L. No. 117-328 (2023). https://www.congress.gov/117/bills/hr2617/BILLS-117hr2617enr.pdf

[9] CHIPS and Science Act of 2022, Pub. L. No. 117-167, 136 Stat. 1370, Divs. A and B (2022). https://www.congress.gov/117/plaws/publ167/PLAW-117publ167.pdf

[10] The White House. (2022). Memorandum M-22-15, Multi-agency research and development priorities for the FY 2024 budget. *The White House*. https://www.whitehouse.gov/wp-content/uploads/2022/07/M-22-15.pdf

[11] NITRD. (2016). *The Federal Big Data Research and Development Plan*. NITRD. https://www.nitrd.gov/pubs/bigdatardstrategicplan.pdf

[12] NSF. (2023). Building the Prototype Open Knowledge Network (Proto-OKN): https://www.nsf.gov/pubs/2023/nsf23571/nsf23571.htm

[13] NITRD Big Data IWG. (2018). *Open Knowledge Network: Summary of the Big Data IWG Workshop October 4-5, 2017*. NITRD. https://www.nitrd.gov/pubs/open-knowledge-network-workshop-report-2018.pdf

[14] Computing Community Consortium. (2019). *Visioning Activity, Artificial Intelligence Roadmap*. Computing Research Association. https://cra.org/ccc/visioning/visioning-activities/2018-activities/artificial-intelligence-roadmap/

[15] Subcommittee on Future Advanced Computing Ecosystem and NITRD High End Computing IWG (2022). *Future Advanced Computing Ecosystem Strategic Plan FY2022 Implementation Roadmap*. NITRD. https://www.nitrd.gov/pubs/FACE-SP-FY22-Implementation-Roadmap.pdf

[16] Fast-Track Action Committee on Advancing Privacy-Preserving Data Sharing and Analytics. (2023). National Strategy to Advance Privacy-Preserving Data Sharing and Analytics. NITRD Subcommittee. https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Strategy-to-Advance-Privacy-Preserving-Data-Sharing-and-Analytics.pdf

[17] Kairouz, P., McMahan, H.B. et al. (2019). Advances and Open Problems in Federated Learning. arXiv (Cornell University).https://arxiv.org/pdf/1912.04977.pdf

[18] *Foundational Research Gaps and Future Directions for Digital Twins*. Nationalacademies.org. Retrieved March 27, 2023, from https://www.nationalacademies.org/our-work/foundational-research-gaps-and-future-directions-for-digital-twins

[19] See 2021 *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence*, which focuses on the anticipated uses and impacts of AI in the year 2030. https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf

[20] GPU is a power-and cost-efficient processor incorporating hundreds of processing cores; this design makes it especially well-suited for inherently parallel applications, including some AI systems.

[21] Neuromorphic computing refers to the ability of the hardware to learn, adapt, and physically reconfigure, taking inspiration from biology or neuroscience.

[22] Borghesi, A., Bartolini, A., Lombardi, M., Milano, M., & Benini, L. (2016). Predictive modeling for job Power consumption in HPC systems. In: Kunkel, J., Balaji, P., Dongarra, J. (eds). High Performance Computing. ISC High Performance 2016. Lecture Notes in Computer Science, vol 9697. Springer, Cham. https://doi.org/10.1007/978-3-319-41321-1_10

[23] These physical limits on computing are called Dennard scaling and lead to high on-chip power densities and the phenomenon called “dark silicon,” where different parts of a chip need to be turned off to limit temperatures and ensure data integrity.

[24] Cocaña-Fernández, A., Ranilla, J. & Sánchez, L. (2015). Energy-efficient allocation of computing node slots in HPC clusters through parameter learning and hybrid genetic fuzzy system modeling. *Journal of Supercomputing, 71*(3), 1163–1174. https://doi.org/10.1007/s11227-014-1320-9

[25] CHIPS and Science Act of 2022, Pub. L. No. 117-167, 136 Stat. 1370, Div. A (2022). https://www.congress.gov/117/plaws/publ167/PLAW-117publ167.pdf

[26] The White House. (2022). *FACT SHEET: President Biden Signs Executive Order to Implement the CHIPS and Science Act of 2022*. https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/25/fact-sheet-president-biden-signs-executive-order-to-implement-the-chips-and-science-act-of-2022/

[27] Ang, J., Chien, A. et al. *Reimagining Codesign for Advanced Scientific Computing: Report for the ASCR Workshop on Reimagining Codesign*. United States. https://doi.org/10.2172/1822199

[28] National Research Council. (2015).*Enhancing the Effectiveness of Team Science*. Committee on the Science of Team Science, N.J. Cooke and M.L. Hilton, Editors. Board on Behavioral, Cognitive, and Sensory Sciences, Division

of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.https://doi.org/10.17226/19007

[29] NITRD. (2016). *The National Artificial Intelligence Research and Development Strategic Plan*. https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf

[30] Laird, J., Ranganath, C., & Gershman, S. (2020). *Future Directions in Human Machine Teaming Workshop*, July 16-17, 2019. Department of Defense. https://rt.cto.mil/basic-research-office-released-the-final-report-from-the-future-directions-in-human-machine-teaming-workshop/

[31] National Academies of Sciences, Engineering, and Medicine. (2022). *Human-AI Teaming: State-of-the-Art and Research Needs*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26355

[32] National Security Commission on Artificial Intelligence (2021). *Final Report, National Security Commission on Artificial Intelligence*. NSCAI. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

[33] Ezer, N., Bruni, S., Cai, Y., Hepenstal, S., Miller, C.A., & Schmorrow, D. (2019). Trust Engineering for Human-AI Teams. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63, 322 - 326. https://www.semanticscholar.org/paper/27dcf6eb28fe279898303c96e83cc74df4ce8a8b

[34] NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography. (n.d.) https://www.ai2es.org/

[35] *Assured Autonomy: Path toward Living with autonomous systems we can trust*. (n.d.). Computing Research Association. https://cra.org/ccc/wp-content/uploads/sites/2/2020/10/Assured-Autonomy-Workshop-Report-Final.pdf

[36] Committee on Human-System Integration Research Topics for the 711th Human Performance Wing of the Air Force Research Laboratory; Board on Human-Systems Integration; Division of Behavioral and Social Sciences and Education; National Academies of Sciences, Engineering, and Medicine. (n.d.). *Read "Human-AI Teaming: State-of-the-Art and Research Needs" at NAP.edu*. https://nap.nationalacademies.org/read/26355/chapter/12

[37] NIST. (2023). AI Risk Management Framework. https://doi.org/10.6028/NIST.AI.100-1

[38] Zhang, D., Clark, J., & Perrault, R. The 2022 AI Index: Industrialization of AI and Mounting Ethical Concerns. Stanford University Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/2022-ai-index-industrialization-ai-and-mounting-ethical-concerns

[39] OSTP. (2022).*Blueprint for an AI Bill of Rights*.https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[40] NIST. (2023). AI Risk Management Framework. https://doi.org/10.6028/NIST.AI.100-1

[41] Mannes, A. (2020). Governance, risk, and artificial intelligence. *AI Magazine, 41*(1), 61–69. https://doi.org/10.1609/aimag.v41i1.5200

[42] Solaiman, I. & Dennison, C. (n.d.) *Process for Adapting Language Models to Society (PALMS) with Values-Targeted Datasets*. OpenAI.com. https://cdn.openai.com/palms.pdf

[43] *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities*. (2021, June 30). U.S. GAO.https://www.gao.gov/products/gao-21-519sp

[44] National Research Council. (2011). *Intelligence Analysis: Behavioral and Social Scientific Foundations*. The National Academies Press. https://doi.org/10.17226/13062

[45] *End-to-End Fairness*. (n.d.). New York University. https://wp.nyu.edu/ml4good/end-to-end-fairness/

[46] Note that an additional complicating factor in evaluation of AI performance in this and other areas is that AI may be biased compared to a hypothetical perfect system but may also perform better than the current real-world system, including human decision makers. Under some ethical frameworks, it may be desired to use such an AI system to replace more biased real-world human decision makers.

[47] Kleinberg, J. (2016, September 19). *Inherent Trade-Offs in the Fair Determination of Risk Scores*. arXiv.org. https://arxiv.org/abs/1609.05807

[48] *Future of Work | NSF*. (n.d.). https://www.nsf.gov/eng/futureofwork.jsp

[49] The White House, (Ed.). (2023, March 27). The Impact of Artificial Intelligence on the Future of Workforces in the European Union and the United States of America. Retrieved March 27, 2023, from https://www.whitehouse.gov/wp-content/uploads/2022/12/TTC-EC-CEA-AI-Report-12052022-1.pdf

[50] CDC. (2021, October 21).*The Role of Artificial Intelligence in the Future of Work.* https://blogs.cdc.gov/niosh-science-blog/2021/05/24/ai-future-of-work/

[51] Information Integrity R&D Interagency Working Group. (2022, December). *Roadmap For Researchers on Priorities Related to Information Integrity Research and Development*. NITRD Subcommittee. https://www.whitehouse.gov/wp-content/uploads/2022/12/Roadmap-Information-Integrity-RD-2022.pdf

[52] Roach, B. (Rapporteur). Computer Science and Telecommunications Board, National Academies of Sciences, Engineering, and Medicine. (2021, May). *Read "Assessing and Improving AI Trustworthiness: Current Contexts and Concerns: Proceedings of a Workshop–in Brief" at NAP.edu*. https://nap.nationalacademies.org/read/26208/chapter/1

[53] National Artificial Intelligence Research Resource Task Force. (2023, January). Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem. Retrieved March 27, 2023, from https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf?linkId=198749977

[54] *Artificial Intelligence and Cybersecurity: A Detailed Technical A Workshop Report*. (2020, June). The Networking and Information Technology Research and Development Program. https://www.nitrd.gov/pubs/AI-CS-Detailed-Technical-Workshop-Report-2020.pdf

[55] The White House. (2023, March 1). National Cybersecurity Strategy. Retrieved March 27, 2023, from https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf

[56] *Artificial Intelligence and Cybersecurity: A Detailed Technical A Workshop Report*. (2020, June). The Networking and Information Technology Research and Development Program. https://www.nitrd.gov/pubs/AI-CS-Detailed-Technical-Workshop-Report-2020.pdf

[57] *Artificial Intelligence and Cybersecurity: A Detailed Technical A Workshop Report*. (2020, June). The Networking and Information Technology Research and Development Program. https://www.nitrd.gov/pubs/AI-CS-Detailed-Technical-Workshop-Report-2020.pdf

[58] *Artificial Intelligence and Cybersecurity: A Detailed Technical A Workshop Report*. (2020, June). The Networking and Information Technology Research and Development Program. https://www.nitrd.gov/pubs/AI-CS-Detailed-Technical-Workshop-Report-2020.pdf

[59] *Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem*. (2023, January). National Artificial Intelligence Initiative. https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf

[60] Wilkinson, M. D. (2016, March 15). *The FAIR Guiding Principles for scientific data management and stewardship*. Nature. https://www.nature.com/articles/sdata201618?error=cookies_not_supported&code=213762ad-2d93-482e-8140-478e2f3762f2

[61] Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435 132 Stat. 5529 (2019). https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

[62] Bagwell, R. (2022, August 17). *EOSDIS Distributed Active Archive Centers (DAAC)*. Earthdata. https://www.earthdata.nasa.gov/eosdis/daacs

[63] OSTP. (2022, August 25). *Increasing Access to the Results of Federally Funded Research*. The White House. Retrieved April 18, 2023, from https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

[64] The White House. (2023). FACT SHEET: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research. https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/

[65] NAIIO. (2022, January 19). *AI Researchers Portal: Data Resources*. Retrieved March 27, 2023, from http://www.ai.gov/ai-researchers-portal/data-resources/

[66] Interagency Council on Statistical Policy. (2023, March 27). *The Standard Application Process.* National Center for Science and Engineering Statistics, NSF. Retrieved March 27, 2023, from https://ncses.nsf.gov/about/standard-application-process

[67] *STRIDES Initiative | Data Science at NIH*. (n.d.). https://datascience.nih.gov/strides

[68] *Big Data Platform Overview*. (2022, April). Defense Information Systems Agency. https://disa.mil/-/media/Files/DISA/News/Events/TechNet-Cyber---April-2022/Big-Data-Platform.ashx

[69] *DASHlink - Sample Flight Data*. (n.d.). https://c3.ndc.nasa.gov/dashlink/projects/85/

[70] NSF. (2023). Building the Prototype Open Knowledge Network (Proto-OKN): https://www.nsf.gov/pubs/2023/nsf23571/nsf23571.htm

[71] *All of Us Research Program | National Institutes of Health (NIH)*. (n.d.). https://allofus.nih.gov/

[72] Heaven, W. D. (2020, November 12). *AI is wrestling with a replication crisis*. MIT Technology Review. https://www.technologyreview.com/2020/11/12/1011944/

[73] National Academy of Sciences, Engineering, and Medicine (Ed.). (2023, July 27). *Assessing and Improving AI Trustworthiness: Current Contexts and Concerns: Proceedings of a Workshop–in Brief*. Retrieved March 27, 2023, from https://nap.nationalacademies.org/read/26208/chapter/1

[74] Grother, P. et al. *Ongoing Face Recognition Vendor Test (FRVT)*. (2023, April 04). National Institute of Standards and Technology. https://pages.nist.gov/frvt/reports/11/frvt_11_report.pdf

[75] *DARPA Open Sources Resources to Aid Evaluation of Adversarial AI Defenses*. (2021, December). Defense Advanced Research Projects Agency. https://www.darpa.mil/news-events/2021-12-21

[76] *Maintaining American Leadership in Artificial Intelligence*. (2019, February 14). Federal Register. https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence

[77] National Defense Authorization Act for Fiscal Year 2021, National Artificial Intelligence Initiative Act, Pub. L. No. 116-617, Div. E, § 5001 (2020). https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf

[78] *ISO/IEC JTC 1 Information Technology*. (n.d.). American National Standards Institute - ANSI. https://www.ansi.org/iso/ansi-activities/iso-iec-jtc-1-information-technology

[79] *ISO/IEC JTC 1/SC 42 Artificial intelligence*. (n.d.). ISO.org. https://www.iso.org/committee/6794475.html

[80] *ImageNet*. (n.d.). https://www.image-net.org/

[81] U.S. Green Building Council. (n.d.). *LEED rating system.* https://www.usgbc.org/leed

[82] *ISO/IEC JTC 1/SC 42 Artificial Intelligence*. (2022, October). ISO and IEC Joint Technical Committee (JTC 1) for Information Technology. https://jtc1info.org/sd-2-history/jtc1-subcommittees/sc-42/

[83] Marcus, G., & Davis, E. *Rebooting AI: Building artificial intelligence we can trust*. 2020. Penguin Random House. ISBN: 9780525566045

[84] *Summary of the Big Data IWG Workshop October 4 - 15, 2017*. (2018, November). NITRD Program. https://www.nitrd.gov/pubs/open-knowledge-network-workshop-report-2018.pdf

[85] UCLA Office of Advanced Research Computing. (n.d.). *Strategy for Resilient Manufacturing Ecosystems through Artificial Intelligence.* https://oarc.ucla.edu/nsf-nist-symposium

[86] The White House. (2022, August 25). FACT SHEET: President Biden Signs Executive Order to Implement the CHIPS and Science Act of 2022. https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/25/fact-sheet-president-biden-signs-executive-order-to-implement-the-chips-and-science-act-of-2022/

[87] Luccioni, A. S., Viguier, S., & Ligozat, A.-L. (2022, November 3). *Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model*. arXiv.org. https://arxiv.org/abs/2211.02001

[88] Balenson, D., Tinnel, L., & Benzel, T. (2015, July 31).*Cybersecurity Experimentation of the Future (CEF): Catalyzing a New Generation of Experimental Cybersecurity Research*. The Road to Tomorrow: Cybersecurity Experimentation of the Future. https://cef.cyberexperimentation.org/application/files/2616/2160/7871/CEF_Final_Report_Bound_20150922.pdf

[89] NITRD Program. (2021, December 17). *AI R&D Testbed Inventory*. https://www.nitrd.gov/apps/ai-rd-testbed-inventory/

[90] *Text REtrieval Conference (TREC) Home Page*. (n.d.). https://trec.nist.gov/

[91] *Face Recognition Vendor Test (FRVT)*. (2020, November 30). NIST. https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt

[92] *Computer and Information Research Scientists: Occupational Outlook Handbook: U.S. Bureau of Labor Statistics*. (2023, February 6). https://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm

[93] *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update*. NITRD Program. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

[94] G20 Young Entrepreneurs' Alliance. *How to accelerate skills acquisition in the age of intelligent technologies*. (n.d.). Accenture.com. https://www.accenture.com/t20180920T094705Z__w__/us-en/_acnmedia/Thought-Leadership-Assets/PDF/Accenture-Education-and-Technology-Skills-Research.pdf

[95] CRA Enrollment Committee Institution Subgroup. (2017, April 3). *Generation CS: Report on CS Enrollment*. Computing Research Association. https://cra.org/data/generation-cs/

[96] *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update*. (2019). NITRD Program. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

[97] Gehlhaus, D., & Mutis, S. (2021). *The U.S. AI workforce: Understanding the supply of AI talent*. Center for Security and Emerging Technology. https://doi.org/10.51593/20200068

[98] *National Security Commission on Artificial Intelligence Final Report*. (2021). The National Security Commission on Artificial Intelligence. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

[99] National Science Board. (2015). Figure 3-32: Foreign-born scientists and engineers employed in S&E occupations, by highest degree level and broad S&E occupational category. https://www.nsf.gov/statistics/2018/nsb20181/assets/901/figures/fig03-32.pdf

[100] President's Council of Advisors on Science and Technology. (June 2020). Recommendations for Strengthening American Leadership in Industries of the Future (p. 11, recommendation 1.5). https://science.osti.gov/-/media/_/pdf/about/pcast/202006/PCAST_June_2020_Report.pdf

[101] The 2021 National Security Commission on Artificial Intelligence final report includes a recommendation to "Win the competition for international talent by tailoring visa and green card policies for digital skills, STEM PhDs, entrepreneurs, and technologists."

[102] *Restoring the Foundation*. (2014). American Academy of Arts and Sciences. https://www.amacad.org/multimedia/pdfs/publications/researchpapersmonographs/AmericanAcad_RestoringtheFoundation_Brief.pdf

[103] *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update*. (2019). NITRD Program. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

[104] *Artificial Intelligence (AI) Research Institutes*. (2021, October 7). National Institute of Food and Agriculture. https://www.nifa.usda.gov/grants/funding-opportunities/artificial-intelligence-ai-research-institutes

[105] *Artificial Intelligence (AI) at NSF | NSF - National Science Foundation*. (n.d.-b). https://www.nsf.gov/cise/ai.jsp

[106] Bessen, J., Impink, S., & Seamans, R. (2022, March). *Ethical AI development: Evidence from AI startups*. Brookings. https://www.brookings.edu/wp-content/uploads/2022/03/Seamans_final-PDF.pdf

[107] Committee on Responsible Computing Research: Ethics and Governance of Computing Research and Its Applications; Computer Science and Telecommunications Board; Division on Engineering and Physical Sciences; National Academies of Sciences, Engineering, and Medicine. (2022). *Read "Fostering Responsible Computing Research: Foundations and Practices" at NAP.edu*. https://nap.nationalacademies.org/read/26507/chapter/2

[108] *Ethics of Artificial Intelligence*. (2023, February 3). UNESCO. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

[109] *Technical AI Standards*. (2023, March 14). NIST. https://www.nist.gov/artificial-intelligence/technical-ai-standards

[110] *Expanding AI Innovation through Capacity Building and Partnerships (ExpandAI): Broadening and diversifying the research community in collaboration with National AI Research Institutes. Program Solicitation NSF 23-506* (2023). NSF – National Science Foundation. https://www.nsf.gov/pubs/2023/nsf23506/nsf23506.htm

[111] *Partnership Development in the Federal Government*. (2019, June). Institute for Defense Analyses (IDA). https://www.ida.org/research-and-publications/publications/all/p/pa/partnership-development-in-the-federal-government

[112] National Academies of Sciences, Engineering, and Medicine. (2020, November 30). *Information Technology Innovation: Resurgence, Confluence, and Continuing Impact*. The National Academies Press. https://nap.nationalacademies.org/catalog/25961/information-technology-innovation-resurgence-confluence-and-continuing-impact

[113] *Industry-University Cooperative Research Centers Program (IUCRC)*. (2020, May 4). NSF - National Science Foundation. https://beta.nsf.gov/funding/opportunities/industry-university-cooperative-research-centers-0

[114] *U.S.C. Title 35 - PATENTS.* (n.d.). https://www.govinfo.gov/content/pkg/USCODE-2011-title35/html/USCODE-2011-title35-partII-chap18.htm

[115] NAIIO. (2021, April 23). *The (NAII)*. National Artificial Intelligence Initiative. https://www.ai.gov

[116] *Federal Prize Competitions*. (2022, April 1). Center for Security and Emerging Technology. https://cset.georgetown.edu/publication/federal-prize-competitions/

[117] Recommendation of the Council on Artificial Intelligence. (2019, May 21). OECD. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449#mainText

[118] Burke, A., Okrent, A. & Hale, K. (2022, January 18). *The State of U.S. Science and Engineering 2022.* The National Science Board. https://ncses.nsf.gov/pubs/nsb20221

[119] Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J., Parli, V., Shoham, Y., Wald, R., Clark, J., and Perrault, R. (2023, April). *The AI Index 2023 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

[120] *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*. (2020, December). The Federal Register. https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government

[121] *CSIRO and the US National Science Foundation*. (n.d.). CSIRO. https://www.csiro.au/en/work-with-us/international/north-america/national-science-foundation

[122] Under Secretary Fernandez Signs Administrative Arrangement with European Commission's Directorate General for Communications Networks, Content, and Technology (DG-CNECT) on Artificial Intelligence. (2023, January 28). U.S. Department of State. https://www.state.gov/under-secretary-fernandez-signs-administrative-arrangement-with-european-commissions-directorate-general-for-communications-networks-content-and-technology-dg-cnect-on-artificial-intellig/

[123] *Global Partnership on Artificial Intelligence*. (n.d.). https://gpai.ai/

[124] U.S. - India Artificial Intelligence (USIAI) Initiative. (n.d.). Indo-U.S. Science and Technology Forum. https://usiai.iusstf.org/

[125] *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*. (2020a). The Federal Register. https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government

[126] *Declaration of the United States of America and the United Kingdom of Great Britain and Northern Ireland on Cooperation in Artificial Intelligence Research and Development: A Shared Vision for Driving Technological Breakthroughs in Artificial Intelligence*. (2020, September 25). U.S. Department of State. https://www.state.gov/declaration-of-the-united-states-of-america-and-the-united-kingdom-of-great-britain-and-northern-ireland-on-cooperation-in-artificial-intelligence-research-and-development-a-shared-vision-for-driving/

[127] The White House. (2021, September 27). *Fact Sheet: Quad Leaders' Summit*. https://www.whitehouse.gov/briefing-room/statements-releases/2021/09/24/fact-sheet-quad-leaders-summit/

[128] The White House. (2021, September 27). *Fact Sheet: Quad Leaders' Summit*. https://www.whitehouse.gov/briefing-room/statements-releases/2021/09/24/fact-sheet-quad-leaders-summit/

[129] *CSIRO and the US National Science Foundation*. (n.d.-b). CSIRO. https://www.csiro.au/en/work-with-us/international/north-america/national-science-foundation

[130] *U.S. Science Envoy Program*. (n.d.). U.S. Department of State. https://www.state.gov/programs-office-of-science-and-technology-cooperation/u-s-science-envoy-program/

[131] The White House. (2021, April 17). *U.S.- Japan Joint Leaders' Statement: "U.S. – JAPAN GLOBAL PARTNERSHIP FOR A NEW ERA."* https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/16/u-s-japan-joint-leaders-statement-u-s-japan-global-partnership-for-a-new-era/

[132] *International Visitor Leadership Program*. (n.d.). Bureau of Educational and Cultural Affairs, U.S. Department of State. https://eca.state.gov/ivlp

[133] *TechCamp*. (n.d.). Bureau of Educational and Cultural Affairs, U.S. Department of State. https://techcamp.america.gov/

[134] U.S. Office of Personnel Management. *Intergovernmental Personnel Act.* https://www.opm.gov/policy-data-oversight/hiring-information/intergovernment-personnel-act/

[135] The White House. (2021e, December 8). *White House Announces Launch of the International Grand Challenges on Democracy-Affirming Technologies for the Summit for Democracy*. https://www.whitehouse.gov/ostp/news-updates/2021/12/08/white-house-announces-launch-of-the-international-grand-challenges-on-democracy-affirming-technologies-for-the-summit-for-democracy/

[136] NITRD Program. (2021, December 17). *AI R&D Testbed Inventory*. https://www.nitrd.gov/apps/ai-rd-testbed-inventory/

[137] NITRD Program. (2021, December 17). *AI Research Program Repository*. https://www.nitrd.gov/apps/ai-research-program-repository/


EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL