THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN: 2019 UPDATE

A Report by the

SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE

of the

NATIONAL SCIENCE & TECHNOLOGY COUNCIL

JUNE 2019

Dear Colleagues,

In his State of the Union address on February 5, 2019, President Trump stressed the importance of ensuring American leadership in the development of emerging technologies, including artificial intelligence (AI), that make up the Industries of the Future. Reflecting this importance, on February 11, 2019, President Trump signed Executive Order 13859, which established the American Artificial Intelligence Initiative. This Initiative is a whole-of-government approach for maintaining American leadership in AI and ensuring that AI benefits the American people and reflects our Nation's values. The first directive in this Executive Order is for Federal agencies to prioritize AI research and development (R&D) in their annual budgeting and planning process. The attached *National AI R&D Strategic Plan: 2019 Update* highlights the key priorities for Federal investment in AI R&D.

Artificial intelligence presents tremendous opportunities that are leading to breakthroughs in improved healthcare, safer and more efficient transportation, personalized education, significant scientific discoveries, improved manufacturing, increased agricultural crop yields, better weather forecasting, and much more. These benefits are largely due to decades of long-term Federal investments in fundamental AI R&D, which have led to new theories and approaches for AI systems, as well as applied research that allows the translation of AI into practical applications.

The landscape for AI R&D is becoming increasingly complex, due to the significant investments that are being made by industry, academia, and nonprofit organizations. Additionally, AI advancements are progressing rapidly. The Federal Government must therefore continually reevaluate its priorities for AI R&D investments, to ensure that investments continue to advance the cutting edge of the field and are not unnecessarily duplicative of industry investments.

In August of 2018, the Administration directed the Select Committee on AI to refresh the 2016 *National AI R&D Strategic Plan*. This process began with the issuance of a Request for Information to solicit public input on ways that the strategy should be revised or improved. The responses to this RFI, as well as an independent agency review, informed this update to the Strategic Plan.

In this Strategic Plan, eight strategic priorities have been identified. The first seven strategies continue from the 2016 Plan, reflecting the reaffirmation of the importance of these strategies by multiple respondents from the public and government, with no calls to remove any of the strategies. The eighth strategy is new and focuses on 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.

While this Plan does not define specific research agendas for Federal agency investments, it does provide an expectation for the overall portfolio for Federal AI R&D investments. This coordinated Federal strategy for AI R&D will help the United States continue to lead the world in cutting-edge advances in AI that will grow our economy, increase our national security, and improve quality of life.

Sincerely,

Michael Kratsios
Deputy Assistant to the President for Technology Policy
June 21, 2019

Table of Contents

Executive Summary

Artificial intelligence (AI) holds tremendous promise to benefit nearly all aspects of society, including the economy, healthcare, security, the law, transportation, even technology itself. On February 11, 2019, the President signed Executive Order 13859, *Maintaining American Leadership in Artificial Intelligence*.[1] This order launched the American AI Initiative, a concerted effort to promote and protect AI technology and innovation in the United States. The Initiative implements a whole-of-government strategy in collaboration and engagement with the private sector, academia, the public, and like-minded international partners. Among other actions, key directives in the Initiative call for Federal agencies to prioritize AI research and development (R&D) investments, enhance access to high-quality cyberinfrastructure and data, ensure that the Nation leads in the development of technical standards for AI, and provide education and training opportunities to prepare the American workforce for the new era of AI.

In support of the American AI Initiative, this *National AI R&D Strategic Plan: 2019 Update* defines the priority areas for Federal investments in AI R&D. This 2019 update builds upon the first *National AI R&D Strategic Plan* released in 2016, accounting for new research, technical innovations, and other considerations that have emerged over the past three years. This update has been developed by leading AI researchers and research administrators from across the Federal Government, with input from the broader civil society, including from many of America's leading academic research institutions, nonprofit organizations, and private sector technology companies. Feedback from these key stakeholders affirmed the continued relevance of each part of the 2016 Strategic Plan while also calling for greater attention to making AI trustworthy, to partnering with the private sector, and other imperatives.

The *National AI R&D Strategic Plan: 2019 Update* establishes a set of objectives for Federally funded AI research, identifying the following eight strategic priorities:

Strategy 1: Make long-term investments in AI research. Prioritize investments in the next generation of AI that will drive discovery and insight and enable the United States to remain a world leader in 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.

Strategy 3: Understand and address the ethical, legal, and societal implications of AI. Research AI systems that incorporate ethical, legal, and societal concerns through technical mechanisms.

Strategy 4: Ensure the safety and security of AI systems. Advance knowledge of how to design AI systems that are reliable, dependable, safe, and trustworthy.

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.

Strategy 6: Measure and evaluate AI technologies through standards and benchmarks. Develop a broad spectrum of evaluative techniques for AI, including technical standards and benchmarks.

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.

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

1 https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/

Introduction to the 2019 National AI R&D Strategic Plan

Artificial intelligence enables computers and other automated systems to perform tasks that have historically required human cognition and what we typically consider human decision-making abilities. Over the past several decades, AI has advanced tremendously and today promises better, more accurate healthcare; enhanced national security; improved transportation; and more effective education, to name just a few benefits. Increased computing power, the availability of large datasets and streaming data, and algorithmic advances in machine learning (ML) have made it possible for AI development to create new sectors of the economy and revitalize industries. As more industries adopt AI's fundamental technologies, the field will continue to drive profound economic impact and quality-of-life improvements worldwide.

These advancements have been driven primarily by Federal investments in AI R&D, the expertise of America's unsurpassed R&D institutions, and the collective creativity of many of America's most visionary technology companies and entrepreneurs.

In 2016 the Federal Government published first *National AI R&D Strategic Plan*, recognizing AI's tremendous promise and need for continued advancement. It was developed to guide the Nation in our AI R&D investments, provide a strategic framework for improving and leveraging America's AI capabilities, and ensure that those capabilities produce prosperity, security, and improved quality of life for the American people for years to come.

The Plan defined several key areas of priority focus for the Federal agencies that invest in AI. These focus areas, or strategies, include: continued long-term investments in AI; effective methods for human-AI collaboration; understanding and addressing the ethical, legal, and societal implications for AI; ensuring the safety and security of AI; developing shared public datasets and environments for AI training and testing; measuring and evaluating AI technologies through standards and benchmarks; and better understanding the Nation's AI R&D

2019 Update	**RFI responses inform the 2019 National AI R&D Strategic Plan**

In September 2018, the National Coordination Office for Networking and Information Technology Research and Development issued a Request for Information (RFI)[2] on behalf of the Select Committee on Artificial Intelligence, requesting input from all interested parties on the 2016 *National Artificial Intelligence Research and Development Strategic Plan*. Nearly 50 responses were submitted by researchers, research organizations, professional societies, civil society organizations, and individuals; these responses are available online.[3]

Many of the responses reaffirmed the analysis, organization, and approach outlined in the 2016 *National AI R&D Strategic Plan*. A significant number of responses noted the importance of investing in the application of AI in areas such as manufacturing and supply chains; healthcare; medical imaging; meteorology, hydrology, climatology, and related areas; cybersecurity; education; data-intensive physical sciences such as high-energy physics; and transportation. This interest in translational applications of AI technologies has certainly increased since the release of the 2016 *National AI R&D Strategic Plan*. Other common themes echoed in the RFI responses were the importance of developing trustworthy AI systems, including fairness, ethics, accountability, and transparency of AI systems; curated and accessible datasets; workforce considerations; and public-private partnerships for furthering AI R&D.

2 https://www.nitrd.gov/news/RFI-National-AI-Strategic-Plan.aspx

3 https://www.nitrd.gov/nitrdgroups/index.php?title=AI-RFI-Responses-2018

workforce needs. That work was prescient: today, countries around the world have followed suit and have issued their own versions of this plan.

In the three years since the *National AI R&D Strategic Plan* was produced, new research, technical innovations, and real-world deployments have progressed rapidly. The Administration initiated this 2019 update to the *National AI R&D Strategic Plan* to address these advancements, including a rapidly evolving international AI landscape.

Notably, this 2019 Update to the *National AI R&D Strategic Plan* is, by design, solely concerned with addressing the *research and development* priorities associated with advancing AI technologies. It does not describe or recommend policy or regulatory actions related to the governance or deployment of AI, although AI R&D will certainly inform the development of reasonable policy and regulatory frameworks.

AI as an Administration Priority

Since 2017, the Administration has addressed the importance of AI R&D by emphasizing its role for America's future across multiple major policy documents, including the *National Security Strategy*,[4] the *National Defense Strategy*,[5] and the FY 2020 R&D Budget Priorities Memo.[6]

In May 2018, the Office of Science and Technology Policy (OSTP) hosted the White House Summit on Artificial Intelligence for American Industry to begin discussing the promise of AI and the policies needed to realize that promise for the American people and maintain U.S. leadership in the age of AI. The Summit convened over 100 senior government officials, technical experts from top academic institutions, heads of industrial research laboratories, and American business leaders.

In his State of the Union address on February 5, 2019, President Trump stressed the importance of ensuring American leadership in the development of emerging technologies, including AI, that make up the Industries of the Future.

On February 11, 2019, the President signed Executive Order 13859, *Maintaining American Leadership in Artificial Intelligence*.[7] This order launched the American AI Initiative, a concerted effort to promote and protect AI technology and innovation in the United States. The Initiative implements a whole-of-government strategy in collaboration and engagement with the private sector, academia, the public, and like-minded international partners. Among other actions, key directives in the Initiative call for Federal agencies to prioritize AI R&D investments, enhance access to high-quality cyberinfrastructure and data, ensure that the Nation leads in the development of technical standards for AI, and provide education and training opportunities to prepare the American workforce for the new era of AI.

Development of the 2019 Update to the *National AI R&D Strategic Plan*

The 2016 *National AI R&D Strategic Plan* recommended that the many Federal agencies tasked with advancing or adopting AI collaborate to identify critical R&D opportunities and support effective coordination of Federal AI R&D activities, both intramural and extramural research. Reflecting the Administration's prioritization of AI, the National Science and Technology Council (NSTC) has established a new framework to implement this recommendation, consisting of three unique NSTC subgroups made up of members from across the Federal R&D agencies to cover (1) senior leadership

---

4 https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf

5 https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

6 https://www.whitehouse.gov/wp-content/uploads/2018/07/M-18-22.pdf

7 https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/

and strategic vision, (2) operational planning and tactical implementation, and (3) research and technical expertise. These subgroups are:

- The Select Committee on AI,[8] consisting of the heads of departments and agencies principally responsible for the government's AI R&D, advises the Administration on interagency AI R&D priorities; considers the creation of Federal partnerships with industry and academia; establishes structures to improve government planning and coordination of AI R&D; identifies opportunities to leverage Federal data and computational resources to support our national AI R&D ecosystem; and supports the growth of a technical, national AI workforce.
- The NSTC Subcommittee on Machine Learning and Artificial Intelligence (MLAI), consisting of agency AI leaders and administrators, serves as the operational and implementation arm of the Select Committee, responsible for fulfilling tasking from the Select Committee; creating and maintaining the *National AI R&D Strategic Plan*; identifying and addressing important policy issues related to AI research, testing, standards, education, implementation, outreach, and related areas; and related activities.
- The AI R&D Interagency Working Group, operating under the NSTC's Networking and Information Technology R&D (NITRD) Subcommittee and consisting of research program managers and technical experts from across the Federal Government, reports to the MLAI Subcommittee; helps coordinate interagency AI R&D programmatic efforts; serves as the interagency AI R&D community of practice; and reports government-wide AI R&D spending through the NITRD Subcommittee's annual Supplement to the President's Budget.

In September 2018, the Select Committee initiated an update to the 2016 Strategic Plan, beginning with an RFI seeking broad community input on whether and how the seven strategies of the 2016 *National AI R&D Strategic Plan* merited revision or replacement (see sidebar). Independently, Federal departments and agencies performing or funding AI R&D undertook their own assessments.

An Overview of the 2019 Update to the 2016 *National AI R&D Strategic Plan*

Together, the Select Committee on AI, the NSTC Subcommittee on Machine Learning and AI, and the AI R&D Interagency Working Group of NITRD reviewed the input regarding the *National AI R&D Strategic Plan.* Each of the original seven focus areas or strategies of the 2016 Plan was reaffirmed by multiple respondents from the public and government, with no calls to remove any one strategy. These strategies, updated in this 2019 Update to the Strategic Plan to reflect the current state of the art, are:

Strategy 1: Make long-term investments in AI research;

Strategy 2: Develop effective methods for human-AI collaboration;

Strategy 3: Understand and address the ethical, legal, and societal implications of AI;

Strategy 4: Ensure the safety and security of AI systems;

Strategy 5: Develop shared public datasets and environments for AI training and testing;

Strategy 6: Measure and evaluate AI technologies through standards and benchmarks; and

Strategy 7: Better understand the national AI R&D workforce needs.

[8] https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf

Many responses to the RFI called for greater Federal Government R&D engagement with the private sector, given the fast rise of privately funded AI R&D, and the rapid adoption of AI by industry. As a result, the 2019 Update incorporates a new, eighth strategy:

Strategy 8: Expand public-private partnerships to accelerate advances in AI.

Feedback from the public and Federal agencies identified a number of specific challenges to further AI development and adoption. These challenges, many of which cut across multiple agencies, provide enhanced insight into ways that this *National AI R&D Strategic Plan* can guide the course of AI R&D in America, and many closely relate to the themes addressed in the 2019 *Executive Order on Maintaining American Leadership in Artificial Intelligence.* Examples include the following:

- *Research at the frontiers.* Even though machine learning has brought phenomenal new capabilities in the past several years, continued research is needed to further push the frontiers of ML, as well as to develop additional approaches to the tough technical challenges of AI (Strategy 1).
- *Positive impact.* As AI capabilities grow, the United States must place increased emphasis on developing new methods to ensure that AI's impacts are robustly positive into the future (Strategies 1, 3, and 4).
- *Trust and explainability.* Truly trustworthy AI requires explainable AI, especially as AI systems grow in scale and complexity; this requires a comprehensive understanding of the AI system by the human user and the human designer (Strategies 1, 2, 3, 4, and 6).
- *Safety and security*. Researchers must devise methods to keep AI systems and the data they use secure so that the Nation can leverage the opportunities afforded by this technology while also maintaining confidentiality and safety (Strategies 4, 5, and 6).
- *Technical standards.* As the Nation develops techniques to expand both AI abilities and assurance, it must test and benchmark them; when the techniques are ready, they should be turned into technical standards for the world (Strategy 6).
- *Workforce capability.* Accomplishing these goals will require growing a skilled AI R&D workforce that is currently limited and in high demand; the United States must be creative and bold in training and acquiring the skilled workforce it needs to lead the world in AI research and applications (Strategy 7).
- *Partnerships.* Advances in AI R&D increasingly require effective partnerships between the Federal Government and academia, industry, and other non-Federal entities to generate technological breakthroughs in AI and to rapidly transition those breakthroughs into capabilities (Strategy 8).
- *Cooperation with allies.* Additionally, the Plan recognizes the importance of international cooperation for successful implementation of these goals, while protecting the American AI R&D enterprise from strategic competitors and adversarial nations.

Structure of this 2019 Update to the 2016 *National AI R&D Strategic Plan*

This updated *National AI R&D Strategic Plan* incorporates the original text from the 2016 version, including the following section on R&D Strategy (except for minor edits) and the original 2016 wording of the first seven strategies. For each strategy, *2019 updates to the 2016 National R&D Strategic Plan are provided in shaded boxes at the top of the original seven strategies; these highlight updated imperatives and/or new focus areas for the strategies.* Text below the shaded boxes is *as it originally appeared* in the 2016 *National AI R&D Strategic Plan*, providing observations and context that remain important today (note that some of the original details may have become out of date in the intervening period). In addition, as noted previously, a new eighth strategy is added in this 2019 Update, on expanding public-private partnerships in AI R&D.

AI R&D Strategy

The research priorities outlined in this AI R&D Strategic Plan focus on areas that industry is unlikely to address on their own, and thus, areas that are most likely to benefit from Federal investment. These priorities cut across all of AI to include needs common to the AI sub-fields of perception, automated reasoning/planning, cognitive systems, machine learning, natural language processing, robotics, and related fields. Because of the breadth of AI, these priorities span the entire field, rather than only focusing on individual research challenges specific to each sub-domain. To implement the plan, detailed roadmaps should be developed that address the capability gaps consistent with the plan.

One of the most important Federal research priorities, outlined in Strategy 1, is for sustained long-term research in AI to drive discovery and insight. Many of the investments by the U.S. Federal Government in high-risk, high-reward[9] fundamental research have led to revolutionary technological advances we depend on today, including the Internet, GPS, smartphone speech recognition, heart monitors, solar panels, advanced batteries, cancer therapies, and much, much more. The promise of AI touches nearly every aspect of society and has the potential for significant positive societal and economic benefits. Thus, to maintain a world leadership position in this area, the United States must focus its investments on high-priority fundamental and long-term AI research.

Many AI technologies will work with and alongside humans, thus leading to important challenges in how to best create AI systems that work with people in intuitive and helpful ways.[10] The walls between humans and AI systems are slowly beginning to erode, with AI systems augmenting and enhancing human capabilities. Fundamental research is needed to develop effective methods for human-AI interaction and collaboration, as outlined in Strategy 2.

AI advancements are providing many positive benefits to society and are increasing U.S. national competitiveness.[11] However, as with most transformative technologies, AI presents some societal risks in several areas, from jobs and the economy to safety, ethical, and legal questions. Thus, as AI science and technology develop, the Federal Government must also invest in research to better understand what the implications are for AI for all these realms, and to address these implications by developing AI systems that align with ethical, legal, and societal goals, as outlined in Strategy 3.

A critical gap in current AI technology is a lack of methodologies to ensure the safety and predictable performance of AI systems. Ensuring the safety of AI systems is a challenge because of the unusual complexity and evolving nature of these systems. Several research priorities address this safety challenge. First, Strategy 4 emphasizes the need for explainable and transparent systems that are trusted by their users, perform in a manner that is acceptable to the users, and can be guaranteed to act as the user intended. The potential capabilities and complexity of AI systems, combined with the wealth of possible interactions with human users and the environment, makes it critically important to invest in research that increases the security and control of AI technologies. Strategy 5 calls on the Federal Government to invest in shared public datasets for AI training and testing to advance the progress of AI research and to enable a more effective comparison of alternative solutions.

Strategy 6 discusses how standards and benchmarks can focus R&D to define progress, close gaps, and drive innovative solutions for specific problems and challenges. Standards and benchmarks are

---

9 "High-risk, high-reward" research refers to visionary research that is intellectually challenging but has the potential to make deeply positive, transformative impacts on the field of study.

10 See *2016 Report of 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/2016-report.

11 J. Furman, "Is This Time Different? The Opportunities and Challenges of Artificial Intelligence," Council of Economic Advisors remarks, New York University: AI Now Symposium, July 7, 2016.

essential for measuring and evaluating AI systems and ensuring that AI technologies meet critical objectives for functionality and interoperability.

Finally, the growing prevalence of AI technologies across all sectors of society creates new pressures for AI R&D experts. Opportunities abound for core AI scientists and engineers with a deep understanding of the technology who can generate new ideas for advancing the boundaries of knowledge in the field. The Nation should take action to ensure a sufficient pipeline of AI-capable talent. Strategy 7 addresses this challenge.

Figure 1 (*updated in this 2019 version of the Plan*) provides a graphical illustration of the overall organization of this AI R&D Strategic Plan. Across the bottom row of boxes are the crosscutting, underlying foundations that affect the development of all AI systems; these foundations are described in Strategies 3-7 and the new Strategy 8. The next layer higher (middle row of boxes) includes many areas of research that are needed to advance AI. These R&D areas (including use-inspired basic research) are outlined in Strategies 1-2.[12] Across the top row of boxes in the graphic are examples of applications that are expected to benefit from advances in AI. Together, these components of the AI R&D Strategic Plan define a high-level framework for Federal investments that can lead to impactful advances in the field and positive societal benefits.

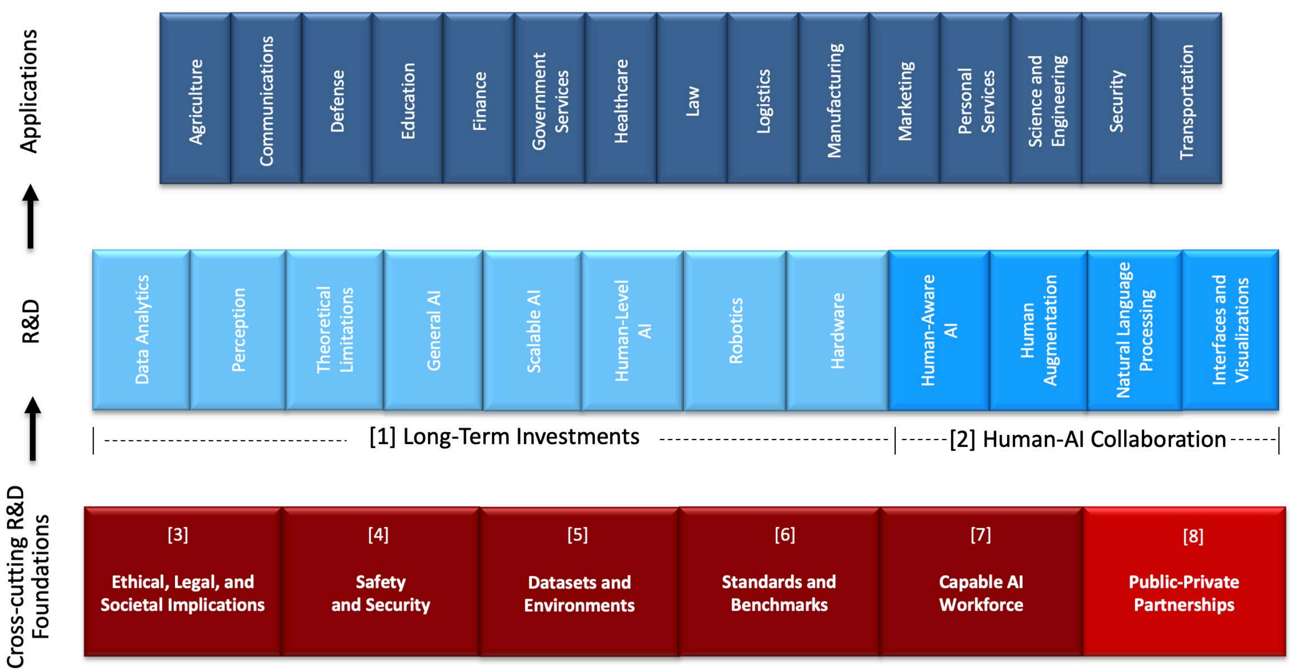

Figure 1. Organization of the AI R&D Strategic Plan (2019 update, to include Strategy 8). A combination of crosscutting R&D foundations (*in the lower row*) are important for all AI research. Many AI R&D areas (*in the middle row*) can build upon these crosscutting foundations to impact a wide array of societal applications (*in the top row*). The numbers in brackets indicate the number of the Strategy in this plan that further develops each topic. The ordering of these strategies does not indicate a priority of importance.

[12] 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 use for society in mind. For example, the fundamental NIH investments in IT are often called use-inspired basic research.

Strategy 1: Make Long-Term Investments in AI Research

2019 Update	**Sustaining long-term investments in fundamental AI research**

Since the release of the 2016 *National AI R&D Strategic Plan*, powerful new capabilities, primarily ML applications to well-defined tasks, have continued to emerge. These capabilities have demonstrated impacts in a diverse array of applications, such as classifying genetic sequences,[20,21] managing limited wireless spectrum resources,[22] interpreting medical images,[23] and grading cancers.[24] These rapid advances required decades of research for the technologies and applications to mature.[25] To maintain this progress in ML to achieve advancements in other areas of AI, and to strive toward the long-term goal of general-purpose AI, the Federal Government must continue to foster long-term, fundamental research in ML and AI. This research will give rise to transformational technologies and, in turn, breakthroughs across all sectors of society.

Much of the current progress in the field has been in specialized, well-defined tasks often driven by statistical ML, such as *classification*, *recognition*, and *regression* (i.e., "narrow AI systems"). Surveys of the

Long-term, fundamental AI research: Recent agency R&D programs

Since the release of the 2016 *National AI R&D Strategic Plan*, a number of agencies have initiated AI R&D programs for Strategy 1:

- NSF has continued to fund foundational research in AI, spanning ML, reasoning and representation, computer vision, computational neuroscience, speech and language, robotics, and multi-agent systems. NSF has launched new joint funding opportunities with other agencies—notably with DARPA in the area of high-performance, energy-efficient hardware for real-time ML[13] and with USDA-NIFA on AI for agricultural science[14]—and with industry.[15,16] In addition, NSF's Harnessing the Data Revolution Big Idea[17] supports research on the foundations of data science, which will serve as a driver of future ML and AI systems.
- DARPA announced in September 2018 a multiyear investment in new and existing programs called the "AI Next" campaign.[18] Key campaign areas include improving the robustness and reliability of AI systems; enhancing the security and resiliency of ML/AI technologies; reducing power, data, and performance inefficiencies; and pioneering the next generation of AI algorithms and applications, such as explainability and commonsense reasoning.
- The *NIH Strategic Plan for Data Science*[19] of September 2018 aims to advance access to data science technology and ML/AI capability for the biomedical research community toward data-driven healthcare research.

13 https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505640&org=NSF
14 https://www.nsf.gov/pubs/2019/nsf19051/nsf19051.jsp
15 https://www.nsf.gov/pubs/2019/nsf19018/nsf19018.jsp
16 https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505651
17 https://www.nsf.gov/cise/harnessingdata/
18 https://www.darpa.mil/work-with-us/ai-next-campaign
19 https://datascience.nih.gov/strategicplan
20 https://ai.googleblog.com/2017/12/deepvariant-highly-accurate-genomes.html
21 https://irp.nih.gov/catalyst/v26i4/machine-learning
22 https://www.spectrumcollaborationchallenge.com/
23 https://news-medical.net/news/20190417/Workshop-explores-the-future-of-artificial-intelligence-in-medical-imaging.aspx
24 https://www.nature.com/articles/nature21056
25 https://www.nitrd.gov/rfi/ai/2018/AI-RFI-Response-2018-Yolanda-Gil-AAAI.pdf

field have noted that long-term investments in fundamental research are needed to continue building on these advances in ML. Further, parallel sustained efforts are required to fully realize the vision of "general-purpose AI"—systems that exhibit the flexibility and versatility of human intelligence in a broad range of cognitive domains.[26,27,28,29]

Emphasis is needed on the development of further ML capabilities to interactively and persistently learn, the connection between perception and attention, and the incorporation of learned models into comprehensive reasoning architectures.[30] Beyond ML, critical research is also needed in other core areas of AI, including in commonsense reasoning and problem solving, probabilistic reasoning, combinatorial optimization, knowledge representation, planning and scheduling, natural language processing, decision making, and human-machine interaction. Advances in these areas will in turn enable collaborative robotics and shared and fully autonomous systems (see Strategy 2). The grand challenge of understanding human intelligence requires significant investments in shared resources and infrastructure.[25] Broad consensus exists for foundational investments in drivers of ML and AI as well, including data provenance and quality, novel software and hardware paradigms and platforms, and the security of AI systems.[31,32] For example, as AI software performs increasingly complex functions in all aspects of daily life and all sectors of the economy, existing software development paradigms will need to evolve to meet software productivity, quality, and sustainability requirements.

Recent Federal investments have prioritized these areas of fundamental ML and AI research (see sidebar) as well as the use of ML and AI across numerous application sectors, including defense, security, energy, transportation, health, agriculture, and telecommunications. Ultimately, AI technologies are critical for addressing a range of long-term challenges, such as constructing advanced healthcare systems, a robust intelligent transportation system, and resilient energy and telecommunication networks.

For AI applications to become widespread, they must be explainable and understandable (see Strategy 3). These challenges are particularly salient for fostering collaborative human-AI relationships (see Strategy 2). Today, the ability to understand and analyze the decisions of AI systems and measure their accuracy, reliability, and reproducibility is limited. Sustained R&D investments are needed to advance trust in AI systems to ensure they meet society's needs and adequately address requirements for robustness, fairness, explainability, and security.

A long-term commitment to AI R&D is essential to continue and expand current technical advances and more broadly ensure that AI enriches the human experience. Indeed, the 2019 *Executive Order on Maintaining American Leadership in Artificial Intelligence* notes:

> Heads of implementing agencies that also perform or fund R&D (AI R&D agencies), shall consider AI as an agency R&D priority, as appropriate to their respective agencies' missions… Heads of such agencies shall take this priority into account when developing budget proposals and planning for the use of funds in Fiscal Year 2020 and in future years. Heads of these agencies shall also consider appropriate administrative actions to increase focus on AI for 2019.

26 https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf
27 http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf
28 https://cra.org/ccc/visioning/visioning-activities/2018-activities/artificial-intelligence-roadmap/
29 https://www.microsoft.com/en-us/research/research-area/artificial-intelligence/
30 https://cra.org/ccc/events/artificial-intelligence-roadmap-workshop-3-learning-and-robotics/
31 https://cra.org/ccc/wp-content/uploads/sites/2/2016/04/AI-for-Social-Good-Workshop-Report.pdf
32 https://openai.com/blog/ai-and-compute/

AI research investments are needed in areas with potential long-term payoffs. While an important component of long-term research is incremental research with predictable outcomes, long-term sustained investments in high-risk research can lead to high-reward payoffs. These payoffs can be seen in 5 years, 10 years, or more. A 2012 National Research Council report emphasizes the critical role of Federal investments in long-term research, noting "the long, unpredictable incubation period—requiring steady work and funding—between initial exploration and commercial deployment."[33] It further notes that "the time from first concept to successful market is often measured in decades." Well-documented examples of sustained fundamental research efforts that led to high-reward payoffs include the World Wide Web and deep learning. In both cases, the basic foundations began in the 1960s; it was only after 30+ years of continued research efforts that these ideas materialized into the transformative technologies witnessed today in many categories of AI.

The following subsections highlight some of these areas. Additional categories of important AI research are discussed in Strategies 2 through 6.

Advancing data-focused methodologies for knowledge discovery

As discussed in the 2016 *Federal Big Data Research and Development Strategic Plan*,[34] many fundamental new tools and technologies are needed to achieve intelligent data understanding and knowledge discovery. Further progress is needed in the development of more advanced machine learning algorithms that can identify all the 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. The veracity of data is particularly challenging when dealing with vast amounts of data, making it difficult for humans to assess and extract knowledge from it. 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 techniques, to create methods for discovering inconsistencies and anomalies in the data, and to develop approaches for incorporating human feedback. Researchers need to explore new methods to enable data and associated metadata to be mined simultaneously.

Many AI applications are interdisciplinary in nature and make use of heterogeneous data. Further investigation of multimodality machine learning is needed to enable knowledge discovery from a wide variety of different types of data (e.g., discrete, continuous, text, spatial, temporal, spatio-temporal, graphs). AI investigators must determine the amount of data needed for training and to properly address large-scale versus long-tail data needs. They must also determine how to identify and process rare events beyond purely statistical approaches; to work with knowledge sources (i.e., any type of information that explains the world, such as knowledge of the law of gravity or of social norms) as well as data sources, integrating models and ontologies in the learning process; and to obtain effective learning performance with little data when big data sources may not be available.

Enhancing the perceptual capabilities of AI systems

Perception is an intelligent system's window into the world. Perception begins with (possibly distributed) sensor data, which comes 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

---

33 National Research Council Computer Science Telecommunications Board, *Continuing Innovation in Information Technology* (The National Academies Press, Washington, D.C., 2012), 11; https://doi.org/10.17226/13427.

34 https://www.nitrd.gov/PUBS/bigdatardstrategicplan.pdf

geometric features, attributes, location, and velocity. 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 longer distances, with higher resolution, and in real time. Perception systems need to be able to integrate data from a variety of sensors and other sources, including the computational cloud, to determine what the AI system is currently perceiving and to allow the prediction of future states. Detection, classification, identification, and recognition of objects remain challenging, especially under cluttered and dynamic conditions. In addition, perception of humans must be greatly improved by using an appropriate combination of sensors and algorithms, so that AI systems can work more effectively with people.[10] A framework for calculating and propagating uncertainty throughout the perception process is needed to quantify the confidence level that the AI system has in its situational awareness and to improve accuracy.

Understanding theoretical capabilities and limitations of AI

While the ultimate goal for many AI algorithms is to address open challenges with human-like solutions, we do not have a good understanding of what the theoretical capabilities and limitations are for AI and the extent to which such human-like solutions are even possible with AI algorithms. Theoretical work is needed to better understand why AI techniques—especially machine learning—often work well in practice. 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, in order 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 up new lines of research on new hardware for AI systems. For example, when invented in the 1960s, Artificial Neural Networks (ANNs) could only be used to solve very simple problems. It only became feasible to use ANNs to solve complex problems after hardware improvements such as parallelization were made, and algorithms were adjusted to make use of the new hardware. Such developments were key factors in enabling today's significant advances in deep learning.

Pursuing research on general-purpose artificial intelligence

AI approaches can be divided into "narrow AI" and "general AI." Narrow AI systems perform individual tasks in specialized, well-defined domains, such as speech recognition, image recognition, and translation. Several recent, highly-visible, narrow AI systems, including IBM Watson and DeepMind's AlphaGo, have achieved major feats.[35,36] Indeed, these particular systems have been labeled "superhuman" because they have outperformed the best human players in Jeopardy! and Go, respectively. But these systems exemplify narrow AI, since they can only be applied to the tasks for which they are specifically designed. Using these systems on a wider range of problems requires a significant re-engineering effort. In contrast, the long-term goal of general AI is to create systems that

---

[35] In 2011, IBM Watson defeated two players considered among the best human players in the Jeopardy! game.

[36] In 2016, AlphaGo defeated the reigning world champion of Go, Lee Se-dol. Notably, AlphaGo combines deep learning and Monte Carlo search—a method developed in the 1980s—which itself builds on a probabilistic method discovered in the 1940s.

exhibit the flexibility and versatility of human intelligence in a broad range of cognitive domains, including learning, language, perception, reasoning, creativity, and planning. Broad learning capabilities would provide general AI systems the ability to transfer knowledge from one domain to another and to interactively learn from experience and from humans. General AI has been an ambition of researchers since the advent of AI, but current systems are still far from achieving this goal. The relationship between narrow and general AI is currently being explored; it is possible that lessons from one can be applied to improve the other and vice versa. While there is no general consensus, most AI researchers believe that general AI is still decades away, requiring a long-term, sustained research effort to achieve it.

Developing scalable AI systems

Groups and networks of AI systems may be coordinated or autonomously collaborate to perform tasks not possible with a single AI system, and may also include humans working alongside or leading the team. The development and use of such multi-AI systems creates significant research challenges in planning, coordination, control, and scalability of such systems. Planning techniques for multi-AI systems must be fast enough to operate and adapt in real time to changes in the environment. They should adapt in a fluid manner to changes in available communications bandwidth or system degradation and faults. Many prior efforts have focused on centralized planning and coordination techniques; however, these approaches are subject to single points of failure, such as the loss of the planner, or loss of the communications link to the planner. Distributed planning and control techniques are harder to achieve algorithmically, and are often less efficient and incomplete, but potentially offer greater robustness to single points of failure. Future research must discover more efficient, robust, and scalable techniques for planning, control, and collaboration of teams of multiple AI systems and humans.

Fostering research on human-like AI

Attaining human-like AI requires systems to explain themselves in ways that people can understand. This will result in a new generation of intelligent systems, such as intelligent tutoring systems and intelligent assistants that are effective in assisting people when performing their tasks. There is a significant gap, however, between the way current AI algorithms work and how people learn and perform tasks. People are capable of learning from just a few examples, or by receiving formal instruction and/or "hints" to performing tasks, or by observing other people performing those tasks. Medical schools take this approach, for example, when medical students learn by observing an established doctor performing a complex medical procedure. Even in high-performance tasks such as world-championship Go games, a master-level player would have played only a few thousand games to train him/herself. In contrast, it would take hundreds of years for a human to play the number of games needed to train AlphaGo. More foundational research on new approaches for achieving human-like AI would bring these systems closer to this goal.

Developing more capable and reliable robots

Significant advances in robotic technologies over the last decade are leading to potential impacts in a multiplicity of applications, including manufacturing, logistics, medicine, healthcare, defense and national security, agriculture, and consumer products. While robots were historically envisioned for static industrial environments, recent advances involve close collaborations between robots and humans. Robotics technologies are now showing promise in their ability to complement, augment, enhance, or emulate human physical capabilities or human intelligence. However, scientists need to make these robotic systems more capable, reliable, and easy-to-use.

Researchers need to better understand robotic perception to extract information from a variety of sensors to provide robots with real-time situational awareness. 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 will allow robots to generalize their skills, perform self-assessment of their current performance, and learn a repertoire of physical movements from human teachers. Mobility and manipulation 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.

Advancing hardware for improved AI

While AI research is most commonly associated with advances in software, the performance of AI systems has been heavily dependent on the hardware upon which it runs. The current renaissance in deep machine learning is directly tied to progress in GPU-based hardware technology and its improved memory,[37] input/output, clock speeds, parallelism, and energy efficiency. Developing hardware optimized for AI algorithms will enable even higher levels of performance than GPUs. One example is "neuromorphic" processors that are loosely inspired by the organization of the brain and, in some cases, optimized for the operation of neural networks.[38]

Hardware advances can also improve the performance of AI methods that are highly data-intensive. Further study of methods to turn on and off data pipelines in controlled ways throughout a distributed system is called for. Continued research is also needed to allow machine learning algorithms to efficiently learn from high-velocity data, including distributed machine learning algorithms that simultaneously learn from multiple data pipelines. More advanced machine learning-based feedback methods will allow AI systems to intelligently sample or prioritize data from large-scale simulations, experimental instruments, and distributed sensor systems, such as Smart Buildings and the Internet of Things (IoT). Such methods may require dynamic I/O 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

While improved hardware can lead to more capable AI systems, AI systems can also improve the performance of hardware.[39] This reciprocity will lead to further advances in hardware performance, since physical limits on computing require novel approaches to hardware designs.[40] AI-based methods could be especially important for improving the operation of high-performance computing (HPC) systems. Such systems consume vast quantities of energy. AI is being used to predict 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

---

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

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

[39] M. Milano and L. Benini, "Predictive Modeling for Job Power Consumption in HPC Systems," In *Proceedings of High Performance Computing: 31st International Conference, ISC High Performance 2016* (Springer Vol. 9697, 2016).

[40] 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 in order to limit temperatures and ensure data integrity.

self-reconfigurable HPC systems that can handle system faults when they occur, without human intervention.[41]

Improved AI algorithms can increase the performance of multi-core systems by reducing data movements between processors and memory—the primary impediment to exascale computing systems that operate 10 times faster than today's supercomputers.[42] In practice, the configuration of executions in HPC systems are never the same, and different applications are executed concurrently, with the state of each different software code evolving independently in time. AI algorithms need to be designed to operate online and at scale for HPC systems.

[41] A. Cocaña-Fernández, J. Ranilla, and L. Sánchez, "Energy-efficient allocation of computing node slots in HPC clusters through parameter learning and hybrid genetic fuzzy system modeling," *Journal of Supercomputing* 71 (2015):1163-1174.

[42] Exascale computing systems can achieve at least a billion billion calculations per second.

Strategy 2: Develop Effective Methods for Human-AI Collaboration

2019 Update	**Developing AI systems that complement and augment human capabilities, with increasing focus on the future of work**

Since the release of the 2016 *National AI R&D Strategic Plan*, national interest has grown in human-AI collaboration. When AI systems complement and augment human capabilities, humans and AI become partners across a range of shared to fully autonomous scenarios. In particular, human-AI collaboration has been elevated as both a challenge and an opportunity in the context of the future of work.

In the past three years, newly established as well as longstanding conferences, workshops, and task forces have prioritized human-AI collaboration broadly. For example, the Conference on Human Computation and Crowdsourcing has grown from a workshop to a major international conference that fosters research in the intersection of AI and human-computer interaction (HCI).[45] In 2018, the Association for the Advancement of Artificial Intelligence selected human-AI collaboration as the emerging topic for its annual conference.[46] In May 2019, the largest conference on human-computer interaction, CHI, included a workshop on "Bridging the Gap Between AI and HCI."[47] The journal *Human-Computer Interaction* put out a call in March 2019 for submissions for a special issue on "unifying human-computer interaction and artificial intelligence."[48]

Human-AI Collaboration: Recent agency R&D programs

Since the release of the 2016 *National AI R&D Strategic Plan*, several agencies have initiated efforts for Strategy 2:

- NSF's Future of Work at the Human-Technology Frontier[43] Big Idea is supporting socio-technical research enabling a future where intelligent technologies collaborate synergistically with humans to achieve broad participation in the workforce and improve the social, economic, and environmental benefits across a range of work settings.
- NOAA (National Oceanographic and Atmospheric Administration) is advancing human-AI collaboration for hurricane, tornado, and other severe weather predictions where the human forecaster and an AI system work together to improve severe weather warning generation and to identify distinct patterns that are precursors to extreme events. Sometimes referred to as "humans above the loop," human forecasters oversee the AI system's predictions and direct the outcomes.
- NIH has ongoing research in natural language processing based on a database of 96.3 million facts extracted from all MEDLINE citations maintained by the National Library of Medicine.
- A 2019 DOE workshop report on Scientific Machine Learning identified priority research directions, major scientific use cases, and the emerging trend that human-AI collaborations will transform the way science is done.[44]

---

43 https://www.nsf.gov/eng/futureofwork.jsp

44 DOE workshop report, *Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence*: https://www.osti.gov/biblio/1478744.

45 Welcome to HCOMP 2019: https://www.humancomputation.com/.

46 AAAI-18 Emerging Topic Human-AI Collaboration: http://www.aaai.org/Conferences/AAAI/2018/aaai18emergingcall.php.

47 Where is the Human? Bridging the Gap Between AI and HCI: CHI 2019 Workshop: https://michae.lv/ai-hci-workshop/.

48 Call: "Unifying Human Computer Interaction and Artificial Intelligence" issue of *Human-Computer Interaction*: https://ispr.info/2019/02/20/call-unifying-human-computer-interaction-and-artificial-intelligence-issue-of-human-computer-interaction/.

In the context of work, conferences have emerged exploring the role of the human, the machine, and their partnership, such as MIT's Computer Science and Artificial Intelligence Lab (CSAIL) and the Initiative on the Digital Economy that launched the Annual AI and the Future of Work Congress.[49,50] As part of *A 20-Year Community Roadmap for Artificial Intelligence Research in the U.S.*,[51] in 2019 the Computing Community Consortium (CCC) held a workshop focused on meaningful interaction between humans and AI systems.[52] Additionally, the CCC operated the Human Technology Frontier task force in 2017-2018 to focus on the potential of technology to augment human performance in, including but not limited to, the workplace, the classroom, and the healthcare system.[53]

The cross-strategy principle in the 2016 *National AI R&D Strategic Plan*, "appropriate trust of AI systems requires explainability, especially as the AI grows in scale and complexity," has seen an R&D call to action in the context of human-AI collaborations. This principle has been identified by a number of professional societies and agencies as a priority area (see sidebar). This research area reflects the intersection of Strategies 2 and 3, as explainability, fairness, and transparency are key principles for AI systems to effectively collaborate with humans. Likewise, the challenge of understanding and designing human-AI ethics and value alignment into systems remains an open research area. In parallel, the private sector has responded with principles for effective human-AI collaboration.[54,55]

As Federal agencies have increased AI investments in the past three years along mission objectives, they have shared a common emphasis on human-machine cognition, autonomy, and agency, such as in decision support, risk modeling, situational awareness, and trusted machine intelligence (see sidebar). Through such R&D investments, research partnerships are growing across a number of axes, bringing together computational scientists; behavioral, cognitive, and psychological scientists; and scientists and engineers from other domains. New collaborations have formed between academic researchers and users of AI systems inside and outside the workplace.

Moving forward, it is critical that Federal agencies continue to foster AI R&D in the open world to promote the design of AI systems that incorporate and accommodate the situations and goals of users so that AI systems and users can work collaboratively in both anticipated and unanticipated circumstances.

While completely autonomous AI systems will be important in some application domains (e.g., underwater or deep space exploration), many other application areas (e.g., disaster recovery and medical diagnostics) are most effectively addressed by a combination of humans and AI systems working together to achieve application goals. This collaborative interaction takes advantage of the complementary nature of humans and AI systems. While effective approaches for human-AI collaboration already exist, most of these are "point solutions" that only work in specific environments using specific platforms toward specific goals. Generating point solutions for every possible application instance does not scale; more work is thus needed to go beyond these point solutions toward more

49 https://futureofwork.csail.mit.edu/.
50 AI and Future of Work Innovation Summit 2019: https://analyticsevent.com/.
51 https://cra.org/ccc/wp-content/uploads/sites/2/2019/03/AI_Roadmap_Exec_Summary-FINAL-.pdf
52 Artificial Intelligence Roadmap Workshop 2 – Interaction: https://cra.org/ccc/events/artificial-intelligence-roadmap-workshop-2-interaction/.
53 https://cra.org/ccc/human-technology-frontier/
54 https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf
55 https://www.partnershiponai.org/about/#our-work

general methods of human-AI collaboration. The tradeoffs must be explored between designing general systems that work in all types of problems, requiring less human effort to build and greater facility for switching between applications, versus building a large number of problem-specific systems that may work more effectively for each problem.

Future applications will vary considerably in the functional role divisions between humans and AI systems, the nature of the interactions between humans and AI systems, the number of humans and other AI systems working together, and how humans and AI systems will communicate and share situational awareness. Functional role divisions between humans and AI systems typically fall into one of the following categories:

1. *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.
2. *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.
3. *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 where a system should respond very rapidly (e.g., in nuclear reactor control rooms).

Achieving effective interactions between humans and AI systems requires additional R&D to ensure that the system design does not lead to excessive complexity, undertrust, or overtrust. The familiarity of humans with AI systems can be increased through training and experience, to ensure that the human has a good understanding of the AI system's capabilities and what the AI system can and cannot do. To address these concerns, certain human-centered automation principles should be used in the design and development of these systems:[56]

1. Employ intuitive, user-friendly design of human-AI system interfaces, controls, and displays.
2. Keep the operator informed. Display critical information, states of the AI system, and changes to these states.
3. Keep the operator trained. Engage in recurrent training for general knowledge, skills, and abilities (KSAs), as well as training in algorithms and logic employed by AI systems and the expected failure modes of the system.
4. Make automation flexible. Deploying AI systems should be considered as a design option for operators who wish to decide whether they want to use them or not. Also important is the design and deployment of adaptive AI systems that can be used to support human operators during periods of excessive workload or fatigue.[57,58]

Many fundamental challenges arise for researchers when creating systems that work effectively with humans. Several of these important challenges are outlined in the following subsections.

---

[56] C. Wickens and J. G. Hollands, "Attention, time-sharing, and workload." In *Engineering, Psychology and Human Performance* (London: Pearson PLC, 1999), 439-479.

[57] https://www.nasa.gov/mission_pages/SOFIA/index.html

[58] https://cloud1.arc.nasa.gov/intex-na/

Seeking new algorithms for human-aware AI

Over the years, AI algorithms have become able to solve problems of increasing complexity. However, there is a gap between the capabilities of these algorithms and the usability of these systems by humans. *Human-aware* intelligent systems are needed that can interact intuitively with users and enable seamless machine-human collaborations. Intuitive interactions include shallow interactions, such as when a user discards an option recommended by the system; model-based approaches that take into account the users' past actions; or even deep models of user intent that are based upon accurate human cognitive models. Interruption models must be developed that allow an intelligent system to interrupt the human only when necessary and appropriate. Intelligent systems should also have the ability to augment human cognition, knowing which information to retrieve when the user needs it, even when they have not prompted the system explicitly for that information. Future intelligent systems must be able to account for human social norms and act accordingly. Intelligent systems can more effectively work with humans if they possess some degree of emotional intelligence, so that they can recognize their users' emotions and respond appropriately. An additional research goal is to go beyond interactions of one human and one machine, toward a "systems-of-systems", that is, teams composed of multiple machines interacting with multiple humans.

Human-AI system interactions have a wide range of objectives. AI systems need the ability to represent a multitude of goals, actions that they can take to reach those goals, constraints on those actions, and other factors, as well as easily adapt to modifications in the goals. In addition, humans and AI systems must share common goals and have a mutual understanding of them and relevant aspects of their current states. Further investigation is needed to generalize these facets of human-AI systems to develop systems that require less human engineering.

Developing AI techniques for human augmentation

While much of the prior focus of AI research has been on algorithms that match or outperform people performing narrow tasks, more work is needed to develop systems that augment human capabilities across many domains. Human augmentation research includes algorithms that work on a stationary device (such as a computer); wearable devices (such as smart glasses); implanted devices (such as brain interfaces); and in specific user environments (such as specially tailored operating rooms). For example, augmented human awareness could enable a medical assistant to point out a mistake in a medical procedure, based on data readings combined from multiple devices. Other systems could augment human cognition by helping the user recall past experiences applicable to the user's current situation.

Another type of collaboration between humans and AI systems involves active learning for intelligent data understanding. In active learning, input is sought from a domain expert and learning is only performed on data when the learning algorithm is uncertain. This is an important technique to reduce the amount of training data that needs to be generated in the first place, or the amount that needs to be learned. Active learning is also a key way to obtain domain expert input and increase trust in the learning algorithm. Active learning has so far only been used within supervised learning; further research is needed to incorporate active learning into unsupervised learning (e.g., clustering, anomaly detection) and reinforcement learning.[59] Probabilistic networks allow domain knowledge to be included in the form of prior probability distributions. General ways of allowing machine learning algorithms to incorporate domain knowledge must be sought, whether in the form of mathematical models, text, or others.

[59] While supervised learning requires humans to provide the ground-truth answers, reinforcement learning and unsupervised learning do not.

Developing techniques for visualization and human-AI interfaces

Better visualization and user interfaces are additional areas that need much greater development to help humans understand large-volume modern datasets and information coming from a variety of sources. Visualization and user interfaces must clearly present increasingly complex data and information derived from them in a human-understandable way. Providing real-time results is important in safety-critical operations and may be achieved with increasing computational power and connected systems. In these types of situations, users need visualization and user interfaces that can quickly convey the correct information for real-time response.

Human-AI collaboration can be applied in a wide variety of environments, and where there are constraints on communication. In some domains, human-AI communication latencies are low and communication is rapid and reliable. In other domains (e.g., NASA's deployment of the rovers Spirit and Opportunity to Mars), remote communication between humans and the AI system has a very high latency (e.g., round trip times of 5-20 minutes between Earth and Mars), thus requiring the deployed platform(s) to operate largely autonomously, with only high-level strategic goals communicated to the platform. These communications requirements and constraints are important considerations for the R&D of user interfaces.

Developing more effective language processing systems

Enabling people to interact with AI systems through spoken and written language has long been a goal of AI researchers. While significant advances have been made, considerable open research challenges must be addressed in language processing before humans can communicate as effectively with AI systems as they do with other humans. Much recent progress in language processing has been credited to the use of data-driven machine learning approaches, which have resulted in successful systems that, for example, successfully recognize fluent English speech in quiet surroundings in real time. These achievements, however, are only first steps toward reaching longer-term goals. Current systems cannot deal with real-world challenges such as speech in noisy surroundings, heavily accented speech, children's speech, impaired speech, and speech for sign languages. The development of language processing systems capable of engaging in real-time dialogue with humans is also needed. Such systems will need to infer the goals and intentions of its human interlocutors, use the appropriate register, style and rhetoric for the situation, and employ repair strategies in case of dialogue misunderstandings. Further research is needed on developing systems that more easily generalize across different languages. Additionally, more study is required on acquiring useful structured domain knowledge in a form readily accessible by language processing systems.

Language processing advances in many other areas are also needed to make interactions between humans and AI systems more natural and intuitive. Robust computational models must be built for patterns in both spoken and written language that provide evidence for emotional state, affect, and stance, and for determining the information that is implicit in speech and text. New language processing techniques are needed for grounding language in the environmental context for AI systems that operate in the physical world, such as in robotics. Finally, since the manner in which people communicate in online interactions can be quite different from voice interactions, models of languages used in these contexts must be perfected so that social AI systems can interact more effectively with people.

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

2019 Update	**Addressing ethical, legal, and societal considerations in AI**

Since the release of the 2016 *National AI R&D Strategic Plan*, R&D activities addressing the ethical, legal, and societal implications of AI system development and deployment have increased. There is a growing realization that AI systems must be "trustworthy," and that AI could transform many sectors of social and economic life, including employment, healthcare, and manufacturing. International organizations such as the Organisation for Economic Co-operation and Development (OECD)[63] and the G7 Innovation Ministers[64] have encouraged R&D to increase trust in and adoption of AI.

The 2016 *National AI R&D Strategic Plan* was prescient in identifying research themes in privacy; improving fairness, transparency, and accountability of AI systems by design; and designing architectures for ethical AI. Research conferences dedicated to fairness, accountability, and transparency in ML and AI systems have flourished.[65] Federal agencies have responded with a variety of new research programs and meetings focused on these critical areas (see sidebar).

Explainability, fairness, and transparency: Recent agency R&D programs

Since the release of the 2016 *National AI R&D Strategic Plan*, a number of agencies have initiated AI R&D programs for Strategy 3:

- DARPA's Explainable AI (XAI) program[60] aims to create a suite of ML techniques that produce more explainable AI systems while maintaining a high level of learning performance (prediction accuracy). XAI will also enable human users to understand, appropriately trust, and effectively manage the emerging generation of AI systems. More generally, DoD is committed to "leading in military ethics and AI safety" as one of five key actions outlined in the strategic approach that guides its efforts to accelerate the adoption of AI systems.[61]
- NSF and Amazon are collaborating[62] to jointly support research focused on AI fairness with the goal of contributing to trustworthy AI systems that are readily accepted and deployed to tackle grand challenges facing society. Specific topics of interest include, but are not limited to, transparency, explainability, accountability, potential adverse biases and effects, mitigation strategies, validation of fairness, and considerations of inclusivity.

60 https://www.darpa.mil/program/explainable-artificial-intelligence
61 "Summary of the 2018 Department of Defense Artificial Intelligence Strategy": https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.
62 https://www.nsf.gov/pubs/2019/nsf19571/nsf19571.htm
63 "OECD Initiatives on AI": http://www.oecd.org/going-digital/ai/oecd-initiatives-on-ai.htm.
64 "G7 Innovation Ministers' Statement on AI": http://www.g8.utoronto.ca/employment/2018-labour-annex-b-en.html.
65 http://www.fatml.org/; https://fatconference.org/; http://www.aies-conference.com/

The 2019 *Executive Order on Maintaining American Leadership in Artificial Intelligence* emphasizes that maintaining American leadership in AI requires a concerted effort to promote advancements in technology and innovation, while protecting civil liberties, privacy, and American values:[1]

> The United States must foster public trust and confidence in AI technologies and protect civil liberties, privacy, and American values in their application in order to fully realize the potential of AI technologies for the American people.

More R&D is needed to develop AI architectures that incorporate ethical, legal, and societal concerns through technical mechanisms such as transparency and explainability. This R&D will require intensive collaboration among technical experts as well as stakeholders and specialists in other fields including the social and behavioral sciences, law, ethics, and philosophy. Since ethical decisions may also be heavily context- or application-dependent, collaboration with domain experts could be required as well. This interdisciplinary approach could be incorporated in the training, design, testing, evaluation, and implementation of AI in the interests of understanding and accounting for AI-induced decisions and actions and mitigating unintended consequences.

Federal agencies should therefore continue to foster the growing community of interest in further R&D of these issues by sponsoring research and convening experts and stakeholders.

When AI agents act autonomously, we expect them to behave according to the formal and informal norms to which we hold our fellow humans. As fundamental social ordering forces, law and ethics therefore both inform and adjudge the behavior of AI systems. The dominant research needs involve both understanding the ethical, legal, and social implications of AI, as well as developing methods for AI design that align with ethical, legal, and social principles. Privacy concerns must also be taken into account; further information on this issue can be found in the *National Privacy Research Strategy*.[66]

As with any technology, the acceptable uses of AI will be informed by the tenets of law and ethics; the challenge is how to apply those tenets to this new technology, particularly those involving autonomy, agency, and control.

As illuminated in "Research Priorities for Robust and Beneficial Artificial Intelligence,"[67]

> In order to build systems that robustly behave well, we of course need to decide what good behavior means in each application domain. This ethical dimension is tied intimately to questions of what engineering techniques are available, how reliable these techniques are, and what trade-offs are made—all areas where computer science, machine learning, and broader AI expertise is valuable.

Research in this area can benefit from multidisciplinary perspectives that involve experts from computer science, social and behavioral sciences, ethics, biomedical science, psychology, economics, law, and policy research. Further investigation is needed in areas both inside and outside of the NITRD-relevant IT domain (i.e., in information technology, as well as in the disciplines mentioned previously) to inform the R&D and use of AI systems and their impacts on society.

The following subsections explore key information technology research challenges in this area.

---

[66] https://www.nitrd.gov/pubs/NationalPrivacyResearchStrategy.pdf

[67] "An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence" (Future of Life Institute): http://futureoflife.org/ai-open-letter/.

Improving fairness, transparency, and accountability by design

Many concerns have been voiced about the susceptibility of data-intensive AI algorithms to error and misuse, and the possible ramifications for gender, age, racial, or economic classes. The proper collection and use of data for AI systems, in this regard, represent an important challenge. Beyond purely data-related issues, however, larger questions arise about the design of AI to be inherently just, fair, transparent, and accountable. Researchers must learn how to design these systems so that their actions and decision-making are transparent and easily interpretable by humans, and thus can be examined for any bias they may contain, rather than just learning and repeating these biases. There are serious intellectual issues about how to represent and "encode" value and belief systems. Scientists must also study to what extent justice and fairness considerations can be designed into the system, and how to accomplish this within the bounds of current engineering techniques.

Building ethical AI

Beyond fundamental assumptions of justice and fairness are other concerns about whether AI systems can exhibit behavior that abides by general ethical principles. How might advances in AI frame new "machine-relevant" questions in ethics, or what uses of AI might be considered unethical? Ethics is inherently a philosophical question while AI technology depends on, and is limited by, engineering. Within the limits of what is technologically feasible, therefore, researchers must strive to develop algorithms and architectures that are verifiably consistent with, or conform to, existing laws, social norms and ethics—clearly a very challenging task. Ethical principles are typically stated with varying degrees of vagueness and are hard to translate into precise system and algorithm design. There are also complications when AI systems, particularly with new kinds of autonomous decision-making algorithms, face moral dilemmas based on independent and possibly conflicting value systems. Ethical issues vary according to culture, religion, and beliefs. However, acceptable ethics reference frameworks can be developed to guide AI system reasoning and decision-making in order to explain and justify its conclusions and actions. A multidisciplinary approach is needed to generate datasets for training that reflect an appropriate value system, including examples that indicate preferred behavior when presented with difficult moral issues or with conflicting values. These examples can include legal or ethical "corner cases," labeled by an outcome or judgment that is transparent to the user.[68] AI needs adequate methods for values-based conflict resolution, where the system incorporates principles that can address the realities of complex situations where strict rules are impracticable.

Designing architectures for ethical AI

Additional progress in fundamental research must be made to determine how to best design architectures for AI systems that incorporate ethical reasoning. A variety of approaches have been suggested, such as a two-tier monitor architecture that separates the operational AI from a monitor agent that is responsible for the ethical or legal assessment of any operational action.[68] An alternative view is that safety engineering is preferred, in which a precise conceptual framework for the AI agent architecture is used to ensure that AI behavior is safe and not harmful to humans.[69] A third method is to formulate an ethical architecture using set theoretic principles, combined with logical constraints

---

[68] A. Etzioni and O. Etzioni, "Designing AI Systems that Obey Our Laws and Values," *Communications of the ACM* 59(9) (2016):29-31.

[69] R. Y. Yampolsky, "Artificial Intelligence Safety Engineering: Why Machine Ethics is a Wrong Approach." In *Philosophy and Theory of Artificial Intelligence,* ed. V.C. Muller (Heidelberg: Springer Verlag, 2013), 389-396.

on AI system behavior that restrict action to conform to ethical doctrine.[70] As AI systems become more general, their architectures will likely include subsystems that can take on ethical issues at multiple levels of judgment, including:[71] rapid response pattern matching rules, deliberative reasoning for slower responses for describing and justifying actions, social signaling to indicate trustworthiness for the user, and social processes that operate over even longer time scales to enable the system to abide by cultural norms. Researchers will need to focus on how to best address the overall design of AI systems that align with ethical, legal, and societal goals.

[70] R. C. Arkin, "Governing Legal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture," Georgia Institute of Technology Technical Report, GIT-GVU-07-11, 2007.

[71] B. Kuipers, "Human-like Morality and Ethics for Robots," AAAI-16 Workshop on AI, Ethics and Society, 2016; https://web.eecs.umich.edu/~kuipers/papers/Kuipers-aaaiws-16.pdf

Strategy 4: Ensure the Safety and Security of AI Systems

2019 Update	**Creating robust and trustworthy AI systems**

Since the 2016 release of the *National AI R&D Strategic Plan*, there has been rapid growth in scientific and societal understanding of AI security and safety. Much of this new knowledge has helped identify new problems: it is more evident now how AI systems can be made to do the wrong thing, learn the wrong thing, or reveal the wrong thing, for example, through adversarial examples, data poisoning, and model inversion, respectively. Unfortunately, technical solutions for these AI safety and security problems remain elusive.

To address all of these problems, the safety and security of AI systems must be considered in all stages of the AI system lifecycle, from the initial design and data/model building, to verification and validation, deployment, operation, and monitoring. Indeed, the notion of "safety (or security) by design" might impart an incorrect notion that these are only concerns of system designers; instead, they must be considered throughout the system lifecycle, not just at the design stage, and so must be an important part of the AI R&D portfolio.

When AI components are connected to other systems or information that must be safe or secure, the AI vulnerabilities and performance requirements (e.g., very low false-positive and false-negative rates, when operating over high volumes of data)

AI safety and security: Recent agency R&D programs

Since the release of the 2016 *National AI R&D Strategic Plan*, a number of agencies have initiated efforts supporting Strategy 4:

- DOT published new Federal guidance for automated vehicles in October 2018 supporting the safe integration of automation into the broad multimodal surface transportation system. *Preparing for the Future of Transportation: Automated Vehicles 3.0*[72] advances DOT's principles for safe integration of automated vehicles. The document also reiterates prior safety guidance, provides new multimodal safety guidance, and outlines a process for working with DOT as this new technology evolves. As of May 2019, fourteen companies had released Voluntary Safety Self-Assessments detailing how they will incorporate safety into their design and testing of automated driving systems.[73]
- In December 2018, IARPA announced two programs on AI security: Secure, Assured, Intelligent Learning Systems (SAILS)[74] and Trojans in Artificial Intelligence (TrojAI).[75] DARPA announced another program in February 2019, Guaranteeing AI Robustness against Deception (GARD).[76] Together, these programs aim to combat a range of attacks on AI systems.
- As noted in Strategy 3, DoD is committed to "leading in military ethics and AI safety" as one of five key actions outlined in the strategic approach that guides its efforts to accelerate the adoption of AI systems.[77]

72 https://www.transportation.gov/av/3
73 https://www.nhtsa.gov/automated-driving-systems/voluntary-safety-self-assessment
74 https://www.iarpa.gov/index.php/research-programs/sails
75 https://www.iarpa.gov/index.php/research-programs/trojai
76 https://www.darpa.mil/news-events/2019-02-06
77 "Summary of the 2018 Department of Defense Artificial Intelligence Strategy": https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

are inherited by the larger systems. These challenges are not static; as AI systems continue to grow in capabilities, they will likely grow in complexity, making it ever harder for correct performance or privacy of information to be verified and validated. This complexity may also make it increasingly difficult to explain decisions in ways that justify high levels of trust from human users (see Strategy 3).

Making AI trustworthy—now and into the future—is a critical issue that requires Federal Government R&D investments (see sidebar), along with collaborative efforts among government, industry, academia, and civil society. Engineering trustworthy AI systems may benefit from borrowing existing practices in safety engineering in other fields that have learned how to account for potential misbehavior of non-AI autonomous or semi-autonomous systems. However, AI-specific problems mean that novel techniques for program analysis, testing, formal verification, and synthesis will be critical to establish that an AI-based system meets its specifications—that is, that the system does exactly what it is supposed to do and no more. These problems are exacerbated in AI-based systems that can be easily fooled, evaded, and misled in ways that can have profound security implications. An emerging research area is adversarial ML, which explores both the analysis of vulnerabilities in ML algorithms as well as algorithmic techniques that yield more robust learning. Well-known attacks on ML include adversarial classifier evasion attacks, where the attacker changes behavior to escape being detected, and poisoning attacks, where training data itself is corrupted. There is growing need for research that systematically explores the space of adversaries that attack ML and other AI-based systems and to design algorithms that provide provable robustness guarantees against classes of adversaries.

Methods must be developed to make safe and secure the creation, evaluation, deployment, and containment of AI, and these methods must scale to match the capability and complexity of AI. Evaluating these methods will require new metrics, control frameworks, and benchmarks for testing and assessing the safety of increasingly powerful systems. Both methods and metrics must incorporate human factors, with safe AI objectives defined by human designers' goals, safe AI operations defined by human users' habits, and safe AI metrics defined by human evaluators' understanding. Producing human-driven and human-understandable methods and metrics for the safety of AI systems will enable policymakers, the private sector, and the public to accurately judge the evolving AI safety landscape and appropriately proceed within it.

Before an AI system is put into widespread use, assurance is needed that the system will operate safely and securely, in a controlled manner. Research is needed to address this challenge of creating AI systems that are reliable, dependable, and trustworthy. As with other complex systems, AI systems face important safety and security challenges due to:[78]

- *Complex and uncertain environments*: In many cases, AI systems are designed to operate in complex environments, with a large number of potential states that cannot be exhaustively examined or tested. A system may confront conditions that were never considered during its design.
- *Emergent behavior*: For AI systems that learn after deployment, a system's behavior may be determined largely by periods of learning under unsupervised conditions. Under such conditions, it may be difficult to predict a system's behavior.
- *Goal misspecification*: Due to the difficulty of translating human goals into computer instructions, the goals that are programmed for an AI system may not match the goals that were intended by the programmer.

[78] J. Bornstein, "DoD Autonomy Roadmap – Autonomy Community of Interest," Presentation at NDIA 16th Annual Science & Engineering Technology Conference, March 2015.

- *Human-machine interactions*: In many cases, the performance of an AI system is substantially affected by human interactions. In these cases, variation in human responses may affect the safety of the system.[79]

To address these issues and others, additional investments are needed to advance AI safety and security,[80] including explainability and transparency, trust, verification and validation, security against attacks, and long-term AI safety and value-alignment.

Improving explainability and transparency

A key research challenge is increasing the "explainability" or "transparency" of AI. Many algorithms, including those based on deep learning, are opaque to users, with few existing mechanisms for explaining their results. This is especially problematic for domains such as healthcare, where doctors need explanations to justify a particular diagnosis or a course of treatment. AI techniques such as decision-tree induction provide built-in explanations but are generally less accurate. Thus, researchers must develop systems that are transparent, and intrinsically capable of explaining the reasons for their results to users.

Building trust

To achieve trust, AI system designers need to create accurate, reliable systems with informative, user-friendly interfaces, while the operators must take the time for adequate training to understand system operation and limits of performance. Complex systems that are widely trusted by users, such as manual controls for vehicles, tend to be transparent (the system operates in a manner that is visible to the user), credible (the system's outputs are accepted by the user), auditable (the system can be evaluated), reliable (the system acts as the user intended), and recoverable (the user can recover control when desired). A significant challenge to current and future AI systems remains the inconsistent quality of software production technology. As advances bring greater linkages between humans and AI systems, the challenge in the area of trust is to keep pace with changing and increasing capabilities, anticipate technological advances in adoption and long-term use, and establish governing principles and policies for the study of best practices for design, construction, and use, including proper operator training for safe operation.

Enhancing verification and validation

New methods are needed for verification and validation of AI systems. "Verification" establishes that a system meets formal specifications, while "validation" establishes that a system meets the user's operational needs. Safe AI systems may require new means of *assessment* (determining if the system is malfunctioning, perhaps when operating outside expected parameters), *diagnosis* (determining the causes for the malfunction), and *repair* (adjusting the system to address the malfunction). For systems operating autonomously over extended periods of time, system designers may not have considered every condition the system will encounter. Such systems may need to possess capabilities for self-assessment, self-diagnosis, and self-repair in order to be robust and reliable.

---

[79] J. M. Bradshaw, R. R. Hoffman, M. Johnson, and D. D. Woods, "The Seven Deadly Myths of Autonomous Systems," *IEEE Intelligent Systems* 28(3)(2013):54-61.

[80] See, for instance: D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mane, "Concrete Problems in AI Safety," 2016, arXiv: 1606.06565v2; S. Russell, D. Dewey, and M. Tegmark, "Research Priorities for Robust and Beneficial Artificial Intelligence," 2016, arXiv: 1602.03506; T. G. Dietterich and E. J. Horvitz, "Rise of Concerns about AI: Reflections and Directions," *Communications of the ACM*, 58(10)(2015); and K. Sotala and R. Yampolskiy, "Responses to catastrophic AGI risk: A survey," *Physica Scripta*, 90(1), 19 December 2014.

Securing against attacks

AI embedded in critical systems must be robust in order to handle accidents but should also be secure to a wide range of intentional cyber attacks. Security engineering involves understanding the vulnerabilities of a system and the actions of actors who may be interested in attacking it. While cybersecurity R&D needs are addressed in greater detail in the NITRD 2016 *Federal Cybersecurity R&D Strategic Plan*,[81] some cybersecurity risks are specific to AI systems. For example, one key research area is “adversarial machine learning” that explores the degree to which AI systems can be compromised by “contaminating” training data, by modifying algorithms, or by making subtle changes to an object that prevent it from being correctly identified (e.g., prosthetics that spoof facial recognition systems). The implementation of AI in cybersecurity systems that require a high degree of autonomy is also an area for further study. One recent example of work in this area is DARPA’s Cyber Grand Challenge that involved AI agents autonomously analyzing and countering cyber attacks.[82]

Achieving long-term AI safety and value-alignment

AI systems may eventually become capable of “recursive self-improvement,” in which substantial software modifications are made by the software itself, rather than by human programmers. To ensure the safety of self-modifying systems, additional research is called for to develop: self-monitoring architectures that check systems for behavioral consistency with the original goals of human designers; confinement strategies for preventing the release of systems while they are being evaluated; value learning, in which the values, goals, or intentions of users can be inferred by a system; and value frameworks that are provably resistant to self-modification.

---

[81] https://www.nitrd.gov/pubs/2016-Federal-Cybersecurity-Research-and-Development-Strategic-Plan.pdf; this is being updated in 2019.

[82] https://archive.darpa.mil/CyberGrandChallenge_CompetitorSite/

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

2019 Update	**Increasing access to datasets and associated challenges**

At the time of the 2016 *National AI R&D Strategic Plan*'s release, publicly available datasets and environments were already playing a critical role in pushing forward AI R&D, particularly in areas such as computer vision, natural language processing, and speech recognition. ImageNet,[84] with more than 14 million labeled objects, along with associated computer vision community challenges (e.g., the ImageNet Large Scale Visual Recognition Challenge[85] that evaluates algorithms for object detection and image classification), have played an especially vital role in the community. As translational applications for ML are being found in myriad application areas such as healthcare, medicine, and smart and connected communities, the need has grown for publicly available datasets in domain-specific areas.

The importance of datasets and models – in particular, those of the Federal Government – is explicitly called out in the 2019 *Executive Order on Maintaining American Leadership in Artificial Intelligence*:[1]

> Heads of all agencies shall review their Federal data and models to identify opportunities to increase access and use by the greater non-Federal AI research community in a manner that benefits that community, while protecting safety, security, privacy, and confidentiality. Specifically, agencies shall improve data and model inventory documentation to enable discovery and usability, and shall

Shared Public Datasets and Environments for AI Training and Testing: Recent agency R&D programs

Since the release of the 2016 *National AI R&D Strategic Plan*, a number of agencies have initiated efforts supporting Strategy 5:

- DOT sponsored the Second Strategic Highway Research Program (SHRP2) Naturalistic Driving Study (NDS),[83] which recorded more than 5.4 million trips taken by more than 3,400 drivers and vehicles. An in-vehicle data acquisition system (DAS) unit gathered and stored data from forward radar, four video cameras, accelerometers, vehicle network information, a geographic positioning system, and an onboard lane tracker. Data from the DAS were recorded continuously while participants' vehicles were operating. Whereas summaries of the NDS data are public, access to the detailed datasets requires qualified research ethics training.
- The VA Data Commons is creating the largest linked medical–genomics dataset in the world with tools for enabling ML and AI, and guided by veterans' preferences. This effort is leveraging applicable NIST standards, laws, and executive orders.
- GSA (General Services Administration) is working to enable the use of cloud computing resources for federally funded AI R&D. Data.gov and code.gov, housed at GSA, contain over 246,000 datasets and code from across agencies and automatically harvest datasets released by agencies.
- The NIH Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) initiative, a partnership with industry-leading cloud service providers, is enabling researcher access to major data assets that are funded across NIH and that are stored in cloud environments.

[83] https://insight.shrp2nds.us/
[84] http://www.image-net.org/
[85] http://www.image-net.org/challenges/LSVRC/

prioritize improvements to access and quality of AI data and models based on the AI research community's feedback.

A new NSTC Subcommittee on Open Science was created in 2018 to coordinate Federal efforts on open and FAIR (findable, accessible, interoperable, and reusable) data. R&D investments will be needed to develop tools and resources that make it easier to identify, use, and manipulate relevant datasets (including Federal datasets), verify data provenance, and respect appropriate use policy. Many of these datasets themselves may be of limited use in an AI context without an investment in labeling and curation. Federal agencies should engage and work with AI stakeholders to ensure that appropriately vetted datasets and models that are released for sharing are ready and fit for use and that they are maintained as standards and norms evolve. Ultimately, development and adoption of best practices and standards in documenting dataset and model provenance will enhance trustworthiness and responsible use of AI technologies.

Since 2016, there have also been increased concerns about data content, such as potential bias (see Strategy 3)[86,87] or private information leakage. The 2016 *National AI R&D Strategic Plan* noted that "dataset development and sharing must … follow applicable laws and regulations and be carried out in an ethical manner." The DOT-supported InSight project provides such carefully structured access to data collected during the Naturalistic Driving Study (see sidebar). The 2016 *National AI R&D Strategic Plan* also noted that new "technologies are needed to ensure safe sharing of data, since data owners take on risk when sharing their data with the research community." For example, CryptoNets[88] allows neural networks to operate over encrypted data, ensuring that data remain confidential, because decryption keys are not needed in neural networks. Researchers have also begun developing new ML techniques that use a differential privacy framework to provide quantifiable privacy guarantees over the used data.[89] At the same time, privacy methods must remain sufficiently explainable and transparent to help researchers correct them and make them safe, efficient, and accurate. Furthermore, AI could reveal discoveries beyond the original or intended scope; therefore, researchers must remain cognizant of the potential dangers in access to data or discoveries by adversaries.

Data alone are of little use without the ability to bring computational resources to bear on large-scale public datasets. The importance of computational resources to AI R&D is called out in the 2019 *Executive Order on Maintaining American Leadership in Artificial Intelligence*:[1]

> The Secretaries of Defense, Commerce, Health and Human Services, and Energy, the Administrator of the National Aeronautics and Space Administration, and the Director of the National Science Foundation shall, to the extent appropriate and consistent with applicable law, prioritize the allocation of high-performance computing resources for AI-related applications through: (i) increased assignment of discretionary allocation of resources and resource reserves; or (ii) any other appropriate mechanisms.

---

86 Emily M. Bender and Batya Friedman, "Data Statements for NLP: Toward Mitigating System Bias and Enabling Better Science," *Transactions of the Association for Computational Linguistics* 6 (2018):587-604.

87 Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science* 356(6334):183-186, 14 Apr 2017.

88 Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin Lauter, Michael Naehrig, John Wernsing, "CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy," *2016 International Conference on Machine Learning* 48:201-210; http://proceedings.mlr.press/v48/.

89 Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, "Deep Learning with Differential Privacy," *23rd ACM Conference on Computer and Communications Security*," 2016: 308-318.

and:

> …the Select Committee, in coordination with the General Services Administration (GSA), shall submit a report to the President making recommendations on better enabling the use of cloud computing resources for federally funded AI R&D.

The need for computational capacity for many AI challenges has been increasing rapidly.[32] Federal funding may provide computational capabilities for Federally-funded research. Some companies and universities, however, may have additional computational demands. Overall, there is a national need to study and invest in shared computational resources to promote AI R&D.

The benefits of AI will continue to accrue, but only to the extent that training and testing resources for AI are developed and made available. The variety, depth, quality, and accuracy of training datasets and other resources significantly affects AI performance. Many different AI technologies require high-quality data for training and testing, as well as dynamic, interactive testbeds and simulation environments. More than just a technical question, this is a significant "public good" challenge, as progress would suffer if AI training and testing is limited to only a few entities that already hold valuable datasets and resources, yet we must simultaneously respect commercial and individual rights and interests in the data. Research is needed to develop high-quality datasets and environments for a wide variety of AI applications and to enable responsible access to good datasets and testing and training resources. Additional open-source software libraries and toolkits are also needed to accelerate the advancement of AI R&D. The following subsections outline these key areas of importance.

Developing and making accessible a wide variety of datasets to meet the needs of a diverse spectrum of AI interests and applications

The integrity and availability of AI training and testing datasets is crucial to ensuring scientifically reliable results. The technical as well as the socio-technical infrastructure necessary to support reproducible research in the digital area has been recognized as an important challenge—and is essential to AI technologies as well. The lack of vetted and openly available datasets with identified provenance to enable reproducibility is a critical factor to confident advancement in AI.[90] As in other data-intensive sciences, capturing data provenance is critical. Researchers must be able to reproduce results with the same as well as different datasets. Datasets must be representative of challenging real-world applications, and not just simplified versions. To make progress quickly, emphasis should be placed on making available already existing datasets held by government, those that can be developed with Federal funding, and, to the extent possible, those held by industry.

The machine learning aspect of the AI challenge is often linked with "big data" analysis. Considering the wide variety of relevant datasets, it remains a growing challenge to have appropriate representation, access, and analysis of unstructured or semi-structured data. How can the data be represented—in absolute as well as relative (context-dependent) terms? Current real-world databases can be highly susceptible to inconsistent, incomplete, and noisy data. Therefore, a number of data preprocessing techniques (e.g., data cleaning, integration, transformation, reduction, and representation) are important to establishing useful datasets for AI applications. How does the data preprocessing impact data quality, especially when additional analysis is performed?

---

[90] Toward this end, in 2016 the Intelligence Advanced Research Projects Activity issued a Request for Information on novel training datasets and environments to advance AI. See https://iarpa.gov/index.php/working-with-iarpa/requests-for-information/novel-training-datasets-and-environments-to-advance-artificial-intelligence.

Encouraging the sharing of AI datasets—especially for government-funded research—would likely stimulate innovative AI approaches and solutions. However, technologies are needed to ensure safe sharing of data, since data owners take on risk when sharing their data with the research community. Dataset development and sharing must also follow applicable laws and regulations and be carried out in an ethical manner. Risks can arise in various ways: inappropriate use of datasets, inaccurate or inappropriate disclosure, and limitations in data de-identification techniques to ensure privacy and confidentiality protections.

Making training and testing resources responsive to commercial and public interests

With the continuing explosion of data, data sources, and information technology worldwide, both the number and size of datasets are increasing. The techniques and technologies to analyze data are not keeping up with the high volume of raw information sources. Data capture, curation, analysis, and visualization are all key research challenges, and the science needed to extract valuable knowledge from enormous amounts of data is lagging behind. While data repositories exist, they are often unable to deal with the scaling up of datasets, have limited data provenance information, and do not support semantically rich data searches. Dynamic, agile repositories are needed.

One example of the kind of open/sharing infrastructure program that is needed to support the needs of AI research is the IMPACT program (Information Marketplace for Policy and Analysis of Cyber-risk & Trust) developed by the Department of Homeland Security (DHS).[91] This program supports the global cyber security risk research effort by coordinating and developing real-world data and information sharing capabilities, including tools, models, and methodologies. IMPACT also supports empirical data sharing between the international cybersecurity R&D community, critical infrastructure providers, and their government supporters. AI R&D would benefit from comparable programs across all AI applications.

Developing open-source software libraries and toolkits

The increased availability of open-source software libraries and toolkits provides access to cutting-edge AI technologies for any developer with an Internet connection. Resources such as the Weka toolkit,[92] MALLET,[93] and OpenNLP,[94] among many others, have accelerated the development and application of AI. Development tools, including free or low-cost code repository and version control systems, as well as free or low-cost development languages (e.g., R, Octave, and Python) provide low barriers to using and extending these libraries. In addition, for those who may not want to integrate these libraries directly, any number of cloud-based machine learning services exist that can perform tasks such as image classification on demand through low-latency web protocols that require little or no programming for use. Finally, many of these web services also offer the use of specialized hardware, including GPU-based systems. It is reasonable to assume that specialized hardware for AI algorithms, including neuromorphic processors, will also become widely available through these services.

Together, these resources provide an AI technology infrastructure that encourages marketplace innovation by allowing entrepreneurs to develop solutions that solve narrow domain problems without requiring expensive hardware or software, without requiring a high level of AI expertise, and permitting rapid scaling-up of systems on demand. For narrow AI domains, barriers to marketplace innovation are extremely low relative to many other technology areas.

[91] https://www.dhs.gov/csd-impact
[92] https://sourceforge.net/projects/weka/
[93] http://mallet.cs.umass.edu
[94] https://opennlp.apache.org

To help support a continued high level of innovation in this area, the U.S. Government can boost efforts in the development, support, and use of open AI technologies. Particularly beneficial would be open resources that use standardized or open formats and open standards for representing semantic information, including domain ontologies when available.

Government may also encourage greater adoption of open AI resources by accelerating the use of open AI technologies within the government itself, and thus help to maintain a low barrier to entry for innovators. Whenever possible, government should contribute algorithms and software to open source projects. Because government has specific concerns, such as a greater emphasis on data privacy and security, it may be necessary for the government to develop mechanisms to ease government adoption of AI systems. For example, it may be useful to create a task force that can perform a “horizon scan” across government agencies to find particular AI application areas within departments, and then determine specific concerns that would need to be addressed to permit adoption of such techniques by these agencies.

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

2019 Update	**Supporting development of AI technical standards and related tools**

The 2016 *National AI R&D Strategic Plan* states that "Standards, benchmarks, testbeds, and their adoption by the AI community are essential for guiding and promoting R&D of AI technologies." In the intervening three years, emphasis on standards and benchmarks has continued to rise in the U.S. and globally. The 2019 *Executive Order on Maintaining American Leadership in Artificial Intelligence* explicitly calls out the importance of such standards:[1]

> …[T]he Secretary of Commerce, through the Director of [NIST], shall issue a plan for Federal engagement in the development of technical standards and related tools in support of reliable, robust, and trustworthy systems that use AI technologies.

With AI innovation potentially impacting all sectors and domains of society, many standards development organizations have new AI-related considerations and work items underway, including activities related to AI ethics and trustworthy AI systems (see Strategy 3). 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, Subcommittee 42 on Artificial Intelligence[95]) to develop standards for AI systems and associated considerations. It is critical that Federal, industry, and academic researchers continue to inform these activities, particularly as AI advances and systems reach into areas such as transportation, health care, and food that align with the missions of government agencies.

Standards, benchmarks, and related tools: Recent agency R&D programs

Since the release of the 2016 *National AI R&D Strategic Plan*, NIST in particular has initiated efforts for Strategy 6:

- NIST is engaged in the standardization program of ISO/IEC JTC 1 SC 42 on Artificial Intelligence.[95] A NIST expert is the convener for the Big Data work effort in SC 42. The U.S. delegation to SC 42 includes NIST and other Federal agency experts, as well as representatives from industry and academia. U.S. input to SC 42 is facilitated by the International Committee for Information Technology Standards (INCITS).
- NIST staff participate in additional AI standards activities through standards organizations, such as the American Society of Mechanical Engineers, IEEE, and ISO/IEC. Their activities cover such topics as computational modeling for advanced manufacturing, ontologies for robotics and automation, personal data privacy, and algorithmic bias.
- NIST experts are raising awareness about the importance of consensus standards for AI in multilateral fora, including bodies such as G20 and G7.[96] NIST brings unique Federal Government expertise that grounds policy discussions in practice, in particular, through close collaboration with the private sector. Similarly, NIST lends its standards and related experience to intergovernmental bilateral discussions.

Since 2016, the surge in AI-related standards activities has outpaced the launch of new AI-focused benchmarks and evaluations, particularly as related to trustworthiness of AI systems. In the

95 https://www.iso.org/committee/6794475.html

96 https://home.treasury.gov/policy-issues/international/g-7-and-g-20

intervening time, however, considerations of fairness and bias in benchmark datasets have become increasingly important, with researchers pursuing new facial recognition datasets that seek to minimize bias. 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. Two such recent activities are MLPerf[97] and DAWNbench.[98]

Assessing, promoting, and assuring all aspects of AI trustworthiness requires measuring and evaluating AI technology performance through benchmarks and standards. Beyond being safe, secure, reliable, resilient, explainable, and transparent, trustworthy AI must preserve privacy while detecting and avoiding inappropriate bias. As AI technologies evolve, so will the need to develop new metrics and testing requirements for validation of these essential characteristics.

Standards, benchmarks, testbeds, and their adoption by the AI community are essential for guiding and promoting R&D of AI technologies. The following subsections outline areas where additional progress must be made.

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 technologies meet critical objectives for functionality and interoperability, and that they perform reliably and safely. Adoption of standards brings credibility to technology advancements and facilitates an expanded interoperable marketplace. One example of an AI-relevant standard that has been developed is P1872-2015 (Standard Ontologies for Robotics and Automation), developed by the Institute of Electrical and Electronics Engineers. 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, as well as provide a foundational basis for the application of AI technologies to robotics. Additional work in AI standards development is needed across all subdomains of AI.

Standards are needed to address:

- *Software engineering*: to manage system complexity, sustainment, security, and to monitor and control emergent behaviors;
- *Performance*: to ensure accuracy, reliability, robustness, accessibility, and scalability;
- *Metrics*: to quantify factors impacting performance and compliance to 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, as well as cybersecurity;
- *Privacy*: to control for the protection of information while being processed, when in transit, or being stored;

---

97 https://mlperf.org/

98 https://dawn.cs.stanford.edu/benchmark/

- *Traceability*: to provide a record of events (their implementation, testing, and completion), and for the curation of data; and
- *Domains*: to define domain-specific standard lexicons and corresponding frameworks.

Establishing AI technology benchmarks

Benchmarks, made up of tests and evaluations, provide quantitative 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 technologies, 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 performance of AI technologies. Standard metrics are needed to define quantifiable measures in order to characterize AI technologies, including but not limited to: accuracy, complexity, trust and competency, risk and uncertainty, explainability, unintended bias, comparison to human performance, and economic impact. It is important to note that benchmarks are data driven. Strategy 5 discusses the importance of datasets for training and testing.

As a successful example of AI-relevant benchmarks, the National Institute of Standards and Technology has developed a comprehensive set of standard test methods and associated performance metrics to assess key capabilities of emergency response robots. The objective is to facilitate quantitative comparisons of different robot models by making use of statistically significant data on robot capabilities that was captured using the standard test methods. These comparisons can guide purchasing decisions and help developers to understand deployment capabilities. The resulting test methods are being standardized though the ASTM International Standards Committee on Homeland Security Applications for robotic operational equipment (referred to as standard E54.08.01).[99] Versions of the test methods are used to challenge the research community through the RoboCup Rescue Robot League competitions,[100] which emphasize autonomous capabilities. Another example is the IEEE Agile Robotics for Industrial Automation Competition (ARIAC),[101] a joint effort between IEEE and NIST,[102] which promotes robot agility by utilizing the latest advances in artificial intelligence and robot planning. A core focus of this competition is to test the agility of industrial robot systems, with the goal of enabling those on the shop floors to be more productive, more autonomous, and requiring less time from shop floor workers.

While these 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.

Increasing the availability of AI testbeds

The importance of testbeds was stated in the *Cyber Experimentation of the Future* 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."[103] Having adequate testbeds is a

[99] 2019 update: The resulting test methods are now standards issued by ASTM International Standards Committee on Homeland Secruity Applications for Response Robots (referred to as E54.09).

[100] http://www.robocup2016.org/en/

[101] http://robotagility.wixsite.com/competition

[102] 2019 update: IEEE is no longer a partner of ARIAC, which is now in its third year.

[103] SRI International and USC Information Sciences Institute, "Cybersecurity Experimentation of the Future (CEF): Catalyzing a New Generation of Experimental Cybersecurity Research," Final Report, July 31, 2015.

need across all areas of AI. The government has massive amounts of mission-sensitive data unique to government, but much of this data cannot be distributed to the outside research community. Appropriate programs could be established for academic and industrial researchers to conduct research within secured and curated testbed environments established by specific agencies. AI models and experimental methods could be shared and validated by the research community by having access to these test environments, affording AI scientists, engineers, and students unique research opportunities not otherwise available.

Engaging the AI community in standards and benchmarks

Government leadership and coordination is needed to drive standardization and encourage its widespread use in government, academia, and industry. The AI community—made up of users, industry, academia, and government—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 in order to strengthen their impact. This coordination is needed to collectively gather user-driven requirements, anticipate developer-driven standards, and promote educational opportunities. User-driven requirements shape the objectives and design of challenge problems and enable technology evaluation. Having community benchmarks focuses 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 also 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 technologies. Promoting and coordinating their 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, as well as to develop best practices for technology applications.

One successful example of a high-impact, community-based, AI-relevant benchmark program is the Text Retrieval Conference (TREC),[104] which was started by NIST in 1992 to provide the infrastructure necessary for large-scale evaluation of information retrieval methodologies. More than 250 groups have participated in TREC, including academic and commercial organizations both large and small. The standard, widely available, and carefully constructed set of data put forth by TREC has been credited with revitalizing research on information retrieval.[105,106] A second example is the NIST periodic benchmark program in the area of machine vision applied to biometrics,[107] particularly face recognition.[108] This began with the Face Recognition Technology (FERET) evaluation in 1993, which provided a standard dataset of face photos designed to support face recognition algorithm development as well as an evaluation protocol. This effort has evolved over the years into the Face

[104] http://trec.nist.gov

[105] E. M. Voorhees and D. K. Harman, *TREC Experiment and Evaluation in Information Retrieval* (Cambridge: MIT Press, 2005).

[106] http://googleblog.blogspot.com/2008/03/why-data-matters.html

[107] http://biometrics.nist.gov

[108] http://face.nist.gov

Recognition Vendor Test (FRVT),[109] involving the distribution of datasets, hosting of challenge problems, and conducting of sequestered technology evaluations. This benchmark program has contributed greatly to the improvement of facial recognition technology. Both TREC and FRVT can serve as examples of effective AI-relevant community benchmarking activities, but similar efforts are needed in other areas of AI.

It is important to note that developing and adopting standards, as well as participating in benchmark activities, comes with a cost. R&D organizations are incentivized when they see significant benefit. Updating acquisition processes across agencies to include specific requirements for AI standards in requests for proposals will encourage the community to further engage in standards development and adoption. Community-based benchmarks, such as TREC and FRVT, 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 providing objective and comparative performance metrics for relevant source selections.

[109] P. J. Phillips, “Improving Face Recognition Technology,” *Computer* 44(3)(2011): 84-96.

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

2019 Update	**Advancing the AI R&D workforce, including those working on AI systems and those working alongside them, to sustain U.S. leadership**

Since the release of the 2016 *National AI R&D Strategic Plan*, the demand for AI researchers and practitioners has grown rapidly. Studies have shown that the number of hiring opportunities is expected to rise into the millions over the next decade. As one data point, the U.S. Bureau of Labor Statistics projects that the number of positions for computer and information scientists and engineers will grow by 19% from 2016 to 2026, almost three times faster than the average for all occupations.[111] Moreover, through 2028, AI researchers are expected contribute to as much as $11.5 trillion of cumulative growth promised by intelligent technologies in the G20 countries alone.[112]

U.S. academic institutions are struggling to keep pace with the explosive growth in student interest and enrollment in AI.[113,114,115] At the same time, industry, with its sustained financial support and access to advanced computing facilities and datasets, exerts a strong pull on academic research and teaching talent.[116]

It is critical to maintain a robust academic research ecosystem in AI that, in collaboration with industry R&D, can continue to deliver tremendous dividends[117] by advancing national health, prosperity, and welfare, and securing the national defense.

National AI R&D workforce: Recent agency activities

Since the release of the 2016 *National AI R&D Strategic Plan*, a number of agencies have initiated efforts supporting Strategy 7:

- Apart from supporting undergraduate and graduate students through standard AI research grants, agencies are prioritizing computational- and data-enabled science and engineering in their graduate fellowship programs. For example, in 2018, DOE added a new track to its Computational Science Graduate Fellowship program. This track supports students pursuing advanced degrees in applied mathematics, statistics, or computer science, and promotes more effective use of high-performance systems, including in the areas of AI, ML, and deep learning.[44,110] Also in 2018, NSF began prioritizing computational and data-enabled science and engineering in a subset of awardees of its Graduate Research Fellowships Program.
- The Census Bureau has created the Statistical Data Modernization (SDM) project to bring its workforce, operations, and technologies up to the current state of the art and set the standard for statistical agencies in today's data-driven society. SDM's workforce transformation component will enable the hiring of new data scientists with expertise in new methods and analytics, including the use of AI methods and tools to process and analyze big data. The workforce transformation will also address the upskilling of the current data science workforce.

---

110 https://www.krellinst.org/csgf/math-cs

111 https://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm

112 https://www.accenture.com/t20180920T094705Z__w__/us-en/_acnmedia/Thought-Leadership-Assets/PDF/Accenture-Education-and-Technology-Skills-Research.pdf

113 https://cra.org/data/generation-cs/

114 https://cra.org/wp-content/uploads/2018/05/2017-Taulbee-Survey-Report.pdf

115 http://web.cs.wpi.edu/~cew/papers/CSareas19.pdf

116 https://www.nitrd.gov/rfi/ai/2018/AI-RFI-Response-2018-Yolanda-Gil-AAAI.pdf

117 https://www.nap.edu/catalog/13427/continuing-innovation-in-information-technology

In the three years since the release of the 2016 *National AI R&D Strategic Plan*, various reports have called for continued support for the development of instructional materials and teacher professional development in computer science at all levels. Emphasis is needed at the K–12 levels to feed the Nation's pipeline of AI researchers over many decades.[118] At the undergraduate level, there is a need to focus on integrating advanced computational skills and methods with domain-specific knowledge from other disciplines, given the growing role of computing across disciplines.[119] Sustained support is also needed at the graduate level, where students are conducting fundamental research in ML and AI. Indeed, the 2019 *Executive Order on Maintaining American Leadership in Artificial Intelligence* requires that:[1]

> Heads of implementing agencies that also provide educational grants shall, to the extent consistent with applicable law, consider AI as a priority area within existing Federal fellowship and service programs … [including] … (A) high school, undergraduate, and graduate fellowship; alternative education; and training programs; (B) programs to recognize and fund early-career university faculty who conduct AI R&D, including through Presidential awards and recognitions; (C) scholarship for service programs; (D) direct commissioning programs of the United States Armed Forces; and (E) programs that support the development of instructional programs and curricula that encourage the integration of AI technologies into courses in order to facilitate personalized and adaptive learning experiences for formal and informal education and training.

More broadly, the need for a firm grounding in computational thinking, including through computer science education, is also noted prominently in the Federal Government's December 2018 five-year strategic plan for science, technology, engineering, and mathematics (STEM) education.[120]

In addition, it is imperative to broaden the participation among groups traditionally underrepresented in computing and related fields.

Finally, the AI R&D workforce will consist of multidisciplinary teams comprising not just computer and information scientists and engineers, but also experts from other fields key to AI and ML innovation and its application, including cognitive science and psychology, economics and game theory, engineering and control theory, ethics, linguistics, mathematics, philosophy, and the many domains in which AI may be applied.

Federal agencies are giving priority to training and fellowship programs at all levels to prepare the workforce with requisite AI R&D skills through apprenticeships, skills programs, fellowships, and course work in relevant disciplines (see sidebar). Such training opportunities target both scientists and engineers who contribute to AI R&D innovations and users of AI R&D who may possess relevant domain knowledge. In the case of the former, long-term Federal investment in AI R&D, as described in Strategy 1, further supports the growth of this workforce, both through training the next generation of researchers and by making faculty positions more attractive to current graduate and postdoctoral students. In the case of the latter, new programs are bringing AI-relevant skills to current and future users of AI systems (see sidebar). Federal agencies must therefore continue to strategically foster expertise in the AI R&D workforce that spans multiple disciplines and skill categories to ensure sustained national leadership.

---

[118] https://github.com/touretzkyds/ai4k12/wiki

[119] https://www.nap.edu/catalog/24926/assessing-and-responding-to-the-growth-of-computer-science-undergraduate-enrollments

[120] https://www.whitehouse.gov/wp-content/uploads/2018/12/STEM-Education-Strategic-Plan-2018.pdf

Attaining the needed AI R&D advances outlined in this strategy will require a sufficient AI R&D workforce. Nations with the strongest presence in AI R&D will establish leading positions in the automation of the future. They will become the frontrunners in competencies like algorithm creation and development; capability demonstration; and commercialization. Developing technical expertise will provide the basis for these advancements.

While no official AI workforce data currently exist, numerous recent reports from the commercial and academic sectors are indicating an increased shortage of available experts in AI. AI experts are reportedly in short supply,[121] with demand expected to continue to escalate.[122] High-tech companies are reportedly investing significant resources into recruiting faculty members and students with AI expertise.[123] Universities and industries are reportedly in a battle to recruit and retain AI talent.[124]

Additional studies are needed to better understand the current and future national workforce needs for AI R&D. Data is needed to characterize the current state of the AI R&D workforce, including the needs of academia, government, and industry. Studies should explore the supply and demand forces in the AI workplace, to help predict future workforce needs. An understanding is needed of the projected AI R&D workforce pipeline. Considerations of educational pathways and potential retraining opportunities should be included. Diversity issues should also be explored, since studies have shown that a diverse information technology workforce can lead to improved outcomes.[125] Once the current and future AI R&D workforce needs are better understood, then appropriate plans and actions can be considered to address any existing or anticipated workforce challenges.

[121] "Startups Aim to Exploit a Deep-Learning Skills Gap," *MIT Technology Review*, January 6, 2016.

[122] "AI talent grab sparks excitement and concern," *Nature*, April 26, 2016.

[123] "Artificial Intelligence Experts are in High Demand," *The Wall Street Journal*, May 1, 2015.

[124] "Million dollar babies: As Silicon Valley fights for talent, universities struggle to hold on to their stars," *The Economist*, April 2, 2016.

[125] J. W. Moody, C. M. Beise, A. B. Woszczynski, and M. E. Myers, "Diversity and the information technology workforce: Barriers and opportunities," *Journal of Computer Information Systems* 43 (2003): 63-71.

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

Strategy 8 is new in 2019 and reflects the growing importance of public-private partnerships enabling AI R&D.

American leadership in science and engineering research and innovation is rooted in the Nation's unique government-university-industry R&D ecosystem. As the American Association of Arts and Sciences has written, "America's standing as an innovation leader" relies upon "establishing a more robust national Government-University-Industry research partnership."[126] Since the release of the 2016 *National AI R&D Strategic Plan*, the Administration has amplified this vision of promoting "sustained investment in AI R&D in collaboration with academia, industry, international partners and allies, and other non-Federal entities to generate technological breakthroughs in AI and related technologies and to rapidly transition those breakthroughs into capabilities that contribute to U.S. economic and national security."[1]

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, multi-billion-dollar sectors of the Nation's economy.[127] Concurrent advances across government, universities, and industry have been mutually reinforcing and have led to an innovative, vibrant AI sector. Many of today's AI systems have been enabled by the American government-university-industry R&D ecosystem (see sidebar).

Advancing the Nation's AI innovation ecosystem, spanning government, universities, and industry
▪ Deep convolutional neural networks have proven to be a key innovation rooted in AI research. Although this modeling approach emerged from early Federal investments in the late 1980s, there were not enough data nor enough computational capabilities available at the time for neural networks to make accurate predictions. Through the combination of a rise in big data, today's data-intensive scientific methods, and conceptual advances in how to structure and optimize the networks, neural networks have re-emerged as a useful way to improve accuracy in AI models. Interactions between academia and the private sector, including government funding, in recent years have helped reduce the error rate in speech recognition systems, enabling innovations such as real-time translation.[126] ▪ Similarly, Federal investments in reinforcement learning in the 1980s and 1990s—an approach rooted in behavioral psychology that involves learning to associate behaviors with desired outcomes—have led to today's deep learning systems. Through interactions across sectors, computers are increasingly learning like humans, without explicit instruction, and reinforcement learning is driving progress in self-driving cars and other forms of automation where machines can hone skills through experience. Reinforcement learning was the key technology underlying AlphaGo, the program that defeated the world's best Go players, which has seen a growing number of victories over professional players since 2016.[126]

Since the release of the 2016 *National AI R&D Strategic Plan*, additional emphasis has been placed on the benefits of public-private partnerships. These benefits include strategically leveraging resources, including facilities, datasets, and expertise, to advance science and engineering innovations;

---

[126] *Restoring the Foundation: The Vital Role of Research in Preserving the American Dream* (American Academy of Arts and Sciences, Cambridge, MA, 2014); https://www.amacad.org/multimedia/pdfs/publications/researchpapersmonographs/AmericanAcad_RestoringtheFoundation_Brief.pdf.

[127] National Research Council Computer Science Telecommunications Board, *Continuing Innovation in Information Technology* (The National Academies Press, Washington, D.C., 2012); https://doi.org/10.17226/13427.

accelerating the transition of these innovations to practice; and enhancing education and training for next-generation researchers, technicians, and leaders. Government-university-industry R&D partnerships bring pressing, real-world challenges faced by industry to university researchers, enabling "use-inspired research"; leverage industry expertise to accelerate the transition of open and published research results into viable products and services in the marketplace for economic growth; and grow research and workforce capacity by linking university faculty and students with industry representatives, industry settings, and industry jobs (see sidebar).[126,128,129,130] These partnerships build upon joint engagements among Federal agencies that enable synergies in areas where agencies' missions intersect. The Nation also benefits from relationships between Federal agencies and international funders who can work together to address key challenges of mutual interest across a range of disciplines.

While there are many structures and mechanisms for public-private partnerships, some common categories for engagement include:

1. *Individual project-based collaborations.* These efforts enable engagement by university researchers with those in other sectors, including Federal agencies, industry, and international organizations, to identify and leverage synergies in areas of mutual interest.
2. *Joint programs to advance open, precompetitive, fundamental research*. Direct partnerships among organizations across sectors enable funding and support for open, precompetitive, fundamental research in areas of mutual interest to the partners. In general, non-Federal partners contributing research resources receive the same intellectual property rights afforded to the U.S. Government by the Bayh-Dole Act.[131]
3. *Collaborations to deploy and enhance research infrastructure*. Collaborations among Federal agencies, industry, and international organizations significantly enhance the potential for developing new and enhancing existing research infrastructure that in turn enables groundbreaking experimentation by researchers. Partners may offer financial and/or in-kind support to develop and/or enhance research infrastructure.
4. *Collaborations to enhance workforce development including broadening participation*. Multisector partnerships set the foundation for rigorous, engaging, and inspiring instructional materials that enhance workforce development and diversity in STEM professions.

In each of these cases, leveraging each partner's strengths for the benefit of all is vitally important to achieving success.

---

[128] Mathematical Sciences Research Institute report, "*Partnerships: A Workshop on Collaborations between the NSF/MPS & Private Foundations,*" 2015; http://library.msri.org/msri/Partnerships.pdf.

[129] Computing Community Consortium, *"The Future of Computing Research: Industry-Academic Collaborations,"* 2016; http://cra.org/ccc/wp-content/uploads/sites/2/2016/06/15125-CCC-Industry-Whitepaper-v4-1.pdf.

[130] Computing Community Consortium, *"Evolving Academia/Industry Relations in Computing Research: Interim Report released by the CCC,"* 2019; https://www.cccblog.org/wp-content/uploads/2019/03/Industry-Interim-Report-w-footnotes.pdf.

[131] https://history.nih.gov/research/downloads/PL96-517.pdf

Advances in AI R&D stand to benefit from all of these types of public-private partnerships. Partnerships can promote open, precompetitive, fundamental AI R&D; enhance access to research resources such as datasets, models, and advanced computational capabilities; and foster researcher exchanges and/or joint appointments between government, universities, and industry to share AI R&D expertise. Partnerships can also promulgate the inherently interdisciplinary nature of AI R&D, which requires convergence between computer and information science, cognitive science and psychology, economics and game theory, engineering and control theory, ethics, linguistics, mathematics and statistics, and philosophy to drive the development and evaluation of future AI systems that are fair, transparent, and accountable, as well as safe and secure. Federal agencies are actively pursuing public-private partnerships to achieve these goals (see *sidebar*).

Federal agencies must therefore continue to pursue and strengthen public-private partnerships in AI R&D to drive technology development and economic growth by leveraging investments and expertise in areas of mutual interest to government, industry, and academia. In doing so, the U.S. Government will capitalize on a uniquely American innovation ecosystem that has transformed nearly every aspect of the Nation's economy and society over the last two decades through novel information technologies.[127]

Public-private partnerships: Recent agency R&D programs

A number of agencies have already initiated public-private partnerships in support of AI R&D:

- The Defense Innovation Unit (DIU)[132] is a DoD organization that solicits commercial solutions capable of addressing DoD needs. The DIU in turn provides pilot contracts, which can include hardware, software, or other unique services. If successful, pilot contracts lead to follow-on contracts between companies and any DoD entity. A key DIU feature is the rapid pace of the pilot and subsequent contracts.
- NSF and the Partnership on AI, a diverse, multistakeholder organization working to better understand AI's impacts, are partnering to jointly support high-risk, high-reward research at the intersection of the social and technical dimensions of AI.[15]
- The DHS Science and Technology Directorate's Silicon Valley Innovation Program (SVIP)[133] looks to harness commercial R&D innovation ecosystems across the Nation and around the world for technologies with government applications. SVIP employs a streamlined application and pitch process; brings government, entrepreneurs, and industry together to find cutting-edge solutions; and co-invests in and accelerates transition to market.
- The Department of Health and Human Services (HHS) piloted the Health Tech Sprint initiative, also known in its first iteration as "Top Health," modeled in part after the Census Bureau's Opportunity Project. This effort created a nimble framework to public-private collaborations around bidirectional data links. It piloted new models for iterating on data release for AI training and testing, and it developed a voluntary incentivization framework for a public–private AI ecosystem.
- The HHS Division of Research, Innovation, and Ventures is part of the Biomedical Advanced Research and Development Authority at the Office of the Assistant Secretary for Preparedness and Response. It oversees an accelerator network and is recruiting a nonprofit partner that can work with private investors to fund innovative technologies and products to solve systemic health security challenges, with AI applications being one area of interest. Accelerators will connect startups and other businesses with product development and business support services.

[132] https://www.diu.mil/
[133] https://www.dhs.gov/science-and-technology/svip

Abbreviations

AFOSR	Air Force Office of Scientific Research
AI	artificial intelligence
DARPA	Defense Advanced Research Projects Agency
DHS	Department of Homeland Security
DoD	Department of Defense
DOE	Department of Energy
DOT	Department of Transportation
FDA	Food and Drug Administration
FRVT	Face Recognition Vendor Test
GPS	Global Positioning System
GPU	graphics processing unit
GSA	General Services Administration
HHS	Department of Health and Human Services
HPC	high-performance computing
IARPA	Intelligence Advanced Research Projects Activity
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IMPACT	Information Marketplace for Policy and Analysis of Cyber-risk & Trust (DHS)
ISO	International Organization for Standardization
IT	information technology
IWG	interagency working group
ML	machine learning
MLAI	Machine Learning and Artificial Intelligence (Subcommittee of the NSTC)
NASA	National Aeronautics and Space Administration
NCO	National Coordination Office for NITRD
NDS	Naturalistic Driving Study (DOT)
NIFA	National Institute of Food and Agriculture (USDA)
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NITRD	Networking and Information Technology Research and Development program
NLM	National Library of Medicine (NIH)
NSF	National Science Foundation
NSTC	National Science and Technology Council
NTIA	National Telecommunications and Information Administration
ODNI	Office of the Director of National Intelligence
OSTP	Office of Science and Technology Policy
R&D	research and development
RFI	Request for Information
STEM	science, technology, engineering, and mathematics
SVIP	Silicon Valley Innovation Program (DHS)
TREC	Text Retrieval Conference
USDA	U.S. Department of Agriculture
VA	U.S. Department of Veterans Affairs
XAI	explainable AI

National Science & Technology Council

Chair

Kelvin Droegemeier, Director, OSTP

Staff

Chloé Kontos, Executive Director, NSTC

Select Committee on Artificial Intelligence

Co-Chairs

Michael Kratsios, Deputy Assistant to the President for Technology Policy (The White House)

France A. Córdova, Director, NSF

Steven Walker, Director, DARPA

Subcommittee on Machine Learning and Artificial Intelligence

Co-Chairs

Lynne Parker, Assistant Director for Artificial Intelligence, OSTP

James Kurose, Assistant Director, Directorate for Computer Information Science and Engineering (CISE), NSF

Charles Romine, Director, Information Technology Laboratory, NIST

Stephen Binkley, Deputy Director, Science Programs, Office of Science, DOE

Executive Secretary

Faisal D'Souza, NITRD NCO

Subcommittee on Networking & Information Technology Research & Development

Co-Chairs

Kamie Roberts, Director, NITRD NCO

James Kurose, Assistant Director, CISE, NSF

Executive Secretary

Nekeia Butler, NITRD NCO

Artificial Intelligence Research & Development Interagency Working Group

Co-Chairs

Jeff Alstott, Program Manager, IARPA Office of the Director of National Intelligence (ODNI)

Henry Kautz, Division Director, CISE Division of Information and Intelligent Systems, NSF

Staff

Faisal D'Souza, Technical Coordinator, NITRD NCO

Strategic Plan Writing Team

Jeff Alstott, IARPA
Gil Alterovitz, VA
Sameer Antani, NIH
Charlotte Baer, NIFA, USDA
Daniel Clouse, ODNI
Faisal D'Souza, NITRD NCO
Kimberly Ferguson-Walter, U.S. Navy
Michael Garris, NIST
Erwin Gianchandani, NSF
Ross Gillfillan, OSTP
Travis Hall, NTIA
Meghan Houghton, NSF
Henry Kautz, NSF
Erin Kenneally, DHS
David Kuehn, DOT
James Kurose, NSF
James Lawton, AFOSR
Steven Lee, DOE
Aaron Mannes, DHS
Lynne Parker, OSTP
Dinesh Patwardhan, FDA
Elham Tabassi, NIST

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 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, homeland 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 (R&D) 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 Select Committee on Artificial Intelligence

The Select Committee on Artificial Intelligence (AI) advises and assists the NSTC to improve the overall effectiveness and productivity of Federal R&D efforts related to 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 R&D activities, including those related to autonomous systems, biometric identification, computer vision, human-computer interactions, machine learning, natural language processing, and robotics. It also advises the Executive Office of the President on interagency AI R&D 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 technical, national AI workforce.

About the Subcommittee on Machine Learning and Artificial Intelligence

The Machine Learning and Artificial Intelligence (MLAI) Subcommittee monitors the state of the art in machine learning (ML) and artificial intelligence 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 Subcommittee reports to the Committee on Technology and the Select Committee on AI. The MLAI Subcommittee also coordinates AI taskings with the Artificial Intelligence Research & Development Interagency Working Group (see below).

About the Subcommittee on Networking & Information Technology Research & Development

The Networking and Information Technology Research and Development (NITRD) Program is the Nation's primary source of Federally funded work on pioneering information technologies (IT) in computing, networking, and software. The NITRD Subcommittee guides the multiagency NITRD Program in its work to provide the R&D foundations for assuring continued U.S. technological leadership and meeting the needs of the Nation for advanced IT. It reports to the NSTC Committee on Science and Technology Enterprise. The Subcommittee is supported by the interagency working groups that report to it and by its National Coordination Office. More information is available at https://www.nitrd.gov/about/.

About the Artificial Intelligence Research & Development Interagency Working Group

The NITRD AI R&D Interagency Working Group (IWG) coordinates Federal R&D in AI; it also supports and coordinates activities tasked by the Select Committee on AI and the NSTC Subcommittee on Machine Learning and Artificial Intelligence. This vital work promotes U.S. leadership and global competitiveness in AI R&D. The NITRD AI R&D IWG spearheaded the update of this National Artificial Intelligence Research and Development Strategic Plan. More information is available at https://www.nitrd.gov/groups/AI.

About this Document

This document includes the original text from the 2016 *National AI R&D Strategic Plan* with updates prepared in 2019 following Administration and interagency evaluation of the 2016 Plan and of community responses to a Request for Information on updating the Plan. The 2016 strategies were broadly determined to be valid going forward with some reemphases and with a call for a new strategy on Private-Public Partnerships in AI. A shaded call-out box has been inserted at the top of each strategy to highlight updated imperatives and/or new focus areas. The 2019 update adds an entirely new Strategy 8 on Private-Public Partnerships in AI.

Published in the United States of America, 2019.

THE WHITE HOUSE