THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN

National Science and Technology Council

Networking and Information Technology Research and Development Subcommittee

October 2016

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 (R&D) enterprise. One of the NSTC's primary objectives is establishing clear national goals for Federal science and technology investments. The NSTC prepares R&D packages aimed at accomplishing multiple national goals. The NSTC's work is organized under five committees: Environment, Natural Resources, and Sustainability; Homeland and National Security; Science, Technology, Engineering, and Mathematics (STEM) Education; Science; and Technology. Each of these committees oversees subcommittees and working groups that are focused on different aspects of science and technology. More information is available at 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. The mission of OSTP is threefold; first, to provide the President and his senior staff with accurate, relevant, and timely scientific and technical advice on all matters of consequence; second, to ensure that the policies of the Executive Branch are informed by sound science; and third, to ensure that the scientific and technical work of the Executive Branch is properly coordinated so as to provide the greatest benefit to society. The Director of OSTP also serves as Assistant to the President for Science and Technology and manages the NSTC. More information is available at www.whitehouse.gov/ostp.

About the Subcommittee on Networking and Information Technology Research and Development

The Subcommittee on Networking and Information Technology Research and Development (NITRD) is a body under the Committee on Technology (CoT) of the National Science and Technology Council (NSTC). The NITRD Subcommittee coordinates multiagency research and development programs to help assure continued U.S. leadership in networking and information technology, satisfy the needs of the Federal Government for advanced networking and information technology, and accelerate development and deployment of advanced networking and information technology. It also implements relevant provisions of the High-Performance Computing Act of 1991 (P.L. 102-194), as amended by the Next Generation Internet Research Act of 1998 (P. L. 105-305), and the America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education and Science (COMPETES) Act of 2007 (P.L. 110-69). For more information, see www.nitrd.gov.

Acknowledgments

This document was developed through the contributions of the members and staff of the NITRD Task Force on Artificial Intelligence. A special thanks and appreciation to additional contributors who helped write, edit, and review the document: Chaitan Baru (NSF), Eric Daimler (Presidential Innovation Fellow), Ronald Ferguson (DoD), Nancy Forbes (NITRD), Eric Harder (DHS), Erin Kenneally (DHS), Dai Kim (DoD), Tatiana Korelsky (NSF), David Kuehn (DOT), Terence Langendoen (NSF), Peter Lyster (NITRD), KC Morris (NIST), Hector Munoz-Avila (NSF), Thomas Rindflesch (NIH), Craig Schlenoff (NIST), Donald Sofge (NRL), and Sylvia Spengler (NSF).

Copyright Information

October 13, 2016

Dear Colleagues:

We are pleased to transmit with this letter the *National Artificial Intelligence Research and Development Strategic Plan* of the NSTC. This Plan was developed by the Artificial Intelligence Task Force, an interagency working group tasked by the NITRD Subcommittee of the NSTC, at the request of the NSTC Subcommittee on Machine Learning and Artificial Intelligence.

Intelligent computer systems have long been a subject of science fiction. Now, we are entering an era in which AI is having broad and deep impacts on our daily lives, ranging from precision medicine to transportation to education and more. In response, on May 3, 2016, the White House announced a series of actions to spur public dialogue on AI, to identify challenges and opportunities related to this emerging technology, to aid in the use of AI for more effective government, and to prepare for the potential benefits and risks of AI. As part of these actions, the White House directed the creation of a national strategy for research and development in artificial intelligence.

This resulting AI R&D Strategic Plan defines a high-level framework that can be used to identify scientific and technological needs in AI, and to track the progress and maximize the impact of R&D investments to fill those needs. It also establishes priorities for Federally-funded R&D in AI, looking beyond near-term AI capabilities toward long-term transformational impacts of AI on society and the world.

This coordinated AI R&D effort across the Federal government will help the United States capitalize on the full potential of AI technologies to strengthen our economy and better our society. The AI R&D Strategic Plan does not, however, define specific research agendas for individual Federal agencies. Instead, agencies will continue to pursue priorities consistent with their missions, capabilities, authorities, and budgets, while coordinating so that the overall research portfolio is consistent with the AI R&D Strategic Plan.

We look forward to continuing this important work with Federal agencies and other key partners, and using this Plan to guide future decisions in AI R&D.

Sincerely,

Bryan Biegel
Director, National Coordination Office for
Networking and Information Technology
Research and Development

James F. Kurose
Assistant Director, Computer and Information
Science and Engineering
National Science Foundation

Co-Chairs, Subcommittee on Networking and Information Technology Research and Development

National Science and Technology Council

Chair
John P. Holdren
Assistant to the President for Science and Technology and Director, Office of Science and Technology Policy

Staff
Afua Bruce
Executive Director
Office of Science and Technology Policy

Subcommittee on Machine Learning and Artificial Intelligence

Co-Chair
Ed Felten
Deputy U.S. Chief Technology Officer
Office of Science and Technology Policy

Co-Chair
Michael Garris
Senior Scientist
National Institute of Standards and Technology
U.S. Department of Commerce

Subcommittee on Networking and Information Technology Research and Development

Co-Chair
Bryan Biegel
Director, National Coordination Office for Networking and Information Technology Research and Development

Co-Chair
James Kurose
Assistant Director, Computer and Information Science and Engineering
National Science Foundation

Networking and Information Technology Research and Development Task Force on Artificial Intelligence

Co-Chair
Lynne Parker
Division Director
Information and Intelligent Systems
National Science Foundation

Co-Chair
Jason Matheny
Director
Intelligence Advanced Research Projects Activity

Members
Milton Corn
National Institutes of Health

William Ford
National Institute of Justice

Michael Garris
National Institute of Standards and Technology

Steven Knox
National Security Agency

Nikunj Oza
National Aeronautics and Space Administration

Robinson Pino
Department of Energy

Gregory Shannon
Office of Science and Technology Policy

Scott Tousley
Department of Homeland Security

John Launchbury
Defense Advanced Research Projects Agency

Richard Linderman
Office of the Secretary of Defense

Faisal D'Souza
Technical Coordinator
National Coordination Office for Networking and Information Technology Research and Development

Contents

Executive Summary

Artificial intelligence (AI) is a transformative technology that holds promise for tremendous societal and economic benefit. AI has the potential to revolutionize how we live, work, learn, discover, and communicate. AI research can further our national priorities, including increased economic prosperity, improved educational opportunities and quality of life, and enhanced national and homeland security. Because of these potential benefits, the U.S. government has invested in AI research for many years. Yet, as with any significant technology in which the Federal government has interest, there are not only tremendous opportunities but also a number of considerations that must be taken into account in guiding the overall direction of Federally-funded R&D in AI.

On May 3, 2016, the Administration announced the formation of a new NSTC Subcommittee on Machine Learning and Artificial intelligence, to help coordinate Federal activity in AI.[1] This Subcommittee, on June 15, 2016, directed the Subcommittee on Networking and Information Technology Research and Development (NITRD) to create a *National Artificial Intelligence Research and Development Strategic Plan*. A NITRD Task Force on Artificial Intelligence was then formed to define the Federal strategic priorities for AI R&D, with particular attention on areas that industry is unlikely to address.

This *National Artificial Intelligence R&D Strategic Plan* establishes a set of objectives for Federally-funded AI research, both research occurring within the government as well as Federally-funded research occurring outside of government, such as in academia. The ultimate goal of this research is to produce new AI knowledge and technologies that provide a range of positive benefits to society, while minimizing the negative impacts. To achieve this goal, this AI R&D Strategic Plan identifies the following priorities for Federally-funded AI research:

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. Rather than replace humans, most AI systems will collaborate with humans to achieve optimal performance. Research is needed to create effective interactions between humans and AI systems.

Strategy 3: Understand and address the ethical, legal, and societal implications of AI. We expect AI technologies to behave according to the formal and informal norms to which we hold our fellow humans. Research is needed to understand the ethical, legal, and social implications of AI, and to develop methods for designing AI systems that align with ethical, legal, and societal goals.

Strategy 4: Ensure the safety and security of AI systems. Before AI systems are in widespread use, assurance is needed that the systems will operate safely and securely, in a controlled, well-defined, and well-understood manner. Further progress in research is needed to address this challenge of creating AI systems that are reliable, dependable, and trustworthy.

Strategy 5: Develop shared public datasets and environments for AI training and testing. The depth, quality, and accuracy of training datasets and resources significantly affect AI performance. Researchers need to develop high quality datasets and environments and enable responsible access to high-quality datasets as well as to testing and training resources.

Strategy 6: Measure and evaluate AI technologies through standards and benchmarks. Essential to advancements in AI are standards, benchmarks, testbeds, and community engagement that guide and

[1] E. Felten, "*Preparing for the Future of Artificial Intelligence,"* White House Office of Science and Technology Policy blog, May 5, 2016, https://www.whitehouse.gov/blog/2016/05/03/preparing-future-artificial-intelligence.

evaluate progress in AI. Additional research is needed to develop a broad spectrum of evaluative techniques.

Strategy 7: Better understand the national AI R&D workforce needs. Advances in AI will require a strong community of AI researchers. An improved understanding of current and future R&D workforce demands in AI is needed to help ensure that sufficient AI experts are available to address the strategic R&D areas outlined in this plan.

The AI R&D Strategic Plan closes with two recommendations:

> ***Recommendation 1: Develop an AI R&D implementation framework to identify S&T opportunities and support effective coordination of AI R&D investments, consistent with Strategies 1-6 of this plan.***
>
> ***Recommendation 2: Study the national landscape for creating and sustaining a healthy AI R&D workforce, consistent with Strategy 7 of this plan.***

Introduction

Purpose of the National AI R&D Strategic Plan

In 1956, researchers in computer science from across the United States met at Dartmouth College in New Hampshire to discuss seminal ideas on an emerging branch of computing called artificial intelligence or AI. They imagined a world in which "machines use language, form abstractions and concepts, solve the kinds of problems now reserved for humans, and improve themselves".[2] This historic meeting set the stage for decades of government and industry research in AI, including advances in perception, automated reasoning/planning, cognitive systems, machine learning, natural language processing, robotics, and related fields. Today, these research advances have resulted in new sectors of the economy that are impacting our everyday lives, from mapping technologies to voice-assisted smart phones, to handwriting recognition for mail delivery, to financial trading, to smart logistics, to spam filtering, to language translation, and more. AI advances are also providing great benefits to our social wellbeing in areas such as precision medicine, environmental sustainability, education, and public welfare.[3]

The increased prominence of AI approaches over the past 25 years has been boosted in large part by the adoption of statistical and probabilistic methods, the availability of large amounts of data, and increased computer processing power. Over the past decade, the AI subfield of machine learning, which enables computers to learn from experience or examples, has demonstrated increasingly accurate results, causing much excitement about the near-term prospects of AI. While recent attention has been paid to the importance of statistical approaches such as deep learning,[4] impactful AI advances have also been made in a wide variety of other areas, such as perception, natural language processing, formal logics, knowledge representations, robotics, control theory, cognitive system architectures, search and optimization techniques, and many others.

The recent accomplishments of AI have generated important questions on the ultimate direction and implications of these technologies: What are the important scientific and technological gaps in current AI technologies? What new AI advances would provide positive, needed economic and societal impacts? How can AI technologies continue to be used safely and beneficially? How can AI systems be designed to align with ethical, legal, and societal principles? What are the implications of these advancements for the AI R&D workforce?

The landscape for AI R&D is becoming increasingly complex. While past and present investments by the U.S. Government have led to groundbreaking approaches to AI, other sectors have also become significant contributors to AI, including a wide range of industries and non-profit organizations. This investment landscape raises major questions about the appropriate role of Federal investments in the development of AI technologies. What are the right priorities for Federal investments in AI, especially regarding areas and timeframes where industry is unlikely to invest? Are there opportunities for industrial and international R&D collaborations that advance U.S. priorities?

---

[2] J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon, "*A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," August 31, 1955,* http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html.

[3] See presentations from subject matter experts at *Artificial Intelligence for Social Good* workshop, June 7, 2016, http://cra.org/ccc/events/ai-social-good/.

[4] Deep learning refers to a general family of methods that use multi-layered neural networks; these methods have supported rapid progress on tasks once believed to be incapable of automation.

In 2015, the U.S. Government's investment in unclassified R&D in AI-related technologies was approximately $1.1 billion. Although these investments have led to important new science and technologies, there is opportunity for further coordination across the Federal government so that these investments can achieve their full potential.[5]

Recognizing the transformative effects of AI, in May 2016, the White House Office of Science and Technology Policy (OSTP) announced a new interagency working group to explore the benefits and risks of AI.[6] OSTP also announced a series of four workshops, held in the May-July 2016 time frame, aimed at spurring public dialogue on AI, and identifying the challenges and opportunities it entails. The outcomes of the workshops are part of a companion public report, *Preparing for the Future of Artificial Intelligence,* released in conjunction with this plan.

In June 2016, the new NSTC Subcommittee on Machine Learning and Artificial Intelligence—which is chartered to stay abreast of advances in AI within the Federal government, the private sector, and internationally, and to help coordinate Federal activities in AI—tasked the NITRD National Coordination Office (NCO) to create the *National Artificial Intelligence Research and Development Strategic Plan*. The Subcommittee directed that this plan should convey a clear set of R&D priorities that address strategic research goals, focus Federal investments on those areas in which industry is unlikely to invest, and address the need to expand and sustain the pipeline of AI R&D talent.

Input to this AI R&D Strategic Plan has come from a wide range of sources, including Federal agencies, public discussions at AI-related meetings, an OMB data call across all Federal agencies who invest in IT-related R&D, the OSTP Request for Information (RFI) that solicited public input about how America can best prepare for an AI future,[7] and information from open publications on AI.

This plan makes several assumptions about the future of AI.[8] First, it assumes that AI technologies will continue to grow in sophistication and ubiquity, thanks to AI R&D investments by government and industry. Second, this plan assumes that the impact of AI on society will continue to increase, including on employment, education, public safety, and national security, as well as the impact on U.S. economic growth. Third, it assumes that industry investment in AI will continue to grow, as recent commercial successes have increased the perceived returns on investment in R&D. At the same time, this plan assumes that some important areas of research are unlikely to receive sufficient investment by industry, as they are subject to the typical underinvestment problem surrounding public goods. Lastly, this plan assumes that the demand for AI expertise will continue to grow within industry, academia, and government, leading to public and private workforce pressures.

Other R&D strategic plans and initiatives of relevance to this AI R&D Strategic Plan include the *Federal Big Data Research and Development Strategic Plan,*[9] the *Federal Cybersecurity Research and Development Strategic Plan,*[10] the *National Privacy Research Strategy,*[11] the *National Nanotechnology*

[5] While NITRD has several working groups that touch on aspects of AI, there is no current NITRD working group focused specifically on coordinating inter-agency AI R&D investments and activities.
[6] E. Felten, "*Preparing for the Future of Artificial Intelligence,"* White House Office of Science and Technology Policy blog, May 5, 2016, https://www.whitehouse.gov/blog/2016/05/03/preparing-future-artificial-intelligence.
[7] WH/OSTP RFI blog post: https://www.whitehouse.gov/blog/2016/06/27/how-prepare-future-artificial-intelligence.
[8] 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.
[9] *Federal Big Data Research and Development Strategic Plan*, May 2016, https://www.nitrd.gov/PUBS/bigdatardstrategicplan.pdf.
[10] *Federal Cybersecurity Research and Development Strategic Plan*, February 2016, https://www.nitrd.gov/cybersecurity/publications/2016_Federal_Cybersecurity_Research_and_Development_Strategic_Plan.pdf.

Initiative Strategic Plan,[12] the National Strategic Computing Initiative*,*[13] the Brain Research through Advancing Innovative Neurotechnologies Initiative*,*[14] and the National Robotics Initiative.[15] Additional strategic R&D plans and strategic frameworks are in the developmental stages, addressing certain sub-fields of AI, including video and image analytics, health information technology, and robotics and intelligent systems. These additional plans and frameworks will provide synergistic recommendations that complement and expand upon this AI R&D Strategic Plan.

Desired Outcome

This AI R&D Strategic Plan looks beyond near-term AI capabilities toward longer-term transformational impacts of AI on society and the world. Recent advances in AI have led to significant optimism about the potential for AI, resulting in strong industry growth and commercialization of AI approaches. However, while the Federal government can leverage industrial investments in AI, many application areas and long-term research challenges will not have clear near-term profit drivers, and thus may not be significantly addressed by industry. The Federal government is the primary source of funding for long-term, high-risk research initiatives, as well as near-term developmental work to achieve department- or agency-specific requirements or to address important societal issues that private industry does not pursue. The Federal government should therefore emphasize AI investments in areas of strong societal importance that are not aimed at consumer markets—areas such as AI for public health, urban systems and smart communities, social welfare, criminal justice, environmental sustainability, and national security, as well as long-term research that accelerates the production of AI knowledge and technologies.

A coordinated R&D effort in AI across the Federal government will increase the positive impact of these technologies, and provide policymakers with the knowledge needed to address complex policy challenges related to the use of AI. A coordinated approach, moreover, will help the United States capitalize on the full potential of AI technologies for the betterment of society.

This AI R&D Strategic Plan defines a high-level framework that can be used to identify scientific and technological gaps in AI and track the Federal R&D investments that are designed to fill those gaps. The AI R&D Strategic Plan identifies strategic priorities for both near-term and long-term support of AI that address important technical and societal challenges. The AI R&D Strategic Plan, however, does not define specific research agendas for individual Federal agencies. Instead, it sets objectives for the Executive Branch, within which agencies may pursue priorities consistent with their missions, capabilities, authorities, and budgets, so that the overall research portfolio is consistent with the AI R&D Strategic Plan.

The AI R&D Strategic Plan also does not set policy on the research or use of AI technologies nor does it explore the broader concerns about the potential influence of AI on jobs and the economy. While these topics are critically important to the Nation, they are discussed in the Council of Economic Advisors report entitled "Is This Time Different? The Opportunities and Challenges of Artificial Intelligence."[8] The

---

[11] *National Privacy Research Strategy*, June 2016, https://www.nitrd.gov/PUBS/NationalPrivacyResearchStrategy.pdf.
[12] *National Nanotechnology Initiative Strategic Plan*, February 2014, http://www.nano.gov/sites/default/files/pub_resource/2014_nni_strategic_plan.pdf.
[13] *National Strategic Computing Initiative Strategic Plan*, July 2016, https://www.whitehouse.gov/sites/whitehouse.gov/files/images/NSCI%20Strategic%20Plan.pdf
[14] Brain Research through Advancing Innovative Neurotechnologies (BRAIN), April 2013, https://www.whitehouse.gov/BRAIN.
[15] National Robotics Initiative*,* June 2011, https://www.whitehouse.gov/blog/2011/06/24/developing-next-generation-robots.

AI R&D Strategic Plan focuses on the R&D investments needed to help define and advance policies that ensure the responsible, safe, and beneficial use of AI.

A Vision for Advancing our National Priorities with AI

Driving this AI R&D Strategic Plan is a hopeful vision of a future world in which AI is safely used for significant benefit to all members of society. Further progress in AI could enhance wellbeing in nearly all sectors of society,[16] potentially leading to advancements in national priorities, including increased economic prosperity, improved quality of life, and strengthened national security. Examples of such potential benefits include:

Increased economic prosperity: New products and services can create new markets, and improve the quality and efficiency of existing goods and services across multiple industries. More efficient logistics and supply chains are being created through expert decision systems.[17] Products can be transported more effectively through vision-based driver-assist and automated/robotic systems.[18] Manufacturing can be improved through new methods for controlling fabrication processes and scheduling work flows.[19]

How is this increased economic prosperity achieved?

- **Manufacturing:** Technological advances can lead to a new industrial revolution in manufacturing, including the entire engineering product life cycle. Increased used of robotics could enable manufacturing to move back onshore.[20] AI can accelerate production capabilities through more reliable demand forecasting, increased flexibility in operations and the supply chain, and better prediction of the impacts of change to manufacturing operations. AI can create smarter, faster, cheaper, and more environmentally-friendly production processes that can increase worker productivity, improve product quality, lower costs, and improve worker health and safety.[21] Machine learning algorithms can improve the scheduling of manufacturing processes and reduce inventory requirements.[22] Consumers can benefit from access to what is now commercial-grade 3-D printing.[23]
- **Logistics:** Private-sector manufacturers and shippers can use AI to improve supply-chain management through adaptive scheduling and routing.[24] Supply chains can become more robust

[16] See the "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.
[17] E. W. T. Ngai, S. Peng, P. Alexander, and K. K. L. Moon, "Decision support and intelligent systems in the textile and apparel supply chain: An academic review of research articles," *Expert Systems with Applications,* 41(2014): 81-91.
[18] J. Fishelson, D. Freckleton, and K. Heaslip,"Evaluation of automated electric transportation deployment strategies: integrated against isolated," *IET Intelligent Transport Systems*, 7 (2013): 337-344.
[19] C. H. Dagli, ed., *Artificial neural networks for intelligent manufacturing*, Springer Science & Business Media, 2012.
[20] D. W. Brin, "Robotics on the Rise", *MHI Solutions*, Q3, 2013, https://dinahwbrin.files.wordpress.com/2013/07/mhi-solutions-robotics.pdf.
[21] "Robotics Challenge Aims to Enhance Worker Safety, Improve EM Cleanup", DOE Office of Environmental Management, August 31, 2016, http://energy.gov/em/articles/robotics-challenge-aims-enhance-worker-safety-improve-em-cleanup-other-em-events-set.
[22] M. J. Shaw, S. Park, and N. Raman, "Intelligent scheduling with machine learning capabilities: the induction of scheduling knowledge," *IIE transactions*, 24.2 (1992): 156-168.
[23] H. Lipson and M. Kurman, *Fabricated: The new world of 3D printing*, John Wiley & Sons, 2013.
[24] M. S. Fox, M. Barbuceanu, and R. Teigen, "Agent-oriented supply-chain management," *International Journal of Flexible Manufacturing Systems,* 12 (2000): 165-188.

to disruption by automatically adjusting to anticipated effects of weather, traffic, and unforeseen events.[25]

- **Finance:** Industry and government can use AI to provide early detection of unusual financial risk at multiple scales.[26] Safety controls can ensure that the automation in financial systems reduces opportunities for malicious behavior, such as market manipulation, fraud, and anomalous trading.[27] They can additionally increase efficiency and reduce volatility and trading costs, all while preventing systemic failures such as pricing bubbles and undervaluing of credit risk.[28]
- **Transportation:** AI can augment all modes of transportation to materially impact safety for all types of travel.[29] It can be used in structural health monitoring and infrastructure asset management, providing increased trust from the public and reducing the costs of repairs and reconstruction.[30] AI can be used in passenger and freight vehicles to improve safety by increasing situational awareness, and to provide drivers and other travelers with real-time route information.[31] AI applications can also improve network-level mobility and reduce overall system energy use and transportation-related emissions.[32]
- **Agriculture:** AI systems can create approaches to sustainable agriculture that are smarter about the production, processing, storage, distribution, and consumption of agricultural products. AI and robotics can gather site-specific and timely data about crops, apply needed inputs (e.g., water, chemicals, fertilizers) only when and where they are needed, and fill urgent gaps in the agricultural labor force.[33]
- **Marketing:** AI approaches can enable commercial entities to better match supply with demand, driving up revenue that funds ongoing private sector development.[34] It can anticipate and identify consumer needs, enabling them to better find the products and services they want, at lower cost.[35]

---

[25] S. K. Kumar, M. K. Tiwari, and R. F. Babiceanu, "Minimisation of supply chain cost with embedded risk using computational intelligence approaches," *International Journal of Production Research,* 48 (2010): 3717-3739.
[26] A. S. Koyuncugil and N. Ozgulbas, "Financial early warning system model and data mining application for risk detection," *Expert Systems with Applications,* 39 (2012): 6238-6253.
[27] K. Golmohammadi and O. R. Zaiane, "Time series contextual anomaly detection for detecting market manipulation in stock market," *IEEE International Conference on Data Science and Advanced Analytics (DSAA),* 2015.
[28]T. Mizuta, K. Izumi and S. Yoshimura, "Price variation limits and financial market bubbles: Artificial market simulations with agents' learning process," *IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, 2013.
[29] J. H. Gillulay and C. J. Tomlin, "Guaranteed safe online learning of a bounded system," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.
[30] J. M. W. Brownjohn, "Structural health monitoring of civil infrastructure," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* 365 (2007): 589-622.
[31] Dia, Hussein, "An agent-based approach to modelling driver route choice behaviour under the influence of real time information," *Transportation Research Part C: Emerging Technologies,* 10 (2002): 331-349.
[32] H. Kargupta, J. Gama, and W. Fan, "The next generation of transportation systems, greenhouse emissions, and data mining," *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010.
[33] H. Hagras, M. Colley, V. Callaghan, and M. Carr-West, "Online learning and adaptation of autonomous mobile robots for sustainable agriculture," *Autonomous Robots*, 13 (2002): 37-52.
[34] T. Di Noia, E. Di Sciascio, F. M. Donini, and M. Mongiello, "A system for principled matchmaking in an electronic marketplace," *International Journal of Electronic Commerce,* 8 (2004): 9-37.
[35] R. H. Guttman, A. G. Moukas, and P. Maes, "Agent-mediated electronic commerce: a survey," *The Knowledge Engineering Review,* 13 (1998): 147-159.

- **Communications:** AI technologies can maximize efficient use of bandwidth and automation of information storage and retrieval.[36] AI can improve filtering, searching, language translation, and summarization of digital communications, positively affecting commerce and the way we live our lives.[37]
- **Science and Technology:** AI systems can assist scientists and engineers in reading publications and patents, refining theories to be more consistent with prior observations, generating testable hypotheses, performing experiments using robotic systems and simulations, and engineering new devices and software.[38]

Improved educational opportunity and quality of life: Lifelong learning can be possible through virtual tutors that develop customized learning plans to challenge and engage each person based on their interests, abilities, and educational needs. People can live healthier and more active lives, using personalized health information tailored and adapted for each individual. Smart homes and personal virtual assistants can save people time and reduce time lost in daily repetitive tasks.

How will AI improve educational opportunities and social wellbeing?

- **Education:** AI-enhanced learning schools can be universally available, with automated tutoring that gauges the development of the student.[16] AI tutors can complement in-person teachers and focus education on advanced and/or remedial learning appropriate to the student.[16] AI tools can foster life-long learning and the acquisition of new skills for all members of society.[16]
- **Medicine:** AI can support bioinformatics systems that identify genetic risks from large-scale genomic studies (e.g., genome-wide association studies, sequencing studies), and predict the safety and efficacy of new pharmaceuticals.[39] AI techniques can allow assessments across multi-dimensional data to study public health issues and to provide decision support systems for medical diagnoses and prescribe treatments.[40] AI technologies are required for the customization of drugs for the individual; the result can be increased medical efficacy, patient comfort, and less waste.[41]
- **Law:** The analysis of law case history by machines can become widespread.[42] The increased sophistication of these processes can allow for a richer level of analysis for assisting the discovery process.[42] Legal discovery tools can identify and summarize relevant evidence; these systems may even formulate legal arguments with increasing sophistication.[42]

---

[36] I. Kushchu, "Web-based evolutionary and adaptive information retrieval," *IEEE Transactions on Evolutionary Computation,* 9 (2005): 117-125.

[37] J. Jin, P. Ji, Y. Liu, and S. C. J. Lim, "Translating online customer opinions into engineering characteristics in QFD: A probabilistic language analysis approach," *Engineering Applications of Artificial Intelligence,* 41 (2015): 115-127.

[38] R. D. King, J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. N. Soldatova, A. Sparkes, K. E. Whelan, and A. Clare, "The Automation of Science", *Science*, 324 (2009): 85-89.

[39] B. Aksu, A. Paradkhar, M. de Matas, O. Ozer, T. Guneri, and P. York, "A quality by design approach using artificial intelligence techniques to control the critical quality attributes of ramipril tablets manufactured by wet granulation," *Pharmaceutical development and technology,* 18 (2013): 236-245.

[40] P. Szolovits, R. S. Patil, and W. B. Schwartz, "Artificial intelligence in medical diagnosis," *Annals of internal medicine,* 108 (1988): 80-87.

[41] J. Awwalu, Jamilu, A. G. Garba, A. Ghazvini, and R. Atuah,"Artificial Intelligence in Personalized Medicine Application of AI Algorithms in Solving Personalized Medicine Problems," *International Journal of Computer Theory and Engineering,* 7 (2015): 439.

[42] T. Bench-Capon, M. Araszkiewicz, K. Ashley, K. Atkinson, F. Bex, F. Borges, D. Bourcier, P. Bourgine, J. Conrad, E. Francesconi, T. Gordon, G. Governatori, J. Leidner, D. Lewis, R. Loui, L. McCarty, H. Prakken, F. Schilder, E. Schweighofer, P. Thompson, A. Tyrrell, B. Verheij, D. Walton, and A. Wyner, "A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law," *Artificial Intelligence and Law,* 20 (2012): 215-319.

- **Personal services:** AI software can make use of knowledge from multiple sources to provide more accurate information for a multitude of uses.[43] Natural language systems can provide intuitive interfaces to technological systems in real-world, noisy environments.[44] Personalized tools can enable automated assistance with individual and group scheduling.[45] Text can be automatically summarized from multiple search outcomes, enhanced across multiple media.[46] AI can enable real-time spoken multi-lingual translation.[47]

Enhanced national and homeland security: Machine learning agents can process large amounts of intelligence data and identify relevant patterns-of-life from adversaries with rapidly changing tactics.[48] These agents can also provide protection to critical infrastructure and major economic sectors that are vulnerable to attack.[49] Digital defense systems can significantly reduce battlefield risks and casualties.[50]

How is enhanced national and homeland security achieved?

- **Security and law enforcement:** Law enforcement and security officials can help create a safer society through the use of pattern detection to detect anomalous behavior in individual actors, or to predict dangerous crowd behavior.[48] Intelligent perception systems can protect critical infrastructure, such as airports and power plants.[49]
- **Safety and prediction:** Distributed sensor systems and pattern understanding of normal conditions can detect when the probability of major infrastructure disruptions increases significantly, whether triggered by natural or man-made causes.[51] This anticipatory capability can help indicate where the problem will be, to adapt operations to forestall disruption as, or even before it happens.[51]

This vision for the positive use of AI, however, requires significant R&D advancements. Many critical and difficult technical challenges remain in all subfields of AI, both in basic science and in areas of application. AI technologies also present risks, such as the potential disruption of the labor market as humans are augmented or replaced by automated systems, and uncertainties about the safety and reliability of AI systems. Subsequent sections of this AI R&D Strategic Plan discuss high-priority, strategic areas of AI R&D investments that will support this vision, while mitigating potential disruption and risk.

---

[43] K. Wei, J. Huang and S. Fu, "A survey of e-commerce recommender systems," *International Conference on Service Systems and Service Management*, 2007.

[44] M. Fleischman and D. Roy, "Intentional context in situated natural language learning," *Proceedings of the Ninth Conference on Computational Natural Language Learning*, Association for Computational Linguistics, 2005.

[45] P. Berry, K. Conley, M. Gervasio, B. Peintner, T. Uribe, and N. Yorke-Smith, "Deploying a personalized time management agent," *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, 2006.

[46] U. Hahn and I. Mani, "The challenges of automatic summarization," *Computer*, 33 (2000): 29-36.

[47] M. Paul, H. Okuma, H. Yamamoto, E. Sumita, S. Matsuda, T. Shimizu, and S. Nakamura,"Multilingual mobile-phone translation services for world travelers," *22nd International Conference on Computational Linguistics: Demonstration Papers*, Association for Computational Linguistics, 2008.

[48] G. Gross, E. Little, B. Park, J. Llinas, and R. Nagi, "Application of multi-level fusion for pattern of life analysis," *18th International Conference on Information Fusion*, 2015.

[49] S. L. P. Yasakethu, J. Jiang, and A. Graziano, "Intelligent risk detection and analysis tools for critical infrastructure protection," *IEEE International Conference on Computer as a Tool* (EUROCON), 2013.

[50] N. G. Siegel and A. M. Madni, "The Digital Battlefield: A Behind-the-Scenes Look from a Systems Perspective," *Procedia Computer Science*, 28 (2014): 799-808.

[51] B. Genge, C. Siaterlis, and G. Karopoulos, "Data fusion-based anomaly detection in networked critical infrastructures," *43rd Annual IEEE/IFIP Conference on Dependable Systems and Networks Workshop (DSN-W),* 2013.

Current State of AI

Since its beginnings, AI research has advanced in three technology waves. The *first wave* focused on handcrafted knowledge, with a strong focus in the 1980s on rule-based expert systems in well-defined domains, in which knowledge was collected from a human expert, expressed in "if-then" rules, and then implemented in hardware. Such systems-enabled reasoning was applied successfully to narrowly defined problems, but it had no ability to learn or to deal with uncertainty. Nevertheless, they still led to important solutions, and the development of techniques that are still actively used today.

The *second wave* of AI research from the 2000s to the present is characterized by the ascent of machine learning. The availability of significantly larger amounts of digital data, relatively inexpensive massively parallel computational capabilities, and improved learning techniques have brought significant advances in AI when applied to tasks such as image and writing recognition, speech understanding, and human language translation. The fruits of these advances are everywhere: smartphones perform speech recognition, ATMs perform handwriting recognition on written checks, email applications perform spam filtering, and free online services perform machine translation. Key to some of these successes was the development of deep learning.

AI systems now regularly outperform humans on specialized tasks. Major milestones when AI first surpassed human performance include: chess (1997),[52] trivia (2011),[53] Atari games (2013),[54] image recognition (2015),[55] speech recognition (2015),[56] and Go (2016).[57] The pace of such milestones appears to be increasing, as is the degree to which the best-performing systems are based on machine learning methods, rather than sets of hand-coded rules.

Such achievements in AI have been fueled by a strong base of fundamental research. This research is expanding and is likely to spur future advances. As one indicator, from 2013 to 2015 the number of Web of Science-indexed journal articles mentioning "deep learning" increased six-fold (Figure 1). The trends also reveal the increasingly global nature of research, with the United States no longer leading the world in publication numbers, or even publications receiving at least one citation (Figure 2).

The U.S. Government has played a key role in AI research, although the commercial sector is also active in AI-related R&D.[58] There has been a sharp increase in the number of patents that use the term "deep learning" or "deep neural net" (Figure 3). From 2013 to 2014, there was a four-fold increase in venture capital directed to AI startups.[59] AI applications are now generating substantial revenues for large businesses.[60] The impact of AI on financial systems is even larger—automated ("algorithmic") trading is responsible for about half of all global financial trading, representing trillions of dollars in transactions.[61]

---

[52] M. Campbell, A. J. Hoane Jr., F-H. Hsu, "Deep Blue," *Artificial Intelligence*, *134 (2002): 57-83.*

[53] "IBM's "Watson" Computing System to Challenge All Time Jeopardy! Champions," news release by Sony Pictures Television, December 14, 2010.

[54] "Asynchronous Methods for Deep Reinforcement Learning," http://arxiv.org/pdf/1602.01783v2.pdf.

[55] "Deep Residual Learning for Image Recognition," http://arxiv.org/abs/1512.03385v1; for human performance, see http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/.

[56] "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin," http://arxiv.org/abs/1512.02595v1.

[57] S. Byford, "Google's AlphaGo AI beats Lee Se-dol again to win Go series 4-1," *The Verge*, March 15, 2016.

[58] "Microsoft, Google, Facebook and more are investing in artificial intelligence: What is their plan and who are the other key players?", *TechWorld*, September 29, 2106, http://www.techworld.com/picture-gallery/big-data/9-tech-giants-investing-in-artificial-intelligence-3629737/.

[59] "Artificial Intelligence Startups See 302% Funding Jump in 2014," CB Insights**,** February 10, 2015.

[60] "The Business of Google," Investopedia, 2016, http://www.investopedia.com/articles/investing/020515/business-google.asp, retrieved October 5, 2016.

[61] B. M. Weller, "Efficient Prices at Any Cost: Does Algorithmic Trading Deter Information Acquisition?", May 21, 2016, http://ssrn.com/abstract=2662254 or http://dx.doi.org/10.2139/ssrn.2662254.

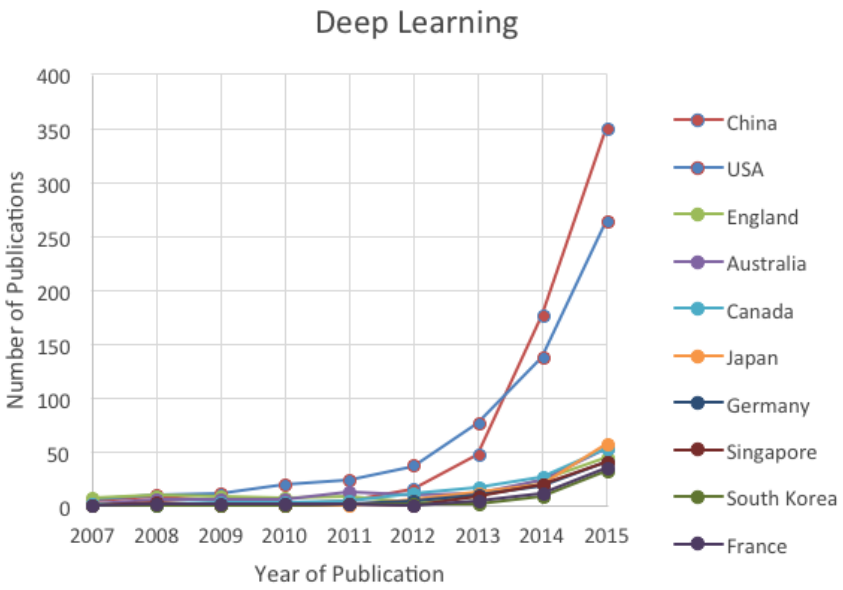


Figure 1: Journal articles mentioning "deep learning" or "deep neural network", by nation.[62]

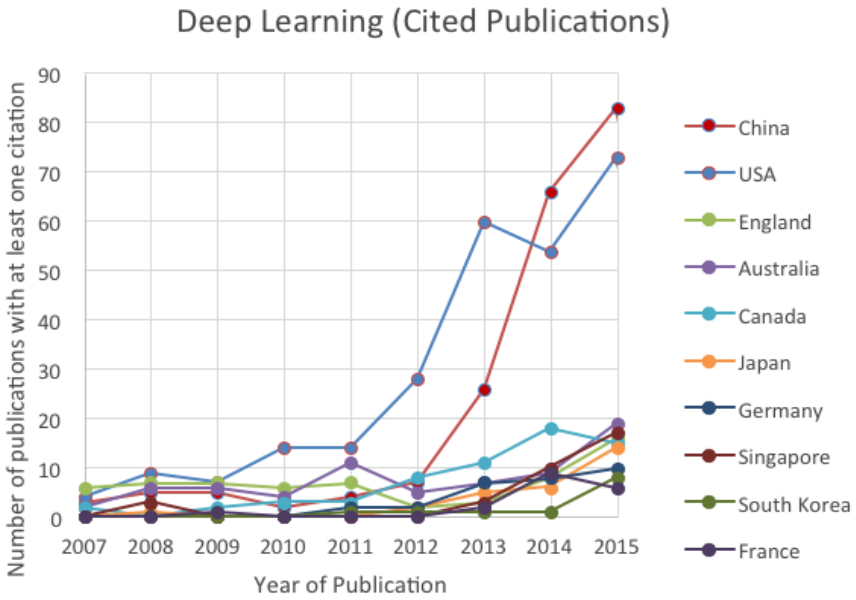


Figure 2: Journal articles cited at least once, mentioning "deep learning" or "deep neural network", by nation.[63]

[62] Data for this figure was obtained from a search of the Web of Science Core Collection for "deep learning" or "deep neural net*", for any publication, retrieved 30 August 2016.

[63] Data for this figure was obtained from a search of the Web of Science Core Collection for "deep learning" or "deep neural net*", limited to publications receiving one or more citations, retrieved 30 August 2016.

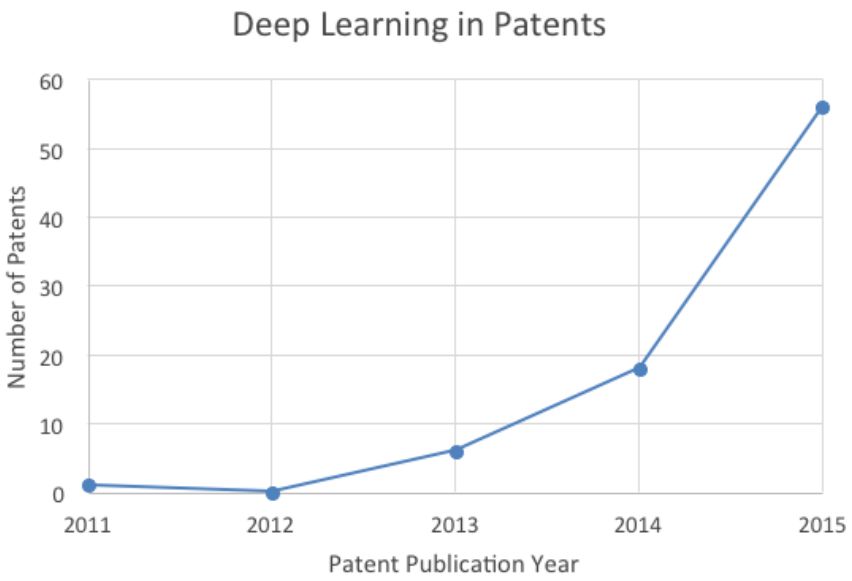


Figure 3: Analysis of number of patents using term "deep learning" or "deep neural net".[64]

Despite progress, AI systems still have their limitations. Virtually all progress has been in "narrow AI" that performs well on specialized tasks; little progress has been made in "general AI" that functions well across a variety of cognitive domains. Even within narrow AI, progress has been uneven. AI systems for image recognition rely on significant human effort to label the answers to thousands of examples.[65] In contrast, most humans are capable of "one-shot" learning from only a few examples. While most machine vision systems are easily confused by complex scenes with overlapping objects, children can easily perform "scene parsing." Scene understanding that is easy for a human is still often difficult for a machine.

The AI field is now in the beginning stages of a possible *third wave*, which focuses on explanatory and general AI technologies. The goals of these approaches are to enhance learned models with an explanation and correction interface, to clarify the basis for and reliability of outputs, to operate with a high degree of transparency, and to move beyond narrow AI to capabilities that can generalize across broader task domains. If successful, engineers could create systems that construct explanatory models for classes of real world phenomena, engage in natural communication with people, learn and reason as they encounter new tasks and situations, and solve novel problems by generalizing from past experience. Explanatory models for these AI systems might be constructed automatically through advanced methods. These models could enable rapid learning in AI systems. They may supply "meaning" or "understanding" to the AI system, which could then enable the AI systems to achieve more general capabilities.

[64] Data for this figure was obtained from a search of the Derwent World Patents Index for "deep learning" or "deep neural net*", retrieved 30 August 2016.
[65] In technical parlance, this refers to *supervised learning.*

R&D Strategy

The research priorities outlined in this AI R&D Strategic Plan focus on areas that industry is unlikely to address 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 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,[16] thus leading to important challenges in how to best create AI systems that work with people in intuitive and helpful ways.[16] 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.[8] However, as with most transformative technologies, AI presents some 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 in order 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 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.[66] Opportunities abound for core AI scientists and engineers with a deep understanding

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

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 4 provides a graphical illustration of the overall organization of this AI R&D Strategic Plan. Across the bottom (in red) are the crosscutting, underlying foundations that affect the development of all AI systems; these foundations are described in Strategies 3-7. The next layer higher (in lighter and medium dark blue) includes many areas of research that are needed to advance AI. These basic research areas (including use-inspired basic research) are outlined in Strategies 1-2.[67] Across the top row of the graphic (in dark blue) are examples of applications that are expected to benefit from advances in AI, as discussed in the Vision section earlier in this document. 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.

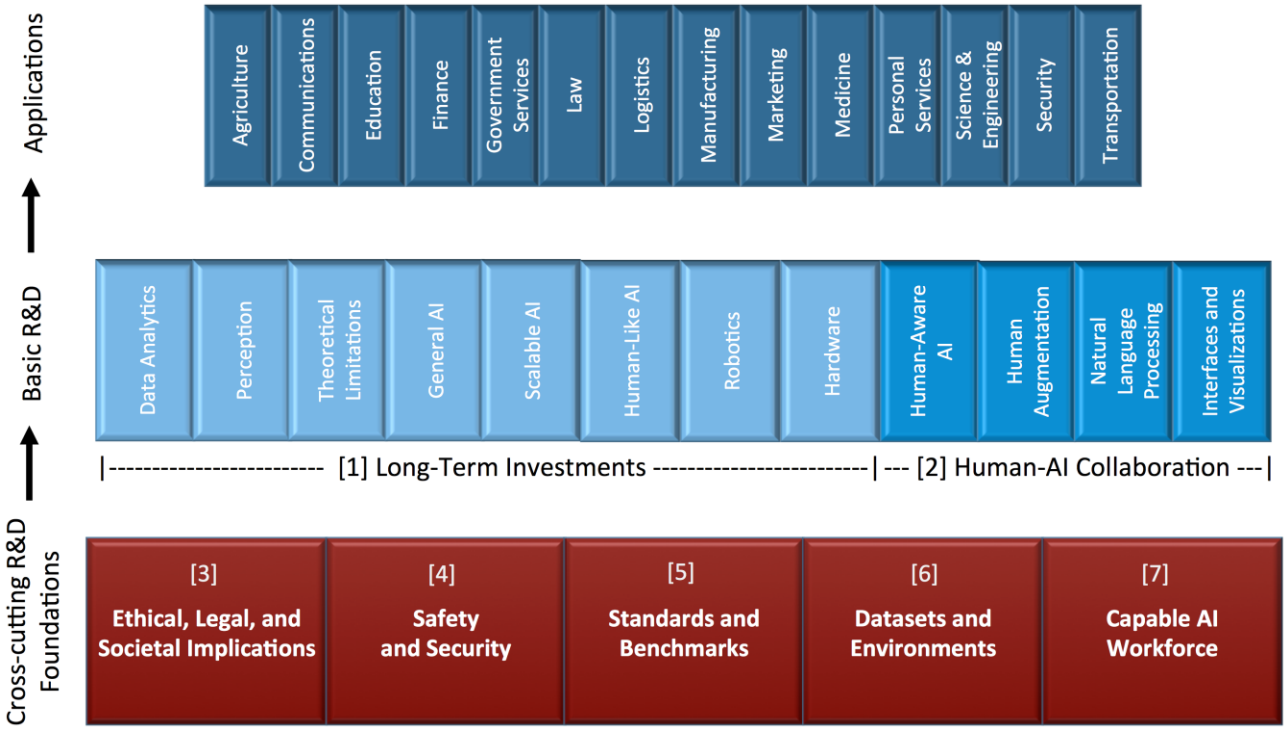


Figure 4. Organization of the AI R&D Strategic Plan. A combination of crosscutting R&D foundations (in the lower red row) are important for all AI research. Many basic AI R&D areas (in lighter and medium dark blue row) can build upon these crosscutting foundations to impact a wide array of societal applications (in top dark blue row). (The small 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.)

Strategy 1: Make Long-Term Investments in AI Research

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

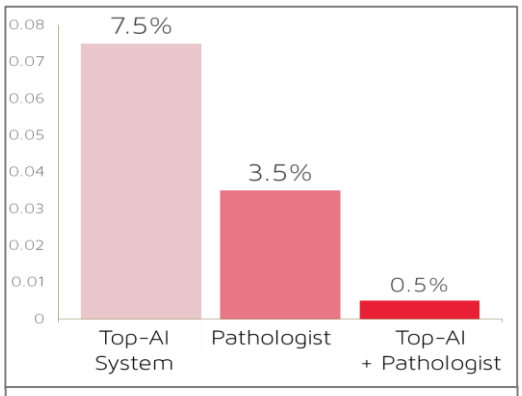
[67] Throughout this document, *basic research* includes both pure basic research and *use-inspired basic research* – the so-called *Pasteur's Quadrant* defined by Donald Stokes in his 1997 book of the same name – referring to basic research that has use for society in mind. For example, the fundamental NIH investments in IT are often called use-inspired basic research.

in 5 years, 10 years, or more. A recent 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."[68] It further notes that "the time from first concept to successful market is often measured in decades".[68] 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.

National Institutes of Health (NIH) grants-supported research

ARTIFICIAL INTELLIGENCE FOR COMPUTATIONAL PATHOLOGY

Image interpretation plays a central role in the pathologic diagnosis of cancer. Since the late 19th century, the primary tool used by pathologists to make definitive cancer diagnoses is the microscope. Pathologists diagnose cancer by manually examining stained sections of cancer tissues to determine the cancer subtype. Pathologic diagnosis using conventional methods is labor-intensive with poor reproducibility and quality concerns. New approaches use fundamental AI research to build tools to make pathologic analysis more efficient, accurate, and predictive. In the 2016 Camelyon Grand Challenge for metastatic cancer detection,[69] the top-performing entry in the competition was an AI-based computational system that achieved an error rate of 7.5%.[70] A pathologist reviewing the same set of evaluation images achieved an error rate of 3.5%. Combining the predictions of the AI system with the pathologist lowered the error rate to down to 0.5%, representing an 85% reduction in error (see image).[71] This example illustrates how fundamental research in AI can drive the development of high performing computational systems that offer great potential for making pathological diagnoses more efficient and more accurate.



AI significantly reduces pathologist error rate in the identification of metastatic breast cancer from sentinel lymph node biopsies.

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 *Federal Big Data Research and Development Strategic Plan,*[9] 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

[68] *Continuing Innovation in Information Technology* (Washington D.C.: The National Academies Press, 2012), page 11.
[69] http://camelyon16.grand-challenge.org/.
[70] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," https://arxiv.org/pdf/1606.05718v1.pdf.
[71] The full results are presented at https://grand-challenge.org/site/camelyon16/results/.

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 multi-modality 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 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.[16] 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.[72,73] 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 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.

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

[73] 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.

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.

NSF-funded Framework on Game Theory for Security

Security is a critical concern around the world, whether it is the challenge of protecting ports, airports and other critical infrastructure; protecting endangered wildlife, forests and fisheries; suppressing urban crime; or security in cyberspace. Unfortunately, limited security resources prevent full security coverage at all times; instead, we must optimize the use of limited security resources. To that end, the "security games" framework—based on basic research in computational game theory, while also incorporating elements of human behavior modeling, AI planning under uncertainty and machine learning—has led to building and deployment of decision aids for security agencies in the United States and around the world.[74] For example, the ARMOR system has been deployed at LAX airport since 2008, the IRIS system for the Federal Air Marshals Service has been in use since 2009, and the PROTECT system for the U.S. Coast Guard since 2011. Typically, given limited security resources (e.g., boats, air marshals, police), and a large number of targets of different values (e.g., different flights, different terminals at an airport), security-games-based decision aids provide a randomized allocation or patrolling schedule that takes into account the weights of different targets and intelligent reaction of the adversary to the different security postures. These applications have been shown to provide a significant improvement in performance of the different security agencies using a variety of metrics, e.g., capture rates, red teams, patrol schedule randomness, and others.[74]

Many types of applications may benefit from game-theoretic approaches to security.

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

[74] M. Tambe, *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*, (Cambridge: Cambridge University Press, 2011).

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,[75] 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,[76] in some cases, optimized for the operation of neural networks.

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.[77] This reciprocity will lead to further advances in hardware performance, since physical limits on computing require novel approaches to hardware designs.[78] 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

[75] 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.

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

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

[78] 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 will need to be turned off in order to limit temperatures and ensure data integrity.

and resource usage, and to make online optimization decisions that increase efficiency; more advanced AI techniques could further enhance system performance. AI can also be used to create self-reconfigurable HPC systems that can handle system faults when they occur, without human intervention.[79]

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.[80] 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.

Strategy 2: Develop Effective Methods for Human-AI Collaboration

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 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

[79] 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.

[80] Exascale computing refers to computing systems that can achieve at least a billion billion calculations per second.

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 the 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:[81]

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.[82, 83]

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.

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

[81] C. Wickens and J. G. Hollands, "Attention, time-sharing, and workload," in *Engineering, Psychology and Human Performance* (London: Pearson PLC, 1999), 439-479.
[82] https://www.nasa.gov/mission_pages/SOFIA/index.html.
[83] https://cloud1.arc.nasa.gov/intex-na/.

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, additional 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.[84] 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.

Developing techniques for visualization and AI-human 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

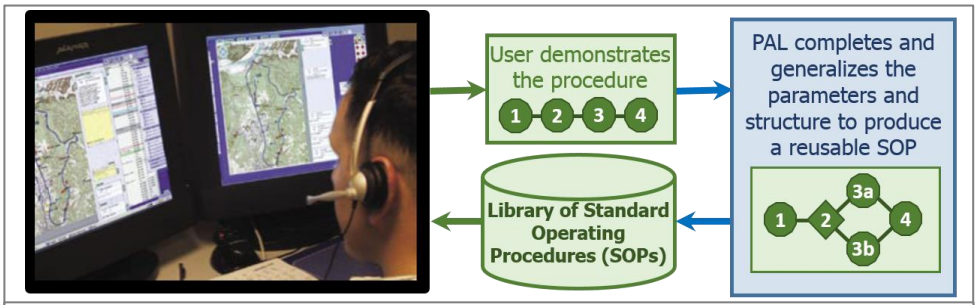
Enabling people to interact with AI systems through spoken and written language has long been a goal

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

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.

DARPA's Personalized Assistant that Learns (PAL) Program Created the Technology that Apple Commercialized as Siri

Computing technology is critical to every aspect of modern life, but the information systems we use daily lack the general, flexible abilities of human cognition. In the Personalized Assistant that Learns (PAL) program,[85] DARPA set about to create cognitive assistants that can learn from experience, reason, and be told what to do via a speech interface. DARPA envisioned PAL technologies making information systems more efficient and effective for users. DARPA and the PAL performers worked with military operators to apply PAL technologies to problems of command and control, and PAL procedure learning technology was integrated in the U.S. Army's Command Post of the Future version Battle Command 10 (see figure) and used around the world.



PAL procedure learning assistant learns reusable procedures by observing one example.

DARPA was also acutely aware of the commercial potential of the PAL technology, especially for mobile applications where speech-based smartphone interaction would be required. DARPA strongly encouraged PAL commercialization and in 2007, in response to DARPA's encouragement, Siri Inc. was created to commercialize PAL technology in a system that could assist a user by managing information and automating tasks through a speech-based interface. In April 2010, Siri Inc. was acquired by Apple, which further developed the technologies to make them an integral part–and the defining feature–of Apple's mobile operating system available on the iPhone and iPad.

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

[85] https://pal.sri.com.

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

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*.

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":

> "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."[86]

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 the disciplines mentioned above) 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.

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.

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

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 multi-disciplinary 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.[87] 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.[87] 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.[88] A third method is to formulate an ethical architecture using set theoretic principles, combined with logical constraints on AI system behavior that restrict action to conform to ethical doctrine.[89] As AI systems become more general, their architectures will likely include subsystems that can take on ethical issues at multiple levels of judgment, including:[90] 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.

Strategy 4: Ensure the Safety and Security of AI Systems

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

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

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

[89] 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.

[90] B. Kuipers, "Human-like Morality and Ethics for Robots", AAAI-16 Workshop on AI, Ethics and Society, 2016.

systems that are reliable, dependable, and trustworthy. As with other complex systems, AI systems face important safety and security challenges due to: [91]

- *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.
- *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. [92]

To address these issues and others, additional investments are needed to advance AI safety and security,[93] 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

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

[92] J. M. Bradshaw, R. R. Hoffman, M. Johnson, and D. D. Woods, "The Seven Deadly Myths of Autonomous Systems," *IEEE Intelligent Systems*, 28, no. 3 (2013): 54-61.

[93] 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, M. Tegmark, 2016, "Research Priorities for Robust and Beneficial Artificial Intelligence," arXiv: 1602.03506; T. G. Dietterich, E. J. Horvitz, 2015, "Rise of Concerns about AI: Reflections and Directions," *Communications of the ACM*, Vol. 58 No. 10; K. S.; R. Yampolsky (19 December 2014), "Responses to catastrophic AGI risk: a survey," *Physica Scripta*, 90 (1).

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.

NASA Ames Research Center - PREDICTING FAILURES BEFORE THEY HAPPEN

NASA Ames Research Center developed a data-driven anomaly detection method called the Inductive Monitoring System (IMS) in 2003 because of inadequacies in model-based methods for anomaly detection. Since then, it has been deployed for system health monitoring applications within NASA including monitoring the space shuttle and International Space Station (ISS), as well as non-NASA applications.

Test launch of Orion Crew Vehicle in 2014, during which the IMS was used to monitor electrical systems.

C-130 Hercules military transport plane, for which predictive software was used to anticipate failures in a valve used to switch air flow between engines.

In 2012, Comprehensive Engineering Management Solutions (CEMSol) licensed IMS, enhanced it, and teamed up with NASA Ames and Lockheed Martin to test it as an Integrated System Health Monitoring system on the Lockheed C-130 Hercules military transport plane. Lockheed Martin invested $70,000 in the test and recovered 10 times that amount almost immediately in reduced maintenance costs and mission delays.[94]

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

[94] "System Health Monitor Predicts Failures Before They Happen," Spinoff 2016, National Aeronautics and Space Administration (NASA), http://spinoff.nasa.gov/Spinoff2016/it_1.html.

cybersecurity R&D needs are addressed in greater detail in the NITRD *Cybersecurity R&D Strategic Plan*, 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.[95]

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.

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

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.[96] 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

[95] https://cgc.darpa.mil.

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

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?

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).[97] 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,[98] MALLET,[99] and OpenNLP,[100] 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

[97] https://www.dhs.gov/csd-impact.
[98] https://sourceforge.net/projects/weka/.
[99] http://mallet.cs.umass.edu.
[100] https://opennlp.apache.org.

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.

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

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 (IEEE). 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;
- *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 (NIST) 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). Versions of the test methods are used to challenge the research community through the RoboCup Rescue Robot League competitions,[101] which emphasize autonomous capabilities. Another example is the IEEE Agile Robotics for Industrial Automation Competition (ARIAC),[102] a joint effort between IEEE and NIST, which

[101] http://www.robocup2016.org/en/.
[102] http://robotagility.wixsite.com/competition.

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:[103] "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." Having adequate testbeds is a 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 their 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 pre-competitive 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

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

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 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.

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

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,[110] with demand expected to continue to escalate.[66] High tech companies are reportedly investing significant resources into recruiting faculty members and students with AI expertise.[111] Universities and industries are reportedly in a battle to recruit and retain AI talent.[112]

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

[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.
[109] P. J. Phillips, "Improving Face Recognition Technology," *Computer,* 44 No. 3 (2011): 84-96.
[110] "Startups Aim to Exploit a Deep-Learning Skills Gap", *MIT Technology Review*, January 6, 2016.
[111] "Artificial Intelligence Experts are in High Demand", *The Wall Street Journal*, May 1, 2015.
[112] "Million dollar babies: As Silicon Valley fights for talent, universities struggle to hold on to their stars", *The Economist*, April 2, 2016.

information technology workforce can lead to improved outcomes.[113] 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.

[113] 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.

Recommendations

The Federal Government in its entirety can support the seven strategic priorities of this Plan and achieve its vision by supporting the following recommendations:

Recommendation 1: Develop an AI R&D implementation framework to identify S&T opportunities and support effective coordination of AI R&D investments, consistent with Strategies 1-6 of this plan.

Federal agencies should collaborate through NITRD to develop an R&D implementation framework that facilitates coordination and progress on the R&D challenges outlined in this plan. This will enable agencies to easily plan, coordinate, and collaborate in support of this strategic plan. The implementation framework should take into account the R&D priorities of each agency, based on their missions, capabilities, authorities, and budget. Based on the implementation framework, funding programs may need to be established for coordinated execution of the national research agenda for AI. To help implement this Strategic Plan, NITRD should consider forming an interagency working group focused on AI, in coordination with existing working groups.

Recommendation 2: Study the national landscape for creating and sustaining a healthy AI R&D workforce, consistent with Strategy 7 of this plan.

A healthy and vibrant AI R&D workforce is important to addressing the R&D strategic challenges outlined in this report. While some reports have indicated a potential growing shortage of AI R&D experts, no official workforce data exists to characterize the current state of the AI R&D workforce, the projected workforce pipeline, and the supply and demand forces in the AI workforce. Given the role of the AI R&D workforce in addressing the strategic priorities identified in this plan, a better understanding is needed for attaining and/or maintaining a healthy AI R&D workforce. NITRD should study how best to characterize and define the current and future AI R&D workforce needs, developing additional studies or recommendations that can ensure a sufficient R&D workforce to address the AI needs of the Nation. As indicated by the outcome of the studies, appropriate Federal organizations should then take steps to ensure that a healthy national AI R&D workforce is created and maintained.

This page intentionally left blank

Acronyms

3-D	Three Dimensional
AI	Artificial Intelligence
ANNs	Artificial Neural Networks
ARIAC	Agile Robotics for Industrial Automation Competition
ARMOR	Assistant for Randomized Monitoring over Routes
ASTM	American Society of the International Association for Testing and Materials
ATM	Automated Teller Machine
BRAIN	Brain Research through Advance Innovative Neurotechnologies
CEMSol	Comprehensive Engineering Management Solutions
COMPETES	America Creating Opportunities to Meaningfully Promote Excellence in Technology Education and Science
CoT	Committee on Technology
DARPA	Defense Advanced Research Projects Agency
DHS	Department of Homeland Security
DoD	Department of Defense
DOE	Department of Energy
DOT	Department of Transportation
FERET	Face Recognition Technology
FRVT	Face Recognition Vendor Test
GPS	Global Positioning System
GPU	Graphics Processing Unit
HPC	High Performance Computing
I/O	Input/Output
IBM	International Business Machines Corporation
IEEE	Institute of Electrical and Electronics Engineers
IMPACT	Information Marketplace for Policy and Analysis of Cyber-risk & Trust
IMS	Inductive Monitoring System
IoT	Internet of Things
IRIS	Intelligent Randomization in International Scheduling
ISS	International Space Station
IT	Information Technology
KSA	Knowledge, Skills and Abilities
LAX	Los Angeles World Airports
MALLET	Machine Learning for Language Toolkit
NASA	National Aeronautics and Space Administration
NCO	National Coordination Office for NITRD
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NITRD	Networking Information Technology Research and Development
NLP	Natural Language Processing
NRL	Naval Research Laboratory
NSF	National Science Foundation

NSTC	National Science and Technology Council
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
PAL	Personalized Assistant that Learns
PROTECT	Port Resilience Operational / Tactical Enforcement to combat Terrorism
R&D	Research and Development
RFI	Request For Information
S&T	Science and Technology
STEM	Science, Technology, Engineering and Mathematics
TREC	Text Retrieval Conference
U.S.	The United States of America