Will Trump Dumb Down

The U.S. Artificial Intelligence Strategy?

Edited by

Michael Erbschloe

Connect with Michael on LinkedIn

FREE eBooks

WHOEVER
WHENEVER
WHEREVER
YOU ARE

INSTANTLY DOWNLOAD THESE MASSIVE BOOK BUNDLES

CLICK ANY BELOW TO ENJOY NOW

3 AUDIOBOOK COLLECTIONS

Classic AudioBooks Vol 1 ▪ Classic AudioBooks Vol 2 ▪ Classic AudioBooks Kids

6 BOOK COLLECTIONS

Sci-Fi ▪ Romance ▪ Mystery ▪ Academic ▪ Classics ▪ Business

Table of Contents

About the Editor

Michael Erbschloe has worked for over 30 years performing analysis of the economics of information technology, public policy relating to technology, and utilizing technology in reengineering organization processes. He has authored several books on social and management issues of information technology that were published by McGraw Hill and other major publishers. He has also taught at several universities and developed technology-related curriculum. His career has focused on several interrelated areas:

- Technology strategy, analysis, and forecasting
- Teaching and curriculum development
- Writing books and articles
- Publishing and editing
- Public policy analysis and program evaluation

Books by Michael Erbschloe

Social Media Warfare: Equal Weapons for All (Auerbach Publications)
Walling Out the Insiders: Controlling Access to Improve Organizational Security (Auerbach Publications)
Physical Security for IT (Elsevier Science)
Trojans, Worms, and Spyware (Butterworth-Heinemann)
Implementing Homeland Security in Enterprise IT (Digital Press)
Guide to Disaster Recovery (Course Technology)
Socially Responsible IT Management (Digital Press)
Information Warfare: How to Survive Cyber Attacks (McGraw Hill)
The Executive's Guide to Privacy Management (McGraw Hill)
Net Privacy: A Guide to Developing & Implementing an e-business Privacy Plan (McGraw Hill)

Introduction

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.

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". 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.

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, 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?

The question remains: will the new administration pursue this approach or dumb down the strategy because of the fear of facts, science, and opposing opinions?

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

(Link: https://www.nitrd.gov/news/national_ai_rd_strategic_plan.aspx)

National Artificial Intelligence Research and Development Strategic Plan

On May 3, 2016, the Obama Administration announced the formation of a new NSTC Subcommittee on Machine Learning and Artificial intelligence, to help coordinate Federal activity in AI. 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 National Artificial Intelligence Research and Development Strategic Plan evaluate progress in AI. Additional research is needed to develop a broad spectrum of evaluative techniques to 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.

This plan makes several assumptions about the future of AI. 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.

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.

AI can increase economic prosperity by the introduction of new products and services which can create new markets, and improve the quality and efficiency of existing goods and services across multiple industries including: Manufacturing, Logistics, Finance, Transportation, Agriculture, Marketing, Communications, and Science and Technology. AI can also improve educational opportunities and social wellbeing and enhanced national and homeland security.

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.

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), trivia (2011), Atari games (2013), image recognition (2015), speech recognition (2015), and Go (2016). 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. 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.

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.

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, thus leading to important challenges in how to best create AI systems that work with people in intuitive and helpful ways. 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. Opportunities abound for core AI scientists and engineers with a deep understanding of the technology who can generate new ideas for advancing the boundaries of knowledge in the field. The Nation should take action to ensure a sufficient pipeline of AI-capable talent. Strategy 7 addresses this challenge.

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.

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.

However, 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.

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.

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 envision 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.

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 system 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.

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.

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.

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.

New methods are also 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.

AI embedded in critical systems must be robust in order to handle accidents, but should also be secure to a wide range of intentional cyber attacks. Security engineering involves understanding the vulnerabilities of a system and the actions of actors who may be interested in attacking it. While cybersecurity R&D needs are addressed in greater detail in the NITRD 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.

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.

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

The importance of testbeds was stated in the Cyber Experimentation of the Future report: "Testbeds are essential so that researchers can use actual operational data to model and run experiments on real-world system[s] … and scenarios in good test environments." 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.

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.

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.

(Link: https://www.nitrd.gov/news/national_ai_rd_strategic_plan.aspx)

Preparing For the Future of Artificial Intelligence

Artificial Intelligence (AI) has the potential to help address some of the biggest challenges that society faces. Smart vehicles may save hundreds of thousands of lives every year worldwide, and increase mobility for the elderly and those with disabilities. Smart buildings may save energy and reduce carbon emissions. Precision medicine may extend life and increase quality of life. Smarter government may serve citizens more quickly and precisely, better protect those at risk, and save money. AI-enhanced education may help teachers give every child an education that opens doors to a secure and fulfilling life. These are just a few of the potential benefits if the technology is developed with an eye to its benefits and with careful consideration of its risks and challenges.

The United States has been at the forefront of foundational research in AI, primarily supported for most of the field's history by Federal research funding and work at government laboratories. The Federal Government's support for unclassified AI R&D is managed through the Networking and Information Technology Research and Development (NITRD) program, and supported primarily by the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation (NSF), the National Institutes of Health (NIH), the Office of Naval Research (ONR), and the Intelligence Advanced Research Projects Activity (IARPA). Major national research efforts such as the National Strategic Computing Initiative, the Big Data Initiative, and the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative also contribute indirectly to the progress of AI research. The current and projected benefits of AI technology are large, adding to the Nation's economic vitality and to the productivity and well-being of its people.

Applications of AI for Public Good

One area of great optimism about AI and machine learning is their potential to improve people's lives by helping to solve some of the world's greatest challenges and inefficiencies. Many have compared the promise of AI to the transformative impacts of advancements in mobile computing. Public- and private-sector investments in basic and applied R&D on AI have already

begun reaping major benefits to the public in fields as diverse as health care, transportation, the environment, criminal justice, and economic inclusion. The effectiveness of government itself is being increased as agencies build their capacity to use AI to carry out their missions more quickly, responsively, and efficiently.

AI and Regulation

AI has applications in many products, such as cars and aircraft, which are subject to regulation designed to protect the public from harm and ensure fairness in economic competition. How will the incorporation of AI into these products affect the relevant regulatory approaches? In general, the approach to regulation of AI-enabled products to protect public safety should be informed by assessment of the aspects of risk that the addition of AI may reduce alongside the aspects of risk that it may increase. If a risk falls within the bounds of an existing regulatory regime, moreover, the policy discussion should start by considering whether the existing regulations already adequately address the risk, or whether they need to be adapted to the addition of AI. Also, where regulatory responses to the addition of AI threaten to increase the cost of compliance, or slow the development or adoption of beneficial innovations, policymakers should consider how those responses could be adjusted to lower costs and barriers to innovation without adversely impacting safety or market fairness.

Currently relevant examples of the regulatory challenges that AI-enabled products present are found in the cases of automated vehicles (AVs, such as self-driving cars) and AI-equipped unmanned aircraft systems (UAS, or "drones"). In the long run, AVs will likely save many lives by reducing driver error and increasing personal mobility, and UAS will offer many economic benefits. Yet public safety must be protected as these technologies are tested and begin to mature. The Department of Transportation (DOT) is using an approach to evolving the relevant regulations that is based on building expertise in the Department, creating safe spaces and test beds for experimentation, and working with industry and civil society to evolve performance-based regulations that will enable more uses as evidence of safe operation accumulates.

Research and Workforce

Government also has an important role to play in the advancement of AI through research and development and the growth of a skilled, diverse workforce. A separate strategic plan for Federally-funded AI research and development is being released in conjunction with this report. The plan discusses the role of Federal R&D, identifies areas of opportunity, and recommends ways to coordinate R&D to maximize benefit and build a highly-trained workforce.

Given the strategic importance of AI, moreover, it is appropriate for the Federal Government to monitor developments in the field worldwide in order to get early warning of important changes arising elsewhere in case these require changes in U.S. policy. The rapid growth of AI has dramatically increased the need for people with relevant skills to support and advance the field. An AI-enabled world demands a data-literate citizenry that is able to read, use, interpret, and communicate about data, and participate in policy debates about matters affected by AI. AI knowledge and education are increasingly emphasized in Federal Science, Technology, Engineering, and Mathematics (STEM) education programs. AI education is also a component of Computer Science for All, the President's initiative to empower all American students from kindergarten through high school to learn computer science and be equipped with the computational thinking skills they need in a technology-driven world.

Economic Impacts of AI

AI's central economic effect in the short term will be the automation of tasks that could not be automated before. This will likely increase productivity and create wealth, but it may also affect particular types of jobs in different ways, reducing demand for certain skills that can be automated while increasing demand for other skills that are complementary to AI. Analysis by the White House Council of Economic Advisors (CEA) suggests that the negative effect of automation will be greatest on lower-wage jobs, and that there is a risk that AI-driven automation will increase the wage gap between less-educated and more-educated workers, potentially increasing economic inequality. Public policy can address these risks, ensuring that workers are retrained and able to succeed in occupations that are complementary to, rather than competing

with, automation. Public policy can also ensure that the economic benefits created by AI are shared broadly, and assure that AI responsibly ushers in a new age in the global economy.

Fairness, Safety, and Governance

As AI technologies move toward broader deployment, technical experts, policy analysts, and ethicists have raised concerns about unintended consequences of widespread adoption. Use of AI to make consequential decisions about people, often replacing decisions made by human-driven bureaucratic processes, leads to concerns about how to ensure justice, fairness, and accountability—the same concerns voiced previously in the Administration's *Big Data: Seizing Opportunities, Preserving Values* report of 2014, as well as the Report to the President on *Big Data and Privacy: A Technological Perspective* published by the President's Council of Advisors on Science and Technology in 2014. Transparency concerns focus not only on the data and algorithms involved, but also on the potential to have some form of explanation for any AI-based determination. Yet AI experts have cautioned that there are inherent challenges in trying to understand and predict the behavior of advanced AI systems.

Use of AI to control physical-world equipment leads to concerns about safety, especially as systems are exposed to the full complexity of the human environment. A major challenge in AI safety is building systems that can safely transition from the "closed world" of the laboratory into the outside "open world" where unpredictable things can happen. Adapting gracefully to unforeseen situations is difficult yet necessary for safe operation. Experience in building other types of safety-critical systems and infrastructure, such as aircraft, power plants, bridges, and vehicles, has much to teach AI practitioners about verification and validation, how to build a safety case for a technology, how to manage risk, and how to communicate with stakeholders about risk.

At a technical level, the challenges of fairness and safety are related. In both cases, practitioners strive to avoid unintended behavior, and to generate the evidence needed to give stakeholders justified confidence that unintended failures are unlikely. Ethical training for AI practitioners and

students is a necessary part of the solution. Ideally, every student learning AI, computer science, or data science would be exposed to curriculum and discussion on related ethics and security topics. However, ethics alone is not sufficient. Ethics can help practitioners understand their responsibilities to all stakeholders, but ethical training should be augmented with technical tools and methods for putting good intentions into practice by doing the technical work needed to prevent unacceptable outcomes.

Global Considerations and Security

AI poses policy questions across a range of areas in international relations and security. AI has been a topic of interest in recent international discussions as countries, multilateral institutions, and other stakeholders have begun to access the benefits and challenges of AI. Dialogue and cooperation between these entities could help advance AI R&D and harness AI for good, while also addressing shared challenges.

Today's AI has important applications in cybersecurity, and is expected to play an increasing role for both defensive and offensive cyber measures. Currently, designing and operating secure systems requires significant time and attention from experts. Automating this expert work partially or entirely may increase security across a much broader range of systems and applications at dramatically lower cost, and could increase the agility of the Nation's cyber-defenses. Using AI may help maintain the rapid response required to detect and react to the landscape of evolving threats.

Challenging issues are raised by the potential use of AI in weapon systems. The United States has incorporated autonomy in certain weapon systems for decades, allowing for greater precision in the use of weapons and safer, more humane military operations. Nonetheless, moving away from direct human control of weapon systems involves some risks and can raise legal and ethical questions.

The key to incorporating autonomous and semi-autonomous weapon systems into American defense planning is to ensure that U.S. Government entities are always acting in accordance with international humanitarian law, taking appropriate steps to control proliferation, and working with partners and Allies to develop standards related to the development and use of such weapon systems. The United States has actively participated in ongoing international discussion on Lethal Autonomous Weapon Systems, and anticipates continued robust international discussion of these potential weapon systems. Agencies across the U.S. Government are working to develop a single, government-wide policy, consistent with international humanitarian law, on autonomous and semi-autonomous weapons.

Preparing for the Future

AI holds the potential to be a major driver of economic growth and social progress, if industry, civil society, government, and the public work together to support development of the technology with thoughtful attention to its potential and to managing its risks.

The U.S. Government has several roles to play. It can convene conversations about important issues and help to set the agenda for public debate. It can monitor the safety and fairness of applications as they develop, and adapt regulatory frameworks to encourage innovation while protecting the public. It can provide public policy tools to ensure that disruption in the means and methods of work enabled by AI increases productivity while avoiding negative economic consequences for certain sectors of the workforce. It can support basic research and the application of AI to public good. It can support development of a skilled, diverse workforce. And government can use AI itself to serve the public faster, more effectively, and at lower cost. Many areas of public policy, from education and the economic safety net, to defense, environmental preservation, and criminal justice, will see new opportunities and new challenges driven by the continued progress of AI. The U.S. Government must continue to build its capacity to understand and adapt to these changes.

As the technology of AI continues to develop, practitioners must ensure that AI-enabled systems are governable; that they are open, transparent, and understandable; that they can work effectively with people; and that their operation will remain consistent with human values and aspirations. Researchers and practitioners have increased their attention to these challenges, and should continue to focus on them.

There is no single definition of AI that is universally accepted by practitioners. Some define AI loosely as a computerized system that exhibits behavior that is commonly thought of as requiring intelligence. Others define AI as a system capable of rationally solving complex problems or taking appropriate actions to achieve its goals in whatever real world circumstances it encounters. Developing and studying machine intelligence can help us better understand and appreciate our human intelligence. Used thoughtfully, AI can augment our intelligence, helping us chart a better and wiser path forward.

In a dystopian vision of this process, these *super-intelligent* machines would exceed the ability of humanity to understand or control. If computers could exert control over many critical systems, the result could be havoc, with humans no longer in control of their destiny at best and extinct at worst. This scenario has long been the subject of science fiction stories, and recent pronouncements from some influential industry leaders have highlighted these fears.

A more positive view of the future held by many researchers sees instead the development of intelligent systems that work well as helpers, assistants, trainers, and teammates of humans, and are designed to operate safely and ethically.

The NSTC Committee on Technology's assessment is that long-term concerns about super-intelligent General AI should have little impact on current policy. The policies the Federal Government should adopt in the near-to-medium term if these fears are justified are almost exactly the same policies the Federal Government should adopt if they are not justified. The best way to build capacity for addressing the longer-term speculative risks is to attack the less

extreme risks already seen today, such as current security, privacy, and safety risks, while investing in research on longer-term capabilities and how their challenges might be managed. Additionally, as research and applications in the field continue to mature, practitioners of AI in government and business should approach advances with appropriate consideration of the long-term societal and ethical questions – in additional to just the technical questions – that such advances portend. Although prudence dictates some attention to the possibility that harmful super-intelligence might someday become possible, these concerns should not be the main driver of public policy for AI.

Machine Learning

Machine learning is one of the most important technical approaches to AI and the basis of many recent advances and commercial applications of AI. Modern machine learning is a statistical process that starts with a body of data and tries to derive a rule or procedure that explains the data or can predict future data. This approach—learning from data—contrasts with the older "expert system" approach to AI, in which programmers sit down with human domain experts to learn the rules and criteria used to make decisions, and translate those rules into software code. An expert system aims to emulate the principles used by human experts, whereas machine learning relies on statistical methods to find a decision procedure that works well in practice.

An advantage of machine learning is that it can be used even in cases where it is infeasible or difficult to write down explicit rules to solve a problem. For example, a company that runs an online service might use machine learning to detect user log-in attempts that are fraudulent. The company might start with a large data set of past login attempts, with each attempt labeled as fraudulent or not using the benefit of hindsight. Based on this data set, the company could use machine learning to derive a rule to apply to future login attempts that predicts which attempts are more likely to be fraudulent and should be subjected to extra security measures. In a sense, machine learning is not an algorithm for solving a specific problem, but rather a more general approach to finding solutions for many different problems, given data about them.

To apply machine learning, a practitioner starts with a historical data set, which the practitioner divides into a *training set* and a *test set*. The practitioner chooses a *model*, or mathematical structure that characterizes a range of possible decision-making rules with adjustable *parameters*. A common analogy is that the model is a "box" that applies a rule, and the parameters are adjustable knobs on the front of the box that control how the box operates. In practice, a model might have many millions of parameters.

The practitioner also defines an *objective function* used to evaluate the desirability of the outcome that results from a particular choice of parameters. The objective function will typically contain parts that reward the model for closely matching the training set, as well as parts that reward the use of simpler rules.

Training the model is the process of adjusting the parameters to maximize the objective function. Training is the difficult technical step in machine learning. A model with millions of parameters will have astronomically more possible outcomes than any algorithm could ever hope to try, so successful training algorithms have to be clever in how they explore the space of parameter settings so as to find very good settings with a feasible level of computational effort.

Once a model has been trained, the practitioner can use the test set to evaluate the accuracy and effectiveness of the model. The goal of machine learning is to create a trained model that will *generalize*—it will be accurate not only on examples in the training set, but also on future cases that it has never seen before. While many of these models can achieve better-than-human performance on narrow tasks such as image labeling, even the best models can fail in unpredictable ways.

Another challenge in using machine learning is that it is typically not possible to extract or generate a straightforward explanation for why a particular trained model is effective. Because trained models have a very large number of adjustable parameters—often hundreds of millions or more—training may yield a model that "works," in the sense of matching the data, but is not

necessarily the simplest model that works. In human decision-making, any opacity in the process is typically due to not having enough information about why a decision was reached, because the decider may be unable to articulate why the decision "felt right." With machine learning, everything about the decision procedure is known with mathematical precision, but there may be simply too much information to interpret clearly.

In recent years, some of the most impressive advancements in machine learning have been in the subfield of deep learning, also known as deep network learning. Deep learning uses structures loosely inspired by the human brain, consisting of a set of units (or "neurons"). Each unit combines a set of input values to produce an output value, which in turn is passed on to other neurons downstream. For example, in an image recognition application, a first layer of units might combine the raw data of the image to recognize simple patterns in the image; a second layer of units might combine the results of the first layer to recognize patterns-of-patterns; a third layer might combine the results of the second layer; and so on.

Deep learning networks typically use many layers—sometimes more than 100— and often use a large number of units at each layer, to enable the recognition of extremely complex, precise patterns in data.

In recent years, new theories of how to construct and train deep networks have emerged, as have larger, faster computer systems, enabling the use of much larger deep learning networks. The dramatic success of these very large networks at many machine learning tasks has come as a surprise to some experts, and is the main cause of the current wave of enthusiasm for machine learning among AI researchers and practitioners.

Autonomy and Automation

AI is often applied to systems that can control physical actuators or trigger online actions. When AI comes into contact with the everyday world, issues of autonomy, automation, and human-machine teaming arise.

Autonomy refers to the ability of a system to operate and adapt to changing circumstances with reduced or without human control. For example, an autonomous car could drive itself to its destination. Despite the focus in much of the literature on cars and aircraft, autonomy is a much broader concept that includes scenarios such as automated financial trading and automated content curation systems. Autonomy also includes systems that can diagnose and repair faults in their own operation, such as identifying and fixing security vulnerabilities.

Automation occurs when a machine does work that might previously have been done by a person. The term relates to both physical work and mental or cognitive work that might be replaced by AI. Automation, and its impact on employment have been significant social and economic phenomena since at least the Industrial Revolution. It is widely accepted that AI will automate some jobs, but there is more debate about whether this is just the next chapter in the history of automation or whether AI will affect the economy differently than past waves of automation have previously.

Human-Machine Teaming

In contrast to automation, where a machine substitutes for human work, in some cases a machine will complement human work. This may happen as a side-effect of AI development, or a system might be developed specifically with the goal of creating a human-machine team. Systems that aim to complement human cognitive capabilities are sometimes referred to as intelligence augmentation.

AI in the Federal Government

The Administration is working to develop policies and internal practices that will maximize the economic and societal benefits of AI and promote innovation. These policies and practices may include:

- investing in basic and applied research and development (R&D);
- serving as an early customer for AI technologies and their applications;
- supporting pilot projects and creating testbeds in real-world settings;
- making data sets available to the public;
- sponsoring incentive prizes;
- identifying and pursuing Grand Challenges to set ambitious but achievable goals for AI;
- funding rigorous evaluations of AI applications to measure their impact and cost-effectiveness; and
- creating a policy, legal, and regulatory environment that allows innovation to flourish while protecting the public from harm.

Using AI in Government to Improve Services and Benefit the American People

One challenge in using AI to improve services is that the Federal Government's capacity to foster and harness innovation in order to better serve the country varies widely across agencies. Some agencies are more focused on innovation, particularly those agencies with large R&D budgets, a workforce that includes many scientists and engineers, a culture of innovation and experimentation, and strong ongoing collaborations with private-sector innovators. Many also have organizations that are specifically tasked with supporting high-risk, high-return research (e.g., the advanced research projects agencies in the Departments of Defense and Energy, as well as the Intelligence Community), and fund R&D across the full range from basic research to advanced development. Other agencies like the NSF have research and development as their primary mission.

But some agencies, particularly those charged with reducing poverty and increasing economic and social mobility, have more modest levels of relevant capabilities, resources, and expertise. For example, while the National Institutes of Health (NIH) has an R&D budget of more than $30 billion, the Department of Labor's R&D budget is only $14 million. This limits the Department

of Labor's capacity to explore applications of AI, such as applying AI-based "digital tutor" technology to increase the skills and incomes of non-college educated workers.

DARPA's "Education Dominance" program serves as an example of AI's potential to fulfill and accelerate agency priorities. DARPA, intending to reduce from years to months the time required for new Navy recruits to become experts in technical skills, now sponsors the development of a digital tutor that uses AI to model the interaction between an expert and a novice.

AI can be a major driver of economic growth and social progress, if industry, civil society, government, and the public work together to support development of the technology, with thoughtful attention to its potential and to managing its risks.

Government has several roles to play. It should convene conversations about important issues and help to set the agenda for public debate. It should monitor the safety and fairness of applications as they develop, and adapt regulatory frameworks to encourage innovation while protecting the public. It should support basic research and the application of AI to public goods, as well as the development of a skilled, diverse workforce. And government should use AI itself, to serve the public faster, more effectively, and at lower cost.

Many areas of public policy, from education and the economic safety net, to defense, environmental preservation, and criminal justice, will see new opportunities and new challenges driven by the continued progress of AI. Government must continue to build its capacity to understand and adapt to these changes.

As the technology of AI continues to develop, practitioners must ensure that AI-enabled systems are governable; that they are open, transparent, and understandable; that they can work effectively with people; and that their operation will remain consistent with human values and

aspirations. Researchers and practitioners have increased their attention to these challenges, and should continue to focus on them.

Developing and studying machine intelligence can help us better understand and appreciate our human intelligence. Used thoughtfully, AI can augment our intelligence, helping us chart a better and wiser path forward.

Recommendations from the report: *Preparing for the Future of Artificial Intelligence*

This section collects all of the recommendations the report, for ease of reference.

Recommendation 1: Private and public institutions are encouraged to examine whether and how they can responsibly leverage AI and machine learning in ways that will benefit society. Social justice and public policy institutions that do not typically engage with advanced technologies and data science in their work should consider partnerships with AI researchers and practitioners that can help apply AI tactics to the broad social problems these institutions already address in other ways.

Recommendation 2: Federal agencies should prioritize open training data and open data standards in AI. The government should emphasize the release of datasets that enable the use of AI to address social challenges. Potential steps may include developing an "Open Data for AI" initiative with the objective of releasing a significant number of government data sets to accelerate AI research and galvanize the use of open data standards and best practices across government, academia, and the private sector.

Recommendation 3: The Federal Government should explore ways to improve the capacity of key agencies to apply AI to their missions. For example, Federal agencies should explore the potential to create DARPA-like organizations to support high-risk, high-reward AI research and its application, much as the Department of Education has done through its proposal to create an "ARPA-ED," to support R&D to determine whether AI and other technologies could significantly improve student learning outcomes.

Recommendation 4: The NSTC MLAI subcommittee should develop a community of practice for AI practitioners across government. Agencies should work together to develop and share standards and best practices around the use of AI in government operations. Agencies should ensure that Federal employee training programs include relevant AI opportunities.

Recommendation 5: Agencies should draw on appropriate technical expertise at the senior level when setting regulatory policy for AI-enabled products. Effective regulation of AI-enabled products requires collaboration between agency leadership, staff knowledgeable about the existing regulatory framework and regulatory practices generally, and technical experts with knowledge of AI. Agency leadership should take steps to recruit the necessary technical talent, or identify it in existing agency staff, and should ensure that there are sufficient technical "seats at the table" in regulatory policy discussions.

Recommendation 6: Agencies should use the full range of personnel assignment and exchange models (e.g. hiring authorities) to foster a Federal workforce with more diverse perspectives on the current state of technology.

Recommendation 7: The Department of Transportation should work with industry and researchers on ways to increase sharing of data for safety, research, and other purposes. The future roles of AI in surface and air transportation are undeniable. Accordingly, Federal actors should focus in the near-term on developing increasingly rich sets of data, consistent with consumer privacy, that can better inform policy-making as these technologies mature.

Recommendation 8: The U.S. Government should invest in developing and implementing an advanced and automated air traffic management system that is highly scalable, and can fully accommodate autonomous and piloted aircraft alike.

Recommendation 9: The Department of Transportation should continue to develop an evolving framework for regulation to enable the safe integration of fully automated vehicles and UAS, including novel vehicle designs, into the transportation system.

Recommendation 10: The NSTC Subcommittee on Machine Learning and Artificial Intelligence should monitor developments in AI, and report regularly to senior Administration leadership about the status of AI, especially with regard to milestones. The Subcommittee should update the list of milestones as knowledge advances and the consensus of experts changes over time. The Subcommittee should consider reporting to the public on AI developments, when appropriate.

Recommendation 11: The Government should monitor the state of AI in other countries, especially with respect to milestones.

Recommendation 12: Industry should work with government to keep government updated on the general progress of AI in industry, including the likelihood of milestones being reached soon.

Recommendation 13: The Federal government should prioritize basic and long-term AI research. The Nation as a whole would benefit from a steady increase in Federal and private-sector AI R&D, with a particular emphasis on basic research and long-term, high-risk research initiatives. Because basic and long-term research especially is areas where the private sector is not likely to invest, Federal investments will be important for R&D in these areas.

Recommendation 14: The NSTC Subcommittees on MLAI and NITRD, in conjunction with the NSTC Committee on Science, Technology, Engineering, and Education (CoSTEM),, should initiate a study on the AI workforce pipeline in order to develop actions that ensure an appropriate increase in the size, quality, and diversity of the workforce, including AI researchers, specialists, and users.

Recommendation 15: The Executive Office of the President should publish a follow-on report by the end of this year, to further investigate the effects of AI and automation on the U.S. job market, and outline recommended policy responses.

Recommendation 16: Federal agencies that use AI-based systems to make or provide decision support for consequential decisions about individuals should take extra care to ensure the efficacy and fairness of those systems, based on evidence-based verification and validation.

Recommendation 17: Federal agencies that make grants to state and local governments in support of the use of AI-based systems to make consequential decisions about individuals should review the terms of grants to ensure that AI-based products or services purchased with Federal grant funds produce results in a sufficiently transparent fashion and are supported by evidence of efficacy and fairness.

Recommendation 18: Schools and universities should include ethics, and related topics in security, privacy, and safety, as an integral part of curricula on AI, machine learning, computer science, and data science.

Recommendation 19: AI professionals, safety professionals, and their professional societies should work together to continue progress toward a mature field of AI safety engineering.

Recommendation 20: The U.S. Government should develop a government-wide strategy on international engagement related to AI, and develop a list of AI topical areas that need international engagement and monitoring.

Recommendation 21: The U.S. Government should deepen its engagement with key international stakeholders, including foreign governments, international organizations, industry, academia, and others, to exchange information and facilitate collaboration on AI R&D.

Recommendation 22: Agencies' plans and strategies should account for the influence of AI on cybersecurity, and of cybersecurity on AI. Agencies involved in AI issues should engage their U.S. Government and private-sector cybersecurity colleagues for input on how to ensure that AI systems and ecosystems are secure and resilient to intelligent adversaries. Agencies involved in cybersecurity issues should engage their U.S. Government and private sector AI colleagues for innovative ways to apply AI for effective and efficient cybersecurity.

Recommendation 23: The U.S. Government should complete the development of a single, government-wide policy, consistent with international humanitarian law, on autonomous and semi-autonomous weapons.

(Link: https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf)

Artificial Intelligence, Automation, and the Economy

Technology is not destiny; economic incentives and public policy can play a significant role in shaping the direction and effects of technological change. Given appropriate attention and the right policy and institutional responses, advanced automation can be compatible with productivity, high levels of employment, and more broadly shared prosperity. In the past, the U.S. economy has adapted to new production patterns and maintained high levels of employment alongside rising productivity as more productive workers have had more incentive to work and more highly paid workers have spent more, supporting this work. But, some shocks have left a growing share of workers out of the labor force. The report on *Artificial Intelligence, Automation, and the Economy* advocates strategies to educate and prepare new workers to enter the workforce, cushion workers who lose jobs, keep them attached to the labor force, and combat inequality. Most of these strategies would be important regardless of AI-driven automation, but all take on even greater importance to the degree that AI is making major changes to the economy.

Accelerating artificial intelligence (AI) capabilities will enable automation of some tasks that have long required human labor. These transformations will open up new opportunities for individuals, the economy, and society, but they have the potential to disrupt the current livelihoods of millions of Americans. Whether AI leads to unemployment and increases in inequality over the long-run depends not only on the technology itself but also on the institutions and policies that are in place. This report examines the expected impact of AI-driven automation on the economy, and describes broad strategies that could increase the benefits of AI and mitigate its costs.

Economics of AI-Driven Automation

Technological progress is the main driver of growth of GDP per capita, allowing output to increase faster than labor and capital. One of the main ways that technology increases productivity is by decreasing the number of labor hours needed to create a unit of output. Labor

productivity increases generally translate into increases in average wages, giving workers the opportunity to cut back on work hours and to afford more goods and services. Living standards and leisure hours could both increase, although to the degree that inequality increases—as it has in recent decades—it offsets some of those gains.

AI should be welcomed for its potential economic benefits. Those economic benefits, however, will not necessarily be evenly distributed across society. For example, the 19th century was characterized by technological change that raised the productivity of lower-skilled workers relative to that of higher-skilled workers. Highly-skilled artisans who controlled and executed full production processes saw their livelihoods threatened by the rise of mass production technologies. Ultimately, many skilled crafts were replaced by the combination of machines and lower-skilled labor. Output per hour rose while inequality declined, driving up average living standards, but the labor of some high-skill workers was no longer as valuable in the market.

In contrast, technological change tended to work in a different direction throughout the late 20th century. The advent of computers and the Internet raised the relative productivity of higher-skilled workers. Routine-intensive occupations that focused on predictable, easily-programmable tasks—such as switchboard operators, filing clerks, travel agents, and assembly line workers—were particularly vulnerable to replacement by new technologies. Some occupations were virtually eliminated and demand for others reduced. Research suggests that technological innovation over this period increased the productivity of those engaged in abstract thinking, creative tasks, and problem-solving and was therefore at least partially responsible for the substantial growth in jobs employing such traits. Shifting demand towards more skilled labor raised the relative pay of this group, contributing to rising inequality. At the same time, a slowdown in the rate of improvement in education, and institutional changes such as the reduction in unionization and decline in the minimum wage, also contributed to inequality—underscoring that technological changes do not uniquely determine outcomes.

Today, it may be challenging to predict exactly which jobs will be most immediately affected by AI-driven automation. Because AI is not a single technology, but rather a collection of technologies that are applied to specific tasks, the effects of AI will be felt unevenly through the economy. Some tasks will be more easily automated than others, and some jobs will be affected more than others—both negatively and positively. Some jobs may be automated away, while for others, AI-driven automation will make many workers more productive and increase demand for certain skills. Finally, new jobs are likely to be directly created in areas such as the development and supervision of AI as well as indirectly created in a range of areas throughout the economy as higher incomes lead to expanded demand.

Recent research suggests that the effects of AI on the labor market in the near term will continue the trend that computerization and communication innovations have driven in recent decades. Researchers' estimates on the scale of threatened jobs over the next decade or two range from 9 to 47 percent. For context, every 3 months about 6 percent of jobs in the economy are destroyed by shrinking or closing businesses, while a slightly larger percentage of jobs are added—resulting in rising employment and a roughly constant unemployment rate. The economy has repeatedly proven itself capable of handling this scale of change, although it would depend on how rapidly the changes happen and how concentrated the losses are in specific occupations that are hard to shift from.

Research consistently finds that the jobs that are threatened by automation are highly concentrated among lower-paid, lower-skilled, and less-educated workers. This means that automation will continue to put downward pressure on demand for this group, putting downward pressure on wages and upward pressure on inequality. In the longer-run, there may be different or larger effects. One possibility is superstar-biased technological change, where the benefits of technology accrue to an even smaller portion of society than just highly-skilled workers. The winner-take-most nature of information technology markets means that only a few may come to dominate markets. If labor productivity increases do not translate into wage increases, then the large economic gains brought about by AI could accrue to a select few. Instead of broadly shared

prosperity for workers and consumers, this might push towards reduced competition and increased wealth inequality.

Historically and across countries, however, there has been a strong relationship between productivity and wages—and with more AI the most plausible outcome will be a combination of higher wages and more opportunities for leisure for a wide range of workers. But the degree that this materializes depends not just on the nature of technological change but importantly on the policy and institutional choices that are made about how to prepare workers for AI and to handle its impacts on the labor market.

Responding to the economic effects of AI-driven automation will be a significant policy challenge for the next Administration and its successors. AI has already begun to transform the American workplace, change the types of jobs available, and reshape the skills that workers need in order to thrive. All Americans should have the opportunity to participate in addressing these challenges, whether as students, workers, managers, technical leaders, or simply as citizens with a voice in the policy debate.

AI raises many new policy questions, which should be continued topics for discussion and consideration by future Administrations, Congress, the private sector, academia, and the public. Continued engagement among government, industry, technical and policy experts, and the public should play an important role in moving the Nation toward policies that create broadly shared prosperity, unlock the creative potential of American companies and workers, and ensure America's continued leadership in the creation and use of AI.

Accelerating AI capabilities will enable automation of some tasks that have long required human labor. Rather than relying on closely-tailored rules explicitly crafted by programmers, modern AI programs can learn from patterns in whatever data they encounter and develop their own rules for how to interpret new information. This means that AI can solve problems and learn with very little human input. In addition, advances in robotics are expanding machines' abilities to interact

with and shape the physical world. Combined, AI and robotics will give rise to smarter machines that can perform more sophisticated functions than ever before and erode some of the advantages that humans have exercised. This will permit automation of many tasks now performed by human workers and could change the shape of the labor market and human activity.

Critically, technology alone will not determine the economic outcomes in terms of growth, inequality or employment. The advanced economies all have had access to similar levels of technology but have had very different outcomes along all of these dimensions because they have had different institutions and policies. But understanding the technological forces is critical to shaping the continued evolution of these policies.

At times, new technologies have raised the productivity and increased employment opportunities for workers with little education, and other times for workers with more. To illustrate the diversity of potential impacts and provide a framework for understanding today, this section discusses historical examples of how innovations affected workers in different ways.

The 19th century was characterized by technological change that raised the productivity of lower-skilled workers and reduced the relative productivity of certain higher-skilled workers. This kind of innovation has been called *unskill-biased technical change*. Highly-skilled artisans who controlled and executed full production processes saw their livelihoods threatened by the rise of mass production technologies that used assembly lines with interchangeable parts and lower-skilled workers. In reaction, some English textile weavers participated in the Luddite Riots of the early 1800s by destroying looms and machinery that threatened to undercut their highly-skilled, highly-paid jobs with lower-wage roles. Ultimately, the protesters' fears came true, and many skilled crafts were replaced by the combination of machines and lower-skill labor. There were also new opportunities for less-skilled workers and output per hour rose. As a result, average living standards could rise, but certain high-skill workers were no longer as valuable in the market.

Technological change tended to work in a different direction throughout the late 20th century. The advent of computers and the internet raised the relative productivity of higher-skilled workers, an example of *skill-biased technical change*. Routine-intensive occupations that focused on predictable, easily-programmable tasks—such as switchboard operators, filing clerks, travel agents, and assembly line workers—have been particularly vulnerable to replacement by new technologies. Some entire occupations were virtually eliminated and demand for others reduced. In routine jobs is largely responsible for recent low labor demand for less educated workers.

Like these past waves of technological advancements, AI-driven automation is setting off labor-market disruption and adjustment. Economic theory suggests that there must be gains from innovations, or they would not be adopted. Market forces alone, however, will not ensure that the financial benefits from innovations are broadly shared.

Today, it may be challenging to predict exactly which jobs will be most immediately affected by AI-driven automation. Because AI is not a single technology, but rather a collection of technologies that are applied to specific tasks, the effects of AI will be felt unevenly through the economy. Some work tasks will be more easily automated than others, and some jobs will be affected more than others.

Some specific predictions are possible based on the current trajectory of AI technology. For example, driving jobs and housecleaning jobs are both jobs that require relatively less education to perform. Advancements in computer vision and related technologies have made the feasibility of fully automated vehicles (AVs), which do not require a human driver, appear more likely, potentially displacing some workers in driving-dominant professions. AVs rely upon, among other things, capabilities of navigating complex environments, analyzing dynamic surroundings, and optimization. Seemingly similar capabilities are required of a household-cleaning robot, for which the operational mandate is less specific (i.e. "clean the house," as opposed to the objective of navigating to a specific destination while following a set of given rules and preserving safety).

And yet the technology that would enable a robot to navigate and clean a space as effectively as a human counterpart appears farther off. In the near to medium term, at least, drivers will probably be impacted more by automation than will housecleaners.

Nevertheless, humans still maintain a comparative advantage over AI and robotics in many areas. While AI detects patterns and creates predictions, it still cannot replicate social or general intelligence, creativity, or human judgment. Of course, many of the occupations that use these types of skills are high-skilled occupations, and likely require higher levels of education. Further, given the current dexterity limits of the robotics that would be needed to implement mass AI-driven automation, occupations that require manual dexterity will also likely remain in demand in the near term.

There are four categories of jobs that might experience direct AI-driven growth in the future. Employment in areas where humans engage with existing AI technologies, develop new AI technologies, supervise AI technologies in practice, and facilitate societal shifts that accompany new AI technologies will likely grow. Current limits on manual dexterity of robots and constraints on generative intelligence and creativity of AI technologies likely mean that employment requiring manual dexterity, creativity, social interactions and intelligence, and general knowledge will thrive. Below are descriptions and potential examples of future employment for each category.

Engagement. Humans will likely be needed to actively engage with AI technologies throughout the process of completing a task. Many industry professionals refer to a large swath of AI technologies as "Augmented Intelligence," stressing the technology's role as assisting and expanding the productivity of individuals rather than replacing human work. Thus, based on the biased-technical change framework, demand for labor will likely increase the most in the areas where humans complement AI-automation technologies. For example, AI technology such as IBM's Watson may improve early detection of some cancers or other illnesses, but a human healthcare professional is needed to work with patients to understand and translate patients'

symptoms, inform patients of treatment options, and guide patients through treatment plans. Shipping companies may also partner workers who pickup and deliver goods over the last 100 feet with AI-enabled autonomous vehicles that move workers efficiently from site to site. In such cases, AI augments what a human is able to do and allows individuals to either be more effective in their specialty task or to operate on a larger scale.

***Development*.** In the initial stages of AI, development jobs are crucial and span multiple industries and skill levels. Most intuitively, there may be a great need for highly-skilled software developers and engineers to put these capacities into practical use in the world. To a certain extent, however, AI is only as good as the data behind it, so there will likely be increased demand for jobs in generating, collecting, and managing relevant data to feed into AI training processes. Applications of AI can range from high-skill tasks such as recognizing cancer in x-ray images to lower-skill tasks such as recognizing text in images. Finally, to an increasing degree, development may include those specializing in the liberal arts and social sciences, such as philosophers with frameworks for ethical evaluations and sociologists investigating the impact of technology on specific populations, who can give input as the new technologies grapple with more social complexities and moral dilemmas.

***Supervision*.** This category encompasses all roles related to the monitoring, licensing, and repair of AI. For example, after the automated vehicle development phase, the need for human registration and testing of such technology to ensure safety and quality control on the roads will still likely exist. As a widespread new technology, AV will require regular repair and maintenance, which may expand mechanic and technician jobs in this space as well. Real-time supervision will also be required in exceptional, marginal, or high-stakes cases, especially those involving morality, ethics, and social intelligence that AI may lack. This might take the form of quality control of recommendations made by AI or online moderation when sensitive subjects are discussed. The capacity for AI-enabled machines to learn is one of the most exciting aspects of the technology, but it may also require supervision to ensure that AI does not diverge from originally intended uses. As machines get smarter and have improved ability to make practical

predictions about the environment, the value of human judgment will increase because it will be the preferred way to resolve competing priorities.

Response to Paradigm Shifts. The technological innovation surrounding AI will likely reshape features of built environment. In the case of AVs, dramatic shifts in the design of infrastructure and traffic laws—which are currently built with the safety and convenience of human drivers in mind—may be needed. The advent of self-driving cars may result in higher demand for urban planners and designers to create a new blueprint for the way the everyday travel landscape is built and used. Paradigm shifts in adjacent fields such as cybersecurity—demanding, for instance, new methods of detecting fraudulent transactions and messages—may also necessitate new occupations and more employment.

AI-driven automation stands to transform the economy over the coming years and decades. The challenge for policymakers will be to update, strengthen, and adapt policies to respond to the economic effects of AI.

Although it is difficult to predict these economic effects precisely with a high degree of confidence, the economic analysis in the previous chapter suggests that policymakers should prepare for five primary economic effects:

- Positive contributions to aggregate productivity growth;
- Changes in the skills demanded by the job market, including greater demand for higher-level technical skills;
- Uneven distribution of impact, across sectors, wage levels, education levels, job types, and locations;
- Churning of the job market as some jobs disappear while others are created; and
- The loss of jobs for some workers in the short-run, and possibly longer depending on policy responses.

There is substantial uncertainty about how strongly these effects will be felt, and how rapidly they will arrive. It is possible that AI will not have large, new effects on the economy, such that the coming years are subject to the same basic workforce trends seen in recent decades—some which are positive, and others which are worrisome and may require policy changes. At the other end of the range of possibilities, the economy might potentially experience a larger shock, with accelerating changes in the job market, and significantly more workers in need of assistance and retraining as their skills are no longer valued in the job market. Given presently available evidence, it is not possible to make specific predictions, so policymakers must be prepared for a range of potential outcomes. At a minimum, some occupations such as drivers and cashiers are likely to face displacement from or restructuring of their current jobs, leading millions of Americans to experience economic hardship in the short-run absent new policies.

Because the effects of AI-driven automation will likely be felt across the whole economy, and the areas of greatest impact may be difficult to predict, policy responses must be targeted to the whole economy. In addition, the economic effects of AI-driven automation may be difficult to separate from those of other factors such as other technological changes, globalization, reduction in market competition and worker bargaining power, and the effects of past public policy choices. Even if it is not possible to determine how much of the current transformation of the economy is caused by each of these factors, the policy challenges raised by the disruptions remain, and require a broad policy response.

(Link: https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED%20AI%20Economy%20Report.pdf)

Acronyms

AAAI	Association for the Advancement of Artificial Intelligence
AGI	Artificial General Intelligence
AI	Artificial Intelligence
APEC	Asia-Pacific Economic Cooperation
BRAIN	Brain Research through Advancing Innovative Neurotechnologies
CALO	Cognitive Agent that Learns and Organizes
CCC	Computing Community Consortium
CCW	Convention on Certain Conventional Weapons
CEA	Council of Economic Advisers
CEO	Chief Executive Officer
CGC	Cyber Grand Challenge (run by DARPA)
CoSTEM	Committee on Science Technology, Engineering, and Education (component of NSTC)
CS	Computer Science
DARPA	Defense Advanced Research Projects Agency
DoD	Department of Defense
DOT	Department of Transportation
FAA	Federal Aviation Administration
FMVSS	Federal Motor Vehicle Safety Standards
IARPA	Intelligence Advanced Research Projects Activity
ICTs	Information and Communication Technologies
IPA	Intergovernmental Personnel Act
LAWS	Lethal Autonomous Weapon Systems
MLAI	Machine Learning and Artificial Intelligence (subcommittee of NSTC)
NAS	National Airspace System
NEC	National Economic Council
NHTSA	National Highway Traffic Safety Administration
NIH	National Institutes of Health
NIPS	Neural Information Processing Systems conference
NITRD	Networking and Information Technology Research and Development (subcommittee of NSTC)
NSF	National Science Foundation
NSTC	National Science and Technology Council
OECD	Organization for Economic Cooperation and Development
OMB	Office of Management and Budget
ONR	Office of Naval Research
OSTP	Office of Science and Technology Policy
R&D	Research and Development
RFI	Request For Information
STEM	Science, Technology, Engineering, and Mathematics
UAS	Unmanned Aerial System
UTM	UAS Traffic Management

FREE eBooks

WHOEVER
WHENEVER
WHEREVER
YOU ARE

INSTANTLY DOWNLOAD THESE MASSIVE BOOK BUNDLES

CLICK ANY BELOW TO ENJOY NOW

3 AUDIOBOOK COLLECTIONS

Classic AudioBooks Vol 1 ▪ Classic AudioBooks Vol 2 ▪ Classic AudioBooks Kids

6 BOOK COLLECTIONS

Sci-Fi ▪ Romance ▪ Mystery ▪ Academic ▪ Classics ▪ Business