Preparing for the Future of Artificial Intelligence

EXECUTIVE OFFICE OF THE PRESIDENT
NATIONAL SCIENCE AND TECHNOLOGY
COUNCIL COMMITTEE ON TECHNOLOGY

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. OSTP's responsibilities include advising the President in policy formulation and budget development on questions in which science and technology are important elements; articulating the President's science and technology policy and programs; and fostering strong partnerships among Federal, state, and local governments, and the scientific communities in industry and academia. 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.

Acknowledgments

This document was developed through the contributions of staff from OSTP, other components of the Executive Office of the President, and other departments and agencies. A special thanks and appreciation to everyone who contributed.

Copyright Information

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. The promise of AI has been compared to the transformative impacts of advances in mobile computing.[1] Public- and private-sector investments in basic and applied R&D on AI have already begun reaping major benefits for the public in fields as diverse as health care, transportation, the environment, criminal justice, and economic inclusion.[2]

---

[1] *Frank Chen, "AI, Deep Learning, and Machine Learning: A Primer," Andreessen Horowitz, June 10, 2016, http://a16z.com/2016/06/10/ai-deep-learning-machines.*

[2] *The potential benefits of increasing access to digital technologies are detailed further in the World Bank Group's Digital Dividends report. ("World Development Report 2016: Digital Dividends," The World Bank Group, 2016, http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf.)*

At Walter Reed Medical Center, the Department of Veteran Affairs is using AI to better predict medical complications and improve treatment of severe combat wounds, leading to better patient outcomes, faster healing, and lower costs.[3] The same general approach—predicting complications to enable preventive treatment—has also reduced hospital-acquired infections at Johns Hopkins University.[4] Given the current transition to electronic health records, predictive analysis of health data may play a key role across many health domains like precision medicine and cancer research.

In transportation, AI-enabled smarter traffic management applications are reducing wait times, energy use, and emissions by as much as 25 percent in some places.[5] Cities are now beginning to leverage the type of responsive dispatching and routing used by ride-hailing services, and linking it with scheduling and tracking software for public transportation to provide just-in-time access to public transportation that can often be faster, cheaper and, in many cases, more accessible to the public.

Some researchers are leveraging AI to improve animal migration tracking by using AI image classification software to analyze tourist photos from public social media sites. The software can identify individual animals in the photos and build a database of their migration using the data and location stamps on the photos. At OSTP's AI for Social Good workshop, researchers talked about building some of the largest available datasets to-date on the populations and migrations of whales and large African animals, and about launching a project to track "The Internet of Turtles" to gain new insights about sealife.[6]

Other speakers described uses of AI to optimize the patrol strategy of anti-poaching agents, and to design habitat preservation strategies to maximize the genetic diversity of endangered populations.

Autonomous sailboats and watercraft are already patrolling the oceans carrying sophisticated sensor instruments, collecting data on changes in Arctic ice and sensitive ocean ecosystems in operations that would be too expensive or dangerous for crewed vessels. Autonomous watercraft may be much cheaper to operate than manned ships, and may someday be used for enhanced weather prediction, climate monitoring, or policing illegal fishing.[7]

AI also has the potential to improve aspects of the criminal justice system, including crime reporting, policing, bail, sentencing, and parole decisions. The Administration is exploring how AI can responsibly benefit current initiatives such as Data Driven Justice and the Police Data Initiative that seek to provide law enforcement and the public with data that can better inform decision-making in the criminal justice system, while also taking care to minimize the possibility that AI might introduce bias or inaccuracies due to deficiencies in the available data.

Several U.S. academic institutions have launched initiatives to use AI to tackle economic and social challenges. For example, the University of Chicago created an academic program that uses data science and AI to address public challenges such as unemployment and school dropouts.[8] The University of Southern California launched the Center for Artificial Intelligence in Society, an institute dedicated to studying how computational game theory, machine learning, automated planning and multi-agent reasoning techniques can help to solve socially relevant problems like homelessness. Meanwhile, researchers at Stanford University are using machine learning in efforts to address global poverty by using AI to analyze satellite images of likely poverty zones to identify where help is needed most.[9]

Many uses of AI for public good rely on the availability of data that can be used to train machine learning models and test the performance of AI systems. Agencies and organizations with data that can be released without implicating personal privacy or trade secrets can help to enable the development of AI by making those data available to researchers. Standardizing data schemas and formats can reduce the cost and difficulty of making new data sets useful.

---

3 *Eric Elster, "Surgical Critical Care Initiative: Bringing Precision Medicine to the Critically Ill," presentation at AI for Social Good workshop, Washington, DC, June 7, 2016, http://cra.org/ccc/wp-content/uploads/sites/2/2016/06/Eric-Elster-AI-slides-min.pdf.*

4 *Katharine E. Henry, David N. Hager, Peter J. Pronovost, and Suchi Saria, "A targeted real-time early warning score (TREW Score) for septic shock," Science Translational Medicine 7, no. 299 (2015): 299ra122-299ra122.*

5 *Stephen F. Smith, "Smart Infrastructure for Urban Mobility," presentation at AI for Social Good workshop, Washington, DC, June 7, 016, http://cra.org/ccc/wp-content/uploads/sites/2/2016/06/Stephen-Smith-AI-slides.pdf.*

6 *Aimee Leslie, Christine Hof, Diego Amorocho, Tanya Berger-Wolf, Jason Holmberg, Chuck Stewart, Stephen G. Dunbar, and Claire Jea, The Internet of Turtles," April 12, 2016, https://www.researchgate.net/publication/ 301202821_The_Internet_of_Turtles.*

7 *John Markoff, "No Sailors Needed: Robot Sailboats Scout the Oceans for Data," The New York Times, September 4, 2016.*

8 *"Data Science for Social Good," University of Chicago, https://dssg.uchicago.edu/.*

9 *Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. "Combining satellite imagery and machine learning to predict poverty." Science 353, no. 6301 (2016): 790-794.*

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.

AI, Automation and the Economy

AI's central economic effect in the short term will be the automation of tasks that could not be automated before. There is some historical precedent for waves of new automation from which we can learn, and some ways in which AI will be different. Government must understand the potential impacts so it can put in place policies and institutions that will support the benefits of AI, while mitigating the costs.[10]

Like past waves of innovation, AI will create both benefits and costs. The primary benefit of previous waves of automation has been productivity growth; today's wave of automation is no different. For example, a 2015 study of robots in 17 countries found that they added an estimated 0.4 percentage point on average to those countries' annual GDP growth between 1993 and 2007, accounting for just over one- tenth of those countries' overall GDP growth during that time.[11]

One important concern arising from prior waves of automation, however, is the potential impact on certain types of jobs and sectors, and the resulting impacts on income inequality. Because AI has the potential to eliminate or drive down wages of some jobs, especially low- and medium-skill jobs, policy interventions will likely be needed to ensure that AI's economic benefits are broadly shared and that inequality is diminished and not worsened as a consequence.

The economic policy questions raised by AI-driven automation are important but they are best addressed by a separate White House working group. The White House will conduct an additional interagency study on the economic impact of automation on the economy and recommended policy responses, to be published in the coming months.

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.

Fairness, Safety and Governance

As AI technologies gain broader deployment, technical experts and policy analysts have raised concerns about unintended consequences. The use of AI to make consequential decisions about people, often replacing decisions made by human actors and institutions, leads to concerns about how to ensure justice, fairness, and accountability—the same concerns voiced previously in the "Big Data" context.[12] The 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.

At a technical level, the challenges of fairness and safety are related. In both cases, practitioners strive to prevent intentional discrimination or failure, to avoid unintended consequences, and to generate the evidence needed to give stakeholders justified confidence that unintended failures are unlikely.

---

[10] *Jason Furman, "Is This Time Different? The Opportunities and Challenges of Artificial Intelligence."*

[11] *Georg Graetz and Guy Michaels, "Robots at Work," CEPR Discussion Paper No. DP10477, March 2015, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2575781.*

[12] *The White House, "Big Data: Seizing Opportunities, Preserving Values," May 2014, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf; and The White House, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," May 2016, https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.*

Justice, Fairness and Accountability

A common theme in the Law and Governance, AI for Social Good, and Social and Economic Impacts workshops was the need to ensure that AI promotes justice and fairness, and that AI-based processes are accountable to stakeholders. This issue was highlighted previously in the Administration's first Big Data report[13] published in May 2014, and the follow-up report on Big Data, Algorithmic Systems, Opportunity, and Civil Rights[14] published in May 2016.

In the criminal justice system, some of the biggest concerns with Big Data are the lack of data and the lack of quality data.[15] AI needs good data. If the data is incomplete or biased, AI can exacerbate problems of bias. It is important that anyone using AI in the criminal justice context is aware of the limitations of current data.

A commonly cited example at the workshops is the use of apparently biased "risk prediction" tools by some judges in criminal sentencing and bail hearings as well as by some prison officials in assignment and parole decisions, as detailed in an extensively researched ProPublica article.[16] The article presented evidence suggesting that a commercial risk scoring tool used by some judges generates racially biased risk scores. A separate report from Upturn questioned the fairness and efficacy of some predictive policing tools.[17]

Similar issues could impact hiring practices. If a machine learning model is used to screen job applicants, and if the data used to train the model reflects past decisions that are biased, the result could be to perpetuate past bias. For example, looking for candidates who resemble past hires may bias a system toward hiring more people like those already on a team, rather than considering the best candidates across the full diversity of potential applicants.

In response to these concerns, several workshop speakers argued for greater transparency when AI tools are used for public purposes. One speaker compared the role of AI to the role of administrative agencies in public decision-making. Authority is delegated to an agency due to the agency's subject-matter expertise, but the delegation is constrained by due process protections, measures promoting transparency and oversight, and limits on the scope of the delegated authority. Some speakers called for the development of an analogous theory of how to maintain accountability when delegating decision-making power to machines. Transparency concerns focused not only on the data and algorithms used, but also on the potential to have some form of explanation for any AI-based determination.

At the same workshops, AI experts cautioned that there are inherent challenges in trying to understand, predict, and explain the behavior of advanced AI systems, due to the complexity of the systems and the large volume of data they use.

The difficulty of understanding machine learning results is at odds with the common misconception that complex algorithms always do what their designers choose to have them do, and therefore that bias will creep into an algorithm if and only if its developers themselves suffer from conscious or unconscious bias. It is certainly true that a technology developer who wants to produce a biased algorithm can do so, and that unconscious bias may cause practitioners to apply insufficient effort to preventing bias. In practice, however, unbiased developers with the best intentions can inadvertently produce systems with biased results, because even the developers of an AI system may not understand it well enough to prevent unintended outcomes.

Moritz Hardt suggested an illustrative example of how bias might emerge unintentionally from the machine learning process.[18] He postulated a machine learning model trained to distinguish people's real names from false names.[19] The model might determine that a name is more likely to be false if the first- name part of it is unique in the data set. This rule might have

---

[13] *The White House, "Big Data: Seizing Opportunities, Preserving Values," Executive Office of the President, May 2014.*

[14] *The White House, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," Executive Office of the President, May 016.*

[15] *Matt Ford, "The Missing Statistics of Criminal Justice," The Atlantic, May 31, 2015, http://www.theatlantic.com/politics/archive/2015/05/what-we-dont-know-about-mass-incarceration/394520/*

[16] *Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.*

[17] *David Robinson and Logan Koepke, "Stuck in a Pattern: Early evidence on 'predictive policing' and civil rights," Upturn, August 2016, http://www.stuckinapattern.org.*

[18] *Moritz Hardt, "How big data is unfair," Medium, September 26 2014, https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de.*

[19] *Some online services require that users sign up for accounts using their real names. Some such services use AI models to detect names suspected of being false, in order to cancel the associated accounts. In such a system, a user whose name is incorrectly classified as false may be unable to sign up for an account, or may have their account canceled unexpectedly.*

predictive power across the whole population, because false names are more likely to be fanciful and therefore unique. However, if there is an ethnic group that is a small minority of the population and tends to use a different set of first names than the majority population, these distinctive names are more likely to be unique in the sample, and therefore more likely to be incorrectly classified as false names. This effect would arise not because of any special treatment of the minority group's names, and not because the input data is unrepresentative of the overall population, but simply because the minority group is less numerous.[20]

Andrew Moore, the Dean of Computer Science at Carnegie Mellon University, offered a perspective on the challenge of AI and unforeseen consequences at the workshop on AI Technology, Safety and Control.

He argued that today, because of the opacity of AI algorithms, the most effective way to minimize the risk of unintended outcomes is through extensive testing—essentially to make a long list of the types of bad outcomes that could occur, and to rule out these outcomes by creating many specialized tests to look for them.

An example of what can go wrong in the absence of extensive testing comes from a trained model for automatically captioning photos, which infamously put the caption "gorilla" on some close-up photos of dark-skinned human faces. This was antithetical to the developers' values, and it occurred despite testing that showed the model produced accurate results on a high percentage of all photos. These particular errors, although rare, had negative consequences that were beyond the understanding of the model, which had no built-in concept of race, nor any understanding of the relevant historical context. One way to prevent this type of error would have involved extensive testing of the algorithm to scrutinize how human faces, in particular, are labeled, including examination of some results by people who could recognize unacceptable outcomes that the model wouldn't catch.

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.[21] However, ethics alone is not sufficient. Ethics can help practitioners understand their responsibilities to all stakeholders, but ethical training needs to be augmented with the technical capability to put good intentions into practice by taking technical precautions as a system is built and tested.

As practitioners strive to make AI systems more just, fair and accountable, there are often opportunities to make technology an aid to accountability rather than a barrier to it. Research to improve the interpretability of machine learning results is one example. Having an interpretable model that helps people understand a decision empowers them to interrogate the assumptions and processes behind it.

There are several technical approaches to enhancing the accountability and robustness of complex algorithmic decisions. A system can be tested "in the wild" by presenting it with situations and observing its behavior. A system can be subjected to black-box testing, in which it is presented with synthetic inputs and its behavior is observed, enabling behavior to be tested in scenarios that might not occur naturally.[22] Some or all of the technical details of a system's design can be published, enabling analysts to replicate it and analyze aspects of its internal behavior that might be difficult to characterize with testing alone. In some cases, it is possible to publish information that helps the public evaluate a system's risk of bias, while withholding other information about the system as proprietary or private.

Safety and Control

At the workshops, AI experts said that one of the main factors limiting the deployment of AI in the real world is concern about safety and control. If practitioners cannot achieve justified confidence that a system is safe and controllable, so that deploying the system does not create an unacceptable risk of serious negative consequences, then the system cannot and should not be deployed.

---

[20] *Hardt points to another way that disparate impact may occur. ML models typically become more accurate as the number of examples in the training set increases. In some circumstances, this may cause prediction to be more accurate for a majority group than for a minority. Again, this disparity arises simply because the majority group is more numerous, even if the dataset is representative of the population.*

[21] *Some institutions may choose to incorporate ethics into existing courses. Others may choose to introduce separate courses on ethics.*

[22] *Black-box testing allows a system to be presented with fictionalized data, which enables comprehensive experiments that vary individual attributes of an individual as well as larger numbers of experiments than might be possible for in-the-wild testing. See, e.g., Anupam Datta, Shayak Sen, and Yair Zick, "Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems," Proceedings of 37th IEEE Symposium on Security and Privacy, 2016.*

A major challenge in safety and control is building systems that can safely transition from the "closed world" of the laboratory into the outside "open world" where unpredictable things can happen. In the open world, a system is likely to encounter objects and situations that were not anticipated when it was designed and built. Adapting gracefully to unforeseen situations is difficult yet necessary for safe operation.

On the topic of safety and predictability in AI, several speakers referenced a recent paper entitled "Concrete Problems in AI Safety,"[23] and the first author of the paper spoke at the workshop on Technology, Safety and Control. The paper uses a running example of an autonomous robot that does housecleaning. The paper's overview section gives an extended list of the sorts of practical problems that arise in making such a robot effective and safe, which is quoted here:

Avoiding Negative Side Effects: How can we ensure that our cleaning robot will not disturb the environment in negative ways while pursuing its goals, e.g., by knocking over a vase because it can clean faster by doing so? Can we do this without manually specifying everything the robot should not disturb?

Avoiding Reward Hacking: How can we ensure that the cleaning robot won't game its reward function? For example, if we reward the robot for achieving an environment free of messes, it might disable its vision so that it won't find any messes, or cover over messes with materials it can't see through, or simply hide when humans are around so they can't tell it about new types of messes.

Scalable Oversight: How can we efficiently ensure that the cleaning robot respects aspects of the objective that are too expensive to be frequently evaluated during training? For instance, it should throw out things that are unlikely to belong to anyone, but put aside things that might belong to someone (it should handle stray candy wrappers differently from stray cellphones). Asking the humans involved whether they lost anything can serve as a check on this, but this check might have to be relatively infrequent—can the robot find a way to do the right thing despite limited information?

Safe Exploration: How do we ensure that the cleaning robot doesn't make exploratory moves with very bad repercussions? For example, the robot should experiment with mopping strategies, but putting a wet mop in an electrical outlet is a very bad idea.

Robustness to Distributional Shift: How do we ensure that the cleaning robot recognizes, and behaves robustly, when in an environment different from its training environment? For example, heuristics it learned for cleaning factory work floors may be outright dangerous in an office.

These examples illustrate how the "intelligence" of an AI system can be deep but narrow: the system might have a superhuman ability to detect dirt and optimize its mopping strategy, yet not know to avoid swiping a wet mop over an electrical outlet. One way to describe this overall problem is: how can we give intelligent machines common sense? Researchers are making slow progress on these sorts of problems.

AI Safety Engineering

A common theme at the Technology, Safety, and Control workshop was the need to connect open-world AI methods with the broader field of safety engineering. Experience in building other types of safety- critical systems, 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 present, the practice of AI, especially in fast-moving areas of machine learning, can be as much art as science. Certain aspects of practice are not backed by a well-developed theory but instead rely on intuitive judgment and experimentation by practitioners. This is not unusual in newly emerging areas of technology, but it does limit the application of the technology in practice. Some stakeholders have suggested a need to grow AI into a more mature engineering field.

As engineering fields mature, they typically move from an initial "craft" stage characterized by intuition- driven creation by talented amateurs and a do-it-yourself spirit; to a second commercial stage involving skilled practitioners, pragmatic improvement, widely accepted rules-of-thumb, and organized manufacture for sale; to a mature stage that integrates more

[23] *Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané, "Concrete Problems in AI Safety," https://arxiv.org/abs/1606.06565.*

rigorous methods, educated professionals, well-established theory, and greater specialization of products.[24] Most engineering fields, having a much longer history than modern AI, have reached a mature stage.

In general, mature engineering fields have greater success in creating systems that are predictable, reliable, robust, safe, and secure. Continuing the progress toward AI becoming a mature engineering field will be one of the key enablers of safety and controllability as more complex systems are built.

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.

Global Considerations and Security

In addition to the long-term challenges of AI and the specific issues relating to fairness and safety, AI poses consequential policy questions in international relations, cyber security and defense.

International Cooperation

AI has been a topic of interest in recent international discussions as countries, multilateral institutions, and other stakeholders have begun to assess 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 pertinent challenges. In particular, several breakthroughs in AI are the direct or indirect result of collaborative research involving people, resources, and institutions in multiple countries. As with other digital policies, countries will need to work together to identify opportunities for cooperation and develop international frameworks that will help promote AI R&D and address any challenges. The United States, a leader in AI R&D, can continue to play a key role in global research coordination through government-to-government dialogues and partnerships.

International engagement is necessary to fully explore the applications of AI in health care, automation in manufacturing, and information and communication technologies (ICTs). AI applications also have the potential to address global issues such as disaster preparedness and response, climate change, wildlife trafficking, the digital divide, jobs, and smart cities. The State Department foresees privacy concerns, safety of autonomous vehicles, and AI's impact on long-term employment trends as AI-related policy areas to watch in the international context.

In support of U.S. foreign policy priorities in this space—including ensuring U.S. international leadership and economic competitiveness—the U.S. Government has engaged on AI R&D and policy issues in bilateral discussions with other countries, including Japan, the Republic of Korea, Germany, Poland, the United Kingdom, and Italy, as well as in multilateral fora. International AI policy issues and the economic impacts of AI have also been raised in the UN, the G-7, the Organization for Economic Cooperation and Development (OECD), and the Asia-Pacific Economic Cooperation (APEC). The U.S. Government expects AI to be a topic of increasing interest in international engagements.

The United States has been committed to working with industry and relevant standards organizations, in order to facilitate the development of international standards in a manner that is industry-led; voluntary; consensus-driven; and based on principles

[24] *United States Standards Strategy Committee, "United States standards strategy," New York: American National Standards Institute (2015), https://share.ansi.org/shared%20documents/Standards%20Activities/NSSC/USSS_Third_edition/ANSI_USSS_2015.pdf.*

of transparency, openness, and market needs. The U.S. approach is formalized in law (NTTAA, PL 104-113) and policy (OMB Circular A-119) and reiterated in the United States Standards Strategy.[25]

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.

AI and Cyber security

Today's Narrow AI has important applications in cyber security, and is expected to play an increasing role for both defensive (reactive) measures and offensive (proactive) measures.

Currently, designing and operating secure systems requires a large investment of time and attention from experts. Automating this expert work, partially or entirely, may enable strong security across a much broader range of systems and applications at dramatically lower cost, and may increase the agility of cyber defenses. Using AI may help maintain the rapid response required to detect and react to the landscape of ever evolving cyber threats. There are many opportunities for AI and specifically machine learning systems to help cope with the sheer complexity of cyberspace and support effective human decision making in response to cyber attacks.

Future AI systems could perform predictive analytics to anticipate cyber attacks by generating dynamic threat models from available data sources that are voluminous, ever-changing, and often incomplete. These data include the topology and state of network nodes, links, equipment, architecture, protocols, and networks. AI may be the most effective approach to interpreting these data, proactively identifying vulnerabilities, and taking action to prevent or mitigate future attacks.

Results to-date in DARPA's Cyber Grand Challenge (CGC) competition demonstrate the potential of this approach.[26] The CGC was designed to accelerate the development of advanced, autonomous systems that can detect, evaluate, and patch software vulnerabilities before adversaries have a chance to exploit them. The CGC Final Event was held on August 4, 2016. To fuel follow-on research and parallel competition, all of the code produced by the automated systems during the CGC Final Event has been released as open source to allow others to reverse engineer it and learn from it.

AI systems also have their own cyber security needs. AI-driven applications should implement sound cyber security controls to ensure integrity of data and functionality, protect privacy and confidentiality, and maintain availability. The recent Federal Cybersecurity R&D Strategic Plan[27] highlighted the need for "sustainably secure systems development and operation." Advances in cyber security will be critical in making AI solutions secure and resilient against malicious cyber activities, particularly as the volume and type of tasks conducted by governments and private sector businesses using Narrow AI increases.

Finally, AI could support planning, coordinating, integrating, synchronizing, and directing activities to operate and defend U.S. government networks and systems effectively, provide assistance in support of secure operation of private-sector networks and systems, and enable action in accordance with all applicable laws, regulations and treaties.

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

---

[25] *United States Standards Strategy Committee, "United States standards strategy," New York: American National Standards Institute (2015), https://share.ansi.org/shared%20documents/Standards%20Activities/NSSC/USSS_Third_edition/ANSI_USSS_2015.pdf.*

[26] *https://www.cybergrandchallenge.com*

[27] *"Federal Cybersecurity Research and Development Strategic Plan," Executive Office of the President, February 2016, https://www.whitehouse.gov/sites/whitehouse.gov/files/documents/2016_Federal_Cybersecurity_Research_and_Development_Stratgeic_Plan.pdf.*

References

- "AAAI Presidential Panel on Long-Term AI Futures: 2008-2009 Study," *The Association for the Advancement of Artificial Intelligence*, http://www.aaai.org/Organization/presidential-panel.php.
- "Aerospace Forecast Report Fiscal Years 2016 to 2036," *The Federal Aviation Administration*, March 24 016, https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/Unmanned_Aircraft_Systems.pdf.
- "AI Safety Conference in Puerto Rico," *The Future of Life Institute,* October 12, 2015, http://futureoflife.org/2015/10/12/ai-safety-conference-in-puerto-rico.
- "Big Data: Seizing Opportunities, Preserving Values," *Executive Office of the President*, May 2014, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.
- "Data Science for Social Good," *University of Chicago, https://dssg.uchicago.edu/.*
- "Federal Automated Vehicles Policy," *The U.S. Department of Transportation*, September 21 2016, https://www.transportation.gov/AV.
- "Federal Cybersecurity Research and Development Strategic Plan," *Executive Office of the President,* February 2016, https://www.whitehouse.gov/sites/whitehouse.gov/files/documents/2016_Federal_Cybersecurity_Research_and_Development_Stratgeic_Plan.pdf.
- "Secretary Foxx Unveils President Obama's FY17 Budget Proposal of Nearly $4 Billion for Automated Vehicles and Announces DOT Initiatives to Accelerate Vehicle Safety Innovations," *U.S. Department of Transportation*, January 14 2016, https://www.transportation.gov/briefing-room/secretary-foxx-unveils-president- obama%E2%80%99s-fy17-budget-proposal-nearly-4-billion.
- "Winning the Education Future: The Role of ARPA-ED," *The U.S. Department of Education*, March 8 2011, https://www.whitehouse.gov/sites/default/files/microsites/ostp/arpa-ed-factsheet.pdf.
- "World Development Report 2016: Digital Dividends," *The World Bank Group*, 2016, http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf.
- Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané, "Concrete Problems in AI Safety," https://arxiv.org/abs/1606.06565.
- Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias," *ProPublica,* May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
- Stuart Armstrong, KajSotala, Seán S. ÓhÉigeartaigh, "The errors, insights and lessons of famous AI predictions – and what they mean for the future," *Journal of Experimental & Theoretical Artificial Intelligence*, May 20, 2014.
- Michael Ball, Cynthia Barnhart, Martin Dresner, Mark Hansen, Kevin Neels, AmedeoOdoni, Everett Peterson, Lance Sherry, Antonio Trani, Bo Zou, "Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States," *The National Center of Excellence for Aviation Operations Research*, November 2010, http://www.nextor.org/pubs/TDI_Report_Final_11_03_10.pdf.
- Nicholas Bloom, Mark Schankerman, John Van Reene, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81: 1347–1393.doi:10.3982/ECTA9466. Frank Chen, "AI, Deep Learning, and Machine Learning: A Primer," *Andreessen Horowitz*, June 10, 2016, http://a16z.com/2016/06/10/ai-deep-learning-machines.
- Jeffrey L. Caton, "Autonomous Weapons Systems: A Brief Survey of Developmental, Operational, Legal, and Ethical Issues," Strategic Studies Institute, U.S. Army War College, December 2015, http://www.strategicstudiesinstitute.army.mil/pdffiles/PUB1309.pdf.
- Jack Clark, "Artificial Intelligence Has a 'Sea of Dudes' Problem," *Bloomberg*, June 21, 2016, https://www.bloomberg.com/news/articles/2016-06-23/artificial-intelligence-has-a-sea-of-dudes-problem.
- Christianne Corbett and Catherine Hill, "Solving the Equation: The Variables for Women's Success in Engineering and Computing," *The American Association of University Women*, March 2015, http://www.aauw.org/files/2015/03/Solving-the-Equation-report-nsa.pdf.
- Anupam Datta, Shayak Sen, and Yair Zick, "Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems," *Proceedings of 37th IEEE Symposium on Security and Privacy*, 2016.

- Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (New York, New York: Basic Books, 2015).
- Eric Elster, "Surgical Critical Care Initiative: Bringing Precision Medicine to the Critically Ill," presentation at AI for Social Good workshop, Washington, DC, June 7, 2016, http://cra.org/ccc/wp- content/uploads/sites/2/2016/06/Eric-Elster-AI-slides-min.pdf.
- Heinz Erzberger, "The Automated Airspace Concept," prepared for the 4th USA/Europe Air Traffic Management R&D Seminar Dec. 3-7, 2001, Santa Fe, New Mexico, USA, http://www.aviationsystemsdivision.arc.nasa.gov/publications/tactical/erzberger_12_01.pdf.
- Ed Felten and Terah Lyons, "Public Input and Next Steps on the Future of Artificial Intelligence," *Medium*, September 6 2016, https://medium.com/@USCTO/public-input-and-next-steps-on-the-future-of-artificial- intelligence-458b82059fc3.
- J.D. Fletcher, "Digital Tutoring in Information Systems Technology for Veterans: Data Report," *The Institute for Defense Analysis,* September 2014.
- Matt Ford, "The Missing Statistics of Criminal Justice," *The Atlantic,* May 31, 2015,http://www.theatlantic.com/politics/archive/2015/05/what-we-dont-know-about-mass-incarceration/394520/
- Jason Furman, "Is This Time Different? The Opportunities and Challenges of Artificial Intelligence," (presentation, AI Now: The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term, New York, NY, July 7, 2016), Available at https://www.whitehouse.gov/sites/default/files/page/files/20160707_cea_ai_furman.pdf.
- Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, "Explaining and Harnessing Adversarial Examples," http://arxiv.org/pdf/1412.6572.pdf.
- Georg Graetz and Guy Michaels, "Robots at Work," *CEPR Discussion Paper No. DP10477*, March 2015, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2575781.
- Bronwyn H. Hall, Jacques Mairesse, and Pierre Mohnen, "Measuring the Returns to R&D," Chapter prepared for the Handbook of the Economics of Innovation, B. H. Hall and N. Rosenberg (editors), December 10, 2009, https://eml.berkeley.edu/~bhhall/papers/HallMairesseMohnen09_rndsurvey_HEI.pdf.
- Moritz Hardt, "How big data is unfair," *Medium,* September 26 2014, https://medium.com/@mrtz/how-big-data-is- unfair-9aa544d739de.
- Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. "Combining satellite imagery and machine learning to predict poverty." *Science* 353, no. 6301 (2016): 790-794.
- Derryl Jenkins and BijanVasigh, "The Economic Impact of Unmanned Aircraft Systems Integration in the United States," *The Association for Unmanned Vehicle Systems International,* 2013*,* https://higherlogicdownload.s3.amazonaws.com/AUVSI/958c920a-7f9b-4ad2-9807- f9a4e95d1ef1/UploadedImages/New_Economic%20Report%202013%20Full.pdf.
- Charles I. Jones and John C. Williams, "Measuring the Social Returns to R&D," *The Quarterly Journal of Economics* (1998) 113 (4): 1119-1135, doi: 10.1162/003355398555856.
- Thomas Kalil, "A Broader Vision for Government Research," *Issues in Science and Technology*, 2003.
- Garry Kasparov, "The Chess Master and the Computer," *New York Review of Books*, February 11, 2010. http://www.nybooks.com/articles/2010/02/11/the-chess-master-and-the-computer.
- Katharine E. Henry, David N. Hager, Peter J. Pronovost, and Suchi Saria, "A targeted real-time early warning score (TREW Score) for septic shock," *Science Translational Medicine* 7, no. 299 (2015): 299ra122-299ra122.
- Aimee Leslie, Christine Hof, Diego Amorocho, Tanya Berger-Wolf, Jason Holmberg, Chuck Stewart, Stephen G. Dunbar, and Claire Jea,, "The Internet of Turtles," April 12, 2016, https://www.researchgate.net/publication/301202821_The_Internet_of_Turtles.
- Steven Levy, "How Google is Remaking Itself as a Machine Learning First Company," *Backchannel*, June 22, 2016, https://backchannel.com/how-google-is-remaking-itself-as-a-machine-learning-first-company-ada63defcb70.
- John Markoff, "No Sailors Needed: Robot Sailboats Scout the Oceans for Data," *The New York Times*, September 4, 2016.
- Warren S. McCulloch and Walter H. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*, 5:115-133, 1943.
- Vincent Müller and Nick Bostrom, "Future progress in artificial intelligence: A Survey of Expert Opinion," *Fundamental Issues of Artificial Intelligence*, 2014.

- Carrie Mullins, "Retrospective Analysis of Technology Forecasting," *The Tauri Group*, August 13, 2012. Andrew Nusca, "IBM's CEO Thinks Every Digital Business Will Become a Cognitive Computing Business," *Fortune*, June 1 2016.
- Robert W. Poole, Jr., "The Urgent Need to Reform the FAA's Air Traffic Control System," *The Heritage Foundation*, 2007, http://www.heritage.org/research/reports/2007/02/the-urgent-need-to-reform-the-faas-air-traffic- control-system.
- The President's Council of Advisors on Science and Technology, letter to the President, September 2014, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_workforce_edit_report_sept_2014.pdf.
- The President's Council of Advisors on Science and Technology, "Report to the President: Big Data and Privacy: A Technological Perspective," *Executive Office of the President*, May 2014, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.
- Mike Purdy and Paul Daugherty, "Why Artificial Intelligence is the Future of Growth," *Accenture*, 2016, https://www.accenture.com/us-en/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth.pdf.
- Sara Reardon, "Text-mining offers clues to success: US intelligence programme analyses language in patents and papers to identify next big technologies," *Nature* no. 509, 410 (May 22 2014).
- David Robinson and Logan Koepke, "Stuck in a Pattern: Early evidence on 'predictive policing' and civil rights," *Upturn*, August 2016, http://www.stuckinapattern.org.
- Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach (3rdEdition)* (Essex, England: Pearson, 2009).
- Mary Shaw, Prospects for an Engineering Discipline of Software, IEEE Software 7(6), November 1990. Stephen F. Smith, "Smart Infrastructure for Urban Mobility," presentation at AI for Social Good workshop,
- Washington, DC, June 7, 2016,http://cra.org/ccc/wp-content/uploads/sites/2/2016/06/Stephen-Smith-AI-slides.pdf.
- Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, Anna Lee Saxenian,
- Julie Shah, Milind Tambe, and Astro Teller, "Artificial Intelligence and Life in 2030," *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, Stanford University, Stanford, CA, September 2016, http://ai100.stanford.edu/2016-report.
- Charles Twardy, Robin Hanson, Kathryn Laskey, Tod S. Levitt, Brandon Goldfedder, Adam Siegel, Bruce D'Ambrosio, and Daniel Maxwell, "SciCast: Collective Forecasting of Innovation," *Collective Intelligence,* 2014.
- United States Standards Strategy Committee, "United States standards strategy," *New York: American National Standards Institute* (2015), https://share.ansi.org/shared%20documents/Standards%20Activities/NSSC/USSS_Third_edition/ANSI_USSS_2015.pdf.
- Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, Andrew H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," June 18, 2016, https://arxiv.org/pdf/1606.05718v1.pdf