United States Senate
Committee on Commerce, Science, and Transportation
Subcommittee on Communications, Technology, Innovation, and the Internet

Digital Decision-Making:
The Building Blocks of Machine Learning and Artificial Intelligence

Written Testimony of
Edward W. Felten
Robert E. Kahn Professor of Computer Science and Public Affairs, Princeton University

December 12, 2017

Chairman Wicker, Ranking Member Schatz, and members of the Committee, thank you for inviting me to speak today about how best to realize the benefits of artificial intelligence.

Artificial Intelligence (AI) and Machine Learning (ML)

Artificial intelligence (AI) and machine learning (ML) have been studied since at least 1950. There has been an unexpected acceleration in technical progress over the last decade, due to three mutually reinforcing factors: the availability of *big data sets*, which are analyzed by *more powerful algorithms*, enabled by *faster computers*. In recent years, machines have met and surpassed human performance on many cognitive tasks, and some longstanding grand challenge problems in AI have been conquered.

Industry has recognized the rise of AI as a technical shift as important as the arrival of the Internet or mobile computing. Companies around the world have invested heavily in AI research and development, and leaders of major companies have described adoption of machine learning as a bet-the-company opportunity.

The strategic importance of AI/ML to the United States goes beyond its economic impact. These technologies will also profoundly affect the future of security issues such as cybersecurity, intelligence analysis, and military affairs.

Fortunately, the United States is currently the world leader in AI/ML research, development, and applications, in both the corporate and academic spheres. Our national lead is not insurmountable, however. Countries around the world are investing

heavily in AI/ML, so our scientists, engineers, and companies need support in their efforts to maintain American leadership.

The Nature of AI/ML Today

The history of AI teaches some important lessons that are useful in considering policy choices.

AI is not a single thing—it is different solutions for different tasks. The greatest progress has been in "narrow AI," which applies a toolbox of specific technical approaches to craft a solution specific to one application or a narrow range of applications. There has been less progress on "general AI," which strives to create a single, all-purpose artificial brain that could address any cognitive challenge and would be as adaptive and flexible as human intelligence. Indeed, there is no clear technical path for achieving general AI, so it appears that for at least the next decade the policy focus should be on the implications of narrow AI.

In a world of narrow AI, there will not be a single moment at which machines surpass human intelligence. Instead, machines may surpass human performance at different times for different cognitive tasks; and humans might retain an advantage on some cognitive tasks for a long time. Even if machines surpass humans in the lab for some task, additional time and effort would need to be invested to translate that advance into practical deployment in the economy.

Successful AI does not think like a human — if it is an intelligence, it is an alien intelligence. Because AI solutions are task-specific and do not directly mimic the human brain, AI systems tend to "think" differently than people. Even when successful, AI systems tend to exhibit a different problem-solving style than humans do. An AI system might handle some extremely complex situations well while failing on cases that seem easy to us. The profound difference between human thinking and AI operation could make human-AI teaming valuable, if the strengths of people and machines can complement each other. At the same time, these differences create challenges in human-AI teaming because the teammates can have trouble understanding each other and predicting their teammates' behavior.

On many cognitive tasks, more engineering effort or more data translates into better AI performance. Many AI systems learn from data. Such systems can be improved by re-engineering them to learn more from the available data or by increasing the amount of data available for training. Either way, devoting more effort to engineering and operating an AI system can improve its performance. Machines are generally worse than humans at learning from experience, but a machine with a very large data set has much more "experience" from which to learn. Using the narrow AI approaches that have been

successful so far, expert AI developers must invest significant effort in applying AI to each specific task.

Benefits of AI/ML

AI is already creating huge benefits, and its potential will only grow as the technology advances further.

For example, AI is a key enabler of precision medicine. AI systems can learn from data about a great many patients, their treatments, and outcomes to enable better choices about how to personalize treatment for the particular needs, history, and genetic makeup of each future patient.

AI is also enabling self-driving cars, which will eventually be much safer than human drivers, saving thousands of American lives every year. Self-driving vehicles will improve mobility for elderly and disabled people who cannot drive and will lower the cost and increase the convenience of transporting people and goods.

Given the tremendous benefits of AI in these and other areas and the likelihood that the technology will be developed elsewhere even if the United States does not lead in AI, it would be counterproductive to try to stop or substantially slow the development and use of AI. We should not ask the industry and researchers to slam on the brakes. Instead, we should ask them to use the steering wheel to guide the direction of AI development in ways that protect safety, fairness, and accountability.

Policies to Support AI Progress

America's leadership in AI has been driven by three factors: our companies, our researchers, and our talented workforce.

American companies recognized the potential of AI early on and have been investing heavily in AI and moving aggressively to hire top talent. This is the area in which our national leadership in AI seems safest, at least in the short run. In the longer run, however, industry must be able to work with world-leading American researchers and workforce to sustain its advantage.

Our lead in research and development is less secure. Federal funding for AI research and development has been relatively flat, even as the importance of the field has dramatically increased. Aggressive hiring by industry has thinned the ranks of the academic researchers and teachers who are needed to train the next generation of leaders. Although industry has carried out and supported a great deal of research, it

cannot and does not cover the full spectrum. The public research community plays an important role in basic research, in research areas such as safety and accountability, and in training young researchers, so investments and policies to support and grow that community are a key enabler of continued American leadership.

The foundations of the future workforce are laid in our K-12 schools. Policies to enhance access to high-quality education for all American children, especially in computing, can grow the number of students who enter higher education eager and able to pursue studies in technical fields such as AI.

The American AI workforce has also been boosted immeasurably over the years by the attractiveness of our universities and industry to the most talented people from around the world. America has been a magnet for talent in AI and other technical fields, and that must continue if we are to retain our leadership. Policies to ensure that America remains an attractive place for foreign-born experts to live, study, work, and start companies are among the most important steps for the future health of our AI enterprise.

Risks and Challenges of AI/ML

The benefits of AI are tempered by some risks and challenges: AI systems may pose safety risks; they may introduce inadvertent bias into decisions; and they may suffer from the kinds of unforeseen consequences brought on by any novel, complex technology. These are very serious issues that require attention from policymakers, AI developers, and researchers.

Much of the criticism of AI/ML systems centers on the risk that adoption of AI/ML will lead inadvertently to biased decisions. There are several ways this could happen. If a system is trained to mimic past human decisions, and those decisions were biased, the system is likely to replicate that bias. If the data used to train a system is derived from one group of people more than another, the result may serve the overrepresented group to the detriment of the underrepresented group. Even with ideal data, statistical artifacts can advantage larger groups to the detriment of smaller ones. Real-world examples of these sorts of biases are well-documented.

The solution is not to stop pursuing AI, but rather to take steps to prevent and mitigate bias. Practitioners should work to improve their data, to ensure that datasets are representative of the population and do not rely on past biased decisions. They should also improve their algorithms by developing and using AI systems that are more resistant to bias, so that even if flaws remain in the data, the system can produce results that are more fair. In both areas, data improvement and algorithm improvement, the

research community is producing promising early results that will improve the anti-bias toolkit available to practitioners. A robust national AI research effort should include studies of algorithmic bias and how to mitigate it.

In considering the risks of bias and accountability in AI, it is important to remember that in most cases the alternative to relying on AI is to rely on human decisions, which are themselves at risk of error, bias, and lack of accountability. In the long run, we will likely rely much more on algorithms to guide decisions, while retaining the human role of determining which goals and criteria should guide each decision.

Accountability, Transparency, and Explainability

The importance of the decisions now made or assisted by AI/ML systems requires that the systems and their operators are accountable to managers, overseers, regulators, and the public. Yet accountability has proven difficult at times due to the complexity of AI systems and current limitations in the theory underlying AI. Improving practical accountability should be an important goal for the AI community.

Transparency is one approach to improve accountability. Disclosing details of a system's code and data can enable outside analysts to study the system and evaluate its behavior and how well the system meets the goals and criteria it is supposed to achieve. Full transparency is often not possible, however. For example, a system's code might include valuable trade secrets that justify withholding aspects of its design, or the data might contain private information about customers or employees that cannot be disclosed.

Even where transparency is possible, it is far from perfect as an accountability mechanism. Outside analysts may have limited practical ability to understand or test a system that is highly complex and meant to operate at very large scale. Indeed, even the designers of a system may struggle to understand the nuances of its operation. Computer science theory says that examining a system beforehand cannot hope to reveal everything the system will do when it is exposed to real-world inputs. So transparency, though useful, is far from a complete solution to the accountability problem.

Another approach to accountability is inspired by the field of safety engineering. The approach is to state clearly which safety, fairness, or compliance properties a system is designed to provide, as well as the operating conditions under which the system is designed to provide those properties. This is backed up with detailed evidence that the system will have the claimed properties, based on a combination of design reviews, laboratory testing, automated analysis tools, and safety monitoring facilities in place during operation. Rather than revealing everything about how the system works, this

approach focuses on specific safety, fairness, or compliance requirements and allows system developers to use the full range of technical tools that exist for ensuring reliable behavior, including the tools that the system developers will already be using internally for quality control.

Much needs to be done to make this approach feasible for routine use. Research can develop and test different approaches to proving behavioral properties of systems. Professionals can convene to develop and pilot best practices and standards. The overarching challenge is to understand how to relate the technical process of engineering for reliable operation to the administrative processes of management, oversight, and compliance.

Regulation and the Role of Government Agencies

There is no need to create special regulations for AI. Where AI is used in sectors or activities that are already regulated, the existing regulations are already protecting the public and regulators need only consider whether and how to adjust the existing regulations to account for changes in practices due to AI.

For example, the Department of Transportation (DOT) and National Highway Traffic Safety Administration (NHTSA) have taken useful steps, under the previous and current Administrations, to clarify how existing safety regulations apply to self-driving vehicles and how Federal safety regulations relate to state vehicle laws. These changes will serve to smooth the adoption of self-driving vehicles which, once they are mature and widely adopted, will save many thousands of lives.

Similarly, the Federal Aviation Administration (FAA) has been striving to adapt aviation regulations to enable safe, commercial use of unmanned aerial systems (UAS, or "drones"), which have benefits in many sectors, such as agriculture. The FAA has taken some steps to increase the flexibility to use UAS commercially, but the interagency process on UAS has been moving slowly. Agencies should be urged to work with the FAA to advance this important process.

Government agencies have important roles to play beyond regulation. For example, the National Institute of Standards and Technology (NIST) and the Department of Commerce can contribute by setting technical standards, codifying best practices in consultation with the private sector, and convening multi-stakeholder discussions, much as they have done in the area of cybersecurity.

All agencies should consider how they might use AI to better accomplish their missions and serve the American people. AI can reduce costs, increase efficiency, and help

agencies better target their use of taxpayer dollars and other limited resources. The National Science and Technology Council's subcommittee on Machine Learning and AI can serve as a focal point for interagency coordination and sharing of ideas and best practices.

With good policy choices and the continued hard work and investment of American companies, researchers, and workers, AI can improve the health and welfare of Americans, boost productivity and economic growth, and make us more secure. Americans currently lead the world in AI. We should not step on the brakes. Instead, we should reach for the accelerator and the steering wheel.

Thank you for the opportunity to testify. I look forward to answering any questions.