United States Government Accountability Office

Report to Congressional Requesters

March 2018

ARTIFICIAL INTELLIGENCE

Emerging Opportunities, Challenges, and Implications

Accessible Version

GAO-18-142SP

GAO

HIGHLIGHTS OF A FORUM

Report to the Committee on Science, Space, and Technology, House of Representatives

March 2018

TECHNOLOGY ASSESSMENT

Artificial Intelligence

Emerging Opportunities, Challenges, and Implications

Why GAO Convened This Forum

Artificial intelligence (AI) holds substantial promise for improving human life and economic competitiveness in a variety of ways and for helping solve some of society's most pressing challenges. At the same time, according to experts, AI poses new risks and could displace workers and widen socioeconomic inequality. To gain a better understanding of the emerging opportunities, challenges, and implications resulting from developments in AI, the Comptroller General of the United States convened the Forum on Artificial Intelligence, which was held on July 6 and 7, 2017, in Washington, D.C.

At the forum, participants from industry, government, academia, and nonprofit organizations considered the potential implications of AI developments in four sectors—cybersecurity, automated vehicles, criminal justice, and financial services. Participants considered policy implications of broadening AI use in the economy and society, as well as associated opportunities, challenges, and areas in need of more research. Following the forum, participants were given the opportunity to review a summary of forum discussions and a draft of this report. Additionally, a draft of this report was reviewed independently by two experts who did not attend the forum. The viewpoints expressed by individuals in the report do not necessarily represent the views of all participants, their organizations, or

View GAO-18-142SP. For more information, contact Timothy M. Persons at (202) 512-6412, personst@gao.gov or James-Christian Blockwood at (202) 512-2639, blockwoodjc@gao.gov.

What the Participants Discussed

Forum participants noted a range of opportunities and challenges related to artificial intelligence (AI), as well as areas needed for future research and for consideration by policymakers. Regarding opportunities, investment in automation through AI technologies could lead to improvements in productivity and economic outcomes, similar to that experienced during previous periods of automation, according to a forum participant. In cybersecurity, AI automated systems and algorithms can help identify and patch vulnerabilities and defend against attacks. Automotive and technology firms use AI tools in the pursuit of automated cars, trucks, and aerial drones. In criminal justice, algorithms are automating portions of analytical work to provide input to human decision makers in the areas of predictive policing, face recognition, and risk assessments. Many financial services firms use AI tools in areas like customer service operations, wealth management, consumer risk profiling, and internal controls.

Selected Questions Regarding the Use of Artificial Intelligence (AI) in Four High-Consequence Sectors

	Selected Questions
Cybersecurity AI applications face threats from cybersecurity attacks, but AI also may be used as a tool for detecting and defending against attacks.	▸ How can autonomous systems be made secure, without stifling innovation? ▸ How useful is a risk-based approach to determining if machine-learning algorithms adhere to legal requirements or ethical norms?
Automated Vehicles Automated vehicles hold promise for increasing driving safety and providing enhanced mobility, but pose challenges for assuring increased safety.	▸ What is the appropriate regulatory framework for automated vehicle safety assurance? ▸ What are the roles of federal, state, and local governments in infrastructure adaptation and addressing issues of liability and enforcement?
Criminal Justice The use of AI in criminal justice may improve the allocation of law enforcement resources and has the potential to reduce crime and jail populations, but also raises concerns about privacy and civil rights violations.	▸ What are the options for assessing accuracy and the potential for bias in AI data and algorithms? ▸ What are solutions for safeguarding privacy in the collection and use of personal information by AI systems?
Financial Services The use of AI in financial services could improve client services and enhance surveillance monitoring, but also poses challenges to ensuring fair lending, attracting and retaining staff with requisite skills, and maintaining hardware and software.	▸ What are the mechanisms to address ethical considerations, tradeoffs, and protections? ▸ How can regulatory sandboxes be used to test new AI products, services, and business models?

Source: GAO Forum on Artificial Intelligence. | GAO-18-142SP

Forum participants also highlighted a number of challenges related to AI. For example, if the data used by AI are biased or become corrupted by hackers, the results could be biased or cause harm. The collection and sharing of data needed to train AI systems, a lack of access to computing resources, and adequate human capital are also challenges facing the development of AI. Furthermore, the widespread adoption of AI raises questions about the adequacy of current laws and regulations. Finally, participants noted the need to develop and adopt an appropriate ethical framework to govern the use of AI in research, as well as explore factors that govern how quickly society will accept AI systems in their daily lives.

After considering the benefits and challenges of AI, forum participants highlighted several policy issues they believe require further attention. In particular, forum participants emphasized the need for policymakers to explore ways to (1) incentivize data sharing, such as providing mechanisms for sharing sensitive information while protecting the public and manufacturers; (2) improve safety and security (e.g., by creating a framework that ensures that the costs and liabilities of providing safety and security are appropriately shared between manufacturers and users); (3) update the regulatory approach that will affect AI (e.g., by leveraging technology to improve and reduce the burden of regulation, while assessing whether desired outcomes are being achieved); and (4) assess acceptable levels of risk and ethical considerations (e.g., by providing mechanisms for assessing tradeoffs and benchmarking the performance of AI systems). As policymakers explore these and other implications, they will be confronted with fundamental tradeoffs, according to forum participants. As such, participants highlighted several areas related to AI they believe warrant further research, including (1) establishing regulatory sandboxes (i.e., experimental safe havens where AI products can be tested); (2) developing high-quality labeled data (i.e., data organized, or labeled, in a manner to facilitate their use with AI to produce more accurate outcomes); (3) understanding the implications of AI on training and education for jobs of the future; and (4) exploring computational ethics and explainable AI, whereby systems can reason without being told explicitly what to do and inspect why they did something, making adjustments for the future.

Implications of Artificial Intelligence (AI) for Policy and Research

Source: GAO Forum on Artificial Intelligence. | GAO-18-142SP

Contents

Table

Figures

Abbreviations

AI	artificial intelligence
AV	automated vehicle
BSA/AML	Bank Secrecy Act and related anti-money laundering
CIP	Customer Identification Program
DARPA	Defense Advanced Research Projects Agency
DOT	Department of Transportation
FBI	Federal Bureau of Investigation
FINRA	Financial Industry Regulatory Authority
FRT	face recognition technology
FST	Forensic Statistical Tool
HAV	highly-automated vehicle
IoT	Internet of Things
NHTSA	National Highway Traffic Safety Administration
NIST	National Institute of Standards and Technology
PERVADE	Pervasive Data Ethics for Computational Research
PIA	Privacy Impact Assessment
SEC	Securities and Exchange Commission
STEM	science, technology, engineering, and math

U.S. GOVERNMENT ACCOUNTABILITY OFFICE
441 G St. N.W.
Washington, DC 20548

Comptroller General
of the United States

March 28, 2018

The Honorable Lamar Smith
Chairman
The Honorable Eddie Bernice Johnson
Ranking Member
Committee on Science, Space, and Technology
House of Representatives

This report represents the results of a Comptroller General forum on recent developments in the area of artificial intelligence (AI)—and key implications regarding potential benefits, challenges to realizing these benefits, and resulting policy implications and research priorities.

The forum agenda (contained in app. I) provided structured forum presentations and discussions to allow each participant to make at least one presentation on his or her area of expertise as it relates to AI and to comment openly and interact with others on any issue. With the assistance of the National Academies, we selected forum participants (listed in app. II) to represent (1) a range of backgrounds, covering academia, government, industry, and nonprofit organizations, and (2) experience and knowledge about the application of AI across varied areas, including cybersecurity, automated vehicles, criminal justice, and financial services, as well as the economic, ethical, legal, and social implications of ongoing and anticipated developments in AI.

Developed under the leadership of our Chief Scientist and Managing Director of Strategic Planning and External Liaison, this report summarizes the ideas and themes that emerged in the forum's discussions and were further developed based on a review of the literature and interviews with subject-matter experts. The viewpoints expressed by individuals in the report do not necessarily represent the views of all participants, their organizations, or GAO.

Appendixes III to V of the report

- list the experts we consulted, in addition to forum participants (app. III);
- reproduce four forward-looking profiles that describe applications of AI in cybersecurity, automated vehicles, criminal justice, and financial services (app. IV), which we sent to forum participants in advance of the July 2017 meeting; and

- explain the scope and methodology of our work (app. V).

Questions may be addressed to Timothy M. Persons, Chief Scientist, at (202) 512-6412 or personst@gao.gov, or James-Christian Blockwood, Managing Director of Strategic Planning and External Liaison, at (202) 512-2639 or blockwoodjc@gao.gov. Contact points for our Office of Congressional Relations and Office of Public Affairs appear on the last page. Major GAO contributors to this report are listed in appendix VI.

Gene L. Dodaro
Comptroller General of the United States

Foreword

According to experts, artificial intelligence (AI) holds substantial promise for improving human life and economic competitiveness in a variety of capacities and for helping to solve some of society's most pressing challenges. At the same time, however, AI poses new risks and has the potential to displace workers in some sectors, requires new skills and adaptability to changing workforce needs, and could exacerbate socioeconomic inequality. In recent years, significant public and private-sector resources have been invested in the development of AI products and services. Accompanying a high demand for people with expertise in AI, investment of human capital in learning about AI techniques and developing skills in AI has been significant.

To gain a better understanding of the broad promise, consequences, and policy considerations resulting from developments in AI, the Comptroller General of the United States convened the Forum on Artificial Intelligence, held on July 6 and 7, 2017, with the assistance of the National Academies, at the Keck Center in Washington, D.C.

In preparation for the forum, we selected four sectors—cybersecurity, automated vehicles, criminal justice and financial services—for more detailed study. We selected these areas to provide variety in the purposes for which AI may be used, and variety in the potential benefits and consequences. Although the consequences and benefits vary across the sectors, each sector can be seen as high-consequence.

Cybersecurity, for example, is a cross-cutting sector, with AI applications across all domains at risk from threats to cybersecurity. At the same time, AI may be used as a tool for countering threats to cybersecurity. Automated vehicles hold the promise for decreasing traffic deaths and providing enhanced mobility but also pose challenges in addressing the explainability of AI decisions. In criminal justice, AI has the potential to better assess risks for recidivism and reduce costs associated with both crime and incarceration, but its use also raises concerns about privacy and civil rights violations. In financial services, benefits of using AI tools could include improvements related to wealth-management activities, access to investment advice, and customer service, but these tools also pose questions around data security and fair lending.

Immediately following this foreword are brief snapshots of the status of AI in each of these sectors, as well as a snapshot illustrating the policy and

research implications of emerging developments in AI. For more details about AI in each of the sectors we examined, see appendix IV.

The forum highlighted that even if capabilities of AI technologies stopped advancing today, the transformation resulting from today's AI will have far-reaching effects across a wide spectrum of human activity, society, and the economy. For policymakers, industry leaders, and citizens, attention will be needed in the United States to ensure the benefits of AI are maximized while simultaneously mitigating the potential harm that may come from any transformation resulting from AI.

Based on forum presentations and discussion, interviews with subject-matter experts, and review of relevant literature, this report provides an introduction and overview of developments in the field of AI, focusing on the challenges, opportunities, and implications of these developments for policy making and research. This report helps clarify the prospects for the near-term future of AI and identifies areas where changes in policy and research may be needed.

All of us at GAO who have worked on this report are grateful to the forum participants (see app. II) and reviewers and interviewees (see app. III) who contributed to our work. We also acknowledge the invaluable support provided by the National Academies.

Timothy M. Persons, Ph.D.
Chief Scientist
U.S. Government Accountability Office

James-Christian B. Blockwood
Managing Director
Strategic Planning and External Liaison
U.S. Government Accountability Office

Snapshot

AI in Cybersecurity

This cybersecurity snapshot summarizes the issues participants discussed at the 2017 Comptroller General Forum on Artificial Intelligence.

Key Policy Areas for Consideration

- Explore options to incentivize both innovation and security in autonomous systems
- Assess the usefulness of a risk-based approach to determining if machine-learning algorithms adhere to legal requirements or ethical norms

Applications

Automated systems and advanced algorithms can help cybersecurity professionals in a variety of ways. For example, these systems can help reduce the time and effort it takes to perform key cybersecurity tasks, such as:

- identifying vulnerabilities,
- patching vulnerabilities,
- detecting attacks, and
- defending against active attacks.

Expert systems remain the most common systems used for cybersecurity, but newer approaches incorporate a combination of machine learning with human expertise to build a predictive model of cyber-attacks. As shown in the figure below, an AI system may use both unsupervised and supervised machine learning to conduct analysis of potential threats.

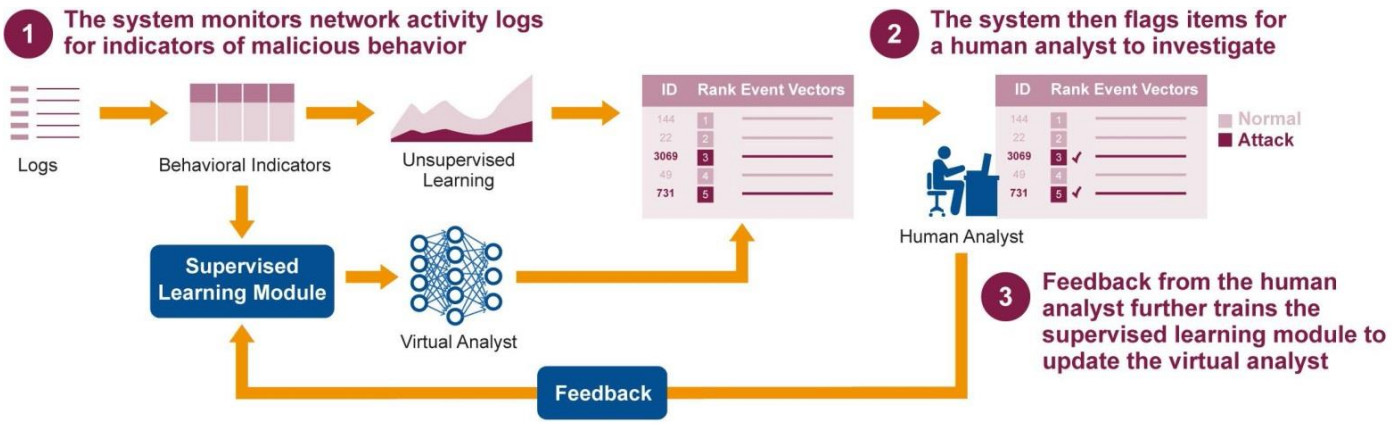
Selected Benefits

- Reducing human workload
- Increasing accuracy in detection of cyber threats
- Processing large amounts of data in short time spans

Selected Challenges

- Depending on human intervention for ongoing operation and periodic maintenance, including the identification and/or verification of attacks
- Addressing ethical and legal concerns of how AI uses personal data
- Addressing AI's own vulnerability to cyber attacks that attempt to maliciously manipulate the AI system's actions
- Countering automated or AI-based attacks

Illustration of machine learning with human feedback for cybersecurity



Source: GAO, adapted from video, Veeramachaneni, Arnaldo et al., *AI2: Training a Big Data Machine to Defend* (https://www.youtube.com/watch?v=b6Hf1O_vpwQ). | GAO-18-142SP

Snapshot

AI in Automated Vehicles

This automated vehicles snapshot summarizes the issues participants discussed at the 2017 Comptroller General Forum on Artificial Intelligence.

Key Policy Areas for Consideration

- Assess the regulatory framework for vehicle safety assurance
- Gain a better understanding of workforce implications of AI, including potential displacement of drivers
- Evaluate the roles of federal, state, and local governments in areas such as infrastructure adaptation, liability, and enforcement
- Assess mechanisms to facilitate data collection, sharing, and collaboration across sectors, including government, industry and others

Applications

Automotive and technology firms use AI tools in the pursuit of automated vehicles, such as automated cars and trucks. While a host of advanced technologies collectively enable vehicle automation, AI technologies provide the "brain" to assess a situation, make a plan, and execute vehicle control decisions. Various automated technologies are available on vehicles today, but fully self-driving automated vehicles that are highly-dependent on AI may be available within the next decade

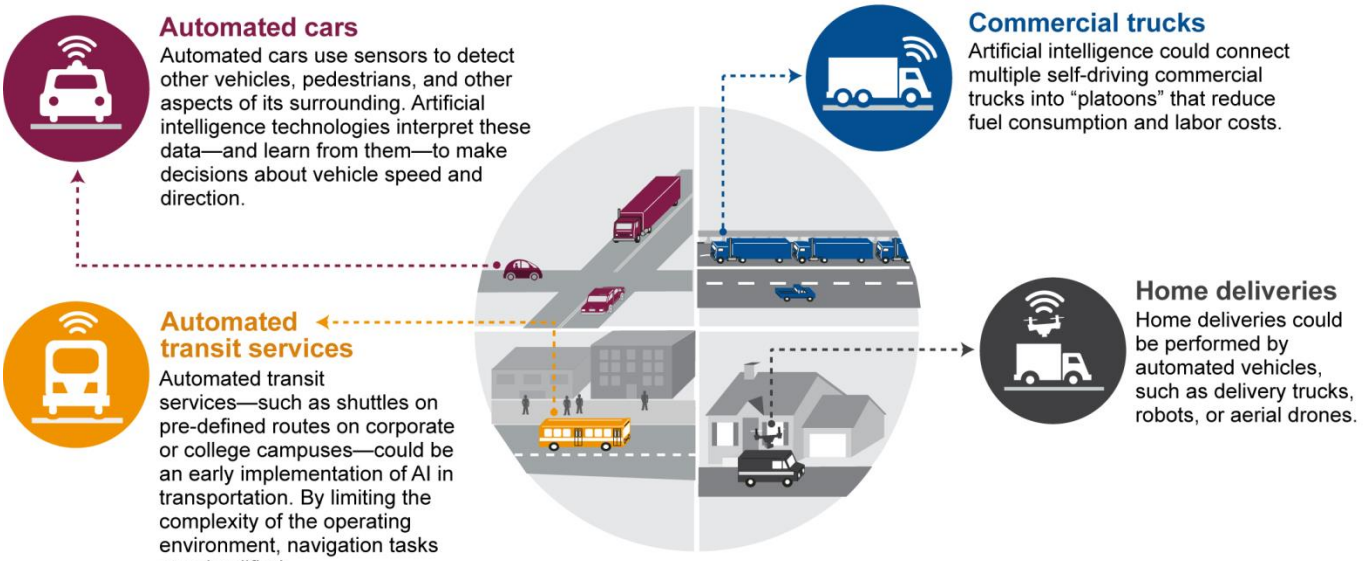
Selected Benefits

- Increasing driving safety
- Improving mobility and access to transportation, including for disadvantaged groups
- Decreasing cost of goods delivery

Selected Challenges

- Managing vehicle safety testing and assurance
- Addressing explainability of AI decisions
- Updating the approach of law enforcement and crash investigations for cases involving AI
- Updating laws and regulations that do not accommodate AI technology

Automated vehicles that rely on AI could be deployed in a variety of applications



Source: Published reports and GAO interviews. | GAO-18-142SP

Snapshot

AI in Criminal Justice

This criminal justice snapshot summarizes the issues participants discussed at the 2017 Comptroller General Forum on Artificial Intelligence.

Key Policy Areas for Consideration

- Assess challenges to ensuring transparency in AI algorithms
- Explore options for assessing accuracy and the potential for bias in AI data and algorithms
- Consider solutions for safeguarding privacy in the collection and use of personal information by AI systems

Applications

There are three early-stage applications of AI in the criminal justice arena. In each application, algorithms are automating portions of analytical work to help provide input to human decision makers. These applications:

- predict where crime is likely to occur to improve allocation of law enforcement resources,
- assist with identification of suspects through face recognition technology, and,
- assess the risk for recidivism when determining how long to sentence individuals convicted of crime.

These three applications are in use across local, state and federal levels of government and across agencies, including law enforcement and the judiciary. Further, as the figure below illustrates, AI can be applied at various stages in the criminal justice process, from before arrest and booking to sentencing and correctional supervision.

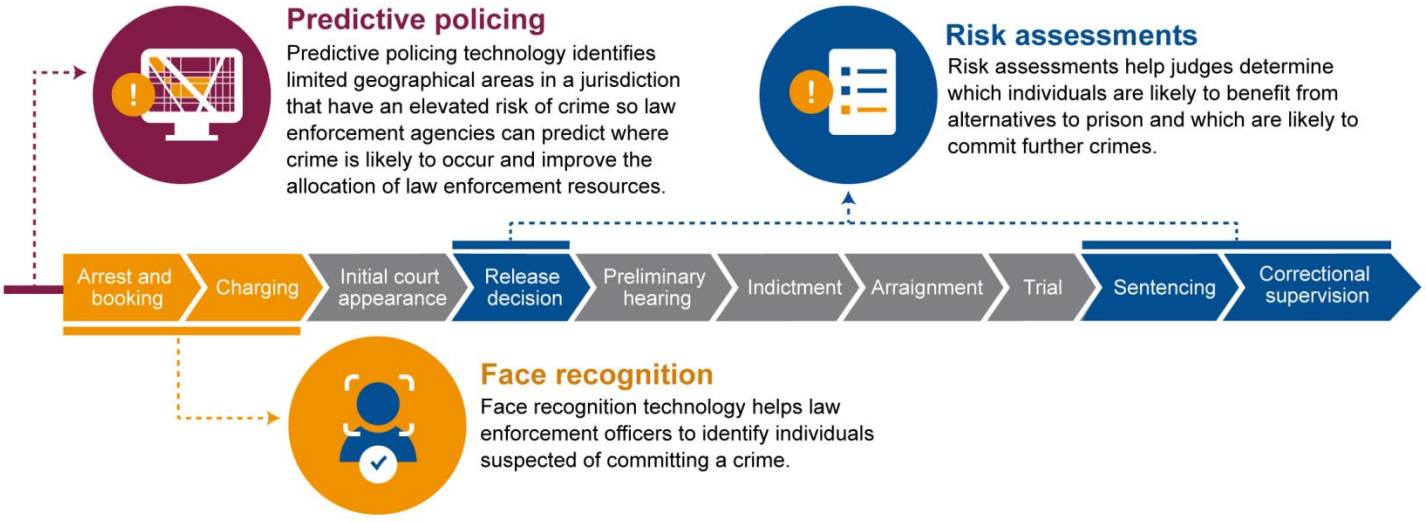
Selected Benefits

- Improving use of limited resources and available data
- Improving identification of criminal suspects
- Potentially reducing crime and jail populations

Selected Challenges

- Addressing fairness and demographic biases
- Ensuring transparency and accuracy of machine learning
- Addressing privacy and civil rights concerns

Use of Artificial Intelligence in Criminal Justice



Source: Published reports and GAO interviews. | GAO-18-142SP

Snapshot

AI in Financial Services

This financial services snapshot summarizes the issues participants discussed at the 2017 Comptroller General Forum on Artificial Intelligence.

Key Policy Areas for Consideration

- Assess options to ensure the safety and security in the sharing of data
- Evaluate mechanisms to address ethical considerations, tradeoffs, and protections
- Assess the impacts of AI on employment, training, and education
- Explore alternative regulatory technology (regtech) approaches and experimental sandboxes

Applications

Many financial services firms (including those in the banking, securities, and insurance industries) have begun to integrate AI tools into their computer systems and operations. Some of these AI tools are helping to augment applications that support functions such as:

- customer service operations (automating call center functions, on-line chatbots, etc.),
- client wealth management (advising financial services professionals or customers directly),
- consumer risk profiling (decisions and rates tied to insurability, lending, etc.), and
- internal controls (monitoring transactions for potential fraud, regulatory compliance, etc.).

The figure below depicts the use of information incorporating AI-based tools that highlight key characteristics of a client, offering insights to the financial services professional about the optional investment strategy to pursue and highlighting high priority clients.

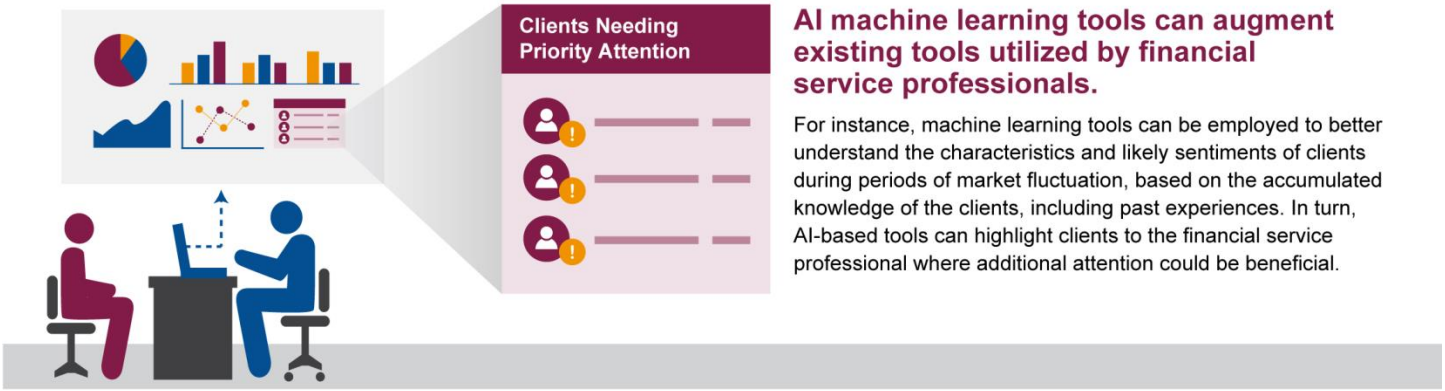
Selected Benefits

- Offering better service and investment strategies to an organization's clients
- Achieving higher productivity in a cost-effective manner
- Enhancing surveillance monitoring by industry entities as well as regulators to better prevent improper market conduct

Selected Challenges

- Ensuring AI-based credit decisions do not violate fair lending laws
- Obtaining complete and appropriately formatted data
- Attracting and retaining staff with requisite data science and machine learning skills
- Maintaining hardware and software

Illustration of machine learning tools used by financial services professionals



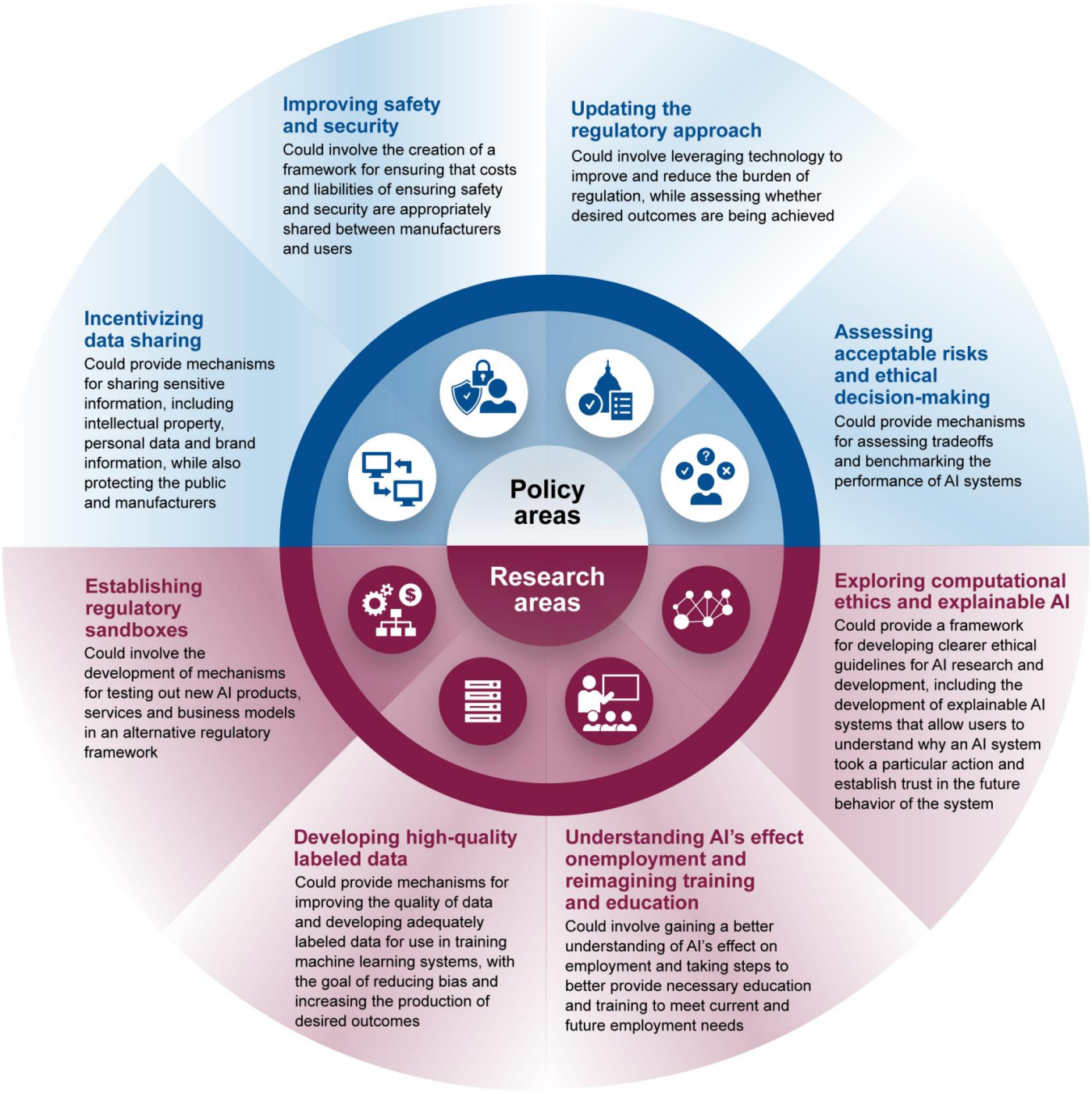
AI machine learning tools can augment existing tools utilized by financial service professionals.

For instance, machine learning tools can be employed to better understand the characteristics and likely sentiments of clients during periods of market fluctuation, based on the accumulated knowledge of the clients, including past experiences. In turn, AI-based tools can highlight clients to the financial service professional where additional attention could be beneficial.

Source: Published reports and GAO interviews. | GAO-18-142SP

Snapshot

Implications of Artificial Intelligence for Policy and Research



Source: GAO Forum on Artificial Intelligence. | GAO-18-142SP

Introduction

Overview of the Evolution and Characteristics of AI

The field of AI was founded on the idea that machines could be used to simulate human intelligence. AI has been defined in a variety of ways, and researchers have also distinguished between narrow and general AI. Narrow AI refers to applications that provide domain-specific expertise or task completion, including today's robotics and applications such as tax-preparation software and on-line "chatbots," which answer questions specific to a product or service. General AI refers to an AI system that exhibits intelligence comparable to that of a human, or beyond, across the range of contexts in which a human might interact. Fictional examples of general AI include the computer H.A.L., from the film *2001: A Space Odyssey*, and Lieutenant Commander Data, from the *Star Trek: The Next Generation* television series.

One conceptualization describes AI as having three distinct waves of development.[1] The first wave of AI comprises expert or rules-based systems, whereby a computer is programmed based on expert knowledge or criteria and produces outputs consistent with this programming. Software programs that do tax preparation or logistics scheduling are examples of expert systems. In second-wave AI systems, statistical or machine learning begins with data and infers rules or decision procedures that accurately predict specified outcomes—on the basis of the data provided. Recent AI advances based on machine learning have been enabled by the increased availability of large data sets and increased computing power. Self-driving automated vehicles are examples of machine-learning systems. Third-wave AI systems—the future of AI, according to one expert—are conceived as combining the strengths of first- and second-wave AI systems, while also being capable of contextual sophistication, abstraction, and explanation.

As reflected in expectations for third wave systems, experts envision that the future of AI will involve contextual awareness, adaptation, and explainable machine-learning algorithms. There has been considerable progress in developing AI that outperforms humans in specific domains,

[1]John Launchbury, *A DARPA Perspective on Artificial Intelligence*, October 3, 2016.

but some observers believe that general AI is unlikely to be achieved for decades. In the interim, it is important to consider what types of policy actions and research may be useful for maximizing benefits and addressing challenges associated with advancements and use of AI.

Overview of the Benefits and Challenges of AI

AI brings with it a number of benefits, according to forum participants. Forum participants said AI may

- improve economic outcomes and increase levels of productivity,
- improve or augment human decision making, and
- provide insights and potential solutions for complex and pressing social and economic problems.

Participants also stressed that there are likely to be benefits related to AI that cannot yet be predicted or may even be hard to imagine.

At the same time, forum participants highlighted a number of challenges that will need to be confronted if the full benefits of AI are to be realized. Such challenges include

- collecting and sharing reliable and high-quality data that are needed to train AI,
- accessing adequate computing resources and requisite human capital,
- ensuring laws and regulations governing systems enabled by AI are adequate and that the application of AI does not infringe on civil liberties, and
- developing an ethical framework to govern the use of AI and ensuring the actions and decisions of AI systems can be adequately explained and accepted by those who interact with such systems.

Overview of the Policy and Research Implications of AI

After considering the benefits and challenges associated with AI, forum participants highlighted several policy issues they believe require further attention. In particular, forum participants emphasized the need for policy-making actions that encompass AI systems to consider:

- incentivizing data sharing,

- improving safety and security,
- updating the regulatory approach, and
- assessing acceptable risks and ethical decisions.

In assessing an acceptable level of risk for an AI system, policymakers and other relevant stakeholders, including those in industry, the social sciences and humanities (e.g., philosophers) will be confronted with fundamental trade-offs. For example, there will likely be difficult questions about what benchmarks the performance of an AI system should be measured against. These questions include whether AI systems will simply be expected to perform at least as well as humans, or whether there will be a higher standard that AI systems exhibit perfect and error-free performance at all times.

Similarly, forum participants also highlighted several areas of research related to AI that they believe warrant further attention. These research efforts include

- establishing experimental regulatory sandboxes,[2]
- developing high-quality labeled data,
- understanding the implications of AI on employment as well as training and education, and
- exploring computational ethics and explainable AI.[3]

Objectives

In this assessment, we summarize the views of experts who participated in the Comptroller General's Forum on Artificial Intelligence. We have also supplemented these views with information from interviews with

[2]A regulatory sandbox, as defined by the United Kingdom Financial Conduct Authority, is "a 'safe space' in which businesses can test innovative products, services, business models and delivery mechanisms without immediately incurring all the normal regulatory consequences of engaging in the activity in question." Financial Conduct Authority, *Regulatory sandbox*, November 2015.

[3]Explainable AI, as conceived by the Defense Advanced Research Projects Agency (DARPA), are "new machine learning systems that [will] have the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future." DARPA, *Explainable Artificial Intelligence (XAI)*, accessed December 20, 2017, https://www.darpa.mil/program/explainable-artificial-intelligence.

other subject-matter experts and relevant literature. More specifically, we explore the following topics and what they mean for the nation:

- How has AI evolved over time, and what are important trends and developments in the relatively near future?
- According to experts, what are the opportunities and future promise, as well as the principal challenges and risks, of AI?
- According to experts, what are the policy implications and research priorities resulting from advances in AI?

Scope and Methodology

We selected forum participants from academia, industry, government, and nonprofit organizations and convened an expert forum with the assistance of the National Academies. The forum was held July 6-7, 2017, with 21 expert participants.[4] The forum agenda (see app. I) was structured in sessions that addressed specific topics, with each session concluding with open discussion among all participants. Each participant gave at least one presentation.

In advance of the forum, we prepared a background reading package, based on interviews with experts and relevant literature, which we distributed to forum participants. The reading package featured a brief overview of the evolution and characteristics of AI, an introduction to the social and economic significance of AI, and four profiles of AI developments in cybersecurity, automated vehicles, criminal justice, and financial services. These profiles are included in appendix IV of this report.

Following the forum, we sent participants an outline of the forum presentations and discussion for their review and comment. This outline was based on a written transcript of forum proceedings and presentations delivered as part of the forum. We also reviewed the forum transcript and interacted with participants afterwards as needed to further develop and better understand points that were raised at the forum. We incorporated feedback from participant comments on the outline and from interactions after the forum, as appropriate. In our report, the use of the term “forum

[4]App. II contains the list of forum participants.

participants" means that more than one participant contributed to the point being made.

Before publication and consistent with our quality-assurance framework, we provided the forum participants with a draft of our report, and incorporated their feedback on that draft as appropriate. As an additional measure of quality assurance, two additional external experts (one with expertise in the technical aspects of AI and another with expertise in the economic implications of AI) who had not participated in the forum reviewed a draft of this report and provided comments that we incorporated as appropriate. Please see appendix V for a more detailed description of our scope and methodology for this report.

We conducted our work from January 2017 through March 2018, in accordance with all sections of GAO's Quality Assurance Framework that are relevant to technology assessments. The framework requires that we plan and perform the engagement to obtain sufficient and appropriate evidence to meet our stated objectives and to discuss any limitations to our work. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

Section I: The Evolution and Characteristics of Artificial Intelligence

The field of Artificial Intelligence (AI) can be traced back to a 1956 workshop organized by John McCarthy, held at Dartmouth College. The workshop's goal was to explore how machines could be used to simulate human intelligence. Disciplines that contribute to AI include computer science, economics, linguistics, mathematics, statistics, evolutionary biology, neuroscience, and psychology, among others. Recent progress and developments in AI have raised new questions about automation and its impacts on the economy. Numerous factors, primarily the trends underlying big data (i.e., increased data availability, storage, and processing power), have contributed to rapid innovation and accomplishments in AI in recent years.[5]

Common AI Technologies Today Include Expert Systems, Machine Learning, Natural-Language Processing, and Computer Vision

Early AI implementation often consisted of expert systems comprising rules within a narrow domain enumerated and programmed by human experts. Expert systems have introduced some degree of productivity gains in recent decades and remain an active area of development.

Machine-learning systems are a central focus in present-day AI innovation. Unlike expert systems, machine-learning algorithms and learning systems are trained against observational or simulated outcomes. Machine learning underpins applications of AI including natural-language processing and computer vision. Examples of natural-language processing include machine translation, as well as personal assistants on smart phones. Computer vision includes algorithms and techniques to classify or understand the content of scenes. These scenes may be recorded by cameras, radar, lasers, or a combination thereof.

[5]For more on trends underlying big data, see, for example, GAO, *Highlights of a Forum: Data and Analytics Innovation: Emerging Opportunities and Challenges*, GAO-16-659SP (Washington, D.C.: Sept. 20, 2016).

Several Definitions and Taxonomies of AI Exist

There is no single universally accepted definition of AI, but rather differing definitions and taxonomies. According to Russell and Norvig (2010), for example, AI is defined as computers or machines that seek to act rationally, think rationally, act like a human, or think like a human.[6] The first of these four approaches is the rational agent, in which AI is designed to achieve goals via perception and taking action as a result. The second approach, thinking rationally, is based on formal logic. In this approach, AI is designed to logically solve problems, make inferences, and optimize outcomes. The third approach, a system designed to behave as a human, is the form of intelligence conceptualized and later popularized as the Turing Test, which involves a test of natural-language processing, knowledge representation, automated reasoning, and learning. The fourth approach, modeling human thinking, is inspired by cognitive science. The research of Nilsson provided a broad definition of AI, as "that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment."[7]

In addition to defining AI overall, researchers have distinguished between narrow and general AI. Narrow AI refers to applications that provide domain-specific expertise or task completion, whereas general AI refers to an AI application that exhibits intelligence comparable to a human, or beyond, across the range of contexts in which humans interact. While there has been considerable progress in developing AI that outperforms humans in specific domains, some observers believe that general AI is unlikely to be achieved for decades in the future.

AI Has Been Conceptualized as Having Three Waves of Development

Another approach to understanding AI is by considering the waves in which the technology has developed, rather than a specific or singular definition. Launchbury (2016) provides a framework that conceptualizes AI as having three waves based on differences in capabilities with respect

[6]Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. (N.J.: Pearson, 2010).

[7]Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievement*. (Cambridge: Cambridge University Press, 2009).

to perceiving, learning, abstracting, and reasoning (see fig. 1).[8] These waves can broadly be described as follows:

- wave 1 – expert or rules-based systems
- wave 2 – statistical learning, perceiving and prediction systems, and
- wave 3 – abstracting and reasoning capability, including explainability.

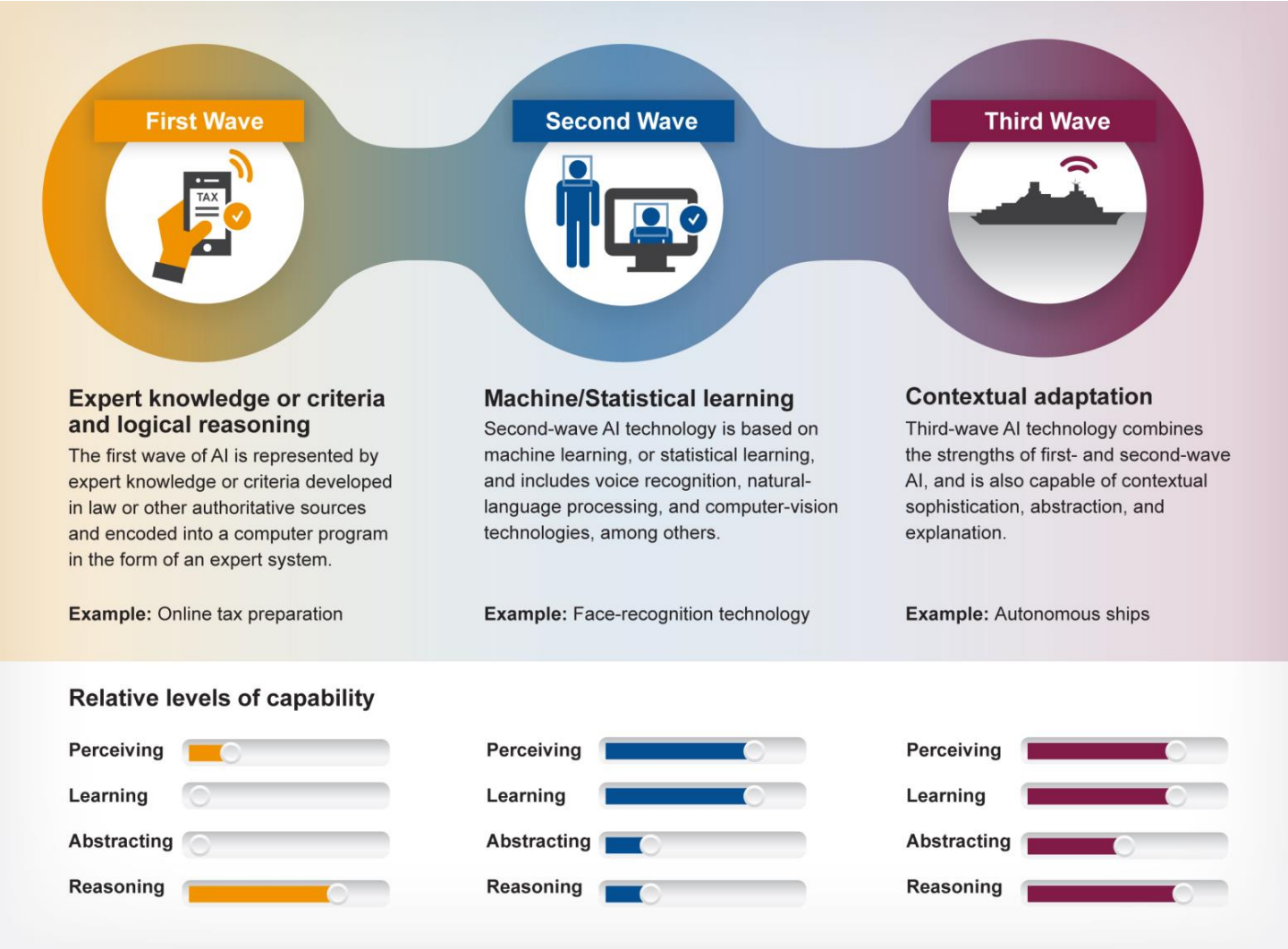
The first wave of AI is represented by expert knowledge or criteria developed in law or other authoritative sources and encoded into a computer algorithm, which is referred to as an expert system. Examples of expert systems include programs that do logistics scheduling or tax preparation. Expert systems are strong with respect to reasoning, as they reflect the logic and rules that are programmed into them. Human tax experts, for example, understand the rules of the tax code, and these rules can be programmed into software that yields a completed tax return based on the inputs provided. First-wave systems continue to yield benefits and are an active area of AI. Expert systems are not strong, however, when it comes to perceiving, learning, or abstracting to a domain outside the one programmed into the system.

Second-wave AI technology is based on machine learning, or statistical learning, and includes natural-language processing (e.g., voice recognition) and computer-vision technologies, among others. In contrast to first-wave systems, second-wave systems are designed to perceive and learn. Second-wave AI systems have nuanced classification and prediction capabilities but no contextual and minimal reasoning capabilities. Examples of second-wave systems include voice-activated digital assistants, applications that assist healthcare workers in selecting appropriate treatment options or making diagnoses, and self-driving automated vehicles.

As reflected in figure 1, third-wave AI technologies combine the strengths of first- and second-wave AI and are also capable of contextual sophistication, abstraction, and explanation. An example of third-wave AI is a ship that can navigate the sea without human intervention for a few months at a time while sensing other ships, navigating sea lanes, and carrying out necessary tasks.

[8]Launchbury, *A DARPA Perspective on Artificial Intelligence*.

Figure 1: The Three Waves of AI.

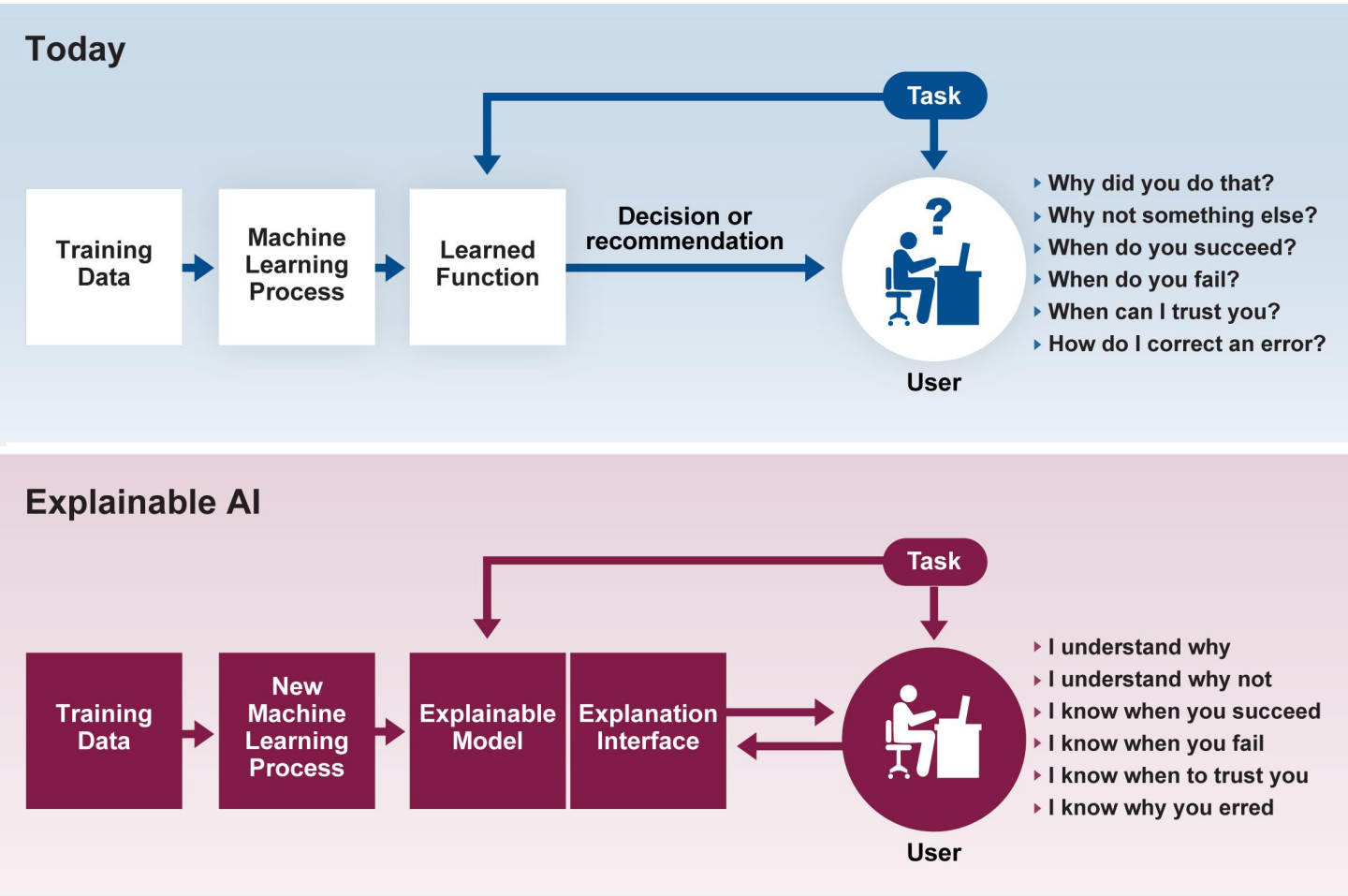


Source: Defense Advanced Research Projects Agency (DARPA) information; Art Explosion (art). | GAO-18-142SP

As described by Launchbury, we are just at the beginning of the third wave of AI, and further research remains before third-wave technologies become prevalent. An important part of third-wave AI will be developing AI systems that are not only capable of adapting to new situations, but also are able to explain to users the reasoning behind these decisions. As illustrated in figure 2, today's machine-learning systems are black-box systems for which users are unable to understand why the system makes a specific decision or recommendation, why a decision may be in error, or how an error can be corrected. The goal of explainable AI is to develop

machine-learning systems that provide an explanation for their decisions and recommendations and allow users to know when, and why, the system will succeed or fail.

Figure 2: Explainable AI Will Provide an Understanding of the Reasoning behind the Decisions or Actions of Machine-Learning Systems



Source: GAO and Defense Advanced Research Projects Agency (DARPA). | GAO-18-142SP

Advances in AI Are Marked by Performance in Game Playing

Advances in AI have often been marked by their performance in playing games. For example, the game Connect Four was solved in 1988 by an expert system that used a strategy described in 1949 for chess. The

world chess champion Garry Kasparov was bested in 1997 by IBM's Deep Blue system. AI approaches, heavily dependent on machine learning, have been refined and applied to increasingly complex games, beating human champions at Jeopardy (IBM's Watson in 2011), Go (Google/DeepMind's AlphaGo in 2016), and poker (Carnegie Mellon University's Libratus in 2017). Learning algorithms may also play in game simulators, rapidly training against game outcomes in lieu of human supervision. This approach was used to train an algorithm to win 49 Atari videogames in 2015.

Machine-Learning Systems Are Credited with Recent Advances in AI

Advances in machine learning in recent years have resulted in systems that are now capable of outperforming humans at some specific tasks. In supervised machine learning, an algorithm is presented data to which labels (or answers) have been assigned. The algorithm then uses the labeled data to identify logical patterns that predict a specified answer to a problem. After a pattern is identified, it can be used to apply to similar problems. In unsupervised machine learning, no labels are given to the inputs, leaving the algorithm to identify structure in the inputs, for example, by clustering similar data. In other words, unsupervised learning is based on grouping like things without a preconceived idea of what to expect. In semisupervised learning, the machine-learning algorithm is provided some data that are labeled with answers and other data that are not labeled. The algorithm then uses the labeled data to determine the pattern and apply labels to the remaining data. Semisupervised learning can be useful in circumstances where there are too many data to be labeled.

Deep neural networks, a subset of machine-learning algorithms, have been trained to classify images, detect objects, identify people from faces, generate text from speech, translate natural languages, and many other tasks. There is some adaptability of deep neural networks for solving adjacent problems, using a technique known as transfer learning. For example, transfer learning was successfully applied from general image recognition to the specific medical-imaging case of diagnosing diabetic eye disease. By initializing the deep neural network on general image-recognition cases, fewer medical-image samples were required for training with performance matching that of human specialists. Such transfer learning may reduce the time and associated cost to apply deep neural networks to adjacent domains.

Section II: Forum Participants Identified Several Benefits of Artificial Intelligence and Challenges to Its Development

The increased adoption of artificial intelligence (AI) will bring with it several benefits, as well as a number of challenges. According to the participants at our forum, both benefits and challenges will need to be carefully considered alongside one another.

In terms of benefits, forum participants said AI may (1) improve economic outcomes and increase levels of productivity, (2) improve or augment human decision making, and (3) provide valuable insights into complex and pressing problems. Participants also stressed that there may be benefits related to AI that cannot yet be predicted or may even be hard to imagine.

In terms of challenges, the participants highlighted several that will eventually need to be addressed. Such challenges include (1) collecting and sharing the data that are needed to train AI, (2) accessing adequate computing resources and requisite human capital, (3) ensuring laws and regulations governing AI are adequate and that the use of AI does not infringe on civil liberties, and (4) developing an ethical framework to govern the use of AI and ensuring the actions and decisions of AI systems can be adequately explained and accepted by those who interact with such systems. Actions to address these challenges could be taken by government, private industry or other nongovernmental actors, or public–private partnerships.

Forum Participants Said AI Could Result in Economic Benefits and Improve Human Decision Making and May Also Provide Insights into Complex and Pressing Problems

Improved economic outcomes and increased levels of productivity. It may be difficult to accurately predict what AI's impact on the economy could be, according to one forum participant. In previous periods, large investments in automation have been highly correlated with improvements in productivity and economic outcomes, which, according to one forum participant, has led some to believe that transformations as a result of AI could have the same outcome.

This same participant noted, however, that no one collects the data needed to measure the impact AI or other types of advanced automation may have on the economy. According to another participant, whatever the effect that AI will have on productivity in particular, and the economy in general, the changes will occur quickly and be difficult to predict.

Estimates of the potential impact AI will have on the economy cover a wide range. In 2016, one global professional-services company, for example, analyzed 12 developed economies and estimated that AI has the potential to boost their labor productivity by up to 40 percent and double their annual economic growth rates by 2035.[9] This company sees AI not just as an enhancer of productivity, but also as a factor of production. More specifically, the company argues AI will not only boost growth by replacing and augmenting human labor, but also because of its potential to create new goods, services, and innovations.

At the same time, one investment bank estimated in January 2017 that by 2025, revenues from AI will range from more than $25 billion to over $40 billion, up from less than $5 billion in 2017.[10] They argue that this growth will result from a confluence of factors, including inexpensive processing power, the expansion of big data, and the competitive needs of businesses across sectors that recognize the need for AI to augment their productivity. Moreover, several venture capital firms are betting on AI's growth potential and have invested significantly in AI startups, with the hopes of selling to one of the large IT companies that are seeking leadership in the AI sector.

AI could also be used to boost productivity by reducing administrative burdens. For example, one recent report highlighted a number of applications where AI could be used by governmental agencies at all levels to answer questions, fill out and search documents, route requests, translate, and draft documents.[11] If AI were used in such applications, the report claims, it could make government more efficient while freeing up time for public employees to build better relationships with the citizens they serve.

[9]Accenture, "Why Artificial Intelligence Is the Future of Growth," September 28, 2016.

[10]Kartik Gada, "Artificial Intelligence: The Ultimate Technological Disruption Ascends," Woodside Capital Partners, January 2017.

[11]Hila Mehr, "Artificial Intelligence for Citizen Services and Government," Ash Center for Democratic Governance and Innovation, Harvard Kennedy School, August 2017.

Improved or augmented human decision making. AI can be used to gather an enormous amount of data and information from multiple locations, characterize the normal operation of a system, and detect abnormalities, much faster than humans can. According to one forum participant, AI is an appropriate technology for the cybersecurity sector because the cyber systems used to provide security generate a vast amount of data, and AI can be used to help determine what are normal conditions and what is abnormal.

> *"By their very nature, the computers can respond much faster than humans can. Humans are very slow. They take hundreds of milliseconds to respond to any kind of information. In hundreds of milliseconds, computers can…do millions of things, literally. And because of that, new kinds of strategies that are gathering information from multiple places all at the same time are possible that weren't before."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Another sector that may benefit from the adoption of AI is the financial sector, where it could be used to improve decision making and, in turn, improve fairness and inclusion for consumers. One participant stated specifically that machine learning could be used to help establish a potential consumer's identity, which is required before they can gain access to banking and credit. Establishing identities for this purpose is especially difficult in some parts of the world, though there is now a massive change underway to collect data for the purposes of identification. For example, over a billion people in India have had their biometrics collected and are participating in the economy through biometric identification. There are, however, issues concerning exactly how the data should be used, understood, and analyzed, according to this same participant. In addition, this participant said that machine learning and credit analytics could be used to collect what is called alternative data to help improve access to credit for individuals who do not meet traditional standards of credit worthiness or who have little or no credit history.

> *"In finance, the impacts [of AI] are overwhelmingly positive for better distribution of assets in financial services, and better affordability. There are hard issues to solve, but the combination of the cell phone and data of AI is the most democratizing thing that has ever happened, ever, in bringing financial services to everyone."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

In addition, AI could be used to create data-informed policy that may help prevent inappropriate or harmful human bias—be it from political pressure or other factors—from creating undesirable results, according to one participant. In the criminal justice sector, for example, some jurisdictions are using AI to assess risks, ranging from determining the likelihood that a defendant will default on bail to the likelihood that a potential parolee

will reoffend. Such systems have been simulated by the National Bureau of Economic Research, which published a working paper in 2017 that lays out the potential benefits of using machine learning to determine whether a defendant should await trial at home or in jail. Specifically, the paper claims that using machine learning in this way could result either in a reduction in crime with no change in jailing rates or in a reduction of pre-trial jailing rates with no increase in crime rates.[12]

"[As] opposed to just being punitive, you could use [artificial intelligence] to understand where [offenders'] needs actually are as opposed to incarcerating [them], which is going to give you a coin-flip's chance that they come back into the system. You could look to alternatives to incarceration that include drug treatment courses and in-home options."

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

However, as another participant at the forum noted, AI is no guarantee of freedom from bias. The participant stressed specifically that if the data being used by AI are biased, the results will be biased as well, and cited recent reporting on racial bias in an algorithm used by courts and parole boards as an example of this type of risk.[13]

AI can help prevent inappropriate or harmful human bias, according to this same participant, if it is carefully used, if the assumptions of the models are thoughtfully considered, and, most importantly, if the outputs of the model are constantly and closely verified. Another participant stated that the baseline is current practice, not perfection, and that the goal should be to become less biased and more accurate, not perfect.

Insights into complex and pressing problems. Some of the participants at our forum believe that AI has the potential to provide insights into—and even help solve—some of the world's most complex and pressing problems. For example, according to one forum participant, university researchers in the United States have successfully used AI to harness the power of social networks to more effectively spread information on preventing the transmission of HIV among homeless youth. More specifically, the researchers collaborated with social workers to better understand the real-life networks that have developed among the more than 6,000 homeless youth in Los Angeles. They then used an algorithm to recruit youths to serve as "peer leaders" and receive training

[12]Jon Kleinberg et al., "Human Decisions and Machine Predictions," National Bureau of Economic Research, February 2017.

[13]Julia Angwin et al., "Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks," *ProPublica,* May 23, 2016.

on preventing HIV. What they found was that the youths selected by the algorithm were far more likely than the youths who were considered most popular to spread the information they received. These same researchers are planning to use AI in a similar way to combat substance abuse and help prevent suicides.

This same participant also described how university researchers are partnering with various groups to better protect wildlife from poachers in Africa. More specifically, the Wildlife Conservation Society is using AI algorithms in Uganda to better predict where poachers might set snares, which enables them to find and remove the snares before they cause harm. This participant stressed that there is an opportunity to use AI for social good in the sense that AI offers new capacity to solve long-standing societal problems that all too often disproportionately affect low-resource communities and developing countries.

> *"In working with the Wildlife Conservation Society, we're predicting where poachers set snares, and so if you can predict where they set snares before they kill animals, the snares can be removed. We've done this work, done six-month trials, removed snares, poachers have been arrested, and hopefully animals have been saved."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Another participant stated that AI could be used to alleviate myriad other problems. As the number of elderly Americans continues to grow, AI could be used to provide medication management, mobility support, housework, meal preparation, and rehabilitation services to a growing number of people who need assistance with day-to-day activities. Examples include the use of robots enabled by AI for medication management and mobility support. This same participant noted that AI could be used to enhance body imaging during surgery and facilitate personalized medical care through genome profiling. Moreover, automated vehicles could help eliminate areas that have a demand for transit but lack access, known as "transit deserts"—while simultaneously cutting the costs of transportation, according to one participant.

> *"So you eliminate transit deserts, you eliminate the people who simply cannot get to jobs, cannot get to education, you eliminate the friction of that, you have goods delivery at maybe a half of the cost that we have today. I think this could be a huge primer on economic productivity and truly improve people's lives."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

In addition, one recent report on the potential applications of AI in the governmental sector quoted an industry representative who claimed that AI could be used to reduce the number and severity of pandemics,

improve food security and sustainable agriculture, and increase public safety by monitoring infrastructure.[14]

There are other complex and pressing problems that may eventually be solved by the adoption of AI. According to one participant, AI could eventually be used to assure regulatory compliance in the financial sector without unnecessary burden on those being regulated. This participant noted that in the United States there are five federal agencies that directly supervise depository institutions, in addition to 50 state regulators and other federal agencies that are involved to some extent. These institutions are, in turn, oftentimes reluctant to enter new sectors or markets because they do not always have a clear picture of how they will be regulated and by whom. This participant argued that we may eventually be able to use AI to issue regulation as computer code, which would increase regulatory efficacy by enabling those being regulated to be automatically compliant. This participant said that the idea of issuing regulations in the form of code is at an early stage, but that it is under active discussion by regulators.

To this participant's knowledge, the most advanced effort in this area is underway in the United Kingdom's Financial Conduct Authority, which employs a unit devoted to developing ideas for improving the regulatory process and modernizing regulatory rules with technology. According to this participant, this unit recently hosted a "tech sprint" that brought together regulators, banks, and technology experts to develop ideas for modernizing the regulatory process, and that the result of this sprint would include developing concepts for moving toward what this unit calls "machine executable regulation." Such an idea could result in potentially issuing regulations in the form of computer code. This participant noted at the forum that issuing regulations in this way could not only cut costs and reduce the burden of complying with outdated financial regulations, but could also help improve the government's ability to more effectively combat money laundering.

[14]Mehr, "Artificial Intelligence for Citizen Services and Government."

Forum Participants Highlighted Several Challenges Associated with AI, Including Barriers to Collecting and Sharing Data and Inadequate Laws and Regulations

Barriers to collecting and sharing data. While not all applications of AI require massive amounts of data, certain applications that use machine learning algorithms do.[15] This can be a problem in sectors where data are not easily aggregated or interpreted or readily available. Such is the case with criminal justice, where the ways in which data are collected and organized varies from jurisdiction to jurisdiction. Such is also true with most vulnerable populations and developing countries, where data have not yet been collected.

> *"Not enough attention has been paid to [areas of social welfare or to developing countries], but these are very important areas where AI could have a very significant impact. And in part, these are areas where we don't have data-rich settings. Data [are] limited, but these are important societal problems to solve."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Some of the participants reminded us throughout the forum that as AI moves from the laboratory into human spaces, and as the problems we ask AI to solve grow in complexity, so too will the data needed to effectively train and test that AI. This may become an especially acute problem if certain firms are able to create "data monopolies"—data rights ownership of economic sectors—and dominate the market with anti-competitive behavior.[16] Certain large information and communication technology firms have already captured tremendous shares of users in their market segments, and as the number of their users increases, these firms will collect even more data. This will allow them to develop an even greater advantage over their competition. There is also a risk, according to experts at our forum, that an adversary could compromise the data being used and result in AI operating in undesirable ways.

> *"Just like optical illusions can trick people, you can build images that will trick the computer and make it so that it will tell you that something is a stop sign with a 100 percent confidence when it's really a yield sign, or vice versa, which could have negative consequences."*

[15]Scott W. Bauguess, "The Role of Big Data, Machine Learning, and AI in Assessing Risks: a Regulatory Perspective," U.S. Securities and Exchange Commission, Keynote Address to OpRisk North America 2017, New York, New York, June 21, 2017.

[16]Kira Radinsky, "Data Monopolists Like Google Are Threatening the Economy," *Harvard Business Review,* March 2, 2015.

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

When data are directly tied to autonomous systems or robots that interact with humans, the threats posed by messy or hacked data are exacerbated, according to one participant. This participant said that, for example, automated vehicles, security systems, and weapons systems that depend on AI could all be easily sabotaged to cause harm. This participant further emphasized that as human decision-makers increasingly rely on AI to make decisions on their behalf, immediate data veracity and integrity becomes ever more important, especially in financial markets and in the legal system.

In addition, we heard from the forum participants that private companies may be unwilling to share their data because they do not want to expose the details of their proprietary technologies or intellectual property. They are also concerned about hacking as more and more things come under control by AI. At the same time, we heard from another participant that there is "openness" among many of those working in AI, such that concerns about proprietary technologies or intellectual property may not be an insurmountable obstacle to innovation.

"There's a lot of openness when it comes to actual algorithms, putting those out there, as far as processing capacity, sharing that or of course making money off of that, that's absolutely there. There are a lot of benchmarks out there, by necessity, where you have labeled data that's been shared so that the industry can benchmark their new algorithms and new approaches against it, so on a fundamental basis there's actually a lot of collaboration, too."

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Lack of access to adequate computing resources and requisite human capital. In addition to enormous feeds of continually updated data, forum participants told us AI researchers and developers need access to storage and processing, both of which are expensive and sometimes difficult to access at the necessary scale. Two participants specifically told us that it is difficult to gain access to high-quality training sets, and one shared a concern that researchers would move away from difficult problems and focus on solving much easier problems with data that are easier to access.

"We have an extraordinary problem, all of us…[with access to individuals with needed skills]. We buy whole companies just to get at the employees. We buy whole faculties just to get at the employees, so there's something there."

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Some forum participants also shared concerns that the accelerated pace of change associated with AI is straining the education and workforce systems' capacity to train and hire individuals with the appropriate skill

sets, leaving many companies struggling to find workers with relevant knowledge, skills, and training. According to one participant, it is not uncommon for large companies to buy smaller companies just to get access to their human capital.

One participant pointed out that in the past, most individuals spent the first part of their lives learning and the second part employing what they learned in their profession. It may, this participant argued, be more effective to provide students with shorter and periodic training instead of the traditional 4-year undergraduate degree. This shift can already be seen in professional schools that are moving to one-year degree programs.

Another participant noted that students who want to work with AI will need to master science, technology, engineering, and math (STEM) disciplines, as well as the social sciences, because the technology that is developed is going to interact with humans. Furthermore, according to another participant, education should be less theoretical and more applied. A third participant added that they believed the jobs of the future will put more of a premium on knowledge of the business process and that successful employees will be those who, when questions arise, know how to find the right answer.

> *"This technology is going to be interacting with humans… and so there's great opportunity, I think, here for human creativity in the arts and to be able to create this technology in a way that is adoptable and acceptable and enhances productivity and trust."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Adequacy of current laws and regulations. The widespread adoption of AI may, according to some forum participants, have implications regarding the adequacy of current laws and regulations. For example, deploying AI requires contractual agreements with the users of the AI, according to one participant, and contracts fall under state law, which could be an issue when state laws create rights through contracts that are not protected by the federal government.

> *"Today, most artificial intelligence products are protected through contractual agreements in which the user agrees to assign rights and relinquish claims for the benefit of the program's operator or creator…. Contracts are creatures of state law. They're not federal law…. [I]f a contract enforced by state law creates rights or a federal scheme does not protect those rights, you easily end up with tricky questions related to federal vs. state power. So a focus on federal schemes alone in trying to understand protection would be very short-sighted."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

One participant noted that current patent and copyright laws provide only limited protection for software and business methods and questioned whether these laws will protect the products created by AI. This same participant also claimed that current patent and copyright laws pose challenges for AI in at least three key areas:

- First, an inventor does not have to reveal very much about their software code to secure a patent or copyright, which may be problematic where public safety, liability, or criminal justice is concerned.
- Second, patent protection lasts 20 years and copyrights created by corporations last 120 years, both of which are considered too long a time horizon for AI, according to this participant. Specifically, because advancements in AI have moved at such an extraordinary pace, this participant argues that data protection for pharmaceuticals may be a better model for protecting innovations in AI. In that context, a brand-name drug company receives 4-5 years of exclusivity in exchange for making its safety and efficacy clinical trial data available to potential competitors. Thus, pharmaceutical companies receive this period of exclusivity, enforced through the context of regulatory approval, in exchange for data openness.
- Third, if AI derives its creative results in part through the collective actions of numerous humans, it is not clear to this participant whether that creativity is attributable solely to the program or to the program's creators.

In addition, one of the participants at the forum raised concerns about ways in which AI could be used to violate civil rights. This participant cautioned, for example, that if law enforcement considers race, class, or gender in AI that is used to assess risk, there is the possibility that a defendant's equal protection rights under the 14th Amendment may be violated, as well as their due process rights under the 5th and 14th Amendments. More specifically, this participant noted that there is a concern that the ways in which input factors for risk-assessment tools are collected may violate a defendant's 5th Amendment right against self-incrimination. This concern has not, however, been well explored by the courts or academics, according to this participant.

Some risk assessment tools use questionnaires and interviews that the defendant needs to take part in for the assessment to function, and there are concerns about whether a defendant could be punished in some way for refusing to speak to the law enforcement or intake officer who administers the questionnaire. At the same time, this participant noted,

some risk assessment tools do not require the defendant to answer questions or be interviewed and therefore do not run the risk of violating the defendant's right not to incriminate themselves.

Another area in which AI may eventually impact the criminal justice system is through the use of probabilistic genotyping in criminal cases. Probabilistic genotyping, according to one criminal justice expert, uses interpretive software to run multiple scenarios—like the risk analysis tools used in finance—to examine trace amounts or complex mixtures of DNA. In recent years, such software has been used to calculate a statistic—or likelihood ratio—of the strength of a DNA match. This contrasts with traditional DNA analysis, which assesses whether a DNA type is present.[17] Currently, probabilistic genotyping does not use artificial intelligence, though according to one forum participant, it may only be a matter of time before relevant datasets are created and made available for use. Even without the use of artificial intelligence, the use of probabilistic genotyping raises transparency and confrontation issues similar to those pertaining to the use of AI in the criminal justice sector. The issue, according to this same participant, is that the source code used in probabilistic genotyping hinders due process because it contains underlying data that are not always well understood and algorithms that are not always fully disclosed to the defense. This participant told us that even though probabilistic genotyping outputs are being used at trial to prove or disprove a defendant's guilt, some defense attorneys have been challenged when they tried to gain access to the genotyping source code, especially if it is proprietary. This raises concerns that using such outputs as evidence in a criminal case can violate a defendant's 6th Amendment right to know their accusers and the nature of the charges and evidence being used against them.

This participant also told us about a case that was heard in a California appeals court that overturned a trial court's decision to side with the defense's request to access the source code,[18] but that more recently, a trial judge in the Southern District of New York found that the defense did have the right to access probabilistic genotyping source code during a Daubert hearing, which is a pretrial hearing where the admissibility of

[17]Jason Tashea, "Defense Lawyers Want to Peek behind the Curtain of Probabilistic Genotyping," *ABA Journal,* December 2017.

[18]People v. Superior Ct. of L.A. Cty. (Chubbs, Real Party in Interest) B258569 (Cal. Ct. App. Div. 4 2015).

expert testimony can be challenged. The judge in that case—United States v. Johnson—lifted a protective order and unsealed the source code behind the probabilistic genotyping tool known as Forensic Statistical Tool (FST), which was created to determine the likelihood that a given defendant's DNA was present in a mixture of multiple individuals' genetic material.[19] FST is owned by New York City's crime lab, and the source code was unsealed in response to a motion by *ProPublica*, which argued there was a public interest in disclosing the code.[20] New York State has since stopped using FST and has begun using another tool, according to this forum participant. Legal fights for access to proprietary DNA analysis software have been and continue to be litigated elsewhere around the country, with mixed results.[21]

Ethical Framework for and Explainability and Acceptance of AI. In addition to new and renewed regulatory and legal implications, the adoption of AI also introduces ethical implications. According to a forum participant, there is a need for a system of computational ethics to help AI choose options that reflect agreed-upon values.

> *"We're going to need to have some kind of computational ethics system. We're not going to be able to anticipate, in advance, all the crazy situations that you're going to have to make complicated decisions."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Moreover, some of the participants at the forum noted that before humans will understand, appropriately trust, and be able to effectively manage AI, an AI application or system needs to explain why it took certain actions and why it valued certain variables more than others.

> *"When these systems take action, they may not be able to explain in real time, I'm shutting down this user's account, I'm disconnecting this machine, I'm disconnecting this computer. But retroactively, when some team comes in to figure out what happened, there's going to need to be an explanation to [the question]... why did you do this?"*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Most of the participants' discussion regarding the "explainability" of AI focused on reasonable assurance of safety. One participant stated that

[19]Order, Doc. 152, United States v. Johnson 15-cr-00565-VEC (S.D.N.Y. 2017).

[20]Lauren Kirchner, "Federal Judge Unseals New York Crime Lab's Software for Analyzing DNA Evidence," *ProPublica,* October 20, 2017.

[21]Tashea, Jason, "Federal Judge Releases DNA Software Source Code That Was Used by New York City's Crime Lab," *ABA Journal,* October 20, 2017.

we will never achieve absolute assurance of safety with automated vehicles, for example, because there cannot be an absolute problem description. Put simply, the automated vehicle will be confronted with a situation in which there is no good outcome, a situation it cannot be trained to foresee, or which is outside of its capabilities. In such instances, the automated vehicle will collect and analyze data, choose a "best" course of action, and will not (with current AI technology) have the ability to constantly generate and consider multiple options. According to another participant, it will be essential for humans to be able to evaluate or analyze the factors the AI weighed before taking a course of action.

> *"Someday, an autonomous car is going to run over a bicycle. And when that happens, we're going to want to be able to inspect the black box and say why did it do that?"*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

A concern related to explainability and social acceptance that one participant shared was whether AI would eventually become competent enough to manipulate humans, especially the most vulnerable among us, like those with dementia or who otherwise need AI to successfully function in society. This same participant stressed that it is a grave misconception to believe that the algorithms used in AI are inherently neutral and trustworthy. Another participant added that sometimes people place too much trust in technology and that their hopes are easily dashed when a technology does not solve their problems. Yet another participant pointed out that some recent research suggests the average person disproportionately holds AI accountable for failure even if the AI's rate of failure was much lower than the baseline level of existing practice.

One forum participant indicated that our society's level of acceptance of AI will depend on how AI affects the economy, the environment, and personal mobility. This same participant noted that the nature of risk affects how we process and accept it, but that most people do not understand how to accurately calculate risk, nor are people very good at appreciating their exposure to it. At the same time, another participant stated that culture will be a determining factor in the degree to which people accept AI. Two other participants added that the average person's level of acceptance of AI will likely depend on how it affects them personally. Lastly, one participant added that in certain areas, our risk tolerance might not be where it needs to be yet, but that in other areas, we may be much more likely to accept the use of AI to reduce risks of injury and death. One such area is industrial agriculture, where accident rates have begun coming down with the introduction of AI.

Section III: Forum Participants Identified Several Cross-Cutting Policy Considerations Related to AI and Several Areas Where More Research Is Needed

After discussing the benefits and challenges associated with AI, the participants at our forum highlighted a number of policies they think will need to be considered, including policies for: (1) incentivizing data sharing, (2) improving safety and security, (3) updating the regulatory approach, and (4) assessing acceptable risks and ethical decisions. They also highlighted several areas they believe deserve more research, including: (1) establishing regulatory sandboxes, (2) developing high-quality labeled data; (3) understanding AI's effect on employment and reimagining training and education, and (4) exploring computational ethics and explainable AI.[22]

Forum Participants Said Policymakers Will Need to Consider Policies to Incentivize Data Sharing, Improve Security, and Update Regulations, among Other Things

Incentivizing data sharing. Concerning data sharing, participants noted that policyholders could further facilitate the sharing of data to improve an industry, including safety outcomes. To address challenges and concerns with data sharing, including concerns that confidential or proprietary business information could be compromised, participants noted that policymakers could further facilitate the sharing of data and proprietary information across the industry. Participants emphasized the need for establishing a "safe space" to protect sensitive information (e.g., intellectual property, privacy, and brand information). Another participant reiterated optimism that government could get the data it needed to properly protect the public, while maintaining proprietary data protections.

Participants noted successful data-sharing efforts through entities such as MITRE and the National Institute of Standards and Technology (NIST). In particular, some participants highlighted data-sharing efforts to improve

[22]For this report on AI, computational ethics is defined as the ethics of AI or moral behavior of artificially intelligent software.

safety outcomes. For instance, one participant mentioned that researchers at MITRE had credited data-sharing efforts in the aviation industry (employing a safe space) with reducing the number of accidents. Another participant emphasized the importance of sharing data to better understand safety outcomes associated with automated vehicles, stating, "[i]f we're going to trust that these vehicles can go out on the road, we need to verify that, in fact, out on the road, they are as safe as we think they are."

Another participant cautioned, however, that for such a safe space to succeed, it will need to start small (e.g., with a few manufacturers) and clearly define the data that are needed and the specific scenarios in which the data will be used. Another participant said that if such safe spaces can be shown as effective in small-scale applications, it will undoubtedly lead to more widespread adoption. This same participant added that protecting and sharing data are essential, because doing so not only helps the manufacturers ensure their products are safe, but also provides valuable information to academics and policymakers. One other participant added that efforts to improve data sharing would also need to consider the treatment of intellectual property.

Participants also emphasized the need to improve the quality of data that feeds AI activity. One participant emphasized that all of the AI tools start with data—training data, testing data, and monitoring data. This participant stressed the importance of addressing problems with data (e.g., whether the data are inadequate, incomplete, or error-filled). Concerning AI, this participant also mentioned that "discussions have to be based on facts, not myths." Another participant emphasized that better data collection requires not only better cooperation, but also a standardization of data-collection definitions, measurement, and methods. According to some of the forum participants, we need to more accurately simulate the real world with training sets and scenarios in a human-space environment, versus a controlled lab, to better prepare for the unpredictability of the real world.

"Human spaces are probably some of the messiest environments you could possibly imagine. It's very difficult to know or plan for things in advance, and a lot of the approaches people take in laboratory settings can be very difficult."

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

Forum participants also highlighted other proposed future data-sharing efforts, citing the benefits of assessing data from multiple sources to improve outcomes. According to one forum participant, the National Science and Technology Council Subcommittee on Machine Learning

and Artificial Intelligence is working collaboratively among federal departments and agencies to promote the sharing of government data to help develop innovative solutions for social good. This sharing may include creating training environments—"safe spaces"—in which sensitive data are protected, among other things.

Another participant noted that in the criminal-justice sector, the federal system could be used as a test bed for various reforms—including data-sharing reforms—because the federal system is unified. This participant argued that if the federal system could find a way to share data related to risk assessments and other areas and show that the data are being utilized in an evenhanded way, the reforms pioneered by the federal system would likely migrate down to the individual state systems. This same participant also claimed that the Bureau of Justice Assistance and the Bureau of Justice Statistics may be the best positioned to initiate any nationwide data standardization and collection projects. Another participant added that the European Union drafted a data protection regulation that has allowed defendants and their attorneys access to the algorithms and underlying data that are being used to decide their future in a court of law and that this may be a good place for policymakers to begin looking at whether such policy is needed in the United States.

Certain forum participants also expressed concerns about limitations in accessing data that could be helpful to researchers. In particular, they maintained that many potentially useful data are guarded by federal agencies that do not provide access to researchers. For instance, one participant said that sometimes agencies make it more difficult to share data because access requires multiple security clearances and nondisclosure agreements. Another participant noted that laws like the Administrative Procedures Act and the Federal Advisory Committee Act may be outdated and preventing federal agencies from sharing data.

Improving safety and security. Participants highlighted challenges and opportunities to enhancing the safety and security of system applications from cyber attacks, including those with AI features. One participant emphasized that if an AI system is running on a computer that is then hacked, the security of the system is only as good as the security of the computer that it is running on. However, protecting against hackers is not something an individual company can do on its own. Rather, efforts to combat hackers need to be industry-wide.

One participant said that the costs of cybersecurity in all forms of network computing are not being shared appropriately and that security breaches

are much costlier than the security measures that are needed to prevent breaches. This participant said that policymakers will need to consider creating some kind of framework that ensures costs—and liabilities—are appropriately shared between manufacturers and users. This participant also noted that those who build cybersecurity systems are not the people who pay for security breaches. One option to improve security, according to this participant, could be to implement a regulatory structure of security ratings akin to crash safety ratings in the automobile industry. Incentives to invest in security may also be created through external factors such as insurance. For instance, the insurance industry would take notice of vehicles that are being stolen at a disproportionate rate. Similarly, in the case of crashes, the liability pressure in the auto industry provides an incentive to invest in security.

Two participants said that policymakers should consider creating a new regulatory structure to better ensure the safety of automated vehicles. One of these participants suggested that computer systems running automated vehicles could be held to a standard similar to the Five-Star Safety Ratings Program, whereby the systems could be tested against penetration by hackers. Further, from an insurance perspective, if an automated vehicle has a defect and causes a crash, the liability will likely fall on the manufacturers, providing incentives to ensure safety and security. The other participant who argued for a new regulatory structure said that it can take up to 7 years for a new standard to be set regarding automobile safety. Instead of continuing with this same system,[23] the federal government could put out an automated vehicle policy that would provide guidance that manufacturers can use to help them better design their automated vehicles. That way, instead of having to address all safety issues, manufacturers would only need to address safety issues related to their design. Then, as best practices develop, the federal government could convert those into standards. This same participant stressed that the federal government should not tell manufacturers how to design their systems or set standards prematurely. Concerning automated vehicles, one participant also emphasized how it was important to have data available to those in academia as well as policymakers in order to make policy decisions.

[23]Federal Motor Vehicle Safety Standards are minimum safety requirements under which manufacturers of motor vehicles and motor vehicle equipment must self-certify conformance.

Updating the regulatory approach. The widespread adoption of AI will have implications for regulators, and lawmakers will need to consider policy options to address these issues, according to multiple forum participants. One participant reinforced the need for regulators to be proactive, including a commitment of resources, because change is occurring so rapidly and in unanticipated ways. Going forward, given the technological advancements, some are advocating changes to the existing regulatory approach and mechanisms used to oversee industries.

For example, one participant explained that regulators of automated vehicles "will not get ahead of the industry" to recommend a way to design AI for these vehicles. As a result, regulating automated vehicles is difficult. Consequently, this participant emphasized that, as a policy matter going forward, a new regulatory structure needs to evolve, and that, accordingly, the federal government should avoid setting standards for automated vehicles prematurely. In addition, data sharing can enhance and validate an understanding of how such vehicles are performing and lead to more informed regulations.

Another interrelated issue raised by a participant about automated vehicles concerned how liability would be regulated. Currently, according to this participant, the manufacturer of the automated vehicle bears all responsibility for crashes, even if these vehicles improve overall public safety. Another participant told us that, absent federal laws addressing AI related to automated vehicles, states have begun passing a patchwork of incongruous and potentially confusing legislation on this issue.

> *"When we look at the challenge of bringing fully safe, largely automated vehicles to the road, we shift risk and return where the automotive manufacturer … now bears all the risk for the vehicle's behavior—all of the risk and none of the benefit. And so we have this public-health situation where industry has contributed to significantly improving public health by virtue of improving the fleet, and taking full responsibility for anything that goes wrong."*

Source: An expert who participated in GAO's Forum on Artificial Intelligence. | GAO-18-142SP

One of the forum participants also said that policymakers should consider allowing financial regulators to explore alternative regulatory approaches and reporting mechanisms, leveraging technology to improve and reduce the burden of regulation. In this regard, one participant discussed the merits of "regtech," that is, linking regulation with technology. This participant said that in such an alternative regulatory channel, those entities being regulated could be afforded the option to submit their regulatory data in a more transparent and real-time manner for review by regulators while reducing other reporting requirements. Implementing a data-intensive regtech approach, where data are reviewed against

understood standards, would allow both regulators and those they are regulating to better understand whether desired outcomes are being achieved. In this regard, another participant highlighted that regulators are beginning to implement AI tools in their market surveillance oversight activities.

According to several participants, policymakers should consider ways to use regtech to reduce the cost associated with and the burden of complying with financial regulations. One point a participant made was that those in a position to act cannot remain risk averse, waiting to see what happens. This participant argued that policymakers need to be proactive and adaptive to harness the true potential of AI, a point which was echoed by others.

This participant also pointed out that technology can be used for good or harm and emphasized that issues of technology are the most important ones facing the regulatory and policy community. However, technology issues have not been placed at the center of the regulatory agenda. This participant observed that technology exists to address many problems in finance, but poor regulation practices have hindered these potential gains. Regulatory structures, according to this participant, are full of gaps and are based on long-standing history and mandates rather than current practices. As a result, this participant believes that the current regulatory framework will not allow innovation and may miss negative changes that enter into the system.

At the same time, one participant emphasized that leveraging technology, including AI tools, could achieve goals of more inclusive finance along with enhanced oversight using data in a manner that reduces burdens on regulated entities. This participant emphasized that more inclusive finance could be achieved in conjunction with a mobile-phone delivery channel. However, this participant also noted that current methods to look at fair lending and the presence of disparate outcomes prevent the industry from trying to serve those with lower income or marginal creditworthiness because of concern over using any other approach than “safe” FICO scores. People can be screened out of access to credit if they have a “thin credit” file (i.e., little or no credit history information) under traditional methods of measuring creditworthiness. Alternatively, other consumer data could provide more robust information about individuals’ creditworthiness.

Another participant noted that other laws and regulations may need to be adapted to account for the fact that humans may not always be behind

decisions that are made by automated systems. For example, this participant discussed laws where intent plays a key role, as is the case in financial market manipulation. If someone programs AI to learn to make money, and it does so in a nefarious way, it is not clear how current laws could be used to prosecute the creator of the AI. In discussing the ethics surrounding automated vehicles, another participant talked about situations in which a human driver would need to judge something, like how fast to drive when children are playing near the road. A human driver in such a situation faces an ethical dilemma where they need to make a judgment that weighs their desire for mobility, a desire for safety, and a desire to adhere to the law. This participant wondered whether laws in the future will account for AI's judgment in the same way.

Some of the participants at the forum also raised concerns about privacy, including ways in which AI could be used by law-enforcement agencies to violate civil liberties, and said that this is an area that needs policy solutions. According to one participant, law-enforcement agencies' use of facial recognition software raises concerns that the people being captured by the software could have their civil rights violated, including the right to freely speak and assemble. Some privacy researchers and advocates have said that such remote biometric identification could have "chilling effects" on human behavior and threaten free speech and freedom to assemble. In a 2011 privacy impact assessment (PIA), the International Justice and Public Safety Network recognized that "[t]he mere possibility of surveillance has the potential to make people feel extremely uncomfortable, cause people to alter their behavior, and lead to self-censorship and inhibition."[24] According to a 2017 report issued by Georgetown Law's Center on Privacy & Technology, no state has passed laws that comprehensively regulate law-enforcement agencies' use of facial recognition software.[25]

Assessing acceptable risks and ethical decision making. According to one participant, policymakers need to decide how they are going to measure, or benchmark, the performance of AI and assess the trade-offs. For instance, what do evaluators compare the performance of AI to? This

[24]The International Justice and Public Safety Network, *Privacy Impact Assessment Report for the Utilization of Facial Recognition Technologies to Identify Subjects in the Field* (June 30, 2011).

[25]Clare Garvie, Alvaro Bedoya, and Jonathan Frankle, Georgetown Law Center on Privacy & Technology, *The Perpetual Line-up: Unregulated Police Face Recognition in America* (Oct. 18, 2016).

participant stressed that the “baseline” is current practice, not perfection—i.e., how humans are performing now, absent AI. Furthermore, this participant continued, we do not have a firm understanding of current practice. On the other hand, as this participant emphasized, “[i]f we have to benchmark [AI] against perfection, as they say, the perfect will be the enemy of the good and we get nowhere.” According to this participant, implementing AI will involve trade-offs. These trade-offs include accuracy, speed of computation, transparency, fairness, and security. Several participants at the forum emphasized that such regulatory questions should be resolved by a variety of stakeholders, including economists, legal scholars, philosophers, and others involved in policy formulation and decision making, and not solely scientists and statisticians.

Similarly, as another participant noted, there is not a clear understanding of how humans drive automobiles—the baseline—which makes it difficult to determine whether automated vehicles really are safer. Some data do exist, according to this participant, but they are proprietary and very difficult to access. To help obtain data sets about the way people drive, another participant suggested exploring public–private partnerships encompassing insurers that have collected data on driving behaviors. Another participant responded that it will be important to understand the context around the data. For example, if data suggest that a driver has slammed on the brakes, then what was happening that caused the driver to do that? What were the external factors? Was the driver paying close attention and slammed on the brakes to avoid a child running into the road?

Another participant said that it is not clear how human drivers and their vehicles affect the roads, bridges, and other transportation infrastructure, nor how automated vehicles, which can operate at all hours of the day, will affect it.

Other policy considerations. In addition to policies for incentivizing data sharing, improving safety and security, updating the regulatory approach, and assessing acceptable risks and ethical decision making, the participants at our forum also pointed out other policy issues. They emphasized that policymakers should consider a variety of other policies that could aid the widespread adoption of AI and mitigate its potential negative consequences.

In implementing AI, for example, one participant said that it should be a requirement that AI developers test for disparate impact before deploying

their technology. This participant noted that such a requirement would be better complied with if the developer was not held liable for the impact. Rather, creating “safe harbors” in conjunction with testing would allow developers an opportunity to seek out input from others to address disparate impacts. Another participant said that it would be desirable to find ways to not only share data, but also best practices associated with using the data, including implementing and testing AI systems.

Another participant highlighted the policy issue of the “information haves and have nots,” or the “digital divide.” This participant noted that many U.S. households, particularly those with lower incomes, do not have access to the Internet and that if people are expected to become more knowledgeable and better trained to work in jobs, such as those that are augmented by AI, then all families need Internet access. In addition, some of the participants said that policymakers will need to consider what to do about the potential displacement of workers, including training programs.

Concerning resources for research, one participant said that there is a gap between private- and public-sector research and that the public sector needs to work toward closing that gap. Otherwise most of the research that is conducted on AI will be to the benefit of the private companies that invest in it.

Forum Participants Highlighted Several Areas Related to AI That Could Benefit from More Research

Establishing regulatory sandboxes. One participant emphasized that in finance there is a worldwide movement to create so-called regulatory sandboxes, where regulators are encouraged to innovate. Innovation can solve some problems, while creating others. Thus a regulatory laboratory can provide a means for letting regulators begin small-scale experimentations and empirical testing of new ideas.

As this participant explained, regulatory sandboxes would provide a safe haven to try new ideas, or to assess the results of alternative regulatory approaches. Financial regulators are trained to be risk averse. However, as discussed earlier by a participant, change is occurring at such a rapid pace that the traditional regulation cannot keep up, requiring a more proactive approach. According to this expert, the current financial regulatory scheme does not encourage innovation. Another participant noted the need for regulatory reform based on competing priorities among

regulators with oversight of a specific AI technology, citing the example of automated vehicles.

Developing high-quality labeled data. One participant emphasized the importance of data collection and how to obtain high-quality labeled data. This encompasses improving the quality of the data during data collection. Another participant we spoke with highlighted the merits of developing adequate labeled data sets. As data become more comprehensive and organized, or labeled, in a manner that facilitates machine learning, AI tools can produce more accurate outcomes.

Understanding AI's effect on employment and reimagining training and education. Other areas of research offered by forum participants encompassed the implications of AI on employment as well as education and training. Some forum participants offered mixed views concerning the impacts associated with AI on employment, while acknowledging the uncertainties. For instance, some forum participants noted that job losses in some areas were likely, while noting the potential for job increases in other areas. Similarly, work from researchers discusses likely declines in employment in certain job categories due to advances in technology along with continued demand in other skill areas.[26] Another participant advocated for research to better understand how jobs were changing. There is no comprehensive federal data source with information on the employment effects AI may have in manufacturing and other segments of the economy. Further, according to two participants, in the absence of a comprehensive data-collection effort, it is unclear which jobs will be created by AI, which jobs may be augmented, or which jobs are likely to be displaced by AI, emphasizing the benefits of more research in this area.

According to a 2017 report on global trends, historically, technological change has initially diminished but then later boosted employment and living standards by enabling new industries and sectors to emerge.[27] These new industries and sectors, according to this report, have created more and better jobs than the ones that were displaced. However,

[26]Carl Benedikt Frey and Michael A. Osborne, "The Future of Employment: How Susceptible Are Jobs to Computerization?," *Technological Forecasting and Social Change*, vol. 114 (2017): 254–280; and James Manyika et al., McKinsey Global Institute, *A Future That Works: Automation, Employment, and Productivity* (January 2017).

[27]Office of the Director of National Intelligence, "Paradox of Progress: Key Global Trends," *Global Trends* (January 2017).

experts have also emphasized that the development and implementation of AI in various sectors of the economy can create hardships for individuals. Certain experts highlighted concerns about the displacement of workers in some sectors and rising socioeconomic inequality.[28] Some executives of companies that have invested heavily in AI share these concerns and have called for various policy responses, such as minimum guaranteed income, taxation of robots, and improved education.[29] Several forum participants shared that, most likely, some jobs will be lost.

One participant added, more specifically, that to address concerns of AI's impact on employment, we will need to collect reliable and granular data. This participant noted a recent report that considered AI that was programmed to perform tasks, rather than jobs. The report found that by 2035, in the most extreme case, 51 percent of the tasks that were studied could be fully automated, though this would correlate to less than 5 percent of the jobs being fully automated.[30] It is possible, this participant continued, that the employees whose jobs become partially automated may see a reduction in their wages, though they may avoid being laid off altogether. This participant suggested that anticipating and identifying these sorts of longer-term trends could provide opportunities for society to respond and adjust accordingly.

Some observers have expressed concern about the dominance of large firms in segments of the information and communication technology sector as a possible threat to competition. A "winner take all" model in certain sectors of the economy can result in various adverse economic impacts, including income inequality, monopoly pricing, and lower employment. Some observers, therefore, advocate procompetitive policies.

The widespread adoption of AI also brings with it a need to reevaluate and reimagine training and education, according to some of the participants. As one participant put it, we have to reimagine training and

[28]Richard Gray, "How long will it take for your job to be automated?," British Broadcasting Corporation (June 19, 2017); Erik Brynjolfsson, "How to Thrive—and Survive—in a World of AI Disruption," *MIT SLOAN Management Review* (Mar. 1, 2017).

[29]Tim Bradshaw, "Tech Leaders at Davos Fret over Effect of AI on Jobs," *Financial Times* (Jan. 20, 2017); Andrew Ng and Neil Jacobstein, "How Artificial Intelligence Will Change Everything," *Wall Street Journal* (Mar. 6, 2017).

[30]Manyika et al., *A Future That Works*.

education and think about the kinds of preparation that people need to participate in these AI developments that are going into effect. Another participant emphasized that as part of this reevaluation, research is needed to determine why the nation's current education system seems ineffective at teaching students to think and adapt, skills that are needed in the future workforce. One participant suggested that "further research is also needed to explore new means to encourage students from low-income backgrounds, women, and minorities to sustain engagement with STEM subjects, as this is where the majority of 21st century jobs will be."

Further, participants emphasized some specific issues for AI research tied to training and education. For instance, as one participant put it, "How do you get humans and AI systems into the same loop? How do you allow them to share the necessary information to make full use of this partnership?" Another participant emphasized the importance of research on "adversarial AI,"[31] encompassing the intersection of machine learning with computer security, which refers to the presence of intelligent and adaptive adversaries that can manipulate data and employ AI to exploit vulnerabilities of algorithms to compromise system security. This participant noted that adversarial AI involves designing algorithms with attackers in mind and that, with respect to cybersecurity, for example, one has to imagine that attackers are going to be clever in all kinds of ways and themselves have access to AI.

Exploring computational ethics and explainable AI. According to one participant, designers of AI are going to have to build frameworks and ethical systems that can reason without being told explicitly what to do. This participant stated that we will have to design systems that are going to operate in environments where we cannot anticipate in advance all the things that could go wrong. If a system makes a mistake, it could have an ability to inspect why it did something. For instance, it could have the ability to compare the expected and actual benefits and costs. Then, adjustments could be made if an incorrect outcome is discovered. Explainable AI and computational ethics are relevant for all places where AI systems are interacting with the physical world.

This same participant emphasized that more research is needed to better understand the trade-offs, costs, and benefits or mistakes related to

[31]For this report, we refer to adversarial AI as the field of research encompassing AI technology and computer security, where AI tools may be employed to either attack and bypass protections of computer systems or protect them from attack and intrusion.

computational ethics. More specifically, one participant stated that there has not been enough government-funded research into how AI can affect society, the economy, national defense, and public safety.

AI researchers are establishing rules of their own governing the use of AI. For example, some groups of technologists, such as the Asilomar AI Principles, OpenAI, and the Partnership on AI, have created sets of ethical considerations.[32] In addition, researchers from six institutions recently formed a group called PERVADE (Pervasive Data Ethics for Computational Research), whose mission is to develop a clearer ethical process for big-data research for use by both universities and private companies.

Some of the participants in our forum expressed concern with such nonbinding solutions. One participant noted that the current and future developers of AI systems may operate by ethical standards or adhere to certain morals or values that may not be compatible with the rest of society or representative of those who will use the AI.

[32]Mehr, “Artificial Intelligence for Citizen Services and Government.”

Appendix I: Forum Agenda

Day One: Thursday, July 6, 2017

8:30—9:00

ARRIVALS, CHECK-IN

9:00—9:20

OPENING SESSION

Welcome: James-Christian Blockwood, Managing Director, Strategic Planning and External Liaison (5 minutes)

Meeting Logistics: Walter Vance (moderator), Assistant Director, Applied Research Methods (5 minutes)

Overview: Timothy M. Persons, Chief Scientist, GAO (10 minutes)

9:20—9:35

SIGNIFICANCE TO THE NATION

Michael Wellman: (15 minutes)

Why is AI important now, and what should we, as a nation, be doing about it?

9:35—10:40

REVIEW OF PROFILES

Timothy M. Persons: (5 minutes)

- Cybersecurity—Discussants: Kathleen Fisher, Uday Veeramachaneni (7 minutes each)
- Automated vehicles—Discussants: Chris Gerdes, Jane Lappin (7 minutes each)
- Criminal justice—Discussants: Jason Tashea, Richard Berk (7 minutes each)

- Financial services—Discussants: Tom Gira, Wes Helms (7 minutes each)

10:40—10:50

FORUM DISCUSSION (ALL)

10:50—11:00

BREAK

11:00—11:10

COMPTROLLER GENERAL REMARKS

The Honorable Gene Dodaro, Comptroller General of the United States (10 minutes)

11:10—11:30

REPORTS FROM THE FIELD

Babak Hodjat (8 minutes)

Jack Clark (8 minutes)

11:30—12:00

FORUM DISCUSSION (ALL)

12:00—1:00

BREAK FOR LUNCH

1:00—1:30

FUTURE PROMISE, BENEFITS

How and why is it important to capture the benefits of AI?

- Jo Ann Barefoot—Finance/civil society benefits (7 minutes)
- Jason Tashea—Criminal justice/civil society benefits (7 minutes)

- Laurel Riek—Human-robot teaming/civil society benefits (7 minutes)
- Robert Seamans—Economic benefits (7 minutes)

1:30—2:15

FORUM DISCUSSION (ALL)

Future Promise and Benefits of AI

2:15—2:30

BREAK

2:30—3:10

CHALLENGES TO CAPTURING THE BENEFITS OF AI

- Natalie Vanatta—Cybersecurity challenges (7 minutes)
- Michael Wagner—Autonomous vehicle/complex software safety (7 minutes)
- Robin Feldman—Patent law, legal challenges (7 minutes)
- Oliver Richard—Economic considerations (7 minutes)
- Michael Garris—Planning for AI research at the national level (7 minutes)

3:10—4:05 FORUM Discussion (ALL)

Challenges to Capturing the Benefits of AI

4:05—4:20

USING AI FOR REAL WORLD IMPACT

Milind Tambe (15 minutes)

4:20—4:30

DAY 1 SUMMARY AND WRAP UP

4:30

DAY 1 ADJOURN

Agenda

Day Two: Friday, July 7, 2017

8:30—8:45

CONVENE AND REVIEW AGENDA

8:45—9:25

MAXIMIZING BENEFITS and ADDRESSING CHALLENGES

What are options for the way forward to address risks and challenges?

- Jane Lappin—Future of automated vehicles (7 minutes)
- Laurel Riek—Robotics technology in daily life (7 minutes)
- Solon Barocas—Accountability and fairness in machine-learning decision making (7 minutes)
- Michael Wellman—Maximizing economic benefits (7 minutes)
- Jack Clark—AI in the greater interest of humanity (7 minutes)

9:25—10:15

FORUM DISCUSSION (ALL)

Exploring Options for the Way Forward

10:15—10:30

BREAK

10:30—11:00

SUMMING UP AI FOR POLICYMAKERS

What key messages should be heard by policymakers and other stakeholders? What research priorities and policy options are needed?

- Kathleen Fisher—Cybersecurity (7 minutes)
- Chris Gerdes—Autonomous vehicles and transportation (7 minutes)
- Richard Berk—Criminal justice (7 minutes)
- Jo Ann Barefoot—Financial services (7 minutes)

11:00—12:00

FORUM DISCUSSION (ALL)

What research priorities and policy options are needed?

12:00—12:10

BREAK

12:10—1:00

SUMMARIZE and WRAP UP (ALL)

What key messages should be heard by policymakers and other stakeholders?

1:00

ADJOURN

Appendix II: List of Forum Participants

Host

Gene L. Dodaro, Comptroller General of the United States

Participants

Jo Ann Barefoot, CEO, Barefoot Innovation Group, former Senior Fellow at the Center for Business & Government, Harvard Kennedy School, Cambridge, MA.

Richard Berk, Professor of Criminology and Statistics, Chair, Department of Criminology, University of Pennsylvania, Philadelphia, PA.

Solon Barocas, Researcher, Microsoft Research Lab, Fairness, Accountability, Transparency, and Ethics in AI group, New York, NY.

James-Christian Blockwood, Managing Director, Strategic Planning and External Liaison, U.S. Government Accountability Office, Washington, DC.

Jack Clark, Director, Strategy & Communications, OpenAI, San Francisco, CA.

Robin Feldman, Director, Institute for Innovation Law, University of California Hastings College of the Law, Hastings, CA.

Kathleen Fisher, Professor and Chair, Computer Science Department, Tufts University, Medford, MA.

Michael D. Garris, Co-chair of the National Science and Technology Council's Subcommittee on Machine Learning and Artificial Intelligence, Senior Scientist, National Institute of Standards and Technology, Gaithersburg, MD.

J. Christian Gerdes, Professor, Mechanical Engineering, Director, Center for Automotive Research, Stanford University, Stanford, CA.

Tom Gira, Executive Vice President of Market Regulation and Transparency Services, Financial Industry Regulatory Authority (FINRA), Washington, D.C.

J. Wesley Helms, North America Cognitive Business Solutions and Financial Services Lead, IBM Watson, Chicago, IL.

Babak Hodjat, Founder and CEO, Sentient Technology, San Francisco, CA.

Jane Lappin, Founder and Co-Chair of the Automated Vehicles Symposium and Director of Public Policy for the Toyota Research Institute, Los Altos, CA.

Timothy M. Persons, Chief Scientist, U.S. Government Accountability Office, Washington, DC.

Oliver Richard, Chief Economist, U.S. Government Accountability Office, Washington, DC.

Laurel Riek, Associate Professor, Computer Science and Engineering, University of California San Diego, San Diego, CA.

Robert Seamans, Associate Professor, Stern School of Business, New York University, New York, NY.

Milind Tambe, Professor, Computer Science and Industrial and Systems Engineering, Co-Director, Center for AI in Society University of Southern California, Los Angeles, CA.

Jason Tashea, Founder and Director, Justice Codes, Criminal justice technology consultant, Research and Evaluation Center, John Jay College of Criminal Justice, New York, NY.

Natalie Vanatta, Deputy Chief of Research and Assistant Professor, Army Cyber Institute, Naval Postgraduate School, West Point, NY.

Uday Veeramachaneni, Co-Founder and CEO, PatternEx, San Jose, CA.

Michael Wagner, Co-Founder and CEO, Edge Case Research, Pittsburgh, PA.

Michael P. Wellman, Professor, Computer Science & Engineering, University of Michigan, Ann Arbor, MI.

Appendix III: List of Other Experts Consulted

This list includes subject-matter experts interviewed in preparation for, or following, the forum, as well as experts who reviewed the draft report. We list regulatory officials from the banking, securities, and insurance industries separately at the end of this appendix. For the list of experts who participated in the Comptroller General forum, see appendix II.

Ignacio Arnaldo, Chief Data Scientist, PatternEx, San Jose, CA.

Susan Athey, Economics of Technology Professor, Graduate School of Business, Stanford University, Stanford, CA.

Alvaro M. Bedoya, Executive Director, Center on Privacy and Technology, Georgetown Law, Washington, DC.

Mark H. Bergstrom, Executive Director, Pennsylvania Commission on Sentencing, State College, PA.

Jeff Brantingham, Co-Founder, PredPol, Santa Cruz, CA.

David Brumley, Professor, Electrical and Computer Engineering Department, and Director, CyLab, Carnegie Mellon University, Pittsburgh, PA.

Erik Brynjolfsson, Professor of Information Technology, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA.

Rachel Carpenter, Co-Founder, Chief Executive Officer, Intrinio, St. Petersburg, FL.

Sandeep Chennakeshu, President, Blackberry Technology Solutions, Waterloo, Ontario.

Michael Chui, Partner, McKinsey Global Institute, San Francisco, CA.

Trevor Darrell, Professor, Department of Computer Engineering and Faculty Director of Partners for Advanced Transportation Technology, University of California–Berkeley, Berkeley, CA.

Colby Dolly, Crime Analysis Supervisor, St. Louis County Police Department, Clayton, MO.

Andrew Ferguson, Assistant Professor, Law School Teaching Services, University of the District of Columbia, Washington, DC.

Joseph French, Co-Founder, Chief Financial Officer and President, Intrinio, St. Petersburg, FL.

Joshua Gans, Jeffrey Skoll Chair in Technical Innovation and Entrepreneurship, Rotman School of Management, University of Toronto, Toronto, Canada.

Claire Garvie, Associate, Center on Privacy and Technology, Georgetown Law, Washington, DC.

Mary Gustanski, Chief Technology Officer, Delphi Technologies, Troy, MI.

John S. Hollywood, Senior Operations Researcher, RAND Corporation, Arlington, VA.

John Launchbury, Chief Scientist, Galois, Portland, OR.

Jennifer Lynch, Senior Staff Attorney, Electronic Frontier Foundation, San Francisco, CA.

Brian Martin, Senior Director of Research and Technology, MorphoTrust USA, Billerica, MA.

Greg Morrisett, Dean, Computer and Information Sciences Department, Cornell University, Ithaca, NY.

Andrew Ng, Co-Founder of Coursera, Adjunct Professor of Computer Science, Stanford University, Stanford, CA.

Sarah Picard-Fritsche, Deputy Director, Research Practice Strategies, Center for Court Innovation, New York, NY.

Gill Pratt, Chief Executive Officer, Toyota Research Institute, Los Altos, CA, and Fellow, Toyota Motor Corporation, Toyota City, Aichi Prefecture, Japan.

Travis Reed, Chief Marketing Officer, PatternEx, San Jose, CA.

Steven E. Shladover, Program Manager, Partners for Advanced Transportation Technology, University of California–Berkeley, Richmond Field Station, Richmond, CA.

Vitaly Shmatikov, Professor, Computer Science Department, Cornell University, Ithaca, NY.

Daniel C. Smith, Senior Regulatory Advisor, Waymo, Mountain View, CA.

Brian Walter, Global Industry Leader, Watson Customer Insights and Cognitive Experience, IBM Watson Financial Services Solutions, Industry Platforms, New York, NY.

Regulatory Officials from the Banking, Securities, and Insurance Industries:

Lazaro Barreiro, Director for Governance, Operational Risk Policy; Bethany Dugan, Deputy Comptroller for Operational Risk; Steven Key, Associate Deputy Comptroller for Bank Supervision Policy; Tom Melo, Associate Deputy Comptroller for Enterprise Governance Operations; Mark Williams, External Audit Coordinator, Enterprise Governance; and Robert Wright, Bank Examiner, Information Technology, Office of the Comptroller of the Currency.

Kathy Bateman, Information Technology Specialist; David Dimitrious, Senior Special Counsel; Dan Fisher, Branch Chief; Gordon Fuller, Counsel; Paula Jenson, Chief Counsel; Jennah Mathieson, Managing Executive; Brice Prince, Special Counsel; Roxanne Ramnauth, Deputy Managing Executive; Joanne Rutkowski, Assistant Director, Assistant Chief Counsel; Steve Samson, Information Technology Manager; Shauna Sappington, Senior Special Counsel; and Tim White, Special Counsel of the Division of Trading and Markets, Securities and Exchange Commission (SEC).

Marco Enriquez, Operations Research Analyst, and Austin Gerig, Assistant Director of the Division of Economic and Risk Analysis, SEC.

Michael Hershaft, Senior Special Counsel; Alan Lenarcic, Branch Chief; Akrivi Mazarakis, Branch Chief; Joseph Murphy, Attorney-Advisor; Jianqi Wang, Quantitative Research Analyst; Elcin Yildirim, Assistant Director, Office of Compliance, Inspections, and Examinations, SEC.

Kristy L. Croushore, Senior Director, Office of Government Affairs; John Kroeper, Executive Vice President; Steve Randich, EVP and Chief Information Officer; and Vincent Saulys, Senior Director, Market Regulation, Financial Industry Regulatory Authority (FINRA).

Denise Matthews, Director of Information Systems, Data Collection and Statistical Analysis; Scott Morris, Chief Technology Officer; Eric Nordman, Director of Regulatory Services; and Brooke Stringer, Government Relations Policy Advisor, National Association of Insurance Commissioners.

Appendix IV: Profiles of AI in Cybersecurity, Automated Vehicles, Criminal Justice, and Financial Services

- Profile 1: AI in Cybersecurity
- Profile 2: AI in Automated Vehicles
- Profile 3: AI in Criminal Justice
- Profile 4: AI in Financial Services

This appendix reproduces four profiles sent to forum participants in a reading package, in advance of the forum, which was held on July 6–7, 2017, at the National Academy of Sciences Keck Center, Washington, D.C.

Profile 1: AI in Cybersecurity[1]

Current and potential uses of AI in cybersecurity

The use of AI in cybersecurity has the potential to increase the ability of organizations to prevent and respond to a widening cyber-attack space. At the same time, deploying expert systems or machine learning tools for cybersecurity faces several challenges. Further, these systems or tools themselves may be vulnerable to new kinds of attacks or manipulation.

Automated systems and advanced algorithms can help cybersecurity professionals in a variety of ways. For example, these systems can help reduce the time and effort it takes to perform key cybersecurity tasks, such as:

- identifying vulnerabilities,
- patching vulnerabilities,
- detecting attacks, and
- defending against active attacks.

Moreover, as expert systems and machine learning techniques advance, they have the potential to improve overall security performance and provide better protection from an increasing number of new and sophisticated cyber threats. These improvements can include reducing false positives and vulnerabilities associated with "alert fatigue."

Researchers and security firms are exploring many ways to use expert systems and machine learning techniques for cybersecurity. The following three examples demonstrate several of the approaches that researchers are currently using.

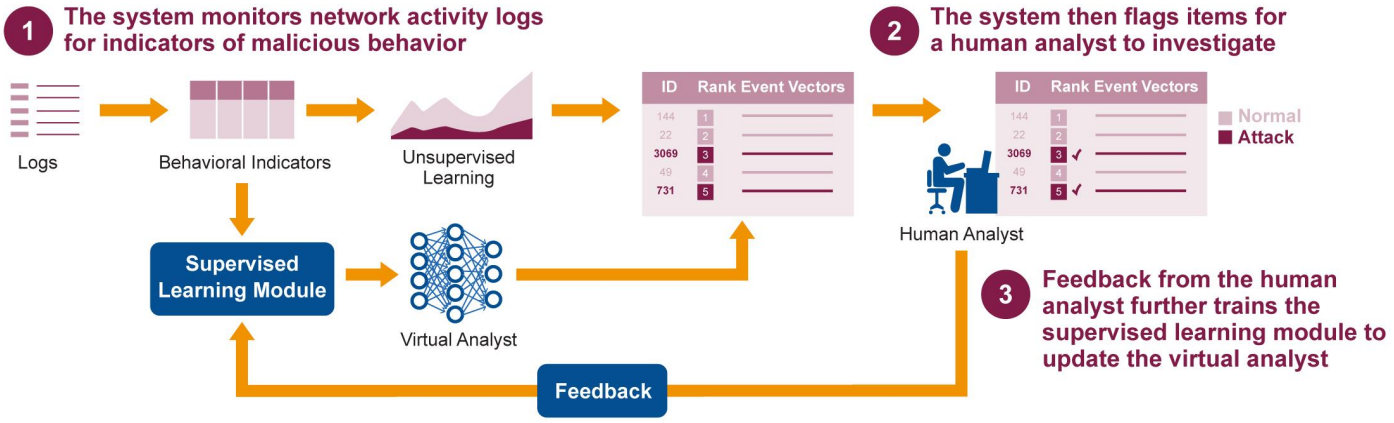
Autonomous exploit detection and repair. One research group developed a system that can find and patch system vulnerabilities without human intervention. Mayhem, the winning system in the Defense Advanced Research Projects Agency (DARPA) 2016 Cyber Grand Challenge, is designed to protect apps (software) from new attacks by hackers. Mayhem works by hardening applications and simultaneously

[1]This cybersecurity profile was designed as a primer, circulated in an advance reading package to participants, and aimed to spur discussion at the 2017 Comptroller General Forum on Artificial Intelligence.

and continuously looking for new bugs[2] that may be exploited by hackers. When new bugs are found, the bot autonomously produces code to protect the software vulnerability. Mayhem is an expert system that performs prescriptive analytics, where machines detect and interact without human intervention. This is in contrast to traditional signature-based intrusion detection systems, which rely on human intervention in anticipating cybersecurity attacks.

Machine learning with human feedback. Another approach incorporates machine learning with human expertise to build a predictive model of cyber attacks. Central to the technology is a recurring cycle of feedback between a human cybersecurity analyst and a continually updated supervised learning module. As shown in figure 3, the AI system uses both unsupervised and supervised machine learning to conduct analysis of potential threats.

Figure 3: Illustration of Machine Learning with Human Feedback for Cybersecurity



Source: GAO, adapted from video, Veeramachaneni, Arnaldo et al., *AI2: Training a Big Data Machine to Defend* (https://www.youtube.com/watch?v=b6Hf1O_vpwQ). | GAO-18-142SP

In this example, the system begins by using an unsupervised learning module to identify anomalies in data that track users' on-line behavior. Behavioral events are ranked according to specified criteria, and a human analyst then labels the events as "normal" or as reflective of a specific type of attack. Labeled events data are fed back into a supervised learning module,[3] which then generates a model that is used to predict

[2]Bugs are coding errors in software that can cause unexpected results.

[3]Given the analyst's feedback, the supervised learning module learns a model that predicts whether a new incoming event is normal or malicious. As more feedback is gathered, the model is constantly refined.

where an attack could occur in the near future. Systems that incorporate machine learning with a human feedback loop are one method being used to improve performance in cybersecurity applications by detecting more events and reducing the number of false positives.[4]

Adversarial machine learning. This approach acknowledges that there are potential vulnerabilities in machine learning-based systems. This has particular relevance in the cybersecurity arena, where adversaries would be highly motivated to exploit cyber defense systems that are based on machine learning algorithms. For example, adversaries could attempt to pollute data that is used to re-train machine learning algorithms. Other attacks might attempt to trick machine learning algorithms by repeatedly testing for and then exploiting blind spots in the learning module of an algorithm. Adversarial machine learning refers to efforts to make a machine learning system robust against such attacks. Specifically, researchers have designed tools to test and strengthen machine learning systems used in cybersecurity applications. These tools take on the role of an adversary as they try to trick machine-learning systems in order to identify weak spots an attack might expose. Once a weak spot has been identified, the software can advise the security analyst to implement changes to the machine-learning system.[5]

Benefits and challenges

The use of artificial intelligence in cybersecurity holds promise in helping an organization protect its data from malicious activity. Sophisticated, automated algorithms can comb through massive volumes of data on users' electronic behavior and network traffic that a human analyst would be unable to accomplish in a similar amount of time. Automated algorithms could also provide faster and more accurate detection of cyber threats, as well as insights to help eliminate underlying vulnerabilities. Moreover, as malevolent actors employ automated tools to launch attacks over an ever-widening attack surface, AI systems can help security professionals to match defensive tools to the means of attack.

[4]Testing of one model showed that combining unsupervised learning with supervised learning detected 85 percent of attacks, or roughly three times more attacks, compared to an algorithm using only unsupervised learning. The supervised learning model also reduced the number of false positives by a factor of five.

[5]In a project titled "Security Evaluation of Machine-Learning Systems," Ben Rubinstein, a professor at the University of Melbourne in Australia, created a software tool that can enable organizations and agencies to test their machine systems' defenses.

Nevertheless, those fighting cyber attacks face numerous challenges. Adversaries in cyberspace are many and range from single individuals and private organizations to governmental organizations and foreign entities. Cyber threats can arise from malicious actors seeking financial, political, or military gain. Fully successful defenses against cyber attacks must block all attempts to infiltrate a system, while an attack only needs to succeed once to be effective. This inherent asymmetry to cybersecurity does not disappear when AI and automated tools are used for defense. Against this backdrop, numerous challenges and risks exist related to the use of expert systems and machine-learning techniques for cybersecurity. These challenges and risks include:

- **Human intervention.** Despite some efforts to build autonomous systems, expert systems and machine-learning approaches still require human interaction with the ongoing operation or periodic refinement of the cybersecurity tool. While the number of human operators may be reduced, having human operators with the right skills will remain a priority, especially if the long-term efficacy of the system is dependent on humans helping to contribute training data to an algorithm or interpreting the algorithm's output.
- **Data privacy.** Machine-learning systems require data, and significant questions arise about the appropriate and legal use of data, especially personal data, in cybersecurity applications. Rules about what personal data is used, transparency about why and how that data is used, guidelines for disclosure of data use, and insight into how conclusions are reached will be important topics to address in the deployment of machine-learning systems.
- **Susceptibility of machine-learning models to leakage of training data.** Machine-learning systems make use of data about individuals across various domains, including purchases, preferences, photographs, health information, and more. Researchers at Cornell University have shown that an adversary can use machine learning to train an inference model to detect whether specific data was used to train a machine-learning model, thereby compromising an individual's personal data.[6]

Automating computer network defense offers many potential gains in terms of efficiency and effectiveness, yet automated systems themselves are susceptible to a range of disruptive and deceptive tactics that might

[6]Reza Shokri et al., "Membership Inference Attacks Against Machine Learning Models,0" accessed June 20, 2017, at http://www.cs.cornell.edu/~shmat/shmat_oak17.pdf.

be difficult to anticipate or quickly identify. These threats are amplified by the ongoing delegation of decision making, sensing, and authentication roles to potentially vulnerable automated systems. Moreover, broader deployment could become riskier as the reliance on autonomous decision-making increases. Risks to deployments can include:

- **Vulnerability to cyber attacks.** When defending directly against a human with clear circumvention goals, machine-learning systems are at risk because at some level they depend on human interaction and training. In essence, adversaries could attempt to create vulnerabilities by influencing automated and human-guided efforts designed to help algorithms learn and improve against cyber attacks.
- **Attack automation.** Technologies used to create advanced, automated attack capabilities are global in nature and commercially available. Sophisticated automated defense systems may become vulnerable to sophisticated automated attack systems.

Policy considerations

In addition to challenges associated with data privacy (noted in previous section), experts with whom we spoke raised policy considerations that included:

- **The challenge of encouraging both innovation and security in autonomous systems.** As expressed by one cybersecurity expert we interviewed, expecting companies to develop safer, more secure products that rely on machine learning or other AI technologies may be unreasonable since it could raise the costs of those products. In addition, the expert noted that advertising products as “safer” may make them a greater target for hackers. A challenge confronting policymakers will be creating appropriate regulations to enhance the security of new technologies that both is proactive and yet does not stifle innovation.
- **Machine-learning algorithms may not be designed to adhere to legal requirements or ethical norms.** Another cybersecurity expert we interviewed noted that machine-learning algorithms are increasingly being used for decision making in a variety of domains that are bound by specific legal requirements, including finance, employment, criminal justice, and healthcare, among others. Algorithmic decision-making practices may need to be assessed to determine whether outcomes are consistent with legal requirements or ethical norms.

Profile 2: AI in Automated Vehicles[7]

Current and potential uses of AI in automated vehicles

Automakers and technology firms are working to develop and deploy automated vehicles. Driving automation technologies—generally a combination of sensing, computational, and other technologies—perform or help perform functions that conventionally are the domain of human judgment and control. For example, these technologies are being designed to assist a driver with specific tasks, such as staying within a travel lane or parking, or to perform all driving tasks without human intervention. Vehicles with some of these technologies are already on the road. While predictions vary, observers estimate that automated vehicles (AV) that perform most driving functions could be five years away, and fully self-driving vehicles may be available in a decade.[8] These vehicles will likely mix with conventional vehicles for decades to come.

The National Highway Traffic Safety Administration (NHTSA) within the Department of Transportation (DOT) adopted the Society of Automotive Engineers' six-tiered framework to describe progressive "levels" of automation incorporated into a vehicle (see table 1). Automation can also be conceptualized across two dimensions: one in which the automated systems are increasingly responsible for vehicle control, and one in which automation increasingly provides assistance for a driver who retains some driving functions.[9]

Table 1: Levels of Driving Automation Adopted by National Highway Traffic Safety Administration

Level	Name	Definition	Example(s)
0	No automation	Human driver controls all aspects of dynamic driving tasks, even when enhanced by warning system	Conventional vehicles

[7]This automated vehicles profile was designed as a primer, circulated in an advance reading package to participants, and aimed to spur discussion at the 2017 Comptroller General Forum on Artificial Intelligence.

[8]For example, see Heineke, Kersten, et al. *Self-driving car technology: When will the robots hit the road?* McKinsey & Company, McKinsey Center for Future Mobility (May 2017). In some limited applications, such as mining, driverless vehicles are already in use.

[9]Such a conceptualization helps clarify that the development of autonomous capabilities may not be linear from 0 to 5. Instead, fully self-driving cars—level 5—could come before level three vehicles, depending on how the challenges—discussed in the next section—for the different technologies are addressed.

Level	Name	Definition	Example(s)
1	Driver assistance	Automation controls one vehicle functions (e.g., steering or speed)	Adaptive cruise control Lane keep assist
2	Partial automation	Automation controls both steering and speed with driver responsible for monitoring and immediate reengagement	Tesla autopilot Audi traffic jam assist
3	Conditional automation	Automation controls both steering and speed and monitors environment; driver may be notified to reengage	Audi traffic jam pilot
4	High automation	Automation performs all aspects of dynamic driving tasks in some driving modes; driver not required to reengage	Closed campus driverless shuttle Driverless valet
5	Full automation	Automation performs all aspects of dynamic driving tasks under all roadway and environmental conditions	Driverless taxi

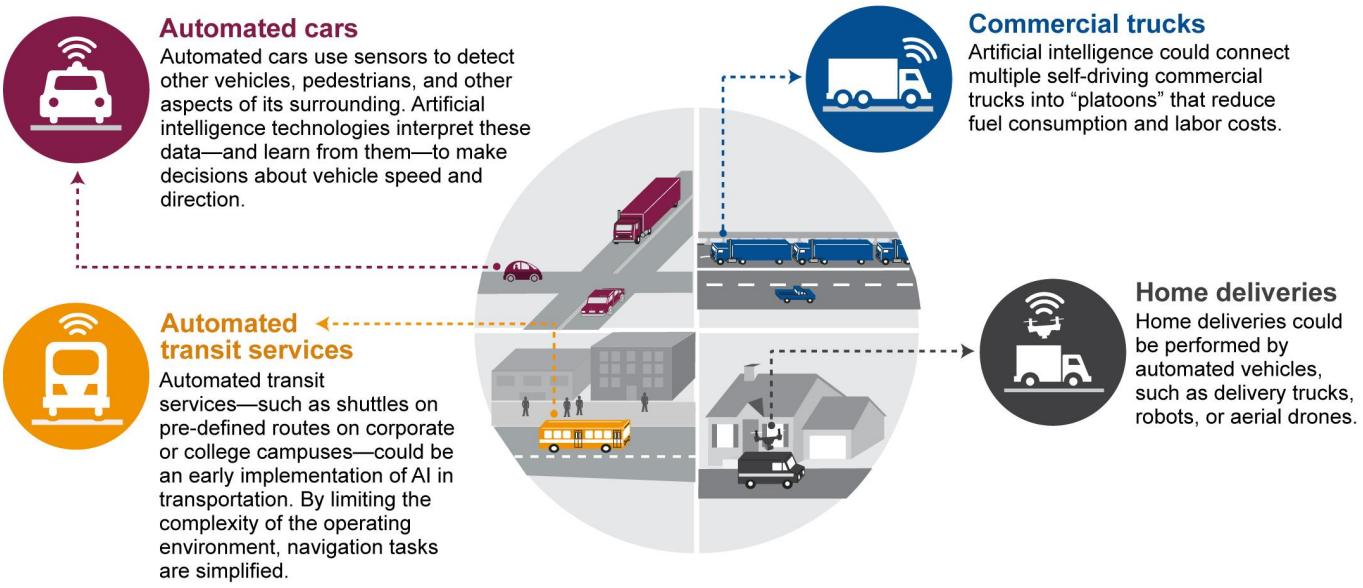
Source: GAO analysis based on U.S. Department of Transportation information as of February 2017 and additional information. | GAO 18 142SP

Artificial intelligence is a fundamental component of vehicle automation. For example, driver assistance and partial automation technologies use advanced sensors to classify and understand the vehicle's surroundings and then respond, if necessary, with minor corrective actions or alert the driver to do so. Highly-automated vehicles (HAVs)—vehicles with automation at levels 3, 4, or 5[10]—are especially likely to rely on artificial intelligence to perform the complex, data-intensive computational and decision-making functions necessary to perceive their surroundings, plan a route, and navigate. For example, AI software can assess data from sensors—such as cameras, radar, and LIDAR—and classify objects as pedestrians about to enter the street or a plastic bag floating in the wind. From this assessment, the system makes a decision and executes an appropriate vehicle response (e.g., yield to the pedestrian, ignore the bag).

The specific approaches to AI technologies under development differ and include advanced expert systems, computer vision, machine learning, or hybrid approaches, to name a few. Therefore, each system will be expected to have different data needs, capabilities, and improvement cycles. Currently, there are no commercially available HAVs, but automotive and technology firms are actively testing and piloting AI-based automated passenger cars and trucks, commercial trucks, and transit systems. For example, some firms are piloting self-driving ride-hailing services, driverless commercial trucking and platooning (in which multiple trucks closely draft in a convoy), and transit services (see fig. 4).

[10]The National Highway Transportation and Safety Administration refers to vehicles at levels 3-5 as highly-automated.

Figure 4: Illustration of Selected Applications of Artificial Intelligence in Vehicle Automation under Development



Source: Published reports and GAO interviews. | GAO-18-142SP

AI-controlled vehicles could greatly reduce human error and offer other potential benefits. By reducing or eliminating the need for a human driver, HAVs have the potential to profoundly change our nation's transportation system. The Department of Transportation contends that automated vehicles could "be the greatest personal transportation revolution since the popularization of the personal automobile nearly a century ago."[11] In an oft-cited statistic, the department observes that 94 percent of crashes can be tied to a human choice or error, and that automated vehicles could reduce or eliminate these errors.[12] Further, because they can use the sensor data and experience of thousands of other vehicles, as well as challenging simulation environments, HAVs could have a learning advantage over human drivers. While individual humans learn from their own narrow experience (and mistakes), machine learning systems can learn from the collective experience of other machines. In addition to the potential to improve safety and the experiential aspects of driving—by, for example, allowing passengers to read rather than drive—HAVs could also

[11]U.S. Department of Transportation, *Federal Automated Vehicles Policy: Accelerating the Next Revolution in Roadway Safety* (Washington, D.C.: Sept. 2016).

[12]S. Singh, *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*, Traffic Safety Facts Crash Stats Report No. DOT HS 812 115 (Washington, D.C.: National Highway Traffic Safety Admin., Feb. 2015).

have broad societal impacts in vehicle ownership, mobility, the environment, energy consumption, land use (e.g., changes in urban parking), and accessibility to transportation.[13] Indeed, some observers believe the cost of ride hailing, if fully automated, could become so low that many people may decide not to own a personal vehicle, thus reducing congestion and air pollution. However, other experts are skeptical and worry that improvements in the driving experience will lead to more miles driven, more congestion, and consequently more air pollution, rather than less. For commercial enterprises such as trucking, industry research suggests that automated trucks could improve productivity, increase freight-system capacity, and alter the federal regulatory environment.[14] By decreasing or eliminating the need for human drivers, freight operations, for example, could be less constrained by hours-of-service rules. Moreover, driverless transit services could increase the availability and efficiency of buses and shuttles, potentially improving access to transportation. Finally, as all of these changes develop, associated industries—from insurance to taxi and truck drivers to automotive manufacturers—may experience market disruptions as new products and firms emerge and compete.

Key policy challenges include safety, vehicle performance, and infrastructure adaptation

Driving is an inherently high-consequence activity; even simple errors can be fatal. Automated vehicles will need to meet public safety and performance expectations if the technologies are to be adopted. Furthermore, there is some debate about the need for infrastructure adaptations to facilitate automated vehicle performance.

- **Safety assurance.** While vehicle developers have commercial incentives to develop vehicles that work safely, automated vehicles also pose challenges for policymakers concerned about public safety. First, policymakers will need to determine how these vehicles will be deemed safe and ready for deployment. Vehicle testing by manufacturers, both on the road and in simulations, can provide some assurance that technologies work properly from a functional

[13]James Anderson et al., *Autonomous Vehicle Technology: A Guide for Policymakers* (Santa Monica, Calif.: RAND, 2016).

[14]Jeffrey Short and Dan Murray, American Transportation Research Institute, *Identifying Autonomous Vehicle Technology Impacts on the Trucking Industry* (Arlington, Va.: Nov. 2016).

perspective (e.g., can follow traffic laws, avoid hazards). Policymakers will need to determine if this assurance is sufficient.[15] Of particular concern might be "edge cases," or situations that are uncommon and may not be experienced in testing (e.g., particularly unusual traffic or road conditions). For partially and conditionally automated vehicles (levels 2 and 3), which require human drivers to be ready to take control of the vehicle if needed, automated systems will need to have effective mechanisms for human re-engagement if these systems are to work safely.[16]

Second, as with conventional vehicles, there is a public interest in identifying safety problems and investigating the causes of crashes. Whereas AI developers might look at these incidents to improve machine learning techniques, regulators and investigators may seek to understand the basis for the AI decision that contributed to the crash to determine liability or fault, and to levy fines. However, with AI systems, the underlying basis for a specific decision might be unknowable or untraceable, making assignment of fault difficult.

Third, automated vehicles and AI systems will need to be secure from malicious cyber-attacks.[17] Hacks are perhaps the most obvious example, but vulnerabilities may even come from less intrusive sources. For example, if a computer vision system can be tricked with spoofed lights or other signals, an automated vehicle (AV) might then respond unsafely.

- **Vehicle performance.** HAV developers will need to define operational standards for how HAVs will interact with other vehicles, pedestrians, cyclists, and the physical space around them. While implementation of these operational standards is a technical question, the public has a stake in the outcome and policymakers may want to contribute to how the industry manages these questions. HAVs will be

[15]All vehicles must meet Federal Motor Vehicle Safety Standards to be sold in the United States. Manufacturers currently self-certify that their vehicles meet these standards. This includes automated vehicles. However, some observers have noted that these standards may not be relevant to some AVs (e.g., the standards assume the presence of a human driver).

[16]Some observers are skeptical that adequate human re-engagement mechanism can be developed, one which incorporates sufficient predictive time horizon to successfully transition control to a distracted human.

[17]See GAO, *Vehicle Cybersecurity: DOT and Industry Have Efforts Underway, but DOT Needs to Define Its Role in Responding to a Real-world Attack*, GAO-16-350 (Washington, D.C.: March 2016).

faced with ethical dilemmas regularly and decisions will need to be made on how much allowance automated vehicles should be given to selectively break traffic laws (e.g., the speed limit) when the circumstances might call for it (e.g., mixing in freeway traffic). Moral questions, too, could merit public discussion and resolution, though some experts consider these questions to be overstated by the media. The human-machine interface also poses challenges. Specifically, passengers in automated vehicles will need some baseline understanding of what to expect. For example, passengers will need to understand if a system is not capable of safe operations at night or in poor weather conditions before using it in those domains.

- **Infrastructure adaptations.** Automated vehicles may call for changes in roadway designs. For example, infrastructure and vehicles could be equipped with communications capabilities that share data to optimize planning or mitigate crashes, provide warnings of upcoming road conditions, or other information. DOT and others foresee different types of connectivity complementing and enhancing vehicle automation.[18] Longer term, if driverless vehicles become the norm, roadway designs currently based on the needs of human drivers could be optimized for lighter weight, less crash-prone, fully-automated vehicles. For example, the width of lanes and pavement thickness could be reduced, increasing capacity and decreasing construction costs, respectively. Conversely, if truck platoons become the norm, the tighter spacing of heavy vehicles could exceed the maximum bridge weight of many of the nation's aging bridges. In urban settings, if ride-sharing becomes dominant, then parking spaces and garages may be greatly reduced. However, vehicle developers are currently pursuing applications with different degrees of reliance on infrastructure, so it is unclear how suited HAVs will be to some driving conditions. For example, HAVs that rely on lane striping and sharp edges to track the roadway may not work on unpaved roads, or roads without markings, limiting deployment in rural areas.[19]

[18]GAO, *Intelligent Transportation Systems: Vehicle-to-Vehicle Technologies Expected to Offer Safety Benefits, but a Variety of Deployment Challenges Exist,* GAO-14-13 (Washington, D.C.; Nov 2013).

[19]It is worth noting that about a third of road miles in the country are unpaved and the fatality rate on rural roads is over twice that on urban roads.

Policy considerations

Policymakers at all levels of government have differing roles to play related to the challenges just discussed. The technical challenges are largely the domain of corporate and academic researchers and developers, but a range of policy questions will likely need to be addressed that stem either from specific technical challenges or the potential socially transformative aspects of AV transportation.

The Department of Transportation—primarily through NHTSA—historically regulates vehicle safety. In September 2016, NHTSA released the Federal Automated Vehicle Policy to help address a wide range of automated vehicle challenges.[20] The policy states that vehicle safety will continue to be a federal responsibility and describes a safety assessment process in which vehicle developers would explain publicly how they have addressed safety issues in 15 areas. NHTSA also has authority to investigate and order recalls of vehicles it determines pose an unreasonable safety risk.[21] Automated vehicles present new technical challenges for the agency in executing these functions because the safety risks associated with HAVs are technologically more complex, and regulators may find it difficult to evaluate the cause of crashes. Automated vehicles also raise questions for federal entities outside DOT on topics such as workforce impacts, wireless spectrum allocation for connected vehicles, and consumer privacy protections.

State, regional, and local governments plan, build, and maintain much of the public infrastructure on which HAVs may eventually travel, but have little control over vehicle development.[22] As a result, they are faced with anticipating and preparing for a transformative technology without a direct means to fully guide it. Public sector owners of infrastructure will need to decide what infrastructure adaptations are in the public interest and when investments are appropriate. For example, current transportation demand models may not provide the foresight needed to inform such decision making. In the face of future uncertainties, transportation planners may

[20]U.S. Department of Transportation, *Federal Automated Vehicles Policy; Accelerating the Next Revolution In Roadway Safety* (Washington, D.C.: Sept. 2016).

[21]Manufacturers are primarily responsible for detecting (and reporting) defects, but NHTSA retains authority to investigate and order recalls on its own.

[22]The federal aid highway program, administered by the Federal Highway Administration, provides about $40 billion each year to states in support of their infrastructure programs.

turn to scenario planning models to help inform decision making today. To the extent public infrastructure investments are warranted, governments at all levels may need to address funding mechanisms that may not be adequate as currently formulated.

In addition to infrastructure, state and local governments traditionally handle a range of automobile-related legal issues such as driver licensing, insurance and liability, enforcement, and to some extent privacy. Automated vehicle technology has the potential to change or upend current divisions among federal, state, and local government roles.[23] For example, as AI technologies that are part of the vehicle take control of driving decisions, the role of the state in issuing driver's licenses arguably changes. States may determine they have some role in helping users to understand the limits and capabilities of HAVs. Further, some states have taken steps to regulate AVs, including directly addressing vehicle performance and testing, leading industry stakeholders to raise concerns about a patchwork of conflicting state regulations. Finally, state law enforcement and first responders will need to know how to interact with HAVs at traffic incidents, work zones, or other events where they are called to assist.

[23]Paul Lewis et al., *Adopting and Adapting: States and Automated Vehicle Policy,* Eno Center for Transportation (Washington, D.C.: June 2017).

Profile 3: AI in Criminal Justice[24]

Current and potential uses of AI in criminal justice

There are three early-stage applications of AI in the criminal justice arena that provide illustrative examples of how machine learning and modeling are being applied. In each application, algorithms are automating portions of analytical work to help provide input to human decision makers. These applications: (1) predict where crime is likely to occur to help target policing, (2) assist with identification of suspects through face recognition technology, and (3) assess the risk for recidivism when determining how long to sentence individuals convicted of crimes, as illustrated in figure 5.[25] These three applications are in use across local, state and federal levels of government and across agencies, including law enforcement and the judiciary.

- **Predictive policing.** Local law enforcement agencies are using predictive policing software to identify likely targets for police intervention and to prevent crime in specific areas. In 2016, one study reported that 20 of the largest 50 police departments in the country were using predictive policing technology and that an additional 11 were exploring options to do so.[26] These law enforcement agencies are doing this by identifying key variables, such as the locations of previous crimes and times of day crimes are generally committed in those locations, and then creating algorithms that forecast where the risk of crime is likely to be high again. These algorithms are typically coded into software, which, in turn, generates maps to display for law enforcement. These maps indicate areas with high risk crime forecasts for law enforcement officers during their patrol in order to direct policing resources to proactively disrupt criminal activity in the specified areas. In its current state, predictive policing uses models to

[24]This criminal justice profile was designed as a primer, circulated in an advance reading package to participants, and aimed to spur discussion at the 2017 Comptroller General Forum on Artificial Intelligence.

[25]According to the Department of Justice's National Institute of Justice, recidivism "refers to a person's relapse into criminal behavior, often after the person receives sanctions or undergoes intervention for a previous crime. Recidivism is measured by criminal acts that resulted in rearrest, reconviction, or return to prison with or without a new sentence during a three-year period following the prisoner's release."

[26]Robinson, David and Koepke, Logan, *Stuck in a Pattern: Early Evidence on "Predictive Policing" and Civil Rights,* ver. 1.2 (Upturn, August 2016).

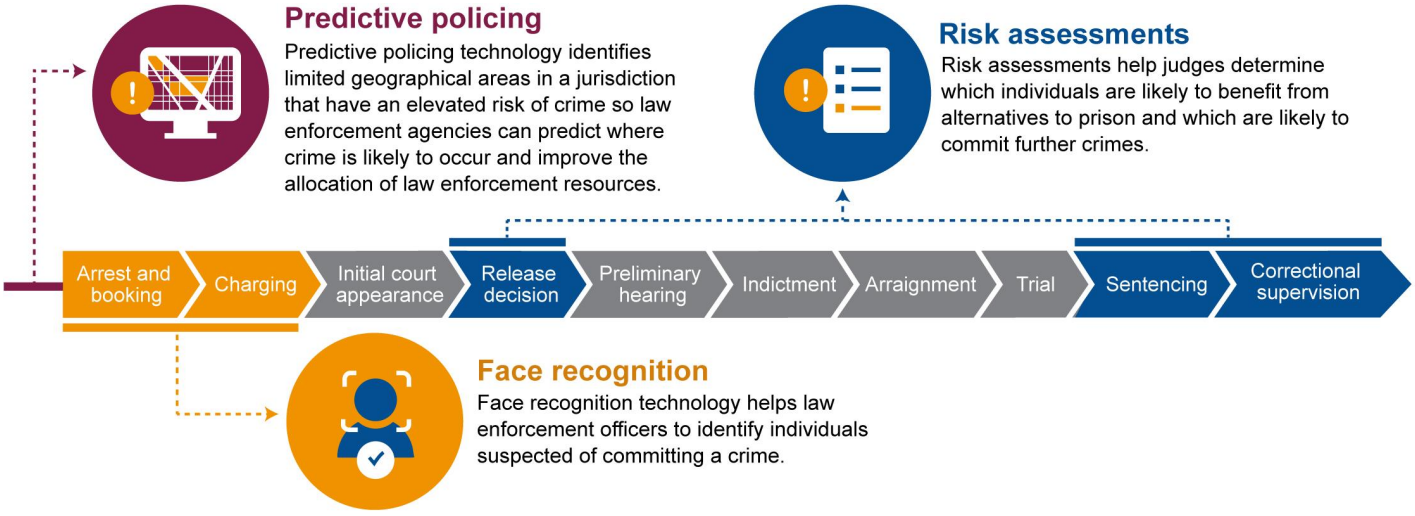
analyze large amounts of historical crime data. According to one AI expert, as more data becomes available, predictive policing software may be able to more accurately forecast crime.

- **Face recognition.** Law enforcement agencies at the state and federal level are also using advanced face recognition technology (FRT) to assist with suspect identification. Although there are no nationwide data available on the use of FRT, a recent report estimates that one in four state and local law enforcement agencies can run face recognition searches of their own databases, run those searches on another agency's face recognition system, or have the option to access such a system. [27] In addition, the same report notes that federal law enforcement agencies, including the U.S. Air Force Office of Special Investigations, the Drug Enforcement Administration, Immigration and Customs Enforcement, and the U.S. Marshals Service have all had access to one or more state or local face recognition systems. FRT automates the comparison of two or more images to determine whether they represent the same individual's face. The technology uses algorithms to first find an individual's face within a photo, then extracts features from the face—distinctive characteristics that can be numerically quantified, like eye position or skin texture—to allow it to examine pairs of faces. The technology then issues a numerical score reflecting the similarity of the features in the two photographs. Rather than producing binary "yes" or "no" answers as to whether or not a match exists, FRT identifies more likely or less likely matches. Law enforcement agencies mainly use face recognition to verify individuals' identities or to identify unknown individuals. For instance, the Federal Bureau of Investigation (FBI) could use FRT to identify a suspect by matching photos or videos from a security camera or smartphone to a database of photos containing mug shots or driver license photos. FRT can help law enforcement agencies identify criminals based on existing databases of photographs. According to officials from one company that provides face recognition applications, these machine learning algorithms are now static after being deployed to the end user but could develop into algorithms that continually update themselves as they are exposed to new data using aspects of AI.
- **Risk assessment in pre-trial release and sentencing.** Judicial officials are using increasingly sophisticated risk assessment tools to help inform pre-trial release and sentencing decisions—including how

[27] Clare Garvie et al., *The Perpetual Line-Up: Unregulated Police Face Recognition In America* (October 18, 2016).

long a new sentence should be and whether or not to release such individuals on parole after serving some of an existing sentence. According to the Marshall Project, there are approximately 60 risk assessment tools in use across the United States that vary in their levels of technological sophistication.[28] These risk assessment tools typically use static variables including the convicted individual's age, employment history, and prior criminal record, to create a score measuring the risk of re-offense. According to experts in this field, the use of machine learning for risk assessment sentencing tools is currently in the early stages of development and implementation. However, experts reported that machine learning has many potential benefits to improve sentencing outcomes. For example, one policy simulation indicated that, when decisions are based on the tool's outputs, jail populations could be reduced by 42 percent with no increase in crime rates, including violent crime.[29]

Figure 5: Use of Artificial Intelligence in Criminal Justice



Source: Published reports and GAO interviews. | GAO-18-142SP

[28]Anna Maria Barry-Jester et al., *The New Science of Sentencing*, The Marshall Project (New York, N.Y.: Aug. 4, 2015).

[29]Jon Kleinberg, et al., *Human Decisions and Machine Predictions*, National Bureau of Economic Research (Cambridge, Mass: Feb. 2017).

Challenges to using AI in criminal justice

Experts cited four main challenges to the further evolution and adoption of AI to support predictive policing, face recognition, and risk assessments in the criminal justice arena: 1) resource constraints, 2) fairness and demographic biases, 3) transparency and accuracy of machine learning, and 4) privacy and civil rights concerns.

- **Resource constraints.** Applications of AI, including predictive policing and FRT, require resources that many law enforcement agencies may not have. For example, because predictive policing maps are constantly changing, an officer must be able to examine maps in real time while on duty and often in the patrol car in order for the results of the predictive policing software to be of optimal use. This requires that officers' mobile terminals have GPS, that there is internet access while officers are in motion, and that any software or baseline functionality needed on their terminals is advanced enough to connect with the predictive policing software. Some law enforcement agencies do not have equipment that meets these requirements and therefore are not positioned to capitalize on the AI technology. Further, AI systems may require law enforcement officers to learn to use new tools, which can require training and time. In addition, in order to create or maintain the software that officers would use, some agencies would have to dedicate existing staff or hire new staff, and they may not have the money to do so. In addition, the staff assigned to interpret the data any predictive policing software would generate would need to have skills in data analysis and some agencies might not have funds to pay for the training.
- **Fairness and demographic biases.** Algorithms used in law enforcement programs may exacerbate racial biases by drawing on data that contains biased information, causing concerns over fairness and demographic biases. The use of data in this way could be used to stigmatize neighborhoods or groups of people. For example, the make-up of a training set used to develop an algorithm for FRT can influence the kinds of photos that an algorithm is most adept at examining. If a training set is skewed toward a certain racial group, the algorithm may be better at identifying members of that group as compared to individuals of other racial groups. In another example, civil rights groups have raised concerns that predictive policing systems are not adequately audited and monitored on an ongoing basis to assess if police are unjustifiably targeting specific neighborhoods. The groups argue that predictive policing algorithms

may lead to biased criminalization of communities of color by further concentrating law enforcement activities in those communities.

- **Transparency and accuracy in machine learning.** Experts have raised concerns about the difficulty of confirming the accuracy of many machine learning technologies due to the lack of transparency from technology companies that manufacture the software and build the algorithms. Further, beyond business practices, there are inherent limitations of many machine learning techniques. For example, many early stage AI approaches to law enforcement are proprietary, and their algorithms are not available to the public. In addition, we and others have raised concerns about limited testing on the systems for accuracy. In 2016, we found that FBI conducted only limited testing to ensure the accuracy of its face recognition capabilities.[30] For example, the agency had not taken steps to determine whether partner law enforcement agencies' FRT systems were sufficiently accurate and did not unnecessarily include photos of innocent people as investigative leads. We recommended that FBI take steps to improve transparency and better ensure that face recognition capabilities are being used in accordance with privacy protection laws and policy requirements and to ensure that FRT systems are sufficiently accurate.
- **Privacy and civil rights concerns.** Privacy and civil rights implications of the use of AI in the criminal justice sector appear to be widespread. For example, researchers and law enforcement officers who participated in a 2009 Department of Justice National Institute of Justice symposium on predictive policing emphasized the need for privacy policies that would ensure the constitutionality of the technique's use.[31] These participants were largely concerned about ensuring that predictive policing technology is compatible with privacy laws and policies. Similarly, a 2016 report from the Georgetown Law Center on Privacy & Technology noted that law enforcement agencies are not taking adequate steps to protect privacy and that law enforcement use of FRT is unregulated and rarely audited for misuse. Its authors urged community leaders to press for FRT policies and legislation that protect privacy, civil liberties, and civil rights.[32] GAO also reported similar issues with FBI's use of FRT, including that it has

[30]GAO, *Face Recognition Technology: FBI Should Better Ensure Privacy and Accuracy*, GAO-16-267 (Washington, D.C.: May 16, 2016).

[31]National Institute of Justice, *Predictive Policing Symposiums* (Jan. 6, 2012).

[32]Garvie et al., 2016.

not taken sufficient steps to oversee the use of the technology and ensure the accuracy of the external databases it used.[33]

Future of AI in criminal justice law enforcement

Though the evidence is limited on how, if at all, AI is enhancing the field of criminal justice, the technology's appeal is broad. As its use and technological capacity expands, there will likely be a number of improvements, including improved accuracy, the ability to use AI in real time and across different forms of media, and more effective use of the technology. In particular, the accuracy of AI predictions may improve as better and more relevant data are collected. For example, predictors of recidivism have been examined over time and include factors like educational attainment and past violent behavior. As prisons automate the collection of inmate data, risk assessment tools designed to predict inmates' future behavior outside the prison setting should have more information to process and statistically assess in order to enhance the risk assessment tools' predictive capacity. Additionally, experts also noted that, as AI capabilities improve, it may be possible to use face recognition to identify individuals through video feeds in real time.

Policy considerations

From a policy perspective, some experts on predictive policing, face recognition, and risk assessments cited the need for federal oversight needed to regulate the use of AI in the criminal justice arena. Experts raised potential policy implications in the following three key areas.

- **Transparency.** Experts agreed that a lack of transparency into the data used by proprietary algorithms can contribute to privacy, bias, and accuracy concerns. As a result, experts contended that enhancing existing federal regulations/policies and/or establishing a federal regulatory body to assess these AI applications' use could have benefits.
- **Privacy.** Federal agency collection and use of personal information is governed primarily by two laws: the Privacy Act of 1974, as amended[34] and the privacy provisions of the E-Government Act of

[33]GAO-16-267.

[34]5 U.S.C § 552a.

2002,[35] both of which were developed prior to the widespread use of machine learning in law enforcement. Participants in a 2009 National Institutes of Justice symposium on predictive policing emphasized the need for privacy policies that would ensure the constitutionality of the newer AI applications' use. One participant recommended using Privacy Impact Assessments (PIAs)[36] in developing such a policy.[37] As noted earlier, in our 2016 report on FBI's use of FRT, we recommended, among other things, that the agency take steps to ensure PIAs are published before using or making changes to a system. [38]

- **Bias and accuracy.** Some experts recommended that the federal government establish a federal regulatory agency to perform independent assessments of the accuracy and potential bias of AI systems. For example, a report from the Georgetown Law Center on Privacy & Technology underscored fairness challenges and suggested that Congress and state legislatures should regulate law enforcement's use of face recognition applications. The authors also noted that the Commerce Department's National Institute of Standards and Technology should create regular tests for algorithmic bias on the basis of race, gender, and age to provide information on algorithm accuracy and bias to decision-makers choosing among multiple vendors.[39] In addition, the RAND Corporation noted that decision assistance systems should be equipped with tools for auditing the causal factors behind key decisions. The report stressed that educating the public about the capabilities of law enforcement algorithms while ensuring that algorithms in use are easily understood would contribute to guarding against inequity and inaccuracy in their application.[40]

[35]Pub. L. No. 107-347, § 208, 116 Stat. 2899 (Dec. 17, 2002).

[36]Subsection 208(b) of E-Government Act of 2002 requires agencies to conduct PIAs that analyze how personal information is collected, stored, shared, and managed in a federal system. Agencies are required to make their PIAs publicly available if practicable.

[37]National Institute of Justice, Predictive Policing Symposiums (Jan. 6, 2012).

[38]GAO-16-267.

[39]Garvie et al., 2016.

[40]Osonde Osaba and William Welser IV, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence* (RAND: 2017).

Profile 4: AI in Financial Services[41]

Current and potential uses of AI in financial services

Many financial services firms (including those in the banking, securities, and insurance industries) have begun to integrate AI tools into their computer systems and operations. Some of these AI tools are helping to augment applications that support functions such as:

- customer service operations (automating call center functions, on-line chatbots,[42] etc.);
- client wealth management (advising financial professionals or customers directly);
- consumer risk profiling (decisions and rates tied to insurability, lending, etc.); and
- internal controls (monitoring transactions for potential fraud, regulatory compliance, etc.).

As firms continue to implement AI tools, financial service regulators are also exploring opportunities to use AI technology to enhance their oversight capabilities. For instance, securities regulators cited current efforts to introduce AI capabilities into their market surveillance tools.

Much of the AI capability cited by industry participants and regulators in financial services encompass machine learning that enhances abilities beyond existing expert systems. Machine learning enhances a computer system's ability to learn from inputs and actual outcomes.

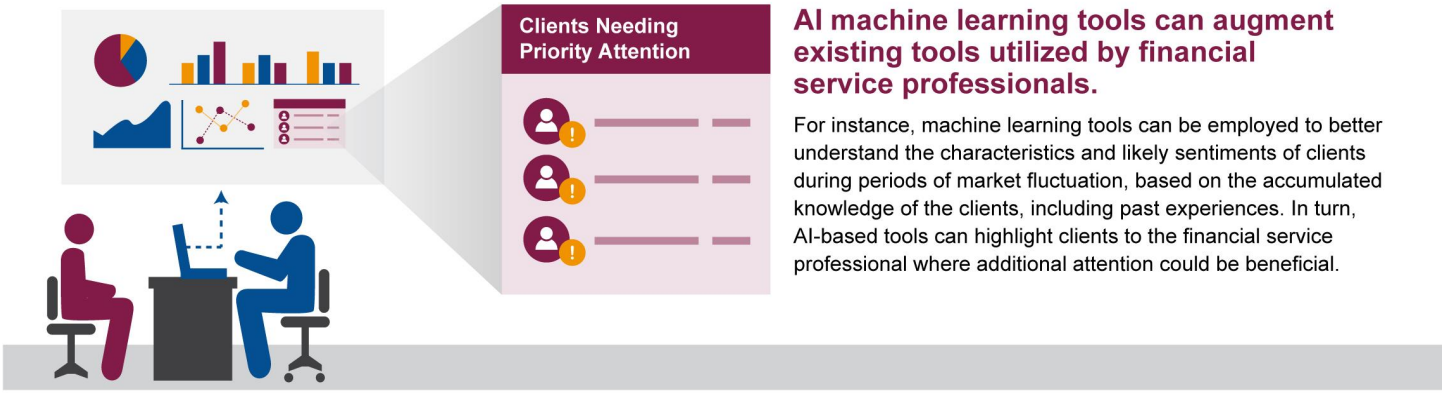
Client service. Numerous AI applications are being used by industry participants in the financial services industries. For instance, in the securities and banking industries, AI tools are being used to better understand clients' investment goals and concerns, and, in turn, customize the investment advice to offer suitable strategies to address the clients' needs. AI-based systems have been developed to augment the tools available to broker-dealers and investment advisors. These

[41]This financial services profile was designed as a primer, circulated in an advance reading package to participants, and aimed to spur discussion at the 2017 Comptroller General Forum on Artificial Intelligence.

[42]A chatbot is a program that interacts directly in a free-form conversation with users via natural language processing.

systems can draw on and analyze information from numerous sources, including communications with a client, recognizing the sentiments of the client related to risk tolerance, and client reactions to conditions in the marketplace, among others. Such tools could be used to identify clients' level of satisfaction and potentially the likelihood of exiting the business relationship based on past experiences with clients involving a similar set of circumstances. Such tools may be designed to interface with the client directly in a client-facing system, such as a robo-advisor, or designed as an advisor-facing system to augment the tools used by a financial services professional to assist in decision making.[43] Figure 6 depicts the use of information incorporating AI-based tools that highlight key characteristics of a client, offering insights to the broker-dealer about the optional investment strategy to pursue and highlighting high priority clients.

Figure 6: Illustration of Machine Learning Tools Used by Financial Service Professionals



Source: Published reports and GAO interviews. | GAO-18-142SP

In the example illustrated in figure 6, algorithms in expert systems can assess characteristics of the client to assist the broker-dealer in developing an investment strategy for the client. Such a system can use facts and trends such as the client's age, income, self-reported level of

[43]A computer system assisting a broker-dealer or investment advisor is characterized as an advisor-facing (or inward-facing) system, aiding the financial services professional inside an organization. In contrast, some organizations offer client-facing systems, such as robo-advisors, which can offer investment advice directly to the client based on input from the client. Fully automated digital wealth management platforms have features that let investors manage their portfolios without direct human interaction and typically collect information on customers and their financial history using online questionnaires. From GAO, *Financial Technology: Information on Subsectors and Regulatory Oversight,* GAO-17-361 (Washington, D.C.: Apr. 19, 2017).

risk tolerance over time, investment goals, investment time horizon, and the performance of different components of the investment portfolio along with the historical performance of various investment options to develop recommended investment and diversification options.

Further, other AI-based tools, as noted by a provider, can allow further refinement of an investment strategy and service to the client by taking into account a multitude of characteristics and circumstances unique to the individual client in the context of a changing environment. AI tools using machine learning features in the example can highlight an immediate need to give attention to a client due to a high likelihood that a client will end the business relationship soon based on numerous information inputs. For instance, the AI-based tools that analyze several aspects of recent communications with the client (e.g. assessing, among other things, the tone, choice of words, emotion, number of words, and context of the conversion) highlight client sentiments that appear to be increasingly negative. Additionally, an assessment of a client relationship using AI tools can take into consideration the accumulated knowledge about the historical nature of conversations unique to this particular client (e.g. common expressions and phrases, typical length of conversations tied to particular inquiries or actions, common tone of the discussion, and past client responses to stress conditions in the marketplace.). Further, machine learning extracts features from outcomes observed empirically. Experts in the firm who have observed successful and unsuccessful outcomes under different scenarios and conditions can share or “label” such knowledge to train the system to “learn” from past experiences.

Analysis of regulatory filings. Firms are also employing AI tools to more expeditiously capture and organize financial information obtained from required Securities and Exchange Commission filings, as filers may present and organize the information in varied formats. Accordingly, machine learning AI tools are being used to identify and organize relevant financial information from unstructured data that is captured from these filings. For instance, as noted by one data service provider, the more quickly and accurately that financial data service providers can identify and compare financial and other relevant data among firms using AI tools, the greater the value added for investors, who may desire to act on such information as expeditiously as possible.

Market oversight. Regulators in the securities industry are also exploring capabilities of AI-based tools to better understand and detect potential manipulation in financial markets. This involves “teaching” the systems various market anomalies to detect potential market manipulation. For

instance, the Financial Industry Regulatory Authority is developing a prototype AI-based system, called the Dynamic Surveillance Platform, which uses supervised machine learning capabilities to learn and detect different patterns of market anomalies to enhance the ability to detect instances of potential illegal manipulation of the securities and options markets. Methods for attempting to manipulate markets have gotten increasingly sophisticated, including schemes to intentionally place bogus buy and/or sell orders with no intention to execute the orders. This prohibited practice is known as spoofing, and occurs in a layering fashion at multiple price points.[44] Through spoofing, the fraudster attempts to influence other market participants by giving other traders a false sense of optimism regarding market supply or demand and then profiting from such market manipulation. With new AI-based tools, as well as future data enhancements to increase the visibility of each trading transaction offered by a new consolidated audit trail being developed, regulators are hopeful that employing machine learning capabilities will help identify future intentional manipulation of the markets.[45]

Due diligence and compliance. AI tools are also being employed by financial institutions in some areas, such as for due diligence and compliance activities, at a faster pace than in other areas. For instance, AI technologies offer promising capabilities to enhance financial institutions' compliance activities associated with the Bank Secrecy Act and related anti-money laundering (BSA/AML) requirements. Under BSA/AML requirements, financial institutions must satisfy the elements of the customer identification and customer due diligence programs—collectively known as "Know Your Customer"—which include having risk-based procedures for verifying the identity of each customer and conducting ongoing monitoring to maintain customer identification and

[44]Spoofing refers to entering an order to entice other participants to join on the same side of the market at a price at which they would not ordinarily trade, and then trading against the other market participants' orders. Layering refers to entering limit orders with the intended effect of moving the market to obtain a beneficial execution on the other side of the market. For instance, in 2015, financial regulators attributed the emergence of a 2010 "flash crash" experienced in financial markets to intentional market manipulative practices that included spoofing and layering. In April 2016, in an effort to help firms identify and halt spoofing activity, the Financial Industry Regulatory Authority (FINRA) began sending monthly supervision report cards to firms for which it identified potential spoofing or layering by the firms or their clients.

[45]On November 15, 2016, the Securities and Exchange Commission (SEC) approved the creation of a single, comprehensive database—a consolidated audit trail—that will enable regulators (including the national securities exchanges and FINRA) to more efficiently and accurately track trading.

identify suspicious transactions.[46] AI-based tools can enhance a financial institution's ability to understand the profile or characteristics of their customers from a variety of sources, as well as the transactions they execute, including remittance transfers to individuals in foreign countries. Remittance transfers—funds sent from individuals in one country to a recipient in another country—may be subject to BSA/AML requirements. Remittances can pose money laundering risks, as funds related to illicit activity may go undetected due to the large volume of transactions or remittance providers' inadequate oversight of the various entities involved.[47] AI tools incorporating machine learning are being implemented to understand the relationships and patterns of such transactions that may emerge between a bank's customer and other individuals, recognizing the potential for criminal activities among these transactions. Conversely, banks have been more reluctant to implement AI tools in the lending arena over concerns related to fair lending requirements, according to regulators.

Risk profiles. In the insurance industry, firms are using AI tools in a variety of ways. As in other industries, applications of AI in the insurance industry include developing risk profiles and marketing opportunities unique to an individual, and further automating call center functions. In the insurance industry, the collection and use of personal information is facilitated through the industry's limited antitrust exemption that allows insurers to pool historic loss information so that they are better able to project future losses and charge actuarially-based prices for their products.[48] This shared loss information encompasses each individual's personal loss history used by insurers to determine rates commensurate with the individual's risk profile. Such a repository of comprehensive data tied to each individual's claims and loss history is well suited for use in AI systems, which can be designed to analyze and correlate individuals' loss history with various characteristics and conditions tied to those

[46]Pursuant to the *USA Patriot Act*, all banks must have a written Customer Identification Program (CIP). The CIP is intended to enable banks to form a reasonable belief that it knows the true identity of each customer.

[47]See GAO, *International Remittances: Money Laundering Risks and Views on Enhanced Customer Verification and Recordkeeping Requirements*, GAO-16-65 (Washington, D.C.: Jan 15, 2016).

[48]The *McCarran–Ferguson Act of 1945*, Pub. L. No. 79-15, 59 Stat. 33 (1945)(codified at 15 U.S.C. §§ 1011-1015), grants insurers limited exemption from federal antitrust laws, allowing them to share and pool loss information to determine the future probability of losses and develop actuarial based prices for insurance products.

individuals. AI tools can use not only an individual's loss history, but also other data sources to create characteristics or a risk profile of the individual, including credit score information and risk factors that insurers deem relevant in assessing risk tied to the individual, from social media sources and the Internet of Things (IoT). Further, insurers can use AI-based tools, incorporating machine learning techniques, to monitor the processing of claims to better detect anomalies and potentially fraudulent activities.

Benefits and challenges

Industry participants and regulators have highlighted both benefits and challenges offered by the use of AI tools in the marketplace. As described earlier, industry participants and regulators offered examples where they have improved, or expect to improve, the performance of their organization using AI-based tools. These benefits include better service to an organization's clients, enhanced surveillance monitoring (by an entity internally as well as externally by financial regulators), and higher productivity achieved in a cost-effective manner.

For instance, one industry participant explained that a firm successfully grew its presence in the marketplace using AI tools to service a growing customer base. Traditionally, the firm established new call centers as part of its business strategy to grow its customer base. By leveraging the capabilities of AI tied to voice recognition and language processing when interacting with customers on voice calls and on-line chatbots, the firm found it could grow faster. It employed an AI-based system, which could determine those tasks it could resolve with a high degree of confidence while recognizing situations where it did not have enough confidence to resolve a customer's inquiry and, in turn, forward the customer to a service representative.

Regulators also shared their optimism for employing AI tools to better detect and prevent improper market conduct, as well as enforce existing laws and regulations in the marketplace. For instance, securities regulators expressed their hope that AI systems will "learn" characteristics of past fraudulent schemes introduced into the marketplace to help detect future schemes and anomalies in the marketplace well before a human could detect them.

At the same time, challenges and growing pains associated with technological advances of AI-based tools also exist. For instance, banking regulators and other industry observers said that banks are

reluctant to move quickly in implementing AI tools for lending operations due to concerns about meeting requirements under existing laws and regulations (e.g., requirements stemming from fair lending laws that prohibit discriminatory practices on lending, whether intentional or not, based on race, gender, color, religion, national origin, marital status, or age). Another challenge is obtaining complete and appropriately formatted data. For instance, incomplete data and data of varying formats must be structured and labeled for training certain machine learning systems. Organizing data is another key challenge that can be addressed with machine learning techniques to cluster and label data in a manner so that it can be analyzed effectively.

Regulators and industry observers also cite other challenges to attracting and retaining staff with requisite data science and machine learning skills and maintaining up-to-date hardware and software. Conversely, industry observers also expressed concern over the potential reduction in employment for tasks that AI-based systems can be taught. Privacy concerns also present challenges to full implementation of AI tools given the sensitivity of personal information and potential abuses (e.g. misuse of personal data, theft or loss of data, and data breaches experienced by private and governmental organizations).

Policy considerations

A common theme voiced by industry participants and regulators centers on privacy issues. As AI-based tools become more prolific and capable, using personal information about individuals and their interactions with society at large raises concerns on the use and protection of such data. The data are derived not only from direct communications with a business that is using AI-based tools to make a business decision impacting the individual, but also from a variety of other sources, including social media and IoT.

Concerning the ability to oversee the financial services industries, financial regulators conveyed that investment advisors still need to adhere to existing laws, regardless of the technology they employ. For instance, securities regulators cautioned that robo-advisors offering investment advice to clients must adhere to suitability requirements tied to investment recommendations.

For regulatory compliance functions, some industry observers advocate exploring alternative regulatory approaches and reporting mechanisms, leveraging the coupling of regulation with technology commonly referred

to as regtech. Such an approach would grant regulators fuller access to information in a manner that reduces the burden and costs on regulated entities. For instance, one observer maintained that the existing regulatory structure and reporting procedures were established in an era absent computer systems and vast quantities of data that can now be readily collected and analyzed using those systems. According to industry observers, financial entities generally would have less burdensome reporting requirements as regulators gain fuller access to underlying data in a more continuous and transparent fashion. In turn, regulators could be better positioned to oversee the financial condition and market conduct behavior of financial entities in a more proactive manner.

Appendix V: Scope and Methodology

This report on the emerging opportunities, challenges, and implications of artificial intelligence (AI) is based primarily on a GAO expert forum, held at the National Academy of Sciences Keck Center in Washington, D.C., on July 6–7, 2017. The report also draws on relevant literature and consultation with subject-matter experts in addition to those participating in the forum. Our methodology included (1) selecting and inviting forum participants, who had a wide range of expertise and views, with the assistance of the National Academy; (2) developing a preforum Reading Package that included four profiles of distinct areas in which AI is being used, namely cybersecurity, criminal justice, automated vehicles, and financial markets, which we sent to participants in advance of the forum; (3) convening and recording the forum and preparing an annotated outline of forum presentations and discussion, based primarily on forum transcripts, which we sent to participants for their review; (4) drafting the report based on consideration of the transcripts, participants' responses to the annotated outline of the forum proceedings, the material developed earlier for the Reading Package, and other relevant literature, as well as consultation with forum participants and other subject-matter experts; and (5) obtaining internal GAO reviews and sending a draft of the report for comment by the forum participants and two additional experts who had not participated in the forum, as an additional measure of quality control.

Selection and Report of Policy Relevant Topics

Based on recognition of ongoing trends and issues described in literature, we identified policy-relevant topics concerning new developments related to advances and increasing implementation of AI, in consultation with our Comptroller General, who decided to convene a GAO Comptroller General forum. We briefed majority and minority staff for the House Committee on Science, Space and Technology, and they wrote a request letter in support of this activity.

Objectives

The objectives of the report for this effort are to synthesize views of participants in the forum held on July 6 and 7, 2017, convened by the Comptroller General, supplemented by the views of other subject-matter experts and relevant literature, concerning the following topics:

- How has AI evolved over time, and what are important trends and developments in the relatively near-term future?
- According to experts, what are the opportunities and future promise, as well as the principal challenges and risks, of AI?
- According to experts, what are the policy implications and research priorities resulting from advances in AI?

Creation of the Preforum Reading Package

To develop common background material and help prepare for the forum, and in consultation with requester staff, we identified four key topic areas (which we refer to as profile areas), namely cybersecurity, automated vehicles, criminal justice, and financial markets. These areas were designed to represent variety in the purposes for which AI may be used, distinct potential benefits and risks, and varying levels of development with respect to the use of AI.

For each of the four profile areas, GAO analysts and specialists (1) conducted preliminary literature reviews and semistructured interviews with experts from industry, government, and academia; (2) created a draft reflecting the experts' opinions and material from literature; and (3) subjected the drafts to review by relevant stakeholders. Comments from stakeholders were incorporated in the profiles as appropriate. See appendix III for a list of the experts whom we consulted (in addition to forum participants).

We shared the four area profiles with forum invitees in advance of the Comptroller General forum and have included them in appendix IV. The reading package was not designed to be comprehensive or definitive; instead, it was developed to provide information that could stimulate discussion among a broad array of experts representing varied perspectives. The profiles also were intended to give readers of this product information about the various ways in which AI applications are currently being used and could be used in future years and to help put forum findings in context.

Lastly, we and National Academy of Sciences staff members talked briefly with additional experts that were not included in the appendix IV listing of experts.

Forum Participant Selection

To prepare for the Comptroller General forum, GAO contracted with the National Academy of Sciences to assist in selecting participants. We met with the National Academy of Sciences to help ensure balance and to help us assess potential conflicts of interest for forum participants, with GAO making final determinations regarding potential conflicts of interest.

In our initial discussions with the National Academy we agreed that forum participants should

- as a group, represent a range of backgrounds, experience, and knowledge in terms of representing (1) academia, business, government, and nonprofit organizations (such as think tanks); (2) experience with AI development across a range of areas; and (3) diverse professional backgrounds and would include, but not be limited to, experts in AI developments in cybersecurity, automated vehicles, criminal justice, and financial markets;
- from an individual perspective, be able to provide diverse perspectives on issues related to (1) economic opportunities; (2) potential impacts on jobs; or (3) concerns about data privacy, civil rights, and liberties.

Our criteria for defining an expert included one or more of the following (1) a significant position in an organization or organizations relevant to the development and implementation of AI, including a university or academic institution, nonprofit organization, business, or government agency; (2) authorship of papers in professional journals or other substantial publications relevant to the development and implementation of AI; and (3) selection to appear on an expert panel or make public presentations relevant to the development and implementation of AI.

To implement final selections for Comptroller General forum participants, first the National Academy identified potential participants based on the criteria listed above. GAO and the National Academy then met again to discuss the National Academy's list of potential participants along with other participants whom GAO felt met the requisite qualifications. This strategy allowed the National Academy an opportunity to independently identify and internally discuss potential invitees before GAO shared suggestions for potential invitees. This strategy was employed to bring increased independence to the selection process, though GAO made final determinations regarding participant selections.

Independence of Forum Participants

To exercise due diligence and to understand forum participants' potential conflicts of interest, we asked all forum participants to sign a form that asked participants about their perspectives and circumstances. Specifically, we asked participants (1) whether their immediate family had any investments or assets that could be affected, in a direct and predictable way, by a decision or action based on the information or opinions they would provide to GAO; (2) whether they or their spouse received any income or hold any organizational positions that could be affected, in a direct and predictable way, by the information or opinions they would provide GAO; (3) whether there were any other circumstances, not addressed in the two previous questions, that could be reasonably viewed by others as affecting participants' point of view on the topics to be discussed. GAO received signed responses from all forum participants to these queries, and five participants identified investments or organizational positions that they believed might be affected by the information they provided GAO. Given the overall balance among the participants (with respect to sector represented and diverse perspectives on issues such as economic impacts and data privacy) and that our report does not make any recommendations with respect to AI policy, we did not find these conflicts to be material to our report.

Comptroller General Forum and Participant Follow-up

As previously noted, the Comptroller General forum was held on July 6–7, 2017, at the National Academy of Sciences Keck Center. The forum agenda (included as app. I) allowed for considerable open discussion and flexibility. Each participant gave at least one presentation during the forum, and presentations were followed by open discussion among all participants. The forum was recorded and the discussion was transcribed.

Following the forum, we sent participants an outline of the forum presentations and discussion for their review and comment. This outline was based on a written transcript of forum proceedings and presentations delivered as part of the forum. We incorporated feedback from participant comments on the outline and postforum interaction as appropriate.

Before publication and consistent with our quality assurance framework, we provided the forum participants with a draft of our report, and incorporated their feedback on that draft as appropriate. In our report, the use of the term "forum participants" means that more than one participant contributed to the point being made. As an additional measure of quality

assurance, two additional external experts (one with expertise in the technical aspects of AI and another with expertise in the economic implications of AI) who had not participated in the forum reviewed a draft of this report and provided comments that we incorporated as appropriate.

GAO reviewed meeting transcripts and interacted with participants after the meeting as needed to (1) better understand or develop first-hand examples of some key points raised by participants in the meeting, or (2) identify references to relevant literature, or both.

Findings and Recommendations

This nonaudit engagement was designed to represent primarily the viewpoints of experts who were selected to participate in the Comptroller General forum. The experts were selected by GAO with assistance from the National Academy of Sciences to help ensure balance and independence, and the forum was designed to help ensure that all significant viewpoints were represented. While we present a summary of the forum and relevant issues, the testimonial evidence of experts in this engagement is not being used to develop GAO recommendations for executive-branch actions or to present matters for congressional consideration.

Disclosure

Forum attendees and other experts were informed that GAO would not directly identify individuals or their affiliations in association with specific comments (without their permission), and this product does not do so.

Agency Comments

We provided a draft of this report to officials at the Department of Defense, the Department of Transportation, the Department of Justice, the Securities and Exchange Commission, the Office of the Comptroller of the Currency, and the National Association of Insurance Commissioners with a request for comments. We incorporated their technical comments into the report, as appropriate.

Quality-Assurance Statement

We conducted our work from January 2017 through March 2018, in accordance with all sections of GAO's Quality Assurance Framework that

are relevant to technology assessments. The framework requires that we plan and perform the engagement to obtain sufficient and appropriate evidence to meet our stated objectives and to discuss any limitations to our work. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

Appendix VI: GAO Contacts and Staff Acknowledgments

GAO Contacts

James-Christian Blockwood, Managing Director, Strategic Planning and External Liaison, (202) 512-2639 or BlockwoodJC@gao.gov

Timothy M. Persons, Chief Scientist, (202) 512-6412 or PersonsT@gao.gov

Other leadership for this project was provided by the following:

Stephen Sanford, Assistant Director, Strategic Planning and External Liaison, and Virginia Chanley, Analyst-in-Charge, Applied Research and Methods

Staff Acknowledgments

Key contributors include the following:

Joy Booth, Assistant Director, Homeland Security and Justice

David Chrisinger, Foresight and Strategic Planning Analyst, Strategic Planning and External Liaison

Cathy Colwell, Assistant Director, Physical Infrastructure

Adam Couvillion, Senior Analyst, Homeland Security and Justice

Philip Farah, Assistant Director, Applied Research and Methods

Jonathan Felbinger, Senior Engineer, Applied Research and Methods

Miriam Hill, Senior Analyst, Homeland Security and Justice

Barry Kirby, Senior Analyst, Financial Markets and Community Investment

Maria McMullen, Visual Communications Analyst, Forensic Audits and Investigative Service

Brian Palmer, Analyst, Information Technology

Dina Shorafa, Analyst, Homeland Security and Justice

Candace Silva-Martin, Analyst, Homeland Security and Justice

John Stambaugh, Senior Analyst, Physical Infrastructure

Walter Vance, Assistant Director, Applied Research and Methods

Image Sources
This section contains credit and copyright information for images and graphics in this product, as appropriate, when that information was not listed adjacent to the image or graphic.
Front cover: GAO.

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's website (https://www.gao.gov). Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to https://www.gao.gov and select "E-mail Updates."

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, https://www.gao.gov/ordering.htm.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

Connect with GAO

Connect with GAO on Facebook, Flickr, Twitter, and YouTube.
Subscribe to our RSS Feeds or E-mail Updates. Listen to our Podcasts.
Visit GAO on the web at https://www.gao.gov.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Website: https://www.gao.gov/fraudnet/fraudnet.htm

Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Orice Williams Brown, Managing Director, WilliamsO@gao.gov, (202) 512-4400,
U.S. Government Accountability Office, 441 G Street NW, Room 7125,
Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, DC 20548

Strategic Planning and External Liaison

James-Christian Blockwood, Managing Director, spel@gao.gov, (202) 512-4707
U.S. Government Accountability Office, 441 G Street NW, Room 7814,
Washington, DC 20548