The Frontiers of
MACHINE LEARNING

2017 Raymond and Beverly Sackler U.S.-U.K. Scientific Forum

NATIONAL ACADEMY OF SCIENCES

THE ROYAL SOCIETY

Foreword

Rapid advances in machine learning—the form of artificial intelligence that allows computer systems to learn from data—are capturing scientific, economic, and public interest. Recent years have seen machine learning systems enter everyday usage, while further applications across healthcare, transportation, finance, and more appear set to shape the development of these fields over the coming decades. The societal and economic opportunities that follow these advances are significant, and nations are grappling with how artificial intelligence might affect society.

There are emerging policy debates in the United States and the United Kingdom about how and where society can make best use of machine learning. As the capabilities of machine learning systems and the range of their applications continue to grow, it is therefore particularly timely for the National Academy of Sciences and the Royal Society to bring together leading figures in these fields.

Since 2008, the Raymond and Beverly Sackler U.S.-U.K. Scientific Forum has brought together thought leaders from a variety of scientific fields to exchange ideas on topics of international scientific concern and to help forge an enduring partnership between scientists in the United States and the United Kingdom.

The forum on "The Frontiers of Machine Learning" took place in the United States on January 31 and February 1, 2017, at the National Academy of Sciences in Washington, D.C. This event brought together leading researchers and policy experts to explore the cutting edges of machine learning research and the implications of technological advances in this field. This report summarizes the high-level discussions at the event focusing on some of the exciting areas of progress in machine learning and the societal debates that follow.

The National Academy of Sciences and the Royal Society share a mission to promote the use of science to benefit society and to inform important policy debates. As Presidents of the National Academy of Sciences and the Royal Society, we are pleased to introduce the latest piece of work supported through the inspired generosity of the Raymond and Beverly Sackler Foundation.

Marcia McNutt
President, National Academy of Sciences

Venki Ramakrishnan
President, Royal Society

Contents

Acknowledgments

The following distinguished scholars served on the planning committee:

Peter Donnelly, FMedSci[1]/FRS,[2] University of Oxford, *Co-Chair*,

Michael Kearns, University of Pennsylvania, *Co-Chair*,

Tom Dietterich, Oregon State University,

Cynthia Dwork, NAS[3]/NAE,[4] Harvard University,

Zoubin Ghahramani, FRS, University of Cambridge,

Nathan Griffiths, University of Warwick,

Sofia Olhede, University College London,

David Vladeck, Georgetown University, and

Bin Yu, NAS, University of California, Berkeley.

The National Academy of Sciences was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists. Its members are drawn from all areas of science, engineering, and medicine. It is the national academy of science in the U.K. The Society's fundamental purpose, reflected in its founding Charters of the 1660s, is to recognize, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

---

[1] FMedSci, Fellow of the Academy of Medical Sciences.
[2] FRS, Fellow of the Royal Society.
[3] NAS, National Academy of Sciences.
[4] NAE, National Academy of Engineering.

Summary

The Raymond and Beverly Sackler U.S.-U.K. Scientific Forum "The Frontiers of Machine Learning" took place on January 31 and February 1, 2017, at the Washington, D.C., headquarters of the National Academy of Sciences. A planning committee of distinguished scholars from the United States and the United Kingdom organized the forum.

The forum drew over 60 in-person attendees and over 500 webcast participants from academia, government, industry, and philanthropy. Participants included industry leaders, machine learning researchers, and experts in privacy and the law, and this report summarizes their high-level interdisciplinary discussions.

The field of machine learning continues to advance at a rapid pace owing to increased computing power, better algorithms and tools, and greater availability of data. Machine learning is now being used in a range of applications, including transportation and developing automated vehicles, healthcare and understanding the genetic basis of disease, and criminal justice and predicting recidivism. As the technology advances, it promises additional applications that can contribute to individual and societal well-being.

Novel technical and societal challenges are arising as machine learning advances. These relate to the ethics of using certain types of data (because of privacy or biased data collection), of managing data in different ways, and of automating certain processes (e.g., life and death decision making). They are also rooted in questions about how humans and machine learning systems interact as well as the societal challenges of adapting to a world in which these systems are increasingly ubiquitous.

Alongside these challenges come exciting opportunities across a range of industries and areas of research, with stimulating cross-disciplinary research taking place in academia and business. A supportive environment for this research—based on open standards and tools, effective education, public confidence through effective engagement, and an understanding of varied users—can help secure the social and economic benefits of machine learning.

This report is a summary of contributions to the forum. It does not reflect the views of the National Academy of Sciences or the Royal Society.

1 Introduction

OVERVIEW OF THE SACKLER FORUM

Machine learning is changing everyday activities through improvements in functions such as voice recognition, object detection, and image perception. The spheres of medicine, transportation, finance, government, and education could be transformed by such innovative technology, as could the way individuals and communities live and work. For example, some tasks typically performed by humans are already being performed by computer systems, and the range and complexity of tasks these systems can perform is growing. This creates novel tensions arising from the legal, social, and ethical implications of machine learning for humans now and in future generations.

To explore these issues in greater depth, the Raymond and Beverly Sackler U.S.-U.K. Scientific Forum "The Frontiers of Machine Learning" was held on January 31 and February 1, 2017, at the Washington, D.C., headquarters of the National Academy of Sciences. A planning committee of distinguished scholars from the United States and the United Kingdom organized the forum.

The forum drew over 60 in-person attendees and over 500 webcast participants from academia, government, industry, and philanthropy. Participants included industry leaders, machine learning researchers, and experts in privacy and the law. They represented strong proponents of widespread adoption of machine learning as well as those with concern for the societal tensions that arise with expanding the use of machine learning. This 2-day meeting included presentations and discussions on the current state of the art in machine learning, its relationship

to other fields, challenges in the field, and key considerations for its future development.

This publication provides a summary of the observations shared by forum participants. It does not reflect a consensus of the participants or the views of the sponsoring organizations. Instead, it is intended to provide perspectives and suggestions from a selection of individuals working at the frontier of machine learning and studying its implications. It examines the evolution of the field and its applications, some of the social and ethical tensions to which advances in machine learning have contributed, and considerations for the future development of the field.

UNDERSTANDING ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, AND ROBOTICS

The terms "artificial intelligence," "machine learning," and "robotics" are often used interchangeably, but important distinctions exist.

While specific definitions vary, artificial intelligence is, generally speaking, any method for programming computers to enable them to carry out tasks or behaviors that would require intelligence if performed by humans. Early work in this field focused on automated reasoning; using these approaches, programs were written as sets of logical statements against which queries were processed by theorem proving and search. For example, computers could be programmed with the rules for board games and then tasked with finding sequences of moves that could defeat an opponent. To obtain the logical rules (or inferences) for such systems, researchers in the field first turned to interviewing experts in the relevant domains (e.g., medical diagnosis, fault diagnosis). When the need to reason under uncertainty became increasingly evident, especially in medicine and in engineering applications, the field embraced probabilistic models as a replacement for systems of logical rules. This in turn required developing methods for probabilistic reasoning. However, in many applications, it was very difficult to build probabilistic models manually and determine their parameters. The field then turned to machine learning and related methods in statistics and pattern recognition to help address this challenge. Instead of interviewing human experts and seeking to codify the steps they take to solve a problem, these systems examine data—that is, examples of how to solve the problem at hand—to find patterns and rules that can drive decision making, such as clustering and classification.

Machine learning draws from a variety of fields, including computer science, statistics, engineering, cognitive science, and neuroscience. Researchers in machine learning develop both the mathematical foundations and the practical applications of systems that learn from data.

Broadly speaking, machine learning can be divided into three

Machine learning draws from a variety of fields, including computer science, statistics, engineering, cognitive science, and neuroscience
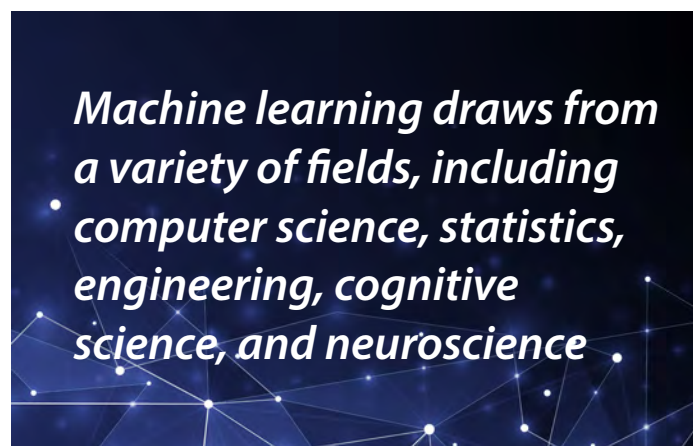
branches: supervised, unsupervised, and reinforcement learning. "Supervised learning" is based on a program that learns from previous examples: a model with specified parameters receives input and makes a prediction based on function estimation. This prediction is then compared to the actual outcome, and the similarity or difference between the two then informs the updates to the model parameters. For example, a supervised learning model can use photos with information about what is shown in those photos to then recognize and identify features in new photos. "Unsupervised learning" can be used where there is a dearth of labeled data on which to train algorithms. An unsupervised machine learning system processes a large amount of data, finding patterns in the data and learning the characteristics that make data points more or less similar to each other. "Reinforcement learning" emphasizes learning by trial and error, using rewards or punishments for success or failure at a task as a means to discover successful ways of behaving in complex environments.

There are countless applications of machine learning in society that help people become more organized and make processes more efficient

The field of machine learning continues to advance at a rapid pace as a result of improved computational resources, better algorithms and tools, increased availability of data to train systems, and the ingenuity of those developing machine learning systems. There are countless applications of machine learning in society that help people become more organized and make processes more efficient—for example, in organizing digital photo collections, managing spam in email platforms, and supporting navigation devices. Machine learning can solve problems related to classification, regression, clustering, dimensionality reduction, semi-supervised learning, and reinforcement learning. These are often referred to as the "canonical" problems in machine learning and are described in Box 1.

The term "artificial intelligence" can often evoke ideas of human-like intelligence in machines, known as "artificial general intelligence." The goal of artificial general intelligence research is to create artificial intelligence systems that have the ability to adapt to a range of tasks and could develop a breadth of capabilities that would match or exceed those of humans. Such systems are different from current artificial intelligence systems that are designed to solve specific target tasks but lack any knowledge or capability beyond those tasks. Some scholars have argued that once systems achieve human-level breadth and capability, they could start improving themselves and therefore rapidly exceed human capabilities. Most researchers maintain,

BOX 1
Canonical Problems in Machine Learning[a]

1. Classification
Classification involves collecting data and assigning them to one of several categories. The task at hand is to predict discrete class labels from input data after a model has been trained on labeled data. Applications of classification include face recognition, image recognition, and medical diagnosis. Typical methods for such tasks include logistic regression, support vector machines, neural networks, random forests, and Gaussian process classifiers.

2. Regression
Regression analyses try to predict continuous quantities from input data. Their applications include financial forecasting and click rate prediction, which impacts a range of applications in Internet advertising. Typical methods to address this task include linear regression, neural networks, and Gaussian processes.

3. Clustering
Clustering is used for analysis in which there are a lot of data that need to be organized in a way to create clusters where similar points are grouped together. This is used in bioinformatics and studies of gene expression, astronomy, document modeling, and network modeling. Typical methods include k-means, Gaussian mixtures, and Dirichlet process mixtures.

4. Dimensionality reduction
Dimensionality reduction is used in applications where the raw data have a high number of dimensions. These approaches map high-dimensional data onto low dimensions, while preserving relevant information, and have a range of applications—for example, in data mining, scientific analysis, and image recognition. Typical methods include principal components analysis, factor analysis, multidimensional scaling, IsoMAP, and Gaussian process latent variable models.

5. Semi-supervised learning
Semi-supervised learning is used in analyses where a large quantity of unlabeled data is available alongside a few data points that have been labeled—for example, semi-supervised learning can be used to combine and learn from each of a small number of annotated images. This approach is particularly useful in applications where labeling data is expensive, such as in drug trials or other studies involving complex experiments. Methods include probabilistic models, graph-based semi-supervised learning, and transductive support vector machines.

6. Reinforcement learning
Reinforcement learning addresses tasks where an agent or a computer program needs to learn to interact with its environment, receiving inputs and making sequential decisions so as to maximize future rewards. It therefore relates to adaptive control and sequential decision making under uncertainty. Agents using reinforcement learning might be physical or virtual, and applications are found in robotics, games, trading, and dialogue systems. Methods in this field include Q-learning, direct-policy methods, and PILCO.

---

[a] The Royal Society, 2017, *Machine Learning: The Power and Promise of Computers That Learn by Example*, https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf.

however, that there is no technical basis for such concerns and that matching human capabilities is not a critical threshold in the development of artificial intelligence. These researchers worry that a focus on developing systems with human-level intelligence detracts from the need to address near-term questions about how to create artificial intelligence systems that behave correctly, safely, and reliably.

THE EVOLUTION OF MACHINE LEARNING APPLICATIONS

Machine learning is a quickly evolving field with numerous application areas. One of the earliest applications of the technology, "information retrieval" involves "finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)."[1] Machine learning provides algorithms and models useful for modern information retrieval. Users benefit from this technology each time they search their e-mail or browse the Web.

"Speech recognition" is the process by which a computer system recognizes spoken words and phrases and translates them into machine-readable format. Though the first speech recognition system was developed in the 1950s, these computer systems could translate only a small number of speech signals into words and phrases. Newer software in modern speech recognition has advanced to the point of understanding and using natural speech patterns (e.g., speed, tone, word choice) that are characteristic of normal conversation. Using layers of neural networks,[2] known as deep learning[3] approaches, the sound waves from speech are converted into accurate plain text. However, it still remains difficult to distinguish one voice from another. Developing computer systems that can understand the meaning of the plain text and make appropriate responses is still challenging.

"Targeted advertising," which became popular in the 1990s, seeks to present advertisements to consumers that increase the likelihood that the consumer will make a certain purchase or perform a certain action. To do this, the companies collect data on their consumers' behaviors or purchase histories and apply machine learning methods to develop predictive models of future customer behavior.

---

[1] C.D. Manning, P. Raghavan, and H. Schütze, 2008, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, U.K.

[2] Neural networks are "computing system[s] made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs." (M. Caudill, 1989, Neural networks primer: Part I, *AI Expert* 2(12): 46-52.) The outputs are equivalent to the sum of the inputs multiplied by their weights. More specifically, the inputs are applied, the actual versus desired output is calculated, the error is propagated backward into the network, and weights are adjusted to minimize the error. Neural networks are trained repeatedly, with inputs adjusted each time, to get closer to a desired output. It remains difficult to predict what will happen with these networks, especially with unknown or untrained input, so they must continually be trained with new samples—simply changing the code is insufficient. Over time the development of more complex, large-scale, multilayered neural networks has allowed researchers to work on problems at scale.

Automated vehicles have the potential to reduce vehicle fatalities, increase mobility, and decrease pollution

In finance, machine learning is used for assessing risk and managing assets. Machine learning has also found applications in bioinformatics where complex biological data (e.g., genetic code) are analyzed using either supervised or unsupervised learning to better understand the ways in which cells function.

Automated vehicles can independently navigate some public roadways by sensing their environment and responding appropriately. Such vehicles learn where to steer and how to locate other vehicles, people, and objects based on a range of machine learning techniques, including using deep neural networks to analyze image and video data. Advances in (1) data handling and communications bandwidth, (2) computing power, (3) learning algorithms, and (4) efficient inference algorithms help make automated vehicles more reliable, and advances in machine learning theory could improve the technology even further. Important challenges remain, such as identifying better metrics to evaluate the severity of driving errors (e.g., causing a head-on collision versus running a red light), improving three-dimensional perception, and functioning under poor weather or lighting conditions.

Automated vehicles have the potential to reduce vehicle fatalities, increase mobility for the elderly and those with disabilities, decrease environmental pollution, and support new models of public transportation. However, questions remain about the likelihood of these

---

[3] Deep learning is a powerful class of machine learning methods that explore data representations using supervised, semi-supervised, or unsupervised learning. An essential characteristic of deep learning is that these methods are not just one function; rather, they are made up of a set of composable subfunctions for model building. Deep learning involves neural network models with algorithmic innovation, larger data sets, larger compute resources, and better software tools, and it is compatible with many variants of machine learning. Deep learning has a number of similarities to the human brain: artificial neural networks act like neurons in the brain to connect various input concepts with potential outputs. E-mail spam filters, Internet search engines, speech recognition devices, translators, photo recognition programs, automatic email repliers, object detection in automated vehicles, and navigation apps all rely on deep learning to simplify tasks for the user. Much modern deep learning demands building differential architectures that can do everything. Although deep learning allows more intricate processing than more basic machine learning approaches, the method is data hungry, compute intensive, deficient in representing uncertainty, easily deceived by oppositional examples, difficult to optimize, and opaque. In order to preserve interpretability and practicality when using deep learning processes, it is important to stress test these models in new environments.

benefits being fully realized. For example, could this technology result in more cars on the road, lengthen commutes, or lead to urban sprawl? Alternatively, might fewer people choose to own cars, leading to increased carpooling and a reduced number of cars on the road? If this is the case, fewer parking lots would be needed across communities, and cities could find new ways to use the space. Any significant change in car ownership also has potential economic impacts; the automotive sector is a large component of many developed economies, and the effects of such disruption are difficult to predict.

Automated vehicle deployment could also affect pedestrians and cyclists, as it changes the commonly accepted—if unwritten—rules of the road. For example, how can a pedestrian make eye contact with a driver before crossing a street when there is no driver in an automated vehicle? Much of the communication among road users and many of the predictions humans make about the behavior of other road users when navigating a vehicle-populated environment is not explicit. Rather, it relies on reading subtle observational cues combined with past experience and knowledge of the environmental norms (e.g., driving behavior in the United States is different from that in India). These questions open interesting avenues of investigation, and researchers are developing models to better understand how humans interact with intelligent devices in potentially dangerous situations. This raises one facet of a broader question about vehicle safety standards: should new standards be developed, or do current road safety standards translate for automated vehicles?

Researchers are developing models to better understand how humans interact with intelligent devices in potentially dangerous situations

Machine learning is also finding new applications in data science itself. There are not yet enough data scientists, statisticians, and machine learning experts to dedicate adequate time to understanding the data, building models, and making predictions. One possible solution to this shortage of experts is the Automatic Statistician project, which is creating systems that are able to automate data analysis.[4] As interest in data science grows, and the sophistication of these tools increases, such projects could help support the use of machine learning in a broad range of research fields or novel applications.

---

[4] The Automatic Statistician, https://www.automaticstatistician.com/index/, accessed June 15, 2017.

2 Machine Learning Challenges

TECHNOLOGICAL CONSIDERATIONS

Different Needs for Different Applications

Much of the excitement generated by machine learning arises from its broad applicability across different subjects or domains. This also presents a challenge for the field: while many approaches can cut across applications, each use case involves different considerations and the construction of different types of systems. The following examples describe how several research areas in biomedicine require customized machine learning techniques. Similar customization is needed in many other fields.

Medical researchers, for example, have begun to use genetics data, medical records, patient registries, activity logs, environmental data, and medical imaging to better understand all aspects of human health. In the field of genomics, modifications in gene regulation and gene function resulting from mutated segments of DNA allow researchers to identify which regions of the genome influence a particular trait associated with a particular disease. Statistical methods related to machine learning are being developed to enable researchers to design medical interventions that would improve the lives of people with particular health issues in the future.

Mobile health interventions that employ a smart phone application or wearable technology to encourage specific behavior, such as increased exercise, are another potential applica-

tion area for machine learning techniques. Two possible schemes for generating notifications to patients are "pull" interventions, which rely on individuals' requests for interventions, and "push" interventions, which employ sensors, self-reports, and computer algorithms to decide when an intervention is needed as well as what intervention might be most appropriate. The shared goal of both types of notifications is to promote healthy behavior and aid in the development of longer-term treatment policies. This is similar to the action-reward structure of reinforcement learning, with the long-term benefit of improved health serving as the reward for the action of responding to the health notifications.

In the study of neuroscience, the ability of machine learning to spot patterns and make predictions makes it a key tool in analyzing the vast quantities of neuroimaging data available today. Functional magnetic resonance imaging (fMRI) is used to track blood flow within the brain to deduce functional or physiological properties. When a patient or research subject examined using an fMRI machine is presented with an external stimulus, the responding brain area will have an increase in blood flow, which will be displayed clearly on an fMRI monitor. Classification can then be applied to a set of fMRI images to determine if patterns exist among participants within brain regions of interest. Neuroimaging applications range from diagnostic tools for neurological disorders to enhanced understanding of neural connectivity between brain regions. There are analogous opportunities for machine learning to help in many other areas of medical imaging.

While genomics, mobile health sciences, and neuroscience are typically categorized within the health domain, the challenges posed by each are domain-specific, and the machine learning techniques required for each are highly specialized. For example, the classification tools most appropriate for understanding neuroscience may not be as effective in genomics research because of substantial differences in the nature of the data, its structure and properties, and the relationships between data and outcome or decision. A similar trend can be seen in other disciplines when the machine learning applications in one subcategory may not work for a tangentially related area. For many applications, combining domain knowledge with machine learning expertise will be essential to create effective outputs.

Collaboration

The domain-specificity of the challenges at hand makes collaboration in machine learning particularly important. Furthermore, the key components of machine learning—expertise, computing power, data, and algorithms—are not concentrated in any one domain, and industry, academia, and government all play significant roles.

In the case of automated vehicles, for example, industry has resources and data that could help academics conduct better research. Government plays an important role as well through research funding and through its involvement in deploying infrastructure technology, such as traffic light sensors, that will affect the behaviors of automated vehicles.

While technical solutions can address the broader social and ethical implications of this technology, collaborations also contribute to addressing these challenges. Opportunities in the

spirit of collaboration include the following:

- Adding narratives to technical writing to help professionals in other disciplines better understand machine learning concepts,

- Committing to producing research that can be understood across disciplines,

- Raising awareness via public debate about potentially controversial machine learning issues, and

- Convening debates among diverse stakeholders.

EDUCATIONAL BARRIERS

Curriculum

Companies are aggressively hiring new talent in machine learning. Although course offerings for the next generation of machine learners are expanding, building a stronger skills pipeline remains key to the health of the field:

- If one hopes to be prepared for the future workforce and the real challenges associated with this level of innovation, data literacy is essential. As early as the elementary school level, students could benefit from greater encouragement to develop a science, technology, engineering, and mathematics skill set.

- In many postsecondary curricula, only traditional mathematics is strongly emphasized, whereas students may benefit from a program that includes mathematics alongside computing and communication skills. Students should also gain an understanding of the ethical challenges associated with the use of data.

Curricular changes will not happen overnight, but conversations among interested stakeholders should begin and plans should be made to fund programs that could improve students' chances of securing relevant work in the future.

Infrastructure

In addition to curriculum changes, modifications to university course materials and course content can help prepare both machine learning experts and those likely to work with machine learning systems in the future:

- Technical programs will need to update the skills and knowledge with which they equip students to work in specific fields.

- For those studying machine learning, an understanding of the methods of social sciences and liberal arts will become important for addressing the legal and societal challenges that the technology presents.

- For researchers and scholars in academia, journal and tenure guidelines often focus too narrowly on recognition and attribution in only the traditional form of published papers; it is important that innovative and applied work in machine learning is also both rewarded and shared.

SOCIETAL ISSUES

Fairness, Privacy, Consent, and Cybersecurity

Algorithmic decision-making systems may place a strain on some democratic values as machine learning enables new uses of data that challenge existing notions of fairness, privacy, and consent.

The advanced analytical capabilities offered by machine learning pose new challenges to managing privacy: in some applications, machine learning will use data containing sensitive information, while in other cases machine learning might create sensitive insights from seemingly mundane data. New privacy-preserving technologies (e.g., de-identification of data, differential privacy, homomorphic encryption) are being explored that can help lessen the risks of privacy breaches while enhancing the benefits of data sharing to society.[5]

The advanced analytical capabilities offered by machine learning pose new challenges to managing privacy

New economic models based on data collection also have the potential to raise privacy concerns. Consumers frequently share their personal data when carrying out everyday tasks online—for example, when online shopping—and may not be aware how much data companies can collect about them based on these transactions. Such cases create trade-offs between privacy and convenience or the desire for certain services. Some researchers suggest that there should be an opt-out feature, which would be a relatively straightforward technical fix, in which the user can select that the site erases all information collected and thus will treat the customer as a new user every time that customer visits. Others propose that individuals should be able to control their personal data and "decide for themselves how to weigh the costs and benefits of the collection, use, or disclosure of their information."[6] In this context, machine

---

[5] The Royal Society, 2017, *Machine Learning: The Power and Promise of Computers That Learn by Example*, https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf.

[6] D.J. Solove, 2013, Introduction: Privacy self-management and the consent dilemma, *Harvard Law Review* 126(7): 1880-1903.

learning also raises questions about consent and the extent to which individuals are able to meaningfully consent to the use of their data, especially if those data are combined or processed in novel ways.

Legislative structures that seek to ensure fairness in the provision of some products and services—and balance data protection with innovation—already exist, and these current "fairness structures," set up primarily for financial security, demonstrate points of tension. For example, the U.S. Equal Credit Opportunity Act protects applicants from discrimination (e.g., race, color, religion, national origin, sex, marital status, age) in a credit transaction. Modern credit scoring systems use machine learning to evaluate creditworthiness, and so a question arises about which factors a model can and should use to make a decision. The U.S. Fair Credit Reporting Act supports the accuracy, fairness, and privacy of consumer information included in consumer reporting agency files. Thus, individuals are granted access to the credit score, as well as to all of the factors that may have adversely affected the score. Because of this, the Fair Credit Reporting Act ensures transparency of factors that led to deriving the credit score. The EU General Data Protection Regulation, which is expected to go into effect in May 2018, protects data that expose racial or ethnic origin; political or religious beliefs; and genetic, biometric, and health data.[7]

This is not the case for all existing regulatory instruments: unlike the previously mentioned regulated scores, the unregulated Alternative Credit Score in the United States can use factors that would normally be prohibited under the Fair Credit Reporting Act because it uses those factors for marketing financial products instead of for determining creditworthiness.[8] Unregulated data includes health data not subject to the Health Insurance Portability and Accountability Act of 1996 (HIPAA),[9] transactional data, historic data sets, and commercial data sets.

In seeking to ensure that fairness and privacy are maintained in such analyses, researchers can apply statistical approaches to identify people for whom privacy or fair information practice principles may be violated, such as when an individual's data are represented in long tails (i.e., a distribution with a large number of occurrences far from the central part of the distribution) or as outliers (i.e., data points that differ from most others) and are therefore easy to identify within a data set. Other approaches to managing issues relating to fairness include the following:

- Fair information practice principles (i.e., best practices),

- New approaches to remediation for those negatively affected, and

[7] Information Commissioner's Office, "Overview of the GDPR," https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/introduction/, accessed June 15, 2017.

[8] Consumer Financial Protection Bureau, "CFPB Explores Impact of Alternative Data on Credit Access for Consumers Who Are Credit Invisible," last update February 16, 2017, https://www.consumerfinance.gov/about-us/newsroom/cfpb-explores-impact-alternative-data-credit-access-consumers-who-are-credit-invisible/.

[9] U.S. Department of Health and Human Services, Health Insurance Portability and Accountability Act of 1996, P.L. 104-191,104th Congress, https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996, accessed June 16, 2017.

- Use-based rules for data (i.e., new approaches to auditing outputs).

Cybersecurity issues also arise in many areas. For example, even though medical data subject to HIPAA are required to be encrypted when stored and protected in other ways, there is still a risk that sensitive information could end up in the wrong hands. Such cyberattacks and data thefts are also a concern for technology developers. All companies need to be vigilant in preparing for or preventing these situations, perhaps by drawing on the expertise of hackers.[10]

Trust, Transparency, and Interpretability

"Artificial Intelligence and Life in 2030" suggests that "well-deployed artificial intelligence prediction tools have the potential to provide new kinds of transparency about data and inferences, and may be applied to detect, remove, or reduce human bias, rather than reinforcing it."[11]

A key barrier to achieving transparency in artificial intelligence tools is the lack of a common language, as varying definitions of "transparency," "explainability," and "interpretability" exist across disciplines. Additionally, there are various lenses through which these concepts can be considered. For example, some view "explainability" as a technical issue, while others may view it as a legal or social issue.

A key barrier to achieving transparency in artificial intelligence tools is the lack of a common language

Another important aspect to bear in mind when considering transparency is that it is not synonymous with trust: trust can be attained without transparency. Conversely, just because an algorithm is transparent does not mean that it is trusted or trustworthy. A user might trust an algorithm without questioning or understanding how the result was attained—just as travelers trust that it is safe to board airplanes every day, despite not knowing precisely how airplanes function. Individuals are also likely to trust certain types of medical treatment, without knowing how or why they work.

Humans and computer systems alike can be trained to provide explanations, but these explanations may be too complex to grasp without specialist knowledge or may, in some cases, be biased or untrue. For example, when a medical patient enters an examination room with a complicated set of symptoms, a medical professional may offer either a simplified explanation

---

[10] National Academy of Sciences and the Royal Society, 2015, *Cybersecurity Dilemmas: Technology, Policy, and Incentives: Summary of Discussions at the 2014 Raymond and Beverly Sackler U.S.-U.K. Scientific Forum*, https://www.nap.edu/catalog/21833/.

[11] Panel on the One Hundred Year Study on Artificial Intelligence, "Artificial Intelligence and Life in 2030," Report of the 2015-2016 Study Panel, Stanford University, Stanford, Calif., September 2016, http://ai100.stanford.edu/2016-report.

intended to put the patient at ease or one that would require a medical degree to understand fully. The patient may feel relief having an explanation, yet the patient may not actually have the most accurate or most clear explanation. Who performs an action or supplies a service can also play a role in the extent to which individuals trust a system.

The possibility of providing biased results or unfair outcomes—intentionally or unintentionally—presents two relevant questions: (1) To what extent is the level of interpretability influenced by the type of machine learning deployed and how can advances in research address this? (2) What type of interpretability is required in different contexts?

The human desire to provide accounts of how decisions are made prompts another important series of questions: When a computer system makes a mistake, who or what should be held accountable? Because computer systems can keep complete logs of their processing, they are in principle capable of producing much more complete and correct explanations. Should computer systems therefore be held to higher standards of explanation than humans?

An additional key factor is the level of explanation required. An explanation provided to an engineer who will debug software in an automated vehicle would be far more technical in nature than an explanation about a vehicle malfunction provided to a user. Related to this is the purpose for which an explanation is being provided. Explanation of failure cases and biases that will be used to improve the computational model is likely to require a different level of accuracy and detail than the explanations needed for legal purposes.

Another component of transparency is consistency—an understanding that similar cases will be treated similarly. Consistency is critical to enabling the users of a learning system to anticipate the future decisions of the system based on explanations of past decisions. Without consistency, it may be more difficult for users to gain trust in the system.

Ethical Challenges in Specific Applications
Ethical and legal concerns stemming from issues of fairness and bias relate to issues of trust and transparency.

One application area in which these questions are most palpable is in the criminal justice system. Scoring systems applied to predict the likelihood of repeat offending are increasingly data-driven, and some make use of machine learning.

These systems offer the hope of reducing bias in criminal sentencing, if the algorithm being used can be designed in a way that supports advanced analysis free of societal assumptions about factors such as race, gender, or socioeconomic status. However, the outputs of these systems reflect the data on which they are trained. If these data embody current societal biases or inequality, then—unless it is explicitly programmed otherwise—a machine learning system will replicate these biases.

There are different approaches to monitoring the biases of such systems, many of which are based on the interpretability or transparency of the system:

- If the data used to create such models are publicly available, then the results can be tested or reproduced;

- Rule-based systems or other types of models may be more interpretable; and

- Fairness adjustments or constraints can be built into models.

Identifying, controlling, and being transparent about bias in the system may be more manageable than seeking to eliminate it. In this sense, although interpretable models may not be as accurate as alternative approaches, being able to understand how a model functions and why it may be imperfect is likely to be helpful. Questions about how best to negotiate a trade-off among interpretability, accuracy, and fairness are likely to endure.

The use of machine learning systems in recidivism scoring also offers a new lens through which to view questions about the concept of unique treatment versus legal precedent. The law says that people need to be treated as individuals with dignity and thus have a right to defend themselves in an effort to shape the decision being made about them. However, the legal system also relies on precedent, which encourages the law to be applied to new cases in a similar way to its application in previous cases.

Scoring via machine learning systems is also used in medicine. For example, there is a scoring system to predict mortality for patients with some illnesses and to recommend medical interventions accordingly.[12] Such systems may improve patient outcomes, but they still require careful management. Some scores will fall within the guidelines of existing governance mechanisms—for example, the Clinical Frailty Scale,[13] which is protected by HIPAA in the United States—which means patients will continue to be able to see and object to such scores.

Living Alongside Machine Learning

In addition to specific challenges arising from the governance of data used in machine learning or the capabilities of current machine learning systems, and the benefits to be gained from such systems, there is a broader suite of questions raised by the increasing pervasiveness of machine learning systems. At a fundamental level, these questions ask how society will change as people

How will society change as people live and work with automated systems?
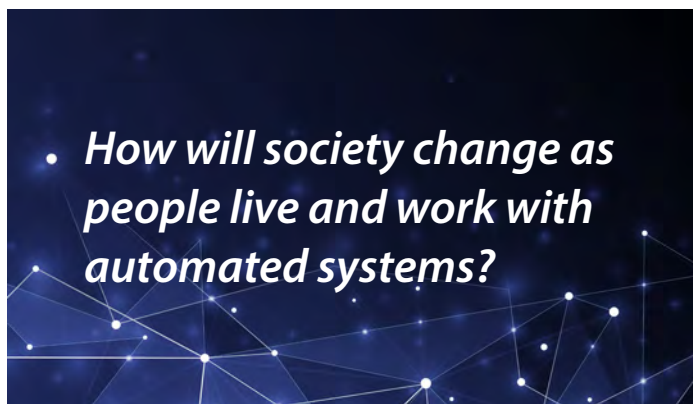
live and work with automated systems, as well as with the new forms of human–computer interaction that could follow.

Concerns about automation and the workforce are already present in debates about machine learning, rooted in fears about humans being replaced by or becoming de-

---

[12] One such example is the use of machine learning to predict outcomes for patients with pneumonia.

[13] The Clinical Frailty Scale is a means of assessing fitness levels of individuals. The scores range from 1 (very fit) to 9 (terminally ill), with higher numbers indicating increased frailness.

Results from public dialogues on machine learning in the U.K. indicate that there is a general preference to have a "human in the loop" for decisions that affect people in a personal or sensitive way—for example, in healthcare
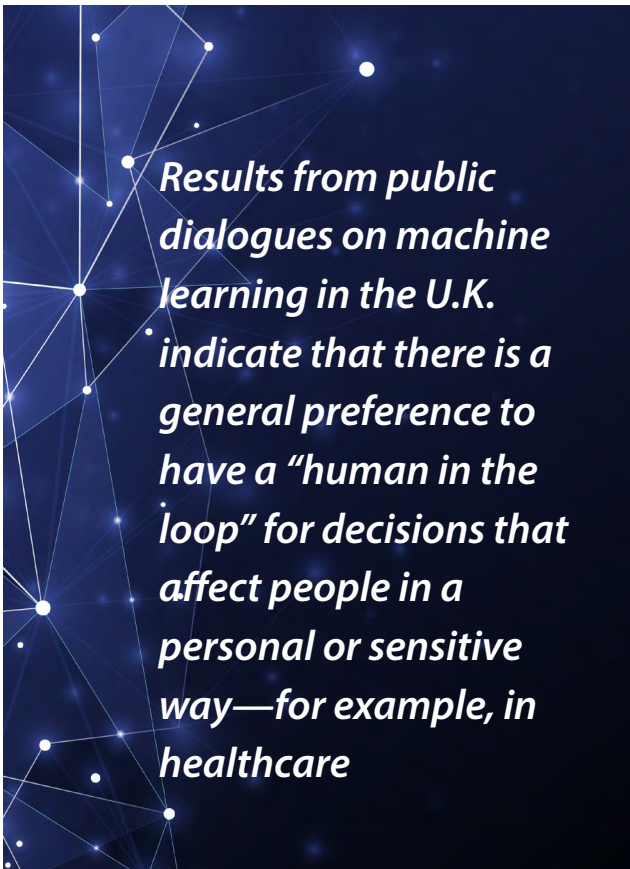
pendent upon machines. In some cases, fears arise about how machine learning could hurt individuals or restrict human experiences by limiting the choices of products, services, and activities available or by reducing interactions between human beings.[14]

Most people are happy to have processes automated that are dirty, dull, or dangerous, but what about other tasks that fall outside these categories? Warfare is an example in which there are continuing ethical debates about the use of algorithms that may make a decision about taking human lives. Results from public dialogues on machine learning in the United Kingdom indicate that there is a general preference to have a "human in the loop" for decisions that affect people in a personal or sensitive way—for example, in healthcare.

There are also broader questions about how the benefits of machine learning can be shared across society and how society can capitalize on the opportunities these technologies present, whether in healthcare, transportation, or education. In seeking to address these questions, there may be lessons from previous technological advances that have transformed the ways people live and work through automation. During the Industrial Revolution, individuals had to adapt to new ways of communicating, traveling, and working. Will humans today be able to adapt similarly to the changes to work and other aspects of life resulting from advances in machine learning? Questions about how people live alongside machine learning will persist, and continuing dialogue will be important in negotiating them.

---

[14] The Royal Society, 2017, *Machine Learning: The Power and Promise of Computers That Learn by Example*.

3 The Future of Machine Learning

UNDERSTANDING THE NEEDS OF DIFFERENT USERS AND USE CASES

As previously discussed, different domains and applications require machine learning systems with different characteristics. For example, the justice system is associated with different pressures than the transportation system. Understanding these different needs and use cases will be important in ensuring the field is able to realize its full potential.

ENGAGEMENT FOR STAKEHOLDERS

As the technical abilities of machine learning systems increase and their societal implications become more pronounced, it will be important for a diverse community of users and developers to help shape both technology development and policy discussions. To push forward the boundaries of machine learning, engineers, data scientists, and computer scientists can benefit from the following perspectives:

- Sociologists could play important roles, including discussing the ethical and societal issues of the technology;

- Psychologists and human factors experts could offer insight on the ways in which humans interact with technology;

- Unions and industrial psychologists could represent the voices of workforce changes that may result from continued development in artificial intelligence;

- Policy makers and regulators could contribute to a dialogue about autonomy and fair information principles; and

- Historians could contribute knowledge about past eras of rapid technological change and inform today's debate.

This interdisciplinary approach should also be reflected in professional training for artificial intelligence and in policy making for the field.

ALGORITHM DEVELOPMENT AND MODEL CONSTRUCTION

Advances in the technical abilities of machine learning algorithms have contributed to the recent success of the field. Algorithms are "encoded procedures for solving a problem by transforming input data into a desired output, based on specified calculations and procedures." [15] They perform specific tasks, and it is important to consider how they will be used and what will be the explicit trade-offs involved in each use case.

Many applications in public policy highlight these trade-offs. For example, when seeking fairness in a model being used in criminal justice and determining acceptable rates of false positives or false negatives, is it worse to incarcerate a low-risk individual (false positive) than to release from prison a high-risk individual (false negative)? If the algorithms are kept "fair" by reducing false negatives and excluding certain predictive criteria based on legally protected personal characteristics, more high-risk individuals may be released or more low-risk individuals may be incarcerated.

In some cases, such as criminal justice, fairness may become even more important than accuracy. It is important that algorithms are developed and conveyed such that the trade-offs are transparent. This will help increase confidence not just in the algorithm but also in the decision-making process. Context is also important in developing the algorithm, if such data can be collected. It may be useful to encode social norms into the machine learning and optimization processes so as to focus more on fairness.

The broader social consequences of the ubiquitous use of algorithms and increasing personalization are also important to consider. This personalization could lead to "algorithmic bubbles" that classify people based on their usual behaviors, beliefs, and actions and preferentially provide them with information consistent with these. This could insulate currently held beliefs from challenge and could divide societies into segments that cease to communicate.

---

[15] T. Gillespie, 2014, "The Relevance of Algorithms," pp. 167-193 in *Media Technologies: Essays on Communication, Materiality, and Society*, eds. T. Gillespie, P.J. Boczkowski, and K.A. Foot, MIT Press, Cambridge, Mass.

GOVERNANCE AND PUBLIC CONFIDENCE

Given the wide range of approaches to machine learning and possible application areas, public confidence in these systems will be key to their continued success.

The term "governance" in this context refers to a diverse set of instruments and behaviors that shape how machine learning is used: it encompasses codes of practice, institutional norms, and individual behaviors, as well as specific policies and direct regulation by the government.

There are already efforts in place for the government to better understand, offer support, and set standards related to artificial intelligence. For example, algorithms for certain tasks already have to follow existing laws, such as being nondiscriminatory. It is important to note that governance is not synonymous with constraint; instead, it can give the confidence and public acceptance that enables new applications to be developed.

While some have called for further governance of machine learning, both the feasibility and desirability of governance approaches based on targeted regulation of machine learning algorithms have been questioned in different settings. "Artificial Intelligence and Life in 2030" suggests, "Attempts to regulate artificial intelligence in general would be misguided since there is no clear definition of artificial intelligence (it is not any one thing), and the risks and considerations are very different in different domains. Instead, policy makers should recognize that to varying degrees and over time, various industries will need distinct, appropriate, regulations that touch on software built using artificial intelligence or incorporating artificial intelligence in some way." [16] The development of best practices by sector may be more useful in steering the development of the field.

This leads to questions about the governance of the technology itself—and those working on its applications. For example, should data collection be self-regulated for devices such as digital assistants? Because these data are considered unregulated, there are no fair reporting laws that currently apply in the United States. In the European Union, the General Data Protection Regulation, which is expected to take effect in May 2018, will address "automated individual decision making, including profiling" and notes that a person "shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her," and that, if such processing is necessary, the person will have "at least the right to obtain human intervention [...] to express his or her point of view and to contest the decision." [17] Furthermore, the regulation requires that if profiling is used, the person should have access to "meaningful information about the logic involved." [18]

---

[16] Panel on the One Hundred Year Study on Artificial Intelligence, "Artificial Intelligence and Life in 2030."

[17] DSB-MIT-SYSTEM, "Article 22 EU GDPR: Automated Individual Decision-making, Including Profiling," http://www.privacy-regulation.eu/en/22.htm, accessed June 16, 2017.

OPEN DISTRIBUTION OF TOOLS AND KNOWLEDGE

It is important to ensure that engineers, designers, business leaders, and inventors have the appropriate knowledge and skills required to make the most out of new technologies

It is important to ensure that engineers, designers, business leaders, and inventors have the appropriate knowledge and skills required to make the most out of new technologies. Sharing knowledge—for example, through educational materials and tools—is essential, as is improving communication across industry, academia, and government. Research from different sectors is often complementary, and providing access to data and real examples is crucial to ensure that learning and best practices spread quickly. Continued funding to support academic research and education in the field of machine learning and related areas can also help move the field forward.

Innovation requires access to data, codes, and other resources. Open sourcing of tools and online learning and encouraging researchers to share their findings publicly can spur increased innovation.

MACHINE LEARNING IN THE FUTURE

As machine learning transforms the world in which we live and improves many aspects of our lives, the legal, social, and ethical implications must be considered. The diverse applications of this technology will likely continue to expand and mature. As they do, collaboration, education, awareness of possible adverse implications of this technology, and creativity in addressing problems are crucial to ensure that the pressing challenges of today are mitigated to allow continued growth in the future.

---

[18] DSB-MIT-SYSTEM, "Article 13 EU GDPR: Information to Be Provided Where Personal Data are Collected from the Data Subject," https://www.privacy-regulation.eu/en/13.htm, accessed June 16, 2017.

Appendix

SACKLER FORUM ON MACHINE LEARNING PROGRAM

Machine learning is at the core of many applications that have become part of daily life, from voice recognition to image perception. These technologies, which a few years ago were performing at noticeably below-human levels, can now out-perform people at some tasks. As the field continues to evolve, machine learning has the potential to play a transformative role across a diverse range of sectors including transportation, medicine, public services, and finance. This forum brought together scientists from the United Kingdom and the United States to explore potential applications for machine learning and discuss the legal and ethical questions that arise as humans and machine learning algorithms interact.

TUESDAY - JANUARY 31, 2017

9:00 AM   Welcome from the National Academy of Sciences and Royal Society
Diane Griffin, Vice President, National Academy of Sciences
Richard Catlow, Foreign Secretary, Royal Society

Welcome from the Co-Chairs
Peter Donnelly, University of Oxford
Michael Kearns, University of Pennsylvania

Session 1: The Frontiers of Machine Learning
The ubiquity of data, accessibility of computing power, and algorithmic advances have driven rapid progress in machine learning over the past five years. Not only does machine learning now underpin many applications that have become part of daily life, the field continues to evolve quickly and has the potential to play a transformative role across a diverse range of sectors. This session explored the frontiers of machine learning, in terms of both cutting-edge technology and near-term applications, and discussed the state of the art of machine learning.

9:15 AM   I Know it's an Idiot but it's MY Artificial Idiot!
Vint Cerf, Google

9:50 AM   Towards Affordable Self-Driving Cars
Raquel Urtasun, University of Toronto

10:25 AM   Probabilistic Machine Learning: Foundations and Frontiers
Zoubin Ghahramani, University of Cambridge

11:00 AM	Break

11:30 AM	Words, Pictures, and Common Sense
	Devi Parikh, Georgia Institute of Technology

12:05 PM	Applied Machine Learning at Google
	Greg Corrado, Google

12:40 PM	Lunch

1:40 PM	Panel Discussion
	Vint Cerf, Google
	Raquel Urtasun, University of Toronto
	Zoubin Ghahramani, University of Cambridge
	Devi Parikh, Georgia Institute of Technology
	Greg Corrado, Google

Session 2: Machine Learning and Society

People and machine learning increasingly interact in a range of contexts. This expansion of machine learning raises legal and ethical questions, re-frames discussions about uses of data, and poses new challenges for the governance of this technology. The social acceptability of different machine learning applications, desirability of automated decision-making processes, adequacy of processes to manage concerns about statistical stereotyping or privacy, and more, will all influence how and where society has confidence in the deployment of machine learning systems. This session explored the societal implications of machine learning and the opportunities and challenges associated with advances in the field.

2:25 PM	Artificial Intelligence and Life in 2030
	Peter Stone, University of Texas at Austin

3:00 PM	Interpretable Machine Learning for Recidivism Prediction
	Cynthia Rudin, Duke University

3:35 PM	Break

4:10 PM	Protecting and Enhancing Our Humanity in an Age of Machine Learning
	Charis Thompson, University of California, Berkeley

4:45 PM   Using Machine Learning in Criminal Justice Risk Assessments
Richard Berk, University of Pennsylvania

5:20 PM   Adjourn for the day

WEDNESDAY - FEBRUARY 1, 2017

9:00 AM   Welcome from the Co-Chairs
Peter Donnelly, University of Oxford
Michael Kearns, University of Pennsylvania

9:05 AM   Privacy and Machine Learning: Promise, Peril, and the Path Forward
Pam Dixon, World Privacy Forum

9:40 AM   Algorithmic Regulation: A Critical Interrogation
Karen Yeung, King's College London

10:15 AM   Break

10:45 AM   Panel Discussion
Peter Stone, University of Texas at Austin
Cynthia Rudin, Duke University
Charis Thompson, University of California, Berkeley
Richard Berk, University of Pennsylvania
Pam Dixon, World Privacy Forum
Karen Yeung, King's College London

11:45 AM   Lunch

Session 3: Machine Learning in Research and Commercial Communities
There are enormous opportunities in machine learning in academia, research labs, and industry. While much of the research and development of machine learning to date has been done in the commercial world, each of these communities will continue advancing this field. Establishing key research challenges and areas of commercial opportunity will therefore be important in moving the frontiers of machine learning forward. This session explored key areas of interest in machine learning in the research and commercial communities.

1:00 PM   Building the Human Wiring Diagram from Linked Genomic and Healthcare Data
Gil McVean, University of Oxford

1:35 PM   Active Optimization and Self-Driving Cars
Jeff Schneider, Carnegie Mellon University and Uber Advanced Technology Center

2:10 PM	Three Principles for Data Science: Predictability, Stability, and Computability
	Bin Yu, University of California, Berkeley
2:45 PM	Experimental Design and Machine Learning Opportunities in Mobile Health
	Susan Murphy, University of Michigan
3:20 PM	Break
3:40 PM	A Deployable Decision Service
	John Langford, Microsoft Research
4:15 PM	Panel Discussion
	Jeff Schneider, Carnegie Mellon University
	Bin Yu, University of California, Berkeley
	Susan Murphy, University of Michigan
	Gil McVean, University of Oxford
	John Langford, Microsoft Research
4:55 PM	Adjourn Meeting

Participants

Richard Berk, University of Pennsylvania
Joanna Bryson, University of Bath
Joaquin Candela, Facebook
Linda Casola, National Academies
Richard Catlow, Royal Society
Vint Cerf, Google
Greg Corrado, Google
Claire Craig, Royal Society
Nello Cristianini, University of Bristol
Kim Dai, U.S. Department of Defense
Thomas Dietterich, Oregon State University
Pam Dixon, World Privacy Forum
Peter Donnelly, University of Oxford
Jon Eisenberg, National Academies
Avi Feller, University of California, Berkeley
Edward Felten, Princeton University
Kay Firth-Butterfield, AI-Austin
Zoubin Ghahramani, University of Cambridge
Fernand Gouveia, British Embassy, Washington, D.C.
Arthur Gretton, University College London
Diane Griffin, Johns Hopkins University
Brian Hall, U.S. Department of Defense
Sabine Hauert, University of Bristol
Kenneth Heafield, University of Edinburgh
Rodney Howard, National Academies
Luke Huan, National Science Foundation
Nick Jennings, Imperial College London
Subbarao Kambhampati, Arizona State University
Behzad Kamgar-Parsi, Office of Naval Research
Michael Kearns, University of Pennsylvania
John Langford, Microsoft Research
Po-Ling Loh, University of Wisconsin, Madison

Gil McVean, University of Oxford
Mitch Mellen, Office of the Director of National Intelligence
Lynette Millett, National Academies
Parsa Mirhaji, Yeshiva University
Tom Mitchell, Carnegie Mellon University
Jessica Montgomery, Royal Society
Miranda Mowbray, Mowbray Ventures
Susan Murphy, University of Michigan
Predrag Neskovic, Office of Naval Research
Regina Nuzzo, Gallaudet University
Susannah Odell, Royal Society
Sofia Olhede, University College London
Devi Parikh, Georgia Institute of Technology
Jerome Pesenti, BenevolentAI
Kate Piblett, British Defence Service
Cynthia Rudin, Duke University
Jeff Schneider, Carnegie Mellon University
Michelle Schwalbe, National Academies
Peter Stone, University of Texas, Austin
Charis Thompson, University of California, Berkeley
Raquel Urtasun, University of Toronto
Andrew H. Van Scyoc, Department of the Navy
Suresh Venkatasubramanian, University of Utah
David Vladek, Georgetown University
Scott Weidman, National Academies
Adrian Weller, University of Cambridge
Patrick Wolfe, University College London
Karen Yeung, King's College London
Catharine Young, British Embassy, Washington, D.C.
Bin Yu, University of California, Berkeley
Rapela Zaman, Royal Society
Jerry Zhu, University of Wisconsin, Madison

FOR FURTHER READING

For more detailed discussion of many of the topics addressed in this document, see the following publications:

The Fourth Industrial Revolution: Proceedings of a Workshop–in Brief, National Academies of Sciences, Engineering, and Medicine, 2017

Information Technology and the U.S. Workforce: Where Are We and Where Do We Go from Here? National Academies of Sciences, Engineering, and Medicine, 2017

Machine Learning: The Power and Promise of Computers That Learn by Example, Royal Society, 2017

Public Views of Machine Learning, Royal Society and Ipsos MORI, 2017

Refining the Concept of Scientific Inference When Working with Big Data: Proceedings of a Workshop, National Academies of Sciences, Engineering, and Medicine, 2017

Continuing Innovation in Information Technology: Workshop Report, National Academies of Sciences, Engineering, and Medicine, 2016

How Modeling Can Inform Strategies to Improve Population Health: Workshop Summary, National Academies of Sciences, Engineering, and Medicine, 2016

Privacy Research and Best Practices: Summary of a Workshop for the Intelligence Community, National Academies of Sciences, Engineering, and Medicine, 2016

Progress and Research in Cybersecurity, Royal Society, 2016

A 21st Century Cyber-Physical Systems Education, National Academies of Sciences, Engineering, and Medicine, 2016

A Look at the Legal Environment for Driverless Vehicles, National Academies of Sciences, Engineering, and Medicine, 2015

Preparing the Workforce for Digital Curation, National Research Council, 2015

Complex Operational Decision Making in Networked Systems of Humans and Machines: A Multidisciplinary Approach, National Research Council, 2014

Emerging and Readily Available Technologies and National Security: A Framework for Addressing Ethical, Legal, and Societal Issues, National Research Council and National Academy of Engineering, 2014

For further information, please contact:
Michelle Schwalbe, National Academies, mschwalbe@nas.edu
Jessica Montgomery, Royal Society, jessica.montgomery@royalsociety.org