You and AI

Conversations about AI technologies and their implications for society

THE ROYAL SOCIETY

SUPPORTED BY DeepMind

You and AI

Conversations about AI technologies and their implications for society

Artificial Intelligence (AI) is the science of making computer systems smart, and an umbrella term for a range of technologies that carry out functions that typically require intelligence in humans. AI technologies already support many everyday products and services, and the power and reach of these technologies are advancing at pace.

The Royal Society is working to support an environment of careful stewardship of AI technologies, so that their benefits can be brought into being safely and rapidly, and shared across society. In support of this aim, the Society's *You and AI* series brought together leading AI researchers to contribute to a public conversation about advances in AI and their implications for society.

What AI can, and cannot, do

The last decade has seen exciting developments in AI – and AI researchers are tackling some fundamental challenges to develop it further

Dr Demis Hassabis FREng FRS
Co-founder, DeepMind

In a lecture exploring the frontiers of AI, Demis Hassabis explained how AI algorithms can learn how to carry out a task, by analysing data and experiencing their environment. Instead of relying on designers to specify a solution to a problem, these algorithms use data to figure out how to solve a problem.

AI research seeks to understand what intelligence is, and then recreate this through computer systems that can automatically perform tasks that require some level of reasoning or intelligence in humans.

In the past, AI research has concentrated on creating detailed rules for how to carry out a task and then developing computer systems that could carry out these rules; researchers and programmers would develop a solution to a problem, codify this into a computer program, and in so doing create a system to implement the solution.

By pre-programming computer systems with rules in this way, computer systems can perform as well as experts in certain, specific tasks. However, while these systems have been successful in highly specific domains – such as playing chess – they have a significant weakness: inflexibility. If something unexpected happens or inputs do not follow a standard pattern, these systems cannot adapt their rules or adjust their approach.

In the last decade, new methods that use learning algorithms have helped create computer systems that are more flexible and adaptive, and Demis Hassabis FRS (co-founder, DeepMind) has been at the forefront of many of these developments.

AI technologies today are a topic of widespread interest, and there is a lot of speculation (and hype) about what AI might or might not achieve in the coming years. However, there remain significant limitations to what AI technologies can do (see Current challenges in AI research). For some, these limitations make AI particularly exciting, as they generate new lines of research that address profound questions about the nature of intelligence.

Image: Dr Demis Hassabis FREng FRS speaking with Professor Andrew Hopper FREng FRS, Treasurer and Vice-President of the Royal Society.

Current challenges in AI research

Unsupervised learning

Many recent advances in AI have been based on techniques called supervised learning or reinforcement learning. In supervised learning, a system is trained using many labelled examples of the correct solutions to a problem, so it can adjust its processes to get the right answer. In reinforcement learning, a system gets a score or reward for getting the right (or wrong) answer. Unsupervised learning deals with situations where systems do not have access to a labelled dataset and there is not an obvious reward system. Humans can learn in this type of environment, but it is challenging to create AI systems that can.

Memory and one-shot learning

Humans are capable of learning very quickly, sometimes from experiencing only one example of a task or problem. So-called 'one shot' learning is challenging for AI systems.

Imagination-based planning

When thinking about how to approach a task, humans can visualise different scenarios and the consequences of acting in different ways in those scenarios. This allows efficient planning for how a situation might play out in real life – thinking about the consequences of different actions before committing to a particular course.

Learning abstract concepts

As well as being able to plan detailed actions in an environment, humans also have a sense of high level concepts like language.

Transfer learning

Transfer learning is the process of taking knowledge from one area and applying it to a new problem that, on the surface, seems different, but which has similar underlying structures.

Language understanding

While humans can learn many different languages, understanding natural language remains a key challenge in AI – AI voice recognition systems can recognise words, but natural language processing to understand meaning at a higher level is more difficult.

Technologies with AI at their heart have the power to change the world

AI technologies already support many everyday products and services, with further applications to come

Professor Dame Julia Slingo DBE FRS
Chair, Cabot Institute, University of Bristol

In an event on AI applications Julia Slingo explained how AI techniques can help create the three-dimensional descriptions of the atmosphere and oceans that meteorologists use to predict local weather patterns.

Understanding the weather and climate

To understand daily weather patterns, meteorologists use AI techniques and simulations that model the Earth's atmosphere, ocean, and land surface, drawing from a combination of the fundamental laws of physics, complex maths, and a large amount of data about previous patterns in order to make predictions about the future. Julia Slingo FRS (Chair, Cabot Institute, University of Bristol) uses advanced statistics to understand the weather. As the predictive power of such techniques improves, such analyses can also help inform policy debates about the likelihood of extreme weather events, such as flooding.

Improving diagnostics and healthcare

One area of significant excitement around the application of AI is in healthcare, and Antonio Criminisi (Principal Researcher, Microsoft Research) is helping drive the development of such applications. One of the most difficult tasks for doctors treating cancer patients at present is in measuring tumours. Producing accurate measurements is important in tracking whether a tumour is growing or shrinking, which indicates how it is responding to treatment, and in creating more effective treatment pathways. AI technologies are helping create new systems to measure tumours, analysing scans and images and assisting medical experts to delineate or segment tumours, and to assess how they are responding to treatments.

Image: A panel discussion featuring Professor Marcus du Sautoy FRS, Professor Dame Julia Slingo FRS, Professor Suzanne Aigrain, Professor Steve Young FREng, and Dr Antonio Criminisi explored current and near-term applications of AI.

Interacting with computers and devices by voice recognition

Advances in AI technologies, such as deep learning, have dramatically increased the accuracy of automatic speech recognition systems that allow virtual personal assistants on smartphones and other devices to recognise and respond to commands. Steve Young (Professor of Information Engineering, University of Cambridge, and Senior Member of Technical Staff – Siri Team, Apple) is helping create these systems, and explained that the technologies that underpin these conversational agents are improving quickly. In future there may be more sophisticated systems that can communicate with users in natural language and that have the authority to execute transactions on behalf of their owner. Unlocking this potential requires further technical advances, for example in learning human conversational skills and improving systems' common sense reasoning. There are also broader questions about the ethics of data use by these systems – learning to answer queries requires access to large pools of data, which could raise privacy concerns – and questions about how society expects such system to answer 'unethical' questions.

Supporting advances in science

AI technologies are already helping scientists observe and understand the universe, extracting insights from large datasets to detect events or areas of interest for further research. In analysing data from the Keppler satellite, for example, AI algorithms can help pick out small signals that indicate the existence of planets from the large amounts of noisy data picked up by the satellite.

In future, such techniques may become even more important as a tool to extract insights from ever-growing datasets. The Square Kilometer Array, for example, will collect huge amounts of data from hundreds of stations, with an expected raw data flow over five times the size of 2015's global internet traffic. Researchers – such as Suzanne Aigrain (Professor of Astrophysics, University of Oxford), who is applying AI to astronomy – will rely on algorithmic systems to process this data, identifying which data to keep for further analysis, by screening for known phenomena or important and unexpected events.

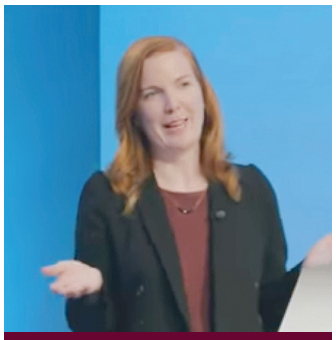
To make best use of these techniques, scientists will need to understand how to effectively implement AI systems in their work. This raises questions about how best to combine subject-specific expertise with technical AI know-how, and how to design systems that produce results, which can be interpreted by their users.

There are big questions about how societies want to use AI

As the power and reach of AI technologies expands, there are new questions and issues for society

Professor Cynthia Dwork
Professor of Computer Science, Harvard University

Cynthia Dwork explored an emerging theory of algorithmic fairness, and the challenges of creating technical solutions to issues of bias and fairness in AI.



Professor Kate Crawford
Principal Researcher, Microsoft Research, and Director, AI Now Institute

Kate Crawford discussed the politics of AI, and the implications of its widening use for social inequalities.

The growing variety and reach of these new applications mean that AI is no longer only a theoretical research domain – it is also an area of research that is applied at scale.

With this scale and influence come questions about its societal implications. Cynthia Dwork (Professor of Computer Science, Harvard) and Kate Crawford (Principal Researcher, Microsoft Research, and Director, AI Now Institute) explored such questions in two events on algorithmic fairness and the politics of AI.

Fairness and bias in AI

The datasets on which AI technologies are trained to carry out a task reflect society, and can contain the biases that were embedded in processes, relationships, or structures at the point of data collection. When this data is used to develop AI, the resulting systems reflect back the social and cultural structures or practices of the past; this means the biases that have shaped society in the past (or shape it today) can form the basis of predictions or recommendations about future action.

This has implications for how AI systems work – or fail to work – for different communities or users. Examples of such failures in recent years have included: women being less likely to be shown adverts for highly paid jobs, companies not making deliveries to poorer neighbourhoods, and racial disparities in approaches to policing or treatment by the justice system.

New research is developing ways of managing bias in data, for example by removing sensitive information – sometimes referred to as 'scrubbing to neutral' – before that data is used to develop AI systems. However, many of these current attempts to remove bias from AI are very narrow, and remain difficult to apply in some of the areas where fairness matters most, which are typically some of the most complex policy areas.

Questions about how to build fair algorithms are the subject of increasing interest in technical communities and ideas about how to create technical 'fixes' to tackle issues of fairness are evolving, but fairness remains a challenging issue. Notions of fairness can relate to groups, and whether different social groups are treated equally or experience similar outcomes, or individual expectations about personal outcomes. People think about fairness in different ways, drawing from ideas about equality of treatment or opportunity, or perceptions of what is just or right.

Technical fixes alone cannot answer these bigger questions about what fairness is or the type of social outcomes societies want AI systems to help create. They require an understanding of the broader forces that influence how and where AI systems are put to use, and the social forces that shape who designs AI systems and for whose benefit.

Thinking about the politics of AI in this way becomes especially important when examining the broader social and economic ramifications of these technologies, such as the impact of AI on work.

AI-enabled automation and work

AI technologies are enabling the automation of a growing range of tasks that are typically carried out by humans. These capabilities have promoted widespread speculation about the impact of AI on work, whether by influencing wage growth or unemployment rates.

AI will have a disruptive effect on work, with some jobs being lost, others being created, and others changing. Drawing lessons from previous waves of technology change and current economic conditions, Joseph Stiglitz ForMemRS (Professor of Economics, Columbia University) explained that technology-enabled changes to work tend to affect lower-paid and lower-qualified workers more than others.

The extent to which AI contributes to a broader, shared prosperity depends on the policies, structures, and institutions in place. Education and skills policy, corporate governance, anti-trust and market policies, labour relations, and employment practices all play a role in shaping how AI is adopted and its influence on work. Ultimately, the extent to which the benefits of AI-enabled economic growth are felt across society will depend on the rules of the game that such policies and institutions set in place, and active consideration is needed now about how society can ensure that the increased use of AI is not accompanied by increased inequality.

Professor Joseph Stiglitz ForMemRS
Professor of Economics, Columbia University

Joseph Stiglitz discussed the impact of AI on work and employment, exploring the potential implications for societal inequalities and the policy responses that could help address these.

Sharing the benefits of AI across society

Dame Wendy Hall DBE FREng FRS
Regius Professor of Computer Science, University of Southampton

During a panel discussion chaired by Professor Jim Al-Khalili FRS, Wendy Hall spoke about how the UK can remain a leading player in the development of AI technologies.

An environment of careful stewardship can help ensure that the benefits of AI are shared across society

AI technologies could support new projects, services, and economic growth that have the potential to bring widespread benefits. As the power and reach of these technologies grow, there are important questions for societies to grapple with about the ways in which AI should – or should not – be used.

To close the series, two panels of experts answered questions from the public about who is – or is not – currently benefitting from AI, and how its benefits can be shared across society.

Wendy Hall FRS (Regius Professor of Computer Science, University of Southampton), Neil Lawrence (Director of Machine Learning, Amazon Research Cambridge), and Ewa Luger (Fellow in Digital Arts and Humanities, University of Edinburgh) considered how AI could benefit communities across the UK, and how the UK can maintain a leading role in leading the development of these technologies internationally.

Then Peter Donnelly FRS (Professor of Statistical Science, University of Oxford), Vivienne Ming (co-founder, Socos Labs), and Suchi Saria (Professor of Machine Learning and Healthcare, Johns Hopkins University) debated who will likely be most affected by AI technologies, how their impacts will be distributed across society, and what type of interventions might be necessary to address concerns about bias, transparency and accountability, and the future of work.

With such significant potential benefits – and such significant potential disruption to lives and livelihoods – it is important there is a public conversation about AI and its societal implications. The Royal Society is working to facilitate an open and robust conversation about the potential and pitfalls of AI, and the lectures, videos, and other materials generated by the You and AI programme aim to support this conversation.

You can find out more at **royalsociety.org**

Image: Panel from left to right: Dr Vivienne Ming, Professor Peter Donnelly FMedSci FRS, Professor Suchi Saria with Professor Brian Cox CBE FRS.

Thank you

The Royal Society would like to thank all those who contributed to these events:

Professor Jim Al-Khalili OBE FRS

Professor Suzanne Aigrain

Professor Chris Bishop FREng FRS

Professor Brian Cox OBE FRS

Dr Antonio Criminisi

Professor Kate Crawford

Professor Sir Peter Donnelly FMedSci FRS

Professor Cynthia Dwork

Dame Wendy Hall DBE FREng FRS

Timandra Harkness

Dr Demis Hassabis CBE FREng FRS

Professor Andrew Hopper CBE FREng FRS

Dr Neil Lawrence

Dr Ewa Luger

Dr Vivienne Ming

Professor Suchi Saria

Professor Marcus du Sautoy OBE FRS

Professor Dame Julia Slingo DBE FRS

Professor Joseph Stiglitz ForMemRS

Professor Steve Young FREng

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society. These priorities are:

- Promoting excellence in science
- Supporting international collaboration
- Demonstrating the importance of science to everyone

For further information
The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T +44 20 7451 2500
W royalsociety.org

Registered Charity No 207043
July 2019 DES5775