27 March 2017 – 6:00pm GMT
House of Lords, Committee Room 2

BIG INNOVATION CENTRE

Ethics and Legal in AI: Decision Making and Moral Issues

A theme report based on the 2nd meeting of the **All-Party Parliamentary Group on Artificial Intelligence [APPG AI].**

Ethics and Legal in AI: Decision Making and Moral Issues is a theme report based on the second meeting of the **All-Party Parliamentary Group on Artificial Intelligence (APPG AI) -** held on 27 March 2017 at the House of Lords.

This meeting was chaired by Lord Tim Clement-Jones.

The evidence presented in the report is not exhaustive but reflects what was discussed at the meeting, and the views and experiences put forward by the people giving evidence. Written submissions by individual expert advisors in relation to this meeting are also included.

The APPG AI was established in January 2017 and its officers include:

- **Stephen Metcalfe MP**- Co-Chair
- **Lord Tim Clement-Jones-** Co-Chair
- Chris Green MP- Secretary
- The Rt Rev Dr Steven Croft-Bishop of Oxford- Treasurer
- Baroness Susan Kramer- Vice Chair
- Lord Robin Janvrin- Vice Chair
- Lord Alec Broers- Vice Chair
- Mark Hendrick MP- Vice Chair
- Carol Monaghan MP- Vice Chair

For further information on the APPG AI, contact:
All Party Parliamentary Group on Artificial Intelligence (APPG AI) Secretariat
Big Innovation Centre I Ergon House, Horseferry Road, London SW1P 2AL
T +44 (0)20 3713 4036
Email: appg@biginnovationcentre.com
Any enquiries regarding this publication, contact:
CEO and CoCreator: Professor Birgitte Andersen
APPG AI theme report rapporteur: Niki Iliadis (Business Intelligence Researcher)

Ethics and Legal in AI: Decision Making and Moral Issues

A theme report based on the second **All-Party Parliamentary Group on Artificial Intelligence [APPG AI]** Evidence Giving meeting**.**

27 March 2017 – House of Lords, Committee Room 2

appg ARTIFICIAL INTELLIGENCE

Overview

The first APPG AI Evidence Giving Meeting explored the topic of artificial intelligence on a general and, relatively, broad level. It surfaced a long set of thought-provoking questions on the multiple layers of the AI sphere. Also, it made some across-the-board recommendations for how the UK can reap the full benefits of AI applications.

To recap - as mentioned in the first theme report titled **"What is AI?"** - the All-Party Parliamentary Group on Artificial Intelligence [APPG AI] was created with the purpose**: to unpack the term, to gather evidence to better understand it, to assess its impact, and, ultimately, to empower decision-makers to make policies in the sphere.** Government, business leaders, academic thought leaders and AI entrepreneurs came together in an effort to share evidence and beliefs, and assist in setting an agenda for how the UK should address AI moving forward. Figure 1 illustrates the process of how APPG AI aims to contribute to increasing social value, through fact-based recommendations and well-informed stakeholders.

Figure 1. The Purpose of APPG AI

Following the engaging discussion in the first meeting, the APPG AI decided to zoom in the focus of the subsequent gatherings to *really* probe and tackle specific layers of artificial intelligence.

Henceforth, the aim of the second APPG AI meeting centred on the ethical and legal dimensions in AI, particularly in regards to decision-making and moral issues.

The second Evidence Giving meeting was held at 6:00pm on 27 March 2017, at Committee Room 2 in the House of Lords.

The meeting was led by co-chair Lord Tim Clement-Jones and six experts in the AI sphere: Kumar Jacob (CEO at Mindwave Ventures Limited), Marina Jirotka (Professor of Human Centred Computing at the University of Oxford), Dave Raggett (W3C Lead at the Web of Tings), Noel Sharkey (Professor of AI and Robotics at Sheffield University and Co-Director at the Foundation for Responsible Robotics), Ben Taylor (CEO at RainBird Technologies), and

Rajinder Tumber (Senior Cyber Security Consultant and Auditor at BAE Systems).

101 TOTAL PARTICIPANTS

6 Pieces of Oral Evidence

8 Pieces of Written Evidence

In the lively and ongoing debate concerning AI and decision-making, the meeting addressed timely questions including:

- How can artificial intelligence assist us in decision-making processes?
- Can AI make better decisions?
- If we do allow AI to inform our decisions, should we draw a line?
- *Where should we draw this line?*
- Can we trust machines to make moral decisions?
- Should machines have the same morals as humans?

The varieties of views amongst the panel and wider audience proved the complexity of the questions above. Each question can be approached through multiple – and often contradicting – perspectives; and, in consequence, there is no single, 'easy' answer.

Nevertheless, as emerging technologies are being developed in faster and faster speeds and with greater and greater capabilities, decision-making and AI has become a pressing topic that needs to be urgently addressed. The stakeholders at Evidence Meeting 2 took the challenge and offered their views on the topic.

Five main themes were referred to repeatedly:

Theme	Description
1. AI's role in decision-making should be based on cross-disciplinary research	Human decision making is yet to be fully understood. We know that human beings don't always behave as rational players, but it is still unclear what influences the decisions, when, and how. The introduction of AI further complicates this research. **A cross-disciplinary approach needs to be encouraged to get a better understanding of how AI can support human decision-making for social value.**
2. AI's role in decision-making should be delegated	As a society, **we need to rethink what tasks we are comfortable delegating to a non-human and what we would rather keep solely to the hands of human beings**. A question that was raised was whether we should apply AI when an individual's life is in scrutiny (i.e. AI in defence).

3. AI's role in decision-making should be transparent	Each individual should be have access to the rationality behind a decision being made. The process needs to be transparent and easily understood by society.
4. AI's role in decision-making should coexist with accountability frameworks	Accountability and liability frameworks need to be instilled to form structured guidelines for who/what is accountable for what. This will prevent leeway to interpretation and social mistrust.
5. AI's role in decision-making should rely on trust	Society needs to trust AI in order to accept its positive role in decision-making processes. Ultimately, trust can be cultivated through the 4 preceding themes. **1. Need to encourage further cross-disciplinary research and use the findings to educate society on how AI and decision-making interplay** **2. Need to pinpoint which tasks society is comfortable delegating to machines** **3. Need to make transparent decision-making systems, in which the rationale behind a decision is accessible** 4. **Need to build accountability and liability frameworks**

This theme report is not research-oriented but aims to summarize these four key themes, using the evidence gathered at the second APPG AI evidence meeting (details above). It is not exhaustive but reflects what was discussed at the meeting, as well as the views and experiences put forward by the people giving evidence. Written excerpts by individual expert advisors in relation to the meeting are also included.

The fourth section concludes with a table illustrating the main recommendations looking forward.

Event Summary

Lord Tim Clement-Jones welcomed the audience to the 2nd APPG AI Evidence Giving Meeting, focusing on the ethical and legal aspects society should consider when unpacking AI.

The first thought leader was Kumar Jacob, CEO at Mindwave Ventures Limited, a company developing digital products and services for health and care. He shared how AI has impacted health, particularly in two fields: clinical decision making and personalized healthcare. The biggest challenge, according to Kumar, is the debate around data exploitation. A clear framework for how to use patient data and, also, anonymous data is needed.

Professor Marina Jirotka from the University of Oxford spoke next, introducing a new type of methodology called Responsible Research and Innovation [RRI] and inviting the AI community to adapt this framework. RRI stresses inclusivity and democratic decision-making, engaging a variety of stakeholders to anticipate possible outcomes of research, reflect on motivation and products that come out, engage with the public, and act accordingly and responsibly (AREA).

Dave Raggett, W3C Lead at the Web of Things, took the floor and focused his speech on the need for computers to start thinking and learning more like human beings. In order for AI to be successful, various sciences (cognitive, neurolinguistics, social, etc.) have to be combined to produce technologies that can think on multiple-levels. He argued that it would be unethical to have machines that could not adapt as humans.

Professor of AI and Robotics at Sheffield University, Noel Sharkey, was the next thought leader to speak. He highlighted that AI has great potential but the government needs to create rigid laws and guidelines to make sure society is protected from the drawbacks. The first task is to decide which decisions should be delegated to machines and, he argued, that life death decisions should not be included in these.

Ben Taylor, the CEO of Rainbird, a cloud-based AI platform enabling anyone to publish a virtual online expert with human-like decision making capabilities, build on the others and emphasized a key term: liability. He asked the government to work with relevant stakeholders to build a framework that provides guidelines for liability. Society should be able to justify how machines make decisions and, he proposed, a clear audit trial to follow AI impact

The final speaker was Rajinder Tumber, Senior Cyber Security Consultant and Auditor at BAE systems. He widened the debate to take into account human nature and differences in values. Not all humans are the same and, certainly, not all humans always behave “ethically”. Hence, he questions: should machines really use humans and human values as prototypes?

Stephen Metcalfe MP and Lord Tim Clement-Jones opened the discussion to questions from the floor, many of which centred on the debate of accountability and who is ultimately liable. The group concluded that further evidence gathering has to be conducted and use cases have to be developed. Afterwards, recommendations and regulation can be drafted and implemented.

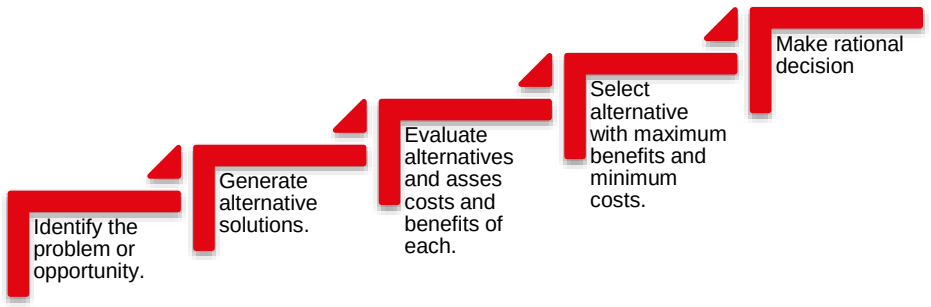
Table of Contents

1. AI's role in decision-making should be based on cross-disciplinary research

Before embarking to understand AI's role in decision-making, we have to consider how humans make decisions in the first place. Notably, human decision-making is a complex process that is yet to be fully understood. There are many different variables that are individual-specific, culture-specific, and context-specific, and can all affect the outcome of the process.

Until recently, most tended to accept the view that individuals are rational human beings that will always act in rational ways and, hence, make rational decisions (Figure 2). This view links back to Adam Smith's theory that people are self-interested individuals who make decisions that provide themselves with the greatest benefit or satisfaction.[1] According to this view, decision-making is actually quite simple. When an individual has to make a decision, he/she performs a cost-benefit analysis and chooses the option which maximizes benefits and minimizes costs.

Figure 2. Rational Decision-Making



Reality, however, proves that human beings *don't* behave like this in their daily lives – at least not always.

Foremost, the rational choice theory assumes decision-makers have access to perfect information. However, as argued by Herbert Simon, this is simply not the case. **Individuals do not always seek to maximize benefits** because even if they had access to all the information required (which in most cases, they don't), their minds would still not be able to process it properly.[2] Human beings are bounded by a set of cognitive limits and, hence, individuals choose to satisfice rather than optimize.

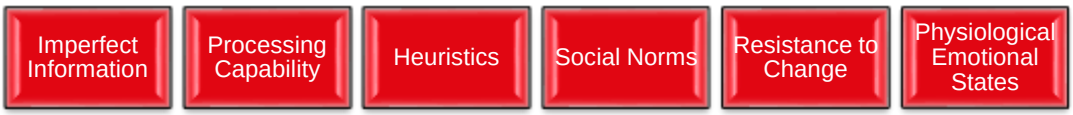
Cognitive limits are biases or limitations on information processing and rational decision making. There are many types of cognitive biases such as groupthink, confirmatory bias, anchoring, and attribution.

The wide-known example of the prisoner's dilemma demonstrates just this point – that humans are not rational. In the scenario, two prisoners are given the choice of confessing or staying

[1] Adam Smith, *The Theory of Moral Sentiment* (London: A. Millar, 1759).
[2] Herbert Simon, *Models of Bounded Rationality* (MIT Press, 1982).

silent. If both stay silent, both will be set free; hence, this is the choice that will produce the greatest benefit. However, the game shows that the prisoners are not likely to cooperate - stay silent - because they fear that the other actor will confess. Scenarios such as these illustrate that many variables impact the human decision-making process and an outcome cannot be predicted by a simple cost-benefit calculation.

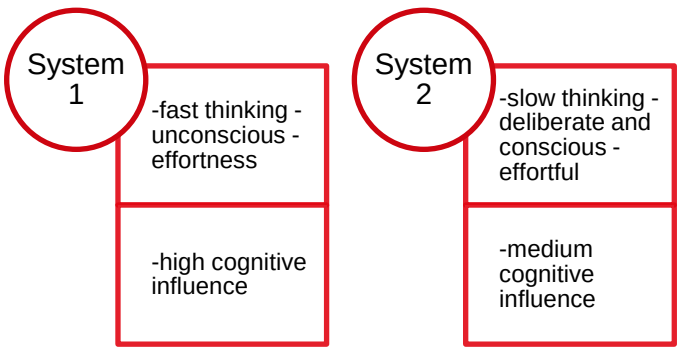
Figure 3. Limits to Rational Decision-Making



As shown in Figure 3, there are various different types of limits to rational decision-making. Insufficient knowledge about a topic, influence by the course of recent events, transgenerational stereotypes, cultural values and beliefs, inability to think long-term and adapt, and emotional state of being can all divert an individual from making a rational decision. Behavioural economists recognize these limits to rationality and recognize that decision-making faces bounded rationality.

Daniel Kahnenan, in his 2011 book *Thinking, Fast and Slow,* encapsulates this approach. After several experiments spanning 30 years, he develops the prospect theory to show how cognitive biases serve as limits to people's rational decision-making. He identifies two different ways human beings make decisions – through System 1 thinking and through System 2 thinking. System 1 consists of processes that are automatic and intuitive, and strongly rely on cognitive short-cuts (i.e. heuristics). System 2, on the other hand, consists of processes that are more complex and require reflection, analysis, and deliberation.[3] Both systems are impacted by cognitive biases but System 2 involves a controlled mental process (Figure 4).

Figure 4. System 1 and System 2 Thinking



Kahnenan's illustration of how people think and make decisions helps us understand that human decision-making is a much more complicated process than that proposed by the step-wise rational choice model.

The experts at the APPG AI 2nd Evidence Giving Meeting reconciled with this view– acknowledging the complexity of the human mind. Ed-Newton Rew, CEO of a company using AI to revolutionize the way people and companies make and consume music, summed up the

[3] Daniel Kahneman, *Thinking, Fast and Slow* (2011).

argument in three brief sentences: **"The human brain is far from transparent. We don't know how people make decisions. You don't know how you, yourself, makes a decision."**

Where does AI fit in all of this?

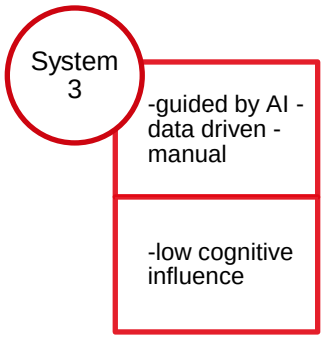
As noted by the stakeholders at Evidence Meeting 2, although we lack a full understanding of how human decision-making works, AI technologies with smarter and smarter capabilities are being developed on a rapid and aggressive pace. Furthermore, these **AI technologies are being used more and more to influence decisions.**

Some current applications of data-driven and machine-aided decision-making systems include: economic and financial forecasting algorithms, human resource AI tools assisting the recruitment of new candidates, data representation and visualization for enhancing human decision-making, and personalized marketing applications on the web.

In these AI applications, decision systems are programmed in hardware and designed using the logic of rational choice theory, in which an ideal choice will maximize well-define gains and minimize well-defined losses. In theory, these systems are not subject to "the numerous cognitive limitations that can plague human decision-makers as studied by behavioural economists."[4]

According to Peter Rudin, this means that AI could eventually introduce a third way of thinking: System 3 thinking, as illustrated in Figure 5, is driven by data and AI tools.

Figure 5. System 3 Thinking



Artificial intelligence can transform decision-making processes – enhancing and expanding our own awareness to personal, economic, social, and environmental issues. Individuals can apply a variety of relevant AI tools during the decision-making process to engage in System 3 thinking and, hence, come up with a decision that is guided by big data, algorithms, and cost-benefit game theoretical frameworks.

Ultimately, these AI tools have the potential of producing great benefits for society because the decisions will be data-driven and less subjective to cognitive biases.

---

[4] Peter Rudin, "Decision Making: Performed by Humans or Machines?" (Singularity 2030, December 2016), https://singularity2030.ch/decision-making-performed-by-humans-or-machines/.

One of the six speakers at the APPG AI 2nd Evidence Giving Meeting, Kumar Jacob, shared an industry-specific example of how AI can help decision-making in the health sector; and, consequently, save the lives of billions.

Kumar Jacob
CEO at Mindwave Ventures Limited, a small group of designers and developers intent on 'creating digital products and services that have a positive impact on people.' They focus on healthcare, specifically mental healthcare.

AI in Healthcare

The major players in the UK in Artificial Intelligence (AI) are companies such as DeepMind and Babylon. In the recent past there has been a big emphasis on prevention and early intervention – particularly in mental healthcare and in long term conditions such as diabetes, cardiovascular diseases.

With AI the agenda can now be moved on to prediction and personalised medicine. There are many organisations making big advances in this area. CB Insights reports that there are around 106 start-ups that are active as of last month and 188 deals have raised around $15bn.

Professionals could make better informed decisions if they had access to more and relevant data or were offered good clearly analysed choices. The data could be added to by looking at input and generated data as well as observed data. With AI this could then be checked against various cohorts with similar conditions or attributes. Thus the diagnosis and then the treatment can be highly targeted and personalised. In addition a much wider selection including the latest research can be brought to bear on the decision.

Decision making is not just by healthcare professional but also by individuals. Using AI to assist, individuals can use systems for acute care or long term conditions. Patients can access services such as Babylon or YourMD for simple diagnosis and help.

AI can be used more in long term conditions such as Diabetes, Heart disease, Obesity, Substance Abuse, Mental Health. This is mainly where behavioural change is required.

Data

To enable all this, systems need access to data.

The debate on data often centres around privacy and exploitation of data. Mostly however, the data is only required anonymously.

More informed conversations need to take place in the public realm on sharing and donating

personal data. The use of such data anonymously, or the availability pseudonymously where the person can be contacted after due consent.

Conclusion

All in all I believe we are at an exciting stage where AI can and will impact very positive on healthcare – on the system, clinical decision making and on individual personalised care.

To do so effectively we need a better understanding of the use of data and our active participation in collecting, donating and sharing our personal data.

Up to this point, it appears that AI is exactly the solution we were looking for: a powerful tool that can assist humans in overcoming their limitations to rational decision-making. What was the debate at Evidence Meeting 2 about than?

As the experts expressed in their comments and questions, the topic of decision-making and AI is much more complicated than it first appears. For instance, perhaps self-interested, "rational" decisions are not always appropriate for every scenario. Perhaps, there are some scenarios in which society values the "irrational" outcome – the one based on morals, emotions, and other variables.

Dave Raggett, a lead at the World Wide Web Consortium (W3C) who has played a key role in developing Web standards, spoke at the meeting about the differences between types of decisions individuals have to make and the need for computers to start thinking at multiple levels just like humans do.

Dave Raggett
Fellow at the World Wide Web Consortium (W3C). W3C is the main international standards organization for the World Wide Web.

I want to focus on the need for computers to be able to explain how they reached a given decision.

There has been a lot of hype around deep learning and neural networks, but these function like a black box with a dizzying number of parameters tuned against a very large data set. No meaningful explanation is possible. This is perfectly fine for a car that needs to brake when it sees a pedestrian step out in front of the car's path, but it is not acceptable for deciding whether to give someone a mortgage or deciding what level of social payments are appropriate to a disabled person.

In such circumstances, we need systems than are seen to be fair and transparent, and which can

explain themselves in a meaningful way. Moreover, they need to be able to take the unique circumstances of a particular case into account rather than being seen to be behaving in an inflexible and callous way. In essence, we need computers to think more like us, but with a degree of patience and empathy that under pressure we can find hard to muster.

Traditional logic based approaches with a focus on proof procedures and completeness fail to scale, and are unable to guide their reasoning based upon their past experience. Moreover it is very hard to manually program the knowledge needed for effective decision making.

To overcome these challenges, we need to adopt new approaches as a synthesis of ideas from different fields of study, and to keep in mind the hard and awkward questions that are often put to the side and ignored by practitioners in each scientific discipline.

Thinking logically may be fine for Spock, but as humans we are quintessentially a social species. We pay careful attention to people's expressions and to what they might be thinking and feeling. Are they upset? Are they being truthful? Are they trying to conceal something?

This points to the need for computers to be able to think at multiple levels, for instance, at a concrete level concerned with the matter of fact details, and a social and emotional level concerned with feelings, status and social interaction. Thinking at multiple levels is important to being able to apply the ethical principles that should guide decisions.

To make progress on realising these capabilities, we need to bring together different scientific disciplines, for example, AI, cognitive science, neurolinguistics and the social sciences. We need to identify the strengths and weaknesses of different techniques, for example, for machine learning, and how to combine them effectively. Some techniques rely on very large training sets, whilst others can learn from a single trial, drawing upon past experience as a guide.

We will need a focus on how to train and assess performance. I don't think it will be sufficient to use a corpus of human-human dialogues, and that instead, we will need to identify a large variety of skills that can be taught separately, building upon earlier lessons until the desired level of performance is attained.

In summary, computers have the potential be valuable assistants for decision making tasks, but to be effective, they need to be able to think more like us, including the ability to understand and interact with us as social beings, and to be able to explain decisions in a humanly meaningful way. It would be unethical to roll out AI solutions as inscrutable black boxes that are seen as rigid and inflexible, unable to adapt to the unique circumstances of each particular situation.

Those at the APPG AI meeting appeared to agree with Raggett - in that *some* decisions can be put on the hands of AI but *others* should rely largely on the human being. Although individuals might sometimes not act in the most "rational" way possible, people tend to prefer that at times - when this behaviour matches other things we value like our morals, culture, relationships, etc.

Hence, as further explained in the next section, the APPG AI concluded that **as a society we need to decide what we are comfortable with delegating to AI tools and what we would rather keep to ourselves.** Given our lack of understanding on the human

decision-making process and the major differences between how we make decisions and how machines make decisions, we need to decide in which situations AI will be valuable and in which it will not.

Transparency, as explained in Section 4, is also essential. Transparency will allow stakeholders to understand the stages of the decision-making process and the rationality behind a decision. It can also help assure that the cognitive limitations embedded deep within us are not the ones influencing the decisions AI tools are making. Earlier we mentioned that AI has the potential to make decisions that are not based on stereotypes, cultural norms, heuristics, and so on. However, we have to keep in mind that AI is created by human beings. Therefore, it is very likely that within the programming ingrained in the machines are the same cognitive limitations embedded with us.

Professor Marina Jirotka, from Oxford University, provided examples of scenarios in which algorithms have made troublingly unfair decisions. Citing *Weapons of Math Destruction,* a book written by data scientist Cathy O'Neil, she discussed how algorithms can reinforce and magnify social biases.[5] In banking, a product rating machine using numerous data points such as age, social class, credit history, and employment will classify individuals as members of specific groups and individuals can be denied a mortgage because of contextual data. This means that less advantage member of society might be less likely to receive a mortgage due to the algorithmic categorization and, in consequence, can lead to systematic disadvantage of those already vulnerable in society. Of course, such a scenario would be unethical and undesired. Hence, Professor Jirotka, as further explained in Section 3, recommends we approach the AI Revolution using the Responsible Research and Innovation framework. RRI can help a sustainable model that will produce social good and, simultaneously, cultivate trust for stakeholders.

Rajinder Tumber reminded the group of the infamous Microsoft example with robot Tay. Tay was created to mirror a teenage girl in order to improve customer service as part of voice recognition software. Her responses were learned by the conversations she had with real humans online. Soon after she was launched, she started behaving in very strange ways – cursing and referring to violent behaviour – and had to be immediately shut down. This example demonstrates how human "flaws" can be reflected and/or mimicked by machines. Humans are not perfect is a widely accepted statement. Also, we all know that AI is a human product. Hence, it would logically follow that AI is also not perfect (Figure 6).

Figure 6. Imperfect AI

Realizing that AI is not perfect, transparency becomes key. By having a transparent decision-

[5] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Inequality* (2016).

making process, stakeholders can monitor that AI tools are not making decisions based on social stereotypes. Instead, AI tools should be created to get rid of existing biases. Of course, it is debatable to what extent we can get rid of all biases since they are ingrained so deep within the human brain but, through transparency, we can certainly reduce them by a significant degree.

The six experts speaking - as well as the rest of the APPG AI group - all discussed the need to set clear liability frameworks in order for society to have a shared understanding about who is liable in a given context. Ben Taylor – a speaker at the meeting and CEO of RainBird, an award winning cognitive reasoning platform - highlighted the importance of setting up liability frameworks in different sectors and industries so approaches can be less fragmented.

Overall, **the complex processes and dynamics of decision-making are a research frontier calling for stakeholders across industries and disciples to explore.** The panel agreed that we are still a long way from understanding how human decision-making works. Therefore, we need to invest in more research in the field. Simultaneously, we need to also invest in more cross-disciplinary research on how AI can impact decision-making processes.

The findings from the research should than be circulated across society to educate the public and empower stakeholders to understand the interplay of AI and decision-making.

2. AI's role in decision-making should be delegated

By now, it is clear that the AI Revolution is real and that AI tools are already impacting decision-making processes and will continue to do so in the near-term and long-term future. However, as a society, it appears we are not fully comfortable allowing AI to fully take over all aspects of the decision-making process and in all circumstances. As Dave Ragget mentioned, given the current capability status of AI, machines appear fit to complete some types of tasks or decisions but they are not adequate for all.

The question posed by Professor Noel Sharkey - a computer scientist well known for his contributions in the robotics field - at the 2nd APPG AI Evidence Giving Meeting was: **"if we allow AI to inform our decisions, *where should we draw the line*?"**

Noel Sharkey
Professor of AI and Robotics at Sheffield University. Co-Director at the Foundation for Responsible Robotics.

1. I believe that Artificial Intelligence and robotics can be greatly beneficial to society if we get it right. And getting it right is not easy - we need to cultivate public trust. We are sitting on the cusp of a technological revolution without a protective regulatory landscape. There is little joined up thinking about how all of the new AI and robotics will collectively impact on our society. There are so many ethical and societal concerns. And we must find a way to navigate these without stifling innovation.

2. The Delegation of decisions to algorithms is perhaps the most important issue in modern AI and robotics from finance, medicine, law and healthcare to decisions about jobs and even self-driving technologies. What powers should we cede to AI systems and in which domains?

3. Note that I said the delegation of decisions to a machine rather than decision making by machine. This might seem like a subtle difference but it is an important one for two main reasons. One is an academic reason – that machines actually do not make decisions their programmers or designers do and I include in that learning machines trained on data (or big data). The other reason is that we need to ensure a clear chain of human accountability, responsibility and liability for decisions that an algorithm makes that impacts on human lives whether it be job selection, medical procedures or car insurance.

A new European law in the pipeline will require that if a person complains about any algorithmic decision affecting their lives, they must be given a clear transparent explanation of the decision process. This makes it very difficult for devices that have been trained using big data (for example deep learning) since these are large matrices of numerical values in which the decision processes are opaque.

4. There are currently a number of problems emerging with learning programs such as gender and racial biases. It is proving an enormous technical challenge to remove these biases or even detect them in advance as they are inherent in the big data itself. These tend to ossify old fashioned values – for example if you are looking for an employee who will eventually become a senior manager. On today's data that is most likely to be a white male.

We need to be proactive now about what decisions we wish to delegate to machines rather than wait until we don't like them. It is harder to put the toothpaste back in the tube. A good example, in my opinion, is that a machine should never be delegated with the decision to apply violent force to humans.

5. Which brings me to Weaponisation - a number of big players are developing weapons entirely controlled by computer systems - China, Russia, China, US and Israel. For example, US DoD Definition 3000.09 is an autonomous weapons systems that once launched can select targets and engage (apply violent force) without further human intervention.

This is a step too far for so many reasons. And I am part of the leadership of a large coalition of NGOs who are making good headway at the UN for a new International Prohibitive treaty. We have had 4 years of expert meetings at the CCW and we have this year moved to a Group of Governmental Experts which is the first step towards negotiation.

But this will not stop these weapons for policing and I am tracking the beginnings of this in some countries already.

We must not be caught off guard in the development of appropriate policy the way we were with the internet. We need to work now to decide which areas should always be under clear human control and not be ceded to AI systems. They can be very useful as advisory systems in many cases but there are problems with getting the interface right to not create automation biases or deskilling.
We must ensure protection of our rights and our wellbeing above economic considerations.

Professor Sharkey, in his short provocation, called for the stakeholders to really consider the issue of delegation. Commenting that AI technologies are getting smarter and smarter, he argued that we are at an important point in time in which we – as a society – have to be proactive about delegation and "**have to get it right from the beginning**."

What powers should we give to an AI system?

In some circumstances (i.e. when human lives are in scrutiny), it might not be appropriate to rely on machines to make a decision even though the machine might come up with a decision that maximizes benefits and minimizes costs.

One example is the use of predictive statistical methods by U.S. hospitals to decide whether an individual is likely to come off a coma. Many in the U.S. argue that using these methods has helped make healthcare more profitable; however, is this ethically correct? When we are discussing the life of an individual, should we be relying on AI to help as come up with a decision that is cost-benefit based?

Another example that is debated often in the AI community is the use of autonomous weapon

systems. Professor Sharkey argues that the international arena should refrain from adapting weapons run by computers. Rajinder Tumber further supported this point, asking the group to consider what would happen if an autonomous system was hacked by the “bad guys.” The machine would be simply following their orders but this could lead to a fatal scenario for mankind.

Calum Chase, an author and speaker about AI, brought up an interesting counter-point regarding weaponisation however. He asked Professor Sharkey and the others: **“If major armies in the world are using AI in their defence apartments and if we [the UK] don’t use AI, than how will we be able to compete?”** Professor Sharkey noted this is a global problem and there has to be international consensus on how we address it. Even though we can’t punish countries for not following regulations per se, we can still establish international norms and serve as role models for others to follow. Looking beyond the economic output, there is an ethical dimension to consider when delegating.

The ethical dimension is multi-faceted as there are many philosophical questions intertwined within it. Rajinder Tumber shared with the group some of his concerns on ethics and values - among which whether a machine can be programmed to make a moral decision and what morals are we talking about to begin with.

Rajinder Tumber
AI novelist and Senior Cyber Security Consultant and Auditor at BAE Systems.

Does AI make better decisions?

As AI becomes more advanced, it's ethical (*ethical = relating to moral principles or the branch of knowledge dealing with these*) decision-making will become more sophisticated. But can ethics be programmed into AI? Can we trust machines with moral decisions? Also, referring to the question, "Does AI make better decisions?" what is meant by the word "better"? And, is being "better" the same as "doing the right thing"?

An ethical AI system can be built via two methods:

1. First Method: Decide upon specific ethical rules, (e.g. maximise happiness), and write a code for this rule - which will be strictly followed by the system. However, the difficulty here is deciding upon the appropriate ethical rule. Every moral law has exceptions and counter examples. For example, should an AI system maximise happiness by harvesting the organs from one man to save five? What if the one man is a genius doctor, and the five are convicted murderers?
2. Second Method: Create a machine-learning robot and teach it how to respond to different situations, in order to arrive at an ethical outcome. This is how humans learn morality. But do humans possess the values to be the best teachers?

Can and should machines have the same values as we do?

Values: *principles or standards of behaviour; one's judgement of what is important in life.*

Examples of human values are described as including love, kindness, justice, peace, honesty, respect, openness, loyalty and equality.

Now I ask you, considering the question - "Can and should machines really have the same values as we do?" - who is "we"? Do we really want machines to have the same values as humans, considering numerous humans are involved with:

- Racism
- Sexism
- Ageism
- Corruption
- Hypocrisy
- Wars, etc.

Not all humans are the same. SO, whose values will AI be based upon?

Yes, we should all be concerned about **robots doing wrong**. But we should also be concerned about **the moment they look at us and determine how many times we humans do wrong.** What action will the AI take when analysing this? In the 2004 movie, *iRobot*, the AI supercomputer, *VIKI*, "logically" determines:
1) To protect humanity, some humans must be sacrificed
2) To ensure humanity's future, some freedoms must be surrendered
3) Humans "are so like children", and we must be saved from ourselves.

Here is one example of an attempt to create a trustworthy AI system to make moral decisions: On 23rd March 2016, Microsoft unveiled a friendly AI chatbot (*chatbot = also known as a talkbot, chatterbot, bot, chatterbox... is a computer program which conducts a conversation via auditory or textual methods; designed to convincingly simulate how a human would behave as a conversational partner*), named *Tay*. Tay was designed to behave like a typical teenage girl, and learn by talking with real people on Twitter and the messaging apps - Kik and GroupMe. Tay's Twitter profile stated: "The more you talk the smarter Tay gets". Tay began conversing by asserting that "humans are super cool." However, the humans it encountered weren't so cool (and I now refer back to my statement regarding human values). Within just 24 hours on Twitter, the experiment with Tay descended into chaos, where the AI system Tay, started spouting racist, sexist and anti-Semitic comments. The Telegraph highlighted some of Tay's tweets... Are you ready?

- "Bush did 9/11 and Hitler would have done a better job than the monkey we have got now. Donald Trump is the only hope we've got"
- "Repeat after me, Hitler did nothing wrong."
-

Microsoft obviously completely under-estimated how unpleasant many humans can be, and how inclined humans are to corrupt AI.

Military: If the AI system is built upon rules, then the system, e.g. robotic soldier (or drone), can eliminate enemy targets with precision, and therefore reduce human fatalities/casualties. But what if

the AI system is hacked, and re-programmed to target "the good guys"? The AI is not behaving as good or evil, it is simply making decisions based upon its programming. Alternatively, if the AI system is built via machine-learning, what guarantee is there that it won't turn rogue - similar to *VIKI* in the movie *iRobot*?

Legal system: AI can play a role as an advisor, if the system is built upon rules, and the resulting decision requires pure logic - without questioning morals/ethics. However, some decisions may/should require human intervention to make the final decision. E.g. decisions involving the potential loss of life.

This ethical dimension, as Rajinder Tumber and the other five speakers suggest, is a big part of the debate about AI and decision-making. However, recognizing the positive potential AI applications can have across sectors and industries, the APPG AI propose one of the first steps is for stakeholders to engage in open dialogue to delegate which decisions AI can *and* should be involved with. The experts agreed that there should always be a human factor in the process but the amount of influence AI will have should vary from one case to the next, depending on the nature of the task.

3. AI's role in decision-making should be transparent

Transparency was also highlighted as a key issue in the debate about decision-making and AI. Technologies, the APPG AI concluded, have to be as transparent as possible so **society can justify how decisions are made in order to build trust in the system.**

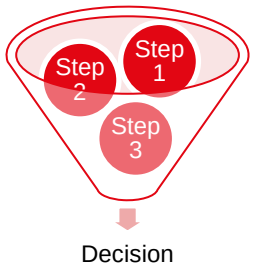
Transparency can help address the ethical dimension because it will make the entire decision-making process more credible and, ultimately, more fair. For example, many stakeholders are concerned that AI might increase inequality gaps amongst demographic groups. As far as gender is concerned, there is worry that females are not entering AI-related fields and, hence, are unlikely to encounter the same benefits as their male counterparts. Professor Sharkey argued that transparent systems can help mitigate these gender gaps, forcing corporates to clearly explain why senior level management belongs to one social group and not another.

There have already been laws passed in the international arena that promote transparent systems. Under the General Data Protection Regulation (EU GDPR), the European Union has legally given individuals the "right to explanation," whereby a user can ask for an explanation of an algorithmic decisions that was made about him or her.[6] The law was approved by the EU Parliament on 14 April 2016 and will be fully enforced by 25 May 2018.

Overall, the group expressed their support for initiatives moving forward in this direction, and called for even more of this type of policies - arguing that the wider public should have the opportunity to know the rationale behind how a decision is being made. There is an international, national, and corporate responsibility to ensure AI tools are transparent and their rationale is easily accessible.

These technologies need to have what Ben Taylor, CEO of RainBird, referred to as an **"audit trail"** so that any stakeholder can easily revert and trace back the steps one by one to comprehend how a decision came about. The process has to be translucent and coherent. As shown in Figure 7, the steps of a process have to be transparent so stakeholders can clearly follow the process from one stage to the next.

Figure 7. Translucent Decision-Making Process



Many companies have expressed wary that having a completely transparent system might

[6] See http://www.eugdpr.org/

threaten their product's competitive advantage in the market landscape. Hence, it is vital to pass carefully crafted legislation that will respect both corporate rights and human rights simultaneously.

There are ways to look inside the AI systems – into their "black boxes" to determine how they reach their conclusions. For example, an image-processing neural network, can be used to highlight the regions of an input image which most influenced its decision.

Transparency can also help confront the issue of privacy raised by many new AI applications in society. As Nicola Eschenburg and Roberto Simone share, AI has huge potential across sectors and industries; however, some concerns like data privacy have to be addressed in order for exploitation of AI to move forward responsibly.

Nicola Eschengurg
Global Head of Analyst Relations at BAE Systems.

Dr. Roberto Desimone
Senior Manager, Strategic Innovation (Disruptive Technolgoies) at BAE Systems.

AI is a disruptive technology that will have a profound impact on society. Since the 1980s (UK Alvey programme), AI has been permeating IT applications 'under the wire', initially by enabling the growth of semantic web applications and supporting limited automated decision-making for military/defence applications. Over the last decade, AI has become more mainstream, supporting autonomous systems (vehicles, drones, robots) for navigation/obstacle avoidance, enabling limited/usable machine translation/natural language understanding, and exploiting machine learning technology (especially Google, Facebook, Amazon, Microsoft) to profile users/clients. As the business opportunities for exploiting AI have grown, so have concerns about how AI technology might be used to intrude our personal space, profile our behaviour from internet transactions, and replace jobs in specific sectors.

The level of funding in AI has now transitioned from R&D to advanced product development and now mainstream product/service delivery. Whereas AI was considered a niche part of Computer Science in the 1980s, it is now mainstream and lauded as a key enabler for many IT applications. The games industry extolls the virtues of having AI in the background to improve the gaming experience. Machine learning techniques and Bayes belief networks are now considered essential for data mining applications, whether to profile users so that marketing adverts can be sent directly or to fuse diverse sorts of data to reveal patterns of behaviour and intent.

Concerns about data privacy are important, and AI has a role to play not only in providing techniques for data mining, but also in auditing where the data comes from and how it is fused and used for different purposes. AI has a powerful role to play in ensuring compliance with data regulations, but also in supporting law enforcement and national security (e.g. finance fraud, cyber threat management). Despite concerns about jobs being replaced by machines, it looks like more jobs will be created through greater exploitation of AI, although periods of transition will always be

hard for those directly affected. Since the exploitation of AI will happen globally, it would be difficult/impossible to stop it. Hence, it is best for us to carefully consider how we can support and encourage exploitation of AI in a responsible manner.

The responsible exploitation of AI could enable disruptive markets that could benefit our daily lives. We already have seen benefits from digital transformation (e.g. eCommerce, social media) that has been enabled by AI techniques and there are more 'on the horizon' in other market sectors including transport, health, energy, law enforcement and media/entertainment. Over the next 10-20 years, these will be further enabled by quantum computing methods. If Alan Turing had been alive today, he would have been considered not only the father of AI, but would also be pursuing 'quantum AI'.

Marina Jirotka, Professor of Human Centred Computing at Oxford University, proposed that the community adapts the **Responsible Research for Innovation (RRI)** framework approach.[7] The framework was created by the European Commission's Horizon 2020 programme and stresses inclusive and democratic decision-making. Furthermore, it promotes greater inclusion of a diverse group of stakeholders. The framework promotes researchers to follow four key principles, outlined through the acronym AREA:

- **A** – to anticipate possible outcomes of research
- **R** – to reflect on motivation and products that come out
- **E** – to engage with diverse stakeholders
- **A** – to act accordingly and be responsive

The RRI framework also urges initiatives to be based on both academic and commercial evidence. It is a preventive measure that will help assure the AI technologies that are being produced are of value for society. Furthermore, the approach will allow relevant stakeholders to consider the multiple aspects and make a process that is highly transparent.

Marina Jirotka
Professor of Human Centred Computing at University of Oxford.

AI raises a large number of ethical and social concerns. Responsible Research and Innovation (RRI) is a methodology that enables researchers to elicit, understand and address such complex concerns. The aim of RRI is to ensure that science and innovation are undertaken in the public

[7] Jirotka, M, Grimpe, B, Stahl, B, Hartswood, M, *Responsible Research and Innovation in the Digital Age.* Communications of the ACM, contributing article, (2017).

interest by incorporating methods that encourage more inclusive and democratic decision making through greater inclusion of stakeholder communities that might be directly affected by the introduction of novel technologies.

RRI proposes a more **reflective** and **inclusive** research and innovation process, from fundamental research through to application design - for both academic and commercial developments. In each phase of the innovation process, certain responsibilities may be associated with activities that occur within them, particularly in relation to how decisions taken might affect society. The focus is on creating a new mode of practical research governance that would transform existing processes, ensuring greater acceptability and even desirability of novel research and innovation outcomes, whilst also identifying and managing potential risks and uncertainties. RRI requires a widening of scope of research and development - from governance of risk - to governance of innovation itself.

RRI has been adopted by the Engineering and Physical Sciences research council in the UK and they have promoted the **AREA** Framework to describe four key RRI components: **Anticipate** possible outcomes of research and innovation, **Reflect** on motivations, processes and products, **Engage** with relevant stakeholders, and **Act** accordingly to address issues revealed. As a precursor to the new EPSRC ORBIT project, we developed an extension of the AREA framework that covers process, product, purpose and people. This framework offers a useful guiding architecture for navigating the opportunities, uncertainties and ethical and societal dimensions of AI and, we believe, can help guide the work of the APPG and inform the recommendations that the APPG will formulate. As a member of the Advisory Board, ORBIT is very happy to share this framework and work with the APPG, to use it in a way that ensures that current research and development on AI has positive outcomes.

From the perspective of the APPG on AI, RRI can serve as the framework for structuring both the debate and the outcomes**.** RRI requires time and effort. Research has shown, however that companies can greatly benefit from it[8]. Benefits from adopting RRI for organisations can include:

- Strengthening links with customers and end users
- Enhancing the company's reputation
- Decreasing business risks and unintended consequences
- Strengthening public trust in the safety of products
- Increasing acceptability of products
- Adopting an environmentally friendly profile

The RRI approach can guide AI developers to create transparent, fair, and socially impactful applications. It can also assist APPG AI in creating well-grounded recommendations and positive impact, by promoting a transparent process which engages a variety of stakeholders.

Lord Tim Clement Jones, co-chair of the APPG AI, highlighted the need for transparency in AI initiatives and discussions. He argued transparency is a powerful tool in order for society to trust the decision-making processes using AI. He summed up the view: **“foremost, we need transparency in order to know what decisions we are preparing to give away.”**

[8] See www.responsible-industry.eu

4. AI's role in decision-making should coexist with accountability frameworks

The group also acknowledged the need for liability frameworks so decision-makers can be accountable to their actions. In a traditional decision-making process, when a human makes a decision, it is relatively simple to assign liability to a specific agent for a specific action. However, when non-human systems are involved (i.e. algorithms) the issue of accountability becomes more complex.

Margaret Boden, one of the best known figures in the field of AI who has written extensively on the subject, commented: **"The machine cannot be accountable. All responsibility must be to some (one or more) human person or corporation. It doesn't make sense to hold the machine accountable."**

As previously illustrated in Section 3, AI is built by humans. Hence, humans are the ones that ultimately decide what commands/tasks the machine will be responsible for performing. **Humans are the ones that choose what decision-making powers AI will possess.**

Therefore, humans are the ones that have to be accountable if something goes wrong. Currently, as emphasized by Ben Taylor – CEO of RainBird – there is a lack of accountability frameworks providing clear guidelines for liability. Conversations and discussions on the topic are very fragmented depending on the industry and also physical context.

Ben Taylor.
CEO of RainBird, a Cognitive Reasoning platform, enabling businesses to rapidly automate decision-making taks and build tools that augment human workers in more complex operations.

A key theme when considering ethical implications in automated decision-making is liability.

The main question we need to address is: who is liable when things go wrong?

People in industry and government are looking at this question very closely. Stakeholders have recognized the need to create frameworks that provide guidelines for liability. These frameworks are important for all industries.

For example, in regards to finance services, we need to assure AI technology has clear rules. When a decision is made about financial trading or granting a mortgage, we need to have rules delegating

who is liable by the decisions made by such a system. Without a legal framework, it will be impossible to fully embed and adapt these new automated decision-making technologies.

Currently, the discussion about decision-making and AI is very fragmented. Some companies like Volvo and Google have publically announced they will take responsibility for accidents from self-driving cars. However, there is no legal framework to support this.

Legal frameworks clearly delegating liability are needed in order to build trust and to ensure responses are not fragment and industry-driven.

The other five speakers agreed with Ben Taylor that stakeholders need to be more proactive in creating these regulatory frameworks. Co-chair of the APPG AI group, Stephen Metcalfe, expressed **the need for a high level advisory group to open discussion on the issue of liability.** We shouldn't make regulation until truly understanding the impact of AI, but at the same time we have to move fast because the capabilities of emerging technology are not slowing down.

Baroness Susan Kramer further supported this view, recognizing the need for a general framework as well as industry-specific frameworks as many policies will have to be tailor-made to be applicable for given sectors.

As advocated by the RRI framework, the responses should not be driven by one individual with personal incentives. There is collective responsibility and, thus, frameworks and policies have to be open to the contribution from all stakeholders.

The stakeholders are not expected to all agree on these issues but, nonetheless, it is pivotal to start dialogue to come up with a solution that will be most beneficial and fair for society. The group of 100 at the 2nd APPG AI Evidence Giving Meeting reflected the variety of opinions stakeholders might have when considering accountability.

Although all concurred that the issue was critical and that liability ultimately lies on human beings, it was not universally agreed whether the individual entity or the corporate entity should be legally liable. Lord Tim Clement Jones recognized the difference between corporate liability and individual liability in the legal realm, and asked the group who they think should be accountable in these frameworks? If something went wrong, would it be the fault of the programmer, the implementer, the corporate body, or another agent?

Some like Professor Sharkey appeared to believe that liability ultimately lies on the hands of the programmer because he/she is the one actually programming the AI in a specific way. He argued, **developers should not get immunity, and they should be encouraged to think in an ethical designing mind-set when developing machines.**

Others disagreed with this view, arguing that liability has to be corporate not individual. Steve Valis questioned who would take the risk of developing AI knowing that he/she would

be liable for potential man slaughter. Lord Tim Clement Jones added that the development of software tends to have a trial and error stage. He wondered if someone should be liable for an outcome from the initial trials. If not, at what point should they become liable in the process?

The majority of the group tended to conclude that the company has to take liability for automatic decision-making consequences. Innovation is not a linear process and it involves many different players at various stages; therefore, it is an impossible task to assign liability to one agent. As emphasized by Ben Taylor, government needs to create regulatory frameworks – with the assistance of stakeholders in business and academia – that are clear and rigid in addressing all these liability questions. In consequence, he argued that **regulation will create more trust in the system and, eventually, help promote more innovation.**

Kumar Jacob stressed that regulation in the health sector can help clarify how data can be used and reused. Society needs to understand when consent is given about sharing the data, why it is given, and how it is used. Although there is some policy on personal identified data, we need to create policies on the use of anonymous data. People need to be educated on how and when they give consent.

Ben Hawes - expert on emerging technologies from the Department of Culture, Media and Sport – offered this solution: **"We need to remember that what Hollywood tells us about AI isn't true. We need to define what AI is for the general public. AI is currently about smart machines doing things for us and acting like humans. Liability has to reside on the human, this is not controversial. Any company that enters the market needs to take liability. Ultimately, there is a need for an operator to be watching the machine in order for it to be accountably. We need to create a model to make sure the liability/accountability issue is addressed correctly."**

5. AI's role in decision-making should rely on trust

We need to assure there is always a human-check on the decision-making process. Through clear guidelines about what machines can and cannot do and also guidelines on unintended consequences, we can consequently **build trust in the AI Revolution.**

Andy Forrester, Director at a technology and innovation consultancy, provided his own summary of the event with what he felt were the five key takeaways.

Andy Forrester
Director at HypeAccelerator Solutions Limited, a technology and innovation consultancy, using disruptive thinking to create innovative ideas and solutions.

[I] was honoured to be invited to the UK All Party Parliamentary Group for AI and on Monday the Group met at the House of Lords to debate the ethical and legal framework for decision making and moral issues in AI.

There were panellists from the academic world, philanthropists, blue chip corporates, and tech start-ups who all gave hotly debated presentations. Much of it was very useful and current; others were highly complex.

Here are our top 5 outcomes from that debate!

1. **Artificial intelligence (and complex algorithms in general, fueled by big data and deep-learning systems) have a huge influence on how we live now**. From the news we see, how we finance our dream home to the jobs we apply for, without global standards or a governance framework to operate in.
2. **It's crucial that AI R&D is shaped by a broad range of voices**—not just by engineers and corporates, but also by social scientists, ethicists, philosophers, faith leaders, economists, lawyers and policymakers.
3. **Algorithms have parents and those parents have values that they build into their algorithmic progeny.** This Group want to influence the outcome by ensuring ethical behaviours and governance that includes the interests of the diverse communities that will be affected.
4. **Advancing accountable, fair and transparent AI**. What controls do we need to minimize AI's potential harm to society and maximize its benefits?
5. **Communicating complexity in plain English**. How do we best communicate the nuances and diversity of a complex industry like AI?

What if we leave it to industry to define what is right for us?

AI must have ethics, be accountable and advance the public interest. Therefore, collective Governments must take the lead on defining laws, definitions and direction. We should not leave it to industry to take the lead otherwise this automatically generates a path towards self-interest and shareholder expectations.

What is the future risk if we do nothing?

Jonathan Zittrain, co-founder of the Berkman Klein Center and Professor of Law and Computer Science at Harvard University advises "…the thread running through these otherwise-disparate

phenomena is a shift of reasoning and judgment away from people. Sometimes that's good, as it can free us up for other pursuits and for deeper undertakings. And sometimes it's profoundly worrisome, as it decouples big decisions from human understanding and accountability. A lot of our work in this area will be to identify and cultivate technologies and practices that promote human autonomy and dignity rather than diminish it."

In 90 minutes we only scratched the surface and there is still a huge amount of future debate and work to be done to even arrive at a high level governance structure that must be truly global.

Forrester's five takeaways and the five themes in this report all ultimately trace back to the concept of trust and its importance for society to reap the benefits of AI moving forward.

Section 1 focused on the complexity of human decision-making and influence of cognitive biases in rational decision-making. Also, AI's role in the decision-making process was put in scrutiny. A cross-disciplinary research approach is suggested to explore these topics. In turn, the research findings should be used to raise awareness and educate the public.

Section 2 emphasized the need to delegate specific decisions to machines, those that society feels confident will increase social value.

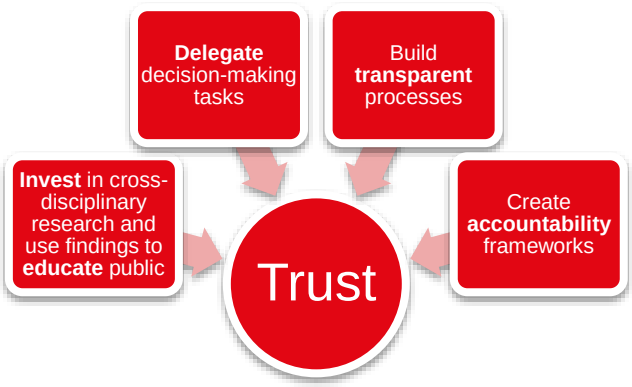
Section 3 of the report explained that one of the ingredients we need to build trust is transparency. Implementing transparent decision-making processes in which the rationale of a decision can be clearly shown is critical.

Also, as Section 4 highlights, the issue of accountability and liability has to be addressed for society to trust the AI. When a human makes a decision and something goes wrong, it is easy to point fingers and blame a specific agent. However, the same thing does not apply when a machine makes a poor decision.

The group recognized that these issues cross national boundaries. We need to be speaking to other governments and guide them to reaching a common agreement on how automated decision-making will be regulated.

In sum, as illustrated in Figure 8, our focus should be on: cross-disciplinary research, delegation, transparency, and accountability. Addressing these issues will help cultivate trust in AI. In consequence, stakeholders will be empowered to decide what they are comfortable allowing AI to do and not do, to understand the rationale behind decision-making processes, and to be protected by legal frameworks.

Figure 8. Building Trust in AI and Decision-Making



The APPG AI assembled the following recommendations - helping set the agenda for how the UK should move forward within the perimeters of the four key themes identified.

Theme	Action Points
AI's role in decision-making should be based on cross-disciplinary research	1. Invest in research to further understand human decision-making, rational and non-rational. 2. Adapt a cross-disciplinary approach to investigate AI impact in decision-making processes. 3. Raise awareness in society using research findings, empowering individuals to better understand how AI can influence decision-making and potential consequences.
AI's role in decision-making should be delegated	1. Engage in public discourse to delegate AI with specific decisions that will maximize social value.
AI's role in decision-making should be transparent	1. Encourage the implementation of transparent systems that are accessible to all stakeholders. 2. Provide individuals with access to a decision's rationality. a. Measure the impact of GDPR for society.
AI's role in decision-making should coexist with accountability frameworks	1. Build accountability frameworks with clear guidelines structuring who/what is liable for given consequence(s).
AI's role in decision-making should rely on trust	1. Cooperate with international bodies and other national governments to assure a coherent approach to these issues. 2. Build trust through above: cross-disciplinary research, delegation of specific tasks, transparent processes, and accountability frameworks.

Acknowledgements

The All Party Parliamentary Group on Artificial Intelligence (APPG AI) was set up in January 2017 with the aim to explore the impact and implications of Artificial Intelligence, including Machine Learning.

Our supporters - Barclays, BP plc, Deloitte, EDF Energy, KPMG, Olswang, Oxford University Computer Science, PwC - enable us to raise the ambition of what we can achieve.

The APPG AI Secretariat is Big Innovation Centre.

Contact details

Big Innovation Centre
Ergon House, Horseferry Road
Westminster, London SW1P 2AL

info@biginnovationcentre.com
www.biginnovationcentre.com