involve people at the heart of decision-making

HOW TO STIMULATE EFFECTIVE PUBLIC ENGAGEMENT ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

MARCH 2019

Contents

1. INTRODUCTION

The technology of artificial intelligence (AI)[1] is already directly impacting the lives of citizens around the world. It is doing this in ways we can see, through identifying people in the photos on our phones or answering our queries in the form of Alexa or Siri. It is also doing this in ways we can't see such as supporting decision-making in a variety of industries and increasingly in public services. Its impact and reach will only grow, potentially affecting increasing aspects of our lives.

There is little transparent debate and few shared answers around the number of important questions these new technologies raise. Public engagement is an effective method to begin to answer and address some of those questions.

There have been several attempts to build public voices into decision-making around AI. For example, the RSA report "Artificial Intelligence: Real Public Engagement" illustrates three key issues that are particularly appropriate for public deliberation and likely to raise a number of ethical questions that could be applied across the technology: Transparency & Explainability; Agency & Accountability and Fairness.

However, this work and the examples we touch on later in this report are one off moments where it's difficult to build effectively on the learning of previous engagement. They can have only limited impact in the wider public policy debate because, however transparent they are, they lack visibility within the wider debate in the field of AI. We are yet to see sustained public engagement around the ethical questions of AI, especially in public services.

Developing this long-term, strategic public debate is important to influence the policy decisions needed to ensure the technology is democratically accountable and ultimately trustworthy from the public's perspective. Establishing sustained engagement poses real challenges given how quickly the technology is developing, its complexity and the levels of uncertainty about its trajectory, potential impacts and the benefits, risks and harms which may result. These challenges are reinforced when so many different actors are involved in designing, commissioning, deploying and critiquing AI systems.

Involve and DeepMind developed a joint project to explore how more sustained public engagement on the ethics of the use of AI in public service delivery might be stimulated. The project brought together a diverse group of stakeholders, each with their own perspectives, understandings and engagements on AI and ethics, to explore this issue through three roundtables.

This report outlines the findings and participant recommendations for working towards sustained public engagement on AI. The production of a resource providing guidance and recommendations on how to support future meaningful public engagement around AI and ethics was identified as a key next step. This report draws together participant recommendations and Involve knowledge where we have identified gaps to scope out some of the components any resource like this might require. The report delivers this through the following sections: **Framing the debate; Governance and decision-making; Communications & language; Practical application of engagement; Moving the agenda forward and Reflections & learning.** Each of the roundtables delved deeper into these different areas. Core questions were posed, and we have captured the participant

[1] When we use the term "AI" throughout the report we are referring to AI, machine-learning and automated decision-making. The Royal Society define AI and machine learning as the following: "If the broad field of artificial intelligence (AI) is the science of making machines smart, then machine learning is a technology that allows computers to perform specific tasks intelligently, by learning from examples."

recommendations, questions and ideas for moving the agenda forward towards sustained public engagement on each area.

Methodology

This work was designed to test whether bringing a diverse group of stakeholders together might lead to new insights and shared action to stimulate robust, sustained and meaningful public engagement in a complex area of scientific and technological innovation. The value of this approach was demonstrated in the breadth of issues surfaced. It did also highlight the investment needed to support stakeholders to identify common ground and shared objectives.

Interest in the roundtables was high and participants were made up of a group of experienced stakeholders ranging from academics in social studies of science, AI business leaders, AI scientists, ethicists, public engagement practitioners and those representing AI in the arts and culture. Invited participants were chosen based on their interest in public engagement with AI and ethics. We worked with contacts in the arts, AI, social science and public engagement fields to map as widely as possible participants that would be interested in this work. We acknowledge (as did participants at the roundtables) there were still gaps in the room in terms of different stakeholder perspectives impacted by this debate.

Participants reviewed the existing debates, learnt about some of the most recent attempts to engage the public in policy debates about AI and started to identify the actions needed to create sustained public engagement in this area.

Please note it is important at this stage to acknowledge that multiple different publics exist when we are talking about public engagement. An essential part of good public engagement is using methods that are suitable for the group of people you are trying to engage with and not using a one-size fits all approach. When we use the term “publics” throughout this report, it is on that basis.

The original purpose and objectives for the work were as follows:

Purpose

- To investigate what meaningful public engagement looks like around AI and ethics and to explore how this can be built into decision-making by researchers, technologists & policymakers.

Objectives

- To create space for a collaborative conversation between a diverse group of stakeholders on the ethical questions raised by AI and various public engagement initiatives designed and launched in response to those questions.
- To provide a forum for stakeholders to build new links and conversations.
- To build a common understanding of how to engage with public perspectives on the ethics of AI.
- To identify further areas of research, collaboration and advocacy around public engagement with the ethics of AI.

Summary of Findings

We found that bringing such a range of stakeholders together threw open the debate and framings of the topic more widely than initially anticipated. Identifying the purpose and focus areas for moving the agenda forward proved challenging due to the multiple angles and perspectives raised. Participants noted that the act of making these diverse perspectives visible in the first place was, nonetheless, a useful output from the work. And it is worth noting, that public perspectives are likely to be even more diverse than those of the stakeholders

involved in these roundtables. Any engagement will both need to deal with the variety of framings the public bring to the questions raised by AI and find ways to make them visible to allow meaningful engagement with their implications.

"These are challenging, but important discussions. Getting around the table with other people in this space is hugely valuable."
Feedback from one participant at the end of the series of roundtables.

The overriding conclusion from the series was that a wide range of stakeholders working around AI have a clear interest and see value in public engagement on AI and ethics. The difficulty faced by stakeholders in agreeing the purpose and possible methods of engagement was a recurring lesson throughout the roundtables too. Without this greater shared understanding of public engagement methodologies, best practices and relevant approaches required, the risk is that different actors attempt to engage the public on different terms, for different purposes and asking different questions. Whilst encouraging more public engagement is no bad thing, the risk arises that the wider the number of framings and questions asked, the harder it is for hard pressed policy makers to make sense of what they are hearing. This could diminish the impact that public perspectives have on policy decisions.

The roundtables made it clear that those making decisions about the design, deployment and use of AI need to work collectively to develop more of a shared understanding of how the public should be involved in their decisions. Participants identified some tensions here that establishing consensus and shared understanding should not be the priority. It was also argued though that even if a shared understanding is not achieved, there is still real value in bringing people together to acknowledge their differences. This underlined further the difficulties faced by stakeholders to define what public engagement is required around AI and ethics. As outlined above, participants identified the first required step to tackle this as the development and sharing of a resource to explain what robust and sustained public engagement could look like which this report has begun to do.

2. Framing the Debate

We shared the words of Kanta Dihal and Stephen Cave from the Centre for the Future of Intelligence with participants. Dihal and Cave highlight the tensions in perspectives and opinions that create the frames and narratives around AI:

"In our era of advanced robotics and artificial intelligence…polarized responses persist, with pundits and the public applauding or warning against each advance."

Kanta Dihal & Stephen Cave

In early conversations, we focused on some of the existing framings used by stakeholders when discussing AI and how these are used by decision-makers to understand public reactions to AI. We drew out the ethical questions raised by these framings that participants identified. These were categorised into the following themes summarised below:

- **Desires:** what are the needs driving our expectations of technologies and how are they being influenced?
- **Perspective:** how do we balance the multitude of competing opinions around this area and the difficulty for individuals and groups to be objective?
- **Fundamental Questions:** have we truly considered the implications of the technology on life as we know it? Do we need to challenge ourselves to explore possibilities outside of our comfort zone?
- **Society:** the questions in this category covered numerous areas but there was commonality around who in society holds responsibility for managing the technology and what impact it will have on society as a whole?

- **Diversity:** who is being represented in the rooms where decisions about the technology are made?
- **Implementation:** when should and when shouldn't we use the technology?
- **Accountability:** who holds overall accountability for the technology and its ethical implications?
- **Transparency & Trust:** to what extent is transparency required to ensure trust in the technology and what should this look like?
- **Language:** what impact can language have on the direction of development of the technology and its implementation?

Each theme captured ethical, social and moral dilemmas that AI poses and they provide a basis for some of the core questions that public engagement could tackle (the full list of questions under each theme is in Appendix A). The framings spanned from very detailed points concerning whether an individual should be able to get an explanation to a decision made about them, through to the nature of the capitalist system in which AI is developing.

These framings and the potential dilemmas AI poses recurred throughout the discussions.

3. Exploring Ways to Engage the Public in Governance

Three core questions were identified by participants for exploring this area further which were:

1. **How & when can mechanisms to involve the public be included within governance?**
2. **How should the government's use of AI systems be mapped and prioritised to inform where engagement can have the most impact?**
3. **What precedents are there for building (or not) public engagement into government decision-making?**

We have then summarised the participant recommendations prompted by these questions into the following themes: focused engagement, accurately framing engagement, mapping the use of AI by government, avoiding duplication of regulation and clarity around ownership of an engagement process.

Focused engagement

Discussing AI broadly leads to a very diffuse and unfocused conversation. Any public engagement needs to be around specifics for example credit scores, facial recognition and specific policy options and it shouldn't just use the broad term "AI". Involve's own experience suggests that the public are more likely to want to be involved in solving specific, relevant problems where technology such as AI might provide part of the solution, rather than in policy debates framed around a specific technology.

Getting the framing right is critical

Any debate about the use of AI quickly reveals its complexity. Participants noted that any public engagement will have to be very clear about the framing of the questions and boundaries of any discussions. Participants identified several areas to consider when thinking about the framing of engagement. Here are some examples but they are not the exclusive framings to be considered when planning a process. Participants raised whether engagement should focus around:

- Global, national or local levels of governance?
- Developing standards vs laws & regulations?
- AI now vs AI potential in future?
- Exploring hopes and fears, and ideal futures?

Map the use of AI by government

Participants recommended that actors in this field, with government, should map where AI is being used in government and by decision-makers so stakeholders can identify where, when, how, who and why to engage around AI. An example raised was whether engagement around AI should be on internal AI systems (like HR) and/or external AI systems (like healthcare, social care).

This prompted a further question as to whether such a mapping might help identify where public engagement might be most useful. For example, the mapping might show that engagement about the specific use of AI is needed or it might show that something else needs to be considered for instance budgeting, resourcing or decision-making in the system more generally instead.

Avoid duplication of regulation

Regulators should consider what the characteristics are of AI that make policy-making in this area any different and identify areas that can be covered by existing and approved regulation. For example, a recommendation was made for decision-makers to consider where the new General Data Protection Regulation can provide governance and regulation for the development and deployment of AI. Another example is the existing legal requirements that could cover some aspects of AI provided by the concept of "Duty of Care" as written about by Will Perrin. These ideas could suggest engagement may not be required in certain areas when looking at future regulation or could enable the engagement to start at a more informed level rather than from a blank sheet.

Ownership

Participants recommended the need to consider how we ensure that the engagement is effective. This discussion raised points suggesting the efficacy of the engagement is influenced by the ownership and driving force behind the process. The following questions were emphasised as key for decision-makers and practitioners of engagement to answer before starting an engagement process:

- Who is responsible for the public engagement?
- Who will act on the engagement and take responsibility for feeding back to the public on the impact of the engagement?

When considering how to build sustained public engagement on AI, recognising and starting to address the questions above will help to create more meaningful processes that will be of more value to decision-makers and therefore have more impact on policy.

4. Communications & Language in Engagement

Participants identified one core question to help unpick how communications and language can impact on the creation of sustained public engagement on AI. The question was:

1. **Is it possible to have a common and shared understanding of AI? How do the understandings of different actors, including the public and government, differ in this area?**

Clarity of language

Several participants requested a Glossary of terms referring to AI to support clear and consistent definitions for public engagement in this area. Participant discussions frequently touched on differing interpretations and use of language around this complex topic which sometimes seemed to be a source of confusion or disagreement. This is something participants acknowledged becomes even more of an issue when trying to communicate this area to the public. The work of Understanding Patient Data (UPD) to support discussions with the public, patients and healthcare professionals about uses of health and care data provides a useful example for how this might be done.

Develop a shared public understanding of AI

Participants made a clear recommendation that those working with AI need to demonstrate possible futures and do more to deconstruct the hype surrounding AI. Public engagement is one method of helping those working with AI to build a shared, factual understanding of AI. It was suggested that such activities could illustrate trade-offs and explain where AI is and isn't in people's lives to increase awareness and understanding.

Government responsibility

The government should be challenged internally and externally to lead on developing and framing this shared public understanding. The roundtables identified that there is a risk of a paternalistic language and framing developing around AI if the government intervenes extensively. However, the view that the government needed to demonstrate clarity on their use and understanding of AI was generally supported. This paper from the AI Now Institute, focused mainly around the US political system, digs into these arguments further looking at a possible algorithmic impact assessment framework to ensure public accountability for government bodies around these new technologies.

Identify which publics

Often people refer to the public as one entity. Even a brief glance at the public debate on any issue demonstrates that this is not true. There are multiple publics and communities with multiple perspectives and views.[2] Any effective engagement needs to start by identifying which publics and communities bring relevant perspectives, views and knowledge before engagement begins. These recommendations mirror developments in other areas of technological development. For example, UPD have been doing lots of work in this area to share good practice and learning from engagement examples and to emphasise the need to recognise the existence of multiple publics. Genome Editing Public Engagement Synergy have focused on similar techniques too by sharing resources to support effective public engagement on another complex issue area. (See the section in this report on Moving the Agenda Forward for more suggestions.)

Importance & influence of communications

The power of language and communications to influence thinking and public acceptability is a critical part of this work. Participants identified additional themes under this area that need to be acknowledged when designing engagement on AI:

- **Communications Purpose:** What is the purpose of public engagement and how can language impact on this? Is it to inform people, or to help policymakers make decisions? A conclusion from this work is that the purpose has to be both to inform the public and policy-makers.
- **Media:** How can media discussions of AI mature to become more informative? How can more of a balance be reached between the AI good news stories and the portrayal of negative consequences of AI?

5. Practical Application of Engagement

This section draws on the discussions about four existing examples of public engagement on AI to answer three core questions.

1. **What could public engagement look like around AI and ethics?**
2. **What existing resources are there to support public engagement?**
3. **What does public engagement look like at different parts of the AI life-cycle?**

As well as planning how to move the agenda forward, one aim of this work was to raise awareness of existing public engagement on AI and ethics. Four public engagement case

[2] See for example: https://webarchive.nationalarchives.gov.uk/20170110123903/http://www.sciencewise-erc.org.uk/cms/our-thinking-2

studies were shared at the second roundtable. These four examples are highlighted below (for more detail, please see Appendix B).

- Royal Society's public views on machine learning,
- Forum for Ethical AI led by the RSA on public engagement with automated decision-making,
- Nesta's public dialogue on AI & Ethics and;
- British Science Association's survey on AI.

Discussions exploring these examples in more detail led to further ideas and recommendations from participants around the design, delivery and scope of possible future public engagement on AI. This included an emphasis on the importance of using creative methods of engagement as well as more traditional processes.

Participants identified the following key recommendations for those looking to deliver public engagement around AI:

Transparent development

Explore engagement through open development of AI applications. Public engagement processes could be implemented during the development phase of AI production instead of post-development. Early engagement would test public acceptability early on and help build public perspectives into the design and development of products.

Participants shared suggested engagement methods for different points in the AI life-cycle:

- Conception - Futures focused engagement
- Design - Scenario-based engagement
- Deployment - Deliberative engagement

Case studies

Share engagement examples and raise awareness of existing public engagement around AI further. The sharing of the case studies at the roundtables really emphasised the importance and support for reviewing and learning from existing engagement projects as a basis for future engagement.

Creative methods

Investigate new and alternative engagement processes and framings that can enable wider reach such as through technology services. One participant suggested an engagement process that could be delivered at scale could be designed in a similar way to how internet cookies request permission from users. An engagement method based on the way permissions for cookies works could capture responses and perspectives of the public on relevant aspects of AI. Further suggestions from participants were examples like Massachusetts Institute of Technology (MIT)'s moral machine; using online platforms; gamified/scenario based methods such as My 2050; or bringing people from the arts into discussions in this area more. (See the section below: Moving the Agenda Forward for more ideas).

Purpose & scope

Many further questions were raised around the practical application of engagement on AI depending on different contexts, sparked particularly from reviewing the case studies that were shared. There was a particular focus in the discussions on the types of questions that should be asked when considering the purpose and scope of engagement:

- **Focus:** Public Engagement can address focused, application level issues (e.g. a health app or policing tool). The following questions should be considered as part of this:
    - Should this technology be deployed at all?
    - Should this technology be used/designed at all?
    - To what extent?

- **Suitability of engagement:** When is engagement not needed/not appropriate?
- **Impact:** What kind of impact or outcome are we imagining public engagement will have?
- **Scale:** How to scale the engagement process? Who should be involved?

6. Moving the Agenda Forward: Identified Gaps

Based on the above findings from the roundtables series we have provided several examples and resources that can help move the agenda forward towards sustained public engagement on AI. This section is informed by the gaps in discussions that emerged. The information is drawn from Involve's experience and the pool of knowledge from roundtable participants who were invited to contribute further suggestions to the report after the initial draft was produced.

The gaps identified were: the need to share high quality engagement standards; explanations for the cost of participatory processes and how to demonstrate the value and purpose of engagement.

Engagement Standards – Traditional & Creative Examples

Throughout the discussions, participants requested further examples of engagement processes and what meaningful engagement looks like and how to define it. We shared Involve's Nine Principles of Effective Deliberative Public Engagement drawn from our experience as engagement practitioners.

Nine Principles for Effective Engagement

1) The process makes a difference
2) The process is transparent
3) The process has integrity
4) The process is tailored to circumstances
5) The process involves the right number and types of people
6) The process treats participants with respect
7) The process gives priority to participants' discussions
8) The process is reviewed and evaluated to improve practice
9) Participants are kept informed

A further example shared by participants is Wellcome's The Art of Health: Exploring Creative Engagement with Health (2016) which outlines creative and alternative engagement ideas that could be considered when designing processes to engage the public around AI.

As mentioned earlier in the report, Genome Editing Public Engagement Synergy (GEPES) – Overview of Resources (2018) is an additional source of information, resources and ideas for public engagement processes when tackling complex topics, in this case, genome editing. The same goes for the earlier highlighted work by UPD which focuses on sharing best practice in the area of public engagement and health.

The University of Cambridge, Centre for the Future of Intelligence and Arm have also developed an award-winning series of publicly accessible short documentaries over the past few years. These are designed as conversation starters and have been shown at a variety of Science Festivals (often followed by discussions with a panel of experts). An attached survey for each has also allowed for more public responses to be captured.

An additional case study shared by a participant was the Global Inquiry into Citizens in the Digital Age. The Fourth Group worked with 10 partners (including SOAS University, Democracy Earth, One Young World and Quartz) to engage with hundreds of people from over 40 countries across all the major continents through 'Open Conversations', focus groups, and in-depth online interviews looking at the impact of tech on people across the world.

What is the cost of engagement?

During the discussions the following question was raised: how do you prioritise public engagement given the opportunity cost (time, resources, etc.)? From our experience, securing the initial investment and internal buy-in for engagement processes is a common challenge and one that will exist in the world of AI too. Involve provides a True Costs of Public Participation (2005) report and a webpage focused on myth busting the expense of engagement processes.

Appendix C in this report captures the top-level arguments from these two pieces. It provides key information for anyone looking to analyse the costs of engagement to help secure internal support for a sustained approach to bringing in the public voice around AI.

What's the potential value & impact of engagement?

Linked to the costs of engagement were several questions raised around why stakeholders should engage the public in the first place. Given the overriding support from participants for public engagement, it felt necessary to share the considered benefits of participation for helping attendees to communicate its value more widely:

- Improved governance
- Improved quality of services, projects and programmes
- Greater capacity building and learning amongst stakeholders
- Appropriate decisions
- Legitimacy / support for decisions
- Accountability to the public, meeting public demand and expectations for involvement.

These are taken from Involve's web pages on the quality and legitimacy of decision-making; costs and benefits of participation and the true costs of public engagement.

(Please see Appendix D for a more detailed outline of these points if required.)

In summary, in response to gaps identified in discussions during this work, we have outlined good practice engagement recommendations; provided detail to explain the costs of engagement processes and highlighted the value of sustained public engagement approaches.

7. Reflections & Learning

To close the series of roundtables, we considered these final questions to reflect on where this work needs to go next:

1) **Should roundtables/engagement processes with diverse groups of stakeholders like this continue?**
2) **What is the core question to ask that group?**
3) **Who are we missing from the room?**
4) **What are the next steps for this work?**

These questions were used as prompts which sparked a wide-ranging discussion on how to explore this agenda further and what the key participant takeaways were from the discussions.

Knowledge of engagement

To establish sustained engagement, stakeholders will need a greater awareness of the purpose of engagement and the methodological standards/practices around engagement including the multiple methods that exist to suit different audiences. This will ensure that the key principles of engagement are shared by all.

It was also recognised in discussions that those advocating participation must talk about engagement with stakeholders and policy-makers as well as delivering direct public

engagement. A real need for participation practitioners to provide wider training for organisations to deliver engagement themselves e.g. “train the trainer” was also highlighted.

Framing

Participants agreed that the use of clear framing of an issue or question is necessary for designing engagement to help focus the scope and purpose of the process, a recurring finding from this work. A key conclusion was also drawn that there are no easy answers to the multiple questions around AI due to the complexity of the topic and that this must be remembered when designing engagement processes. For instance, consensus must not always be sought and it’s the richness of the varied discussions and thinking that can provide new solutions to these complex challenges.

A further recommendation also flagged the need to avoid dealing with more systemic challenges for instance public trust in decisions, through the topic of AI. Examples such as this need to be approached through the bigger question of why some feel democracy isn’t working.

Diversity of participants

A repeated sentiment raised throughout discussions, but particularly when looking at the practical application of engagement, was the consideration of the range of people around the table. Participants recommended a UK wide conversation to prevent conversations only happening in an AI echo chamber in the South East.

Reflecting on the roundtables: further questions to consider

- **Bigger Picture:** How useful is it to have AI as the primary focus for public engagement? Does it shut down alternative social solutions by focusing engagement on AI? Should a more problem focused approach be taken, with consideration given to AI when it offers a potentially viable solution to a key problem faced by society?
- **Control:** Why is engagement necessary? Does it relate back to the human need for control?

Learning

This project was designed to tackle the issue of a lack of sustained public engagement around the use of AI, in particular in public services. The plan was to stimulate robust, sustained and meaningful public engagement in a complex area of scientific and technological innovation amongst a diverse stakeholder group. We wanted to test a new approach to public engagement design, focusing on bringing stakeholders together first before engaging directly with the public.

Involve captured its learning from developing and delivering this process and the learning from participants and their experience throughout the project. Our first reflection would be to ask if the framing of the roundtables was too broad. It proved challenging to focus discussions, especially around the emphasis on the use of AI in public service delivery and we therefore would recommend a more focused scope for stakeholder engagement. This links with feedback from participants who identified the value in the frames that have come out of this work but acknowledge they still need more specific discussion.

“The diversity of voices in the room was excellent for bringing a range of perspectives. However, it meant the conversations struggled to find focus.”

Feedback from one participant after the first roundtable.

Another reflection is that we could have supported the process of moving the agenda forward further with an additional roundtable where we could have discussed tactics and best practices around possible engagement methodologies to take conversations to date to the next level. To combat this, we have invited participants to contribute to the report directly to help continue the conversation and push for further traction on the report’s recommendations. One participant working on an “Earning Trust in Technology Governance” project has already

offered to explore the relevant findings from this work in that project with others interested in this area as one possible next step from the roundtables.

A final reflection to note is around the value of building new connections and creating a forum for discussion for those who have been leading public engagement work in their respective organizations. This is a recurring and significant benefit identified in feedback from participants throughout the series.

8. Conclusion

These roundtables were designed to bring diverse stakeholders together to discuss public engagement around the use of AI in public services. The report highlights the support for engagement from these stakeholders combined with the challenges they face or questions they hold around effective and meaningful engagement.

The lack of consensus on the scope and purpose of engagement demonstrates a real need for those making key decisions about the design, deployment and use of AI to work collectively together to develop more of a shared understanding of how the public should be involved in their decisions.

We now look to address the challenge of picking up the actions and recommendations identified in this report along with, we hope, the multiple participants involved in this work.

9. APPENDIX

Appendix A - Findings from Roundtable 1

The first roundtable was designed to explore the existing frames and narratives used by stakeholders to understand public perspectives around AI. From these discussions we have pulled out the ethical, moral and social dilemmas and questions raised by participants and categorised them thematically. These are outlined below:

Desires
- Ethics of expectation: what influence are pre-emptive and predictive technologies having on our desires?
- What needs are driving tech. development?

Fundamental Questions
- Redefining work without changes to wage capitalism?
- Is there a metaethics of capitalism under this?
- How can we ethically do things ethically? (Metaethics)

Society
- If tech. companies are seen as being part of the problem should they be responsible to solve the problem? Should there also be space for society to debate technology?
- What should the balance be between individual vs collective agency/rights?
- What are the ethical implications for current generations to consider for future generations? E.g. “AI Now vs AI Later”
- Are we ignoring values? E.g. those we can’t measure. Whose values and ethics are being talked about? E.g. international ethics.
- Is it fair or right to buy advantage using AI?

Perspective
- Is anybody pure? Are we ever truly free from influence?
- How can individual perspective and standpoint be taken into consideration in the debate if there are so many different opinions?

Diversity
- How do intersectional inequalities (e.g. Race, Gender, etc.) influence the debates?
- Diversity of views – who is in the room/discussion and allowed access to the debate? Who would like access to the debate?
- Is the whole community represented?

Implementation
- When do we not use it?
- What are the ethical implications of using the technology and what are the ethical implications of not using technology?

Accountability
- Who is responsible for the ethics?
- Who deals with the consequences?
- Who owns the technology?
- Who benefits from it? And who doesn’t? And what is the interaction with inequalities?
- Who has agency?
- Should we be inviting the public to engage if we aren’t listening to them?
- As individuals, can accountability enable us to trust and believe in the decision-making process? Is transparency too much work for individuals?
- Who owns the ethical conversation? Who should?
- Should power be more easily accessible?

Transparency & Trust
- Open Society – what types of power emerge and what impact will they have? Do we need checks & balances for potential power shifts? Who sets these checks & balances?
- What generates trust in corporate decisions?
- Should we focus on trust & accountability not just transparency?
- What is transparency? Who decides what level of transparency is needed?
- Should an individual have the ability to get an explanation about a decision made about themselves?
- What are the trade-offs in the debate e.g. accuracy vs explanation?

Language
- What is the social, moral, political context/frame?
- Good for who? What is “good”?
- Whose interest is at the end of the narrative?
- What is the impact of the language used e.g. is “ownership” the correct framing?
- Fairness: what is fair?
- Where should the balance be between specificity of accounts of AI vs “Catch All” explanations?

Appendix B - Public Engagement Examples

The four examples of public engagement reviewed are:

1. Royal Society's public views on machine learning,
2. Forum for Ethical AI led by the RSA on public engagement with automated decision-making and;
3. Nesta's public dialogue on AI & Ethics.
4. British Science Association's survey on AI.

Here is a short summary of each example. Presentations from the organisations below were provided at Roundtable 2.

1. Public views on machine learning – Royal Society

The Royal Society launched a project on machine learning in November 2015, aiming to increase awareness of this technology, and highlight the opportunities and challenges it presents.

The project focused on the current and near-term (5-10 years) applications of machine learning, investigating the potential of this technology and the barriers to realising this potential, and was supported by a broad programme of public, policy, research, and industry engagement.

The UK public was a key audience for this project, and public engagement an integral part of the programme of work. This engagement included a public events programme that reached over 15,000 people, and a public dialogue with Ipsos MORI that explored public awareness of, and attitudes towards, machine learning.

2. Forum for Ethical AI - RSA

The RSA is facilitating a series of citizens' juries to deliberate on the ethical use of AI for automated decision-making.

The question being explored is: Under what conditions, if any, is it appropriate to use automated decision systems?

Jurors met and heard from expert witnesses in May and June 2018 to produce a set of conditions in response to the question. A final event will be held on October 13 in order to carry out a deep dive on one of these conditions – explainability – in partnership with the Alan Turing Institute and the Information Commissioners' Office. This event will also give jurors a final opportunity to have a discussion with, or make recommendations to, a wider group of stakeholders.

3. Public dialogue on AI & Ethics – NESTA

This one-day public dialogue explored how members of the public might approach some of the biggest questions and the values underpinning their decisions on AI and ethics.

A hypothetical but realistic health case study was used to demonstrate how algorithms can inform decision-making in the public sector. Using a game-like approach, participants (23 public participants and one table of policy-makers working in this field) worked together to develop paper algorithms, and in the process explored what health data their algorithm should use and what decisions it should make.

The aim of this dialogue was to contribute to wider public debate on controversial new technologies; help policymakers to understand public priorities in these areas and directly inform research strategies/ethical frameworks.

4. *Survey on artificial intelligence - British Science Association*

The British Science Association (BSA) conducted a survey for British Science Week 2016 to see how the public thought robotics and artificial intelligence will affect society and culture. The online survey, which had over 2,000 responses, was conducted by YouGov on behalf of the British Science Association.

The survey found that 60 per cent of respondents thought the use of robots or programmes equipped with artificial intelligence will lead to fewer jobs within ten years. It also found 36 per cent of the public believe that the development of AI poses a threat to the long-term survival of humanity.

Appendix C - Costs of public participation

The analysis of the costs and risks of participation is far less detailed, but includes the following taken from Involve's True Costs of Public Participation report:

- Monetary costs, including staff time (paid and unpaid), staff expenses, external staff / consultants, fees to participants, participants' expenses, training for staff and participants, administration, venue hire, other event costs (e.g. refreshments, equipment), newsletters, leaflets, monitoring and evaluation fees.
- Non-monetary costs, including time contributed by participants, and skills needed for the new approach (taking time from other work).
- Risks, including risks to reputation (from bad participatory practice), stress, uncertainty and conflict.
- Direct costs
    - Staff costs- extra and different staff may be needed for support. Training staff. Staff time that cannot be spent on their usual work. "heavy time commitments". Staff costs are likely to increase if you bring in external expertise to run or advise the project.
    - Event costs
    - Publicity
    - Exhibitions, reports, leaflets, websites etc.
- Other costs
    - When people are consulted, they may oppose the initiative which may generate costs in managing the next steps of the engagement process.
- Risks rather than costs - if the process is managed well these risks could translate into benefits rather than costs, it depends on the quality of the process
    - The cumulative effects on multiple forms of participation can be a cost, in the form of 'over-consulting' and 'engagement fatigue' (Newburn and Jones 2002, 52).
    - A problem is the amount of time required of staff to find the additional resources required to make participation work properly (Aycrigg 1998, 18).
    - In times of resource constraint voluntary effort can be seen as an infinite resource and overexploited.
    - May be unpredictable and therefore difficult and potentially costly to manage
    - The drive to create common ground may lead to sub-optimal outcomes
    - Other risks include reputations, failure to deliver on promised outcomes, uncertainty, relationships.

Appendix D – Benefits of Public Engagement

The benefits of participation are often considered to be:

- **Improved governance**, including increased democratic legitimacy for institutions because of close links with citizens, improved reputations for public bodies, increased opportunities for active citizenship, and greater accountability of public bodies because of more effective information dissemination and better dialogue.
- **Greater social cohesion etc**, including bringing diverse and sometimes hostile communities together, bringing 'hard to reach' and 'disadvantaged' groups into discussions, building relationships within and between different communities and social groups ('bonding' and 'bridging' social capital), strengthening and creating new networks that enable different interests to work together as a result of building more positive relationships based on a better knowledge of each other, and increased equality of access to policy and decision-making processes.
- **Improved quality of services, projects and programmes**, including ensuring public service investment is based more on people's expressed needs, reducing management and maintenance costs by reducing vandalism and misuse as a result of engendering a sense of ownership, enabling faster and easier decisions (e.g. on new developments or protective designations) by reducing conflict between different parties and increasing trust through better communications, and enabling people to share in the responsibility for improving their own quality of life (e.g. health and well-being, or the local environment).
- **Greater capacity building and learning**, including raising awareness and increasing understanding of public institutions and the way they work, enabling citizens to better access the services they need, and to understand the boundaries and limitations of different public bodies, building confidence and optimism among citizens who then go on to other civic activities or learning, supporting the voluntary and community sectors by recognising their vital role in building the capacity of community and specific interest groups (especially disadvantaged and excluded groups), and increasing the skills among the staff running participation and those taking part (especially interpersonal skills).
- These points are taken from pages on Involve's website covering costs and benefits of participation and the true costs of public engagement.

Participation delivers the following additional benefits:

- **Appropriate decisions-** more accurate and representative information about the needs, priorities and capabilities of local people – better feedback on existing programmes and users.
- **Legitimacy / support for decisions –** Participation can allow support to develop with stakeholders before the decision is formally taken which reduces the need for the 'decide and defend' approach. Having a say can also aid implementation as there is a feeling of ownership over the results which can lead to less conflict.
- **Accountability to the public –** Participation can build on the formal systems of accountability by enabling citizens to hold elected representatives and others more directly accountable through face-to-face discussions.
- **Inclusion and cohesion –** When carefully designed and implemented participation can create institutions that can enable marginalised and often excluded groups to be brought into the decision-making process. Those who are usually excluded from

decision-making processes may have relevant new information of knowledge to contribute to improve the quality of a decision.

- **Meeting public demand and expectations for involvement –** even the most traditional institutions have long recognised the need to meet public demand for involvement.
- These are taken from Involve's webpage on the quality and legitimacy of decision-making.

Acknowledgements

We would like to thank DeepMind and their Ethics & Society Programme for the support to deliver this project. Particular thanks go to Brittany Smith from their Ethics & Society Programme for her contribution to the project throughout and to this report.

We would also like to thank all of the participants at the roundtables, and particularly the speakers and contributors for the time they took preparing and delivering helpful provocations at the roundtables and in this report:

Alvin Carpio, The Fourth Group

Annette Mees, Royal Opera House

Benedict Dellot, RSA

Beth Singler, University of Cambridge

Diane Beddoes, Deliberate Thinking

Hilary Sutcliffe, Society Inside

Jess Montgomery, The Royal Society

Jon Fitzmaurice, British Science Association

Oliver Buckley, DCMS

Paula Boddington, Cardiff University

Reema Patel, Ada Lovelace Institute

Tom Saunders, UKRI

Authors

Lizzie Adams

Simon Burall

Date of Publication

March 2019