# stimulating effective public debate on the ethics of artificial intelligence

##

## Drawing Out the Ethical Questions & Reviewing Existing Engagement:

## Briefing paper to prompt discussion at DeepMind & Involve roundtable 2

### Introduction

In response to the number of different public engagement projects that have launched recently on topics like automated decision-making and machine learning, we designed these roundtables to support those actively engaging on these topics. The purpose and objectives for the roundtables are outlined below:

#### Purpose

* To investigate what meaningful public engagement looks like around artificial intelligence (AI) and ethics and 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.

From Involve’s experience, we know when bringing such a range of stakeholders together, it is often difficult to achieve consensus on the role that public engagement should play to inform decision-making. The first roundtable in this process explored the various ways AI and ethics are talked about in order to expose the numerous narratives that different actors use.

This highlighted the diverse range of perspectives and emphasised the value of supporting a wide range of stakeholders to engage with the public effectively. For information, our definition of effective and meaningful public engagement throughout this paper and this process is taken from Involve’s [Nine Principles of Effective Deliberative Public Engagement](https://www.involve.org.uk/resources/knowledge-base/what-are-qualities-good-participatory-process/nine-principles-effective).

The aim of the second roundtable is to build on our understanding of these narratives and use them to review and analyse previous public engagement processes in this area. This will allow us to focus the discussion towards the next steps required for further meaningful public engagement on the use of AI in public service delivery.

This paper summarises the findings from roundtable 1, concentrating on the ethical dimensions raised. It then outlines a summary of three recent public engagement processes on this subject and highlights some existing examples of AI in public service delivery.

The paper is designed as a prompt for discussions at roundtable 2 and is not a verbatim account of roundtable 1, nor a full analysis of existing public engagement processes on AI and ethics.

### 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:

**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?

**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)

**Desires**

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

**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?

**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?

**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?

**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 explain?

**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?

**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?

The themes above highlight the breadth of the discussions at the roundtable, which also covered key actors in the system and the actions they should be taking. We have concentrated on the ethical questions raised to provide a basis for furthering the discussion on where the public voice should be included. Below are some public engagement processes that have tried to address this challenge in different ways – some focusing on AI in public services and others on the use of AI in wider sectors.

### Existing Public Engagement Processes

Several public engagement processes have already taken place on the subject of AI. We want to build on and raise awareness with senior stakeholders about this existing work and demonstrate a range of processes to discuss alongside the exposed narratives from roundtable 1. The aim is to incorporate outcomes from this into future decision-making for meaningful public engagement on AI and ethics.

The examples of public engagement we will review are:

1. Royal Society’s [public views on machine learning,](https://royalsociety.org/~/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf)
2. [Forum for Ethical AI](https://www.thersa.org/discover/publications-and-articles/reports/artificial-intelligence-real-public-engagement) led by the RSA on public engagement with automated decision-making and;
3. Nesta’s [public dialogue on AI & Ethics](https://www.involve.org.uk/our-work/our-projects/practice/artificial-intelligence-what-do-public-really-think-about-its).
4. British Science Association’s [survey on AI](https://www.britishscienceassociation.org/News/rise-of-artificial-intelligence-is-a-threat-to-humanity).

Here is a short summary of each example. To inform the next stage of our discussions, our next roundtable will have presentations from the organisations below and we will explore the design, delivery and outcomes of each project.

#### 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.

#### 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.

#### 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.

#### 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.

### Public Service Delivery & AI

The ethical questions arising in the research and deployment of AI are often highly contextual, and the public interest varies accordingly. The next roundtable will concentrate on how public engagement can inform policy making *on the use of AI in public service delivery*.

Therefore, the final section of this briefing paper provides some examples of diverse uses of AI or possible uses of AI in public services to act as potential case studies or discussion points at the second roundtable:

* **Health:**
	+ [Automated Image Diagnosis](http://www.bsa-org.com/bsa-ai/) and its potential to minimise human error and improve decision-making.
	+ The implementation of AI for [health research](https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/transforming-healthcare.html) to speed up research processes for new treatments and medicines.
* **Justice:**
	+ Durham Police’s [use of algorithms](https://www.cam.ac.uk/research/features/helping-police-make-custody-decisions-using-artificial-intelligence) in predicting likelihood of suspects re-offending.
	+ Swansea Police’s use of [Automated Facial Recognition](https://www.south-wales.police.uk/en/advice/facial-recognition-technology/) technology to identify potential suspects from CCTV images.
* **Education:**
	+ The development of [tailored education pathways](https://www.nesta.org.uk/blog/what-does-artificial-intelligence-really-mean-our-education-system/) using AI for more suitable learning opportunities.
	+ The potential for AI to increase [access to education](https://www.forbes.com/sites/bernardmarr/2018/07/25/how-is-ai-used-in-education-real-world-examples-of-today-and-a-peek-into-the-future/#30a5bb23586e) in particular through translation tools helping remove language barriers to education.

We encourage you to bring your own examples to feed into the discussions as well.

### Key Questions

Finally, we invite you to consider the following questions:

* How could public engagement processes inform your current work?
* How can policymakers gain insight into what people are already saying (or not) about AI and where did these views come from?
* Public engagement is often associated with a large price tag. Investment of time and money is required to make this work – but is it cheaper than failed initiatives, lack of buy in and a limited range of perspectives informing developments?
* How can professional stakeholders commit to a new architecture for developing a debate that no one can control through the use of public engagement?

*Author: Lizzie Adams*