# stimulating effective public debate on the ethics of artificial intelligence

##

## Exploring the multiple debates about AI and ethics:

## Briefing paper to prompt discussion at DeepMind & Involve roundtable 1

*“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](https://www.nature.com/articles/d41586-018-05773-y.epdf?author_access_token=zCLBurJLeXcAl6FFp9BSzdRgN0jAjWel9jnR3ZoTv0P3GZvi7nQK9vVPgInrkgE50-CnynbjPLK2DxxUpViq_Zh3a9r2CwHFKeFFlFO9Y2bIgfUBwV5r6HHx6CTc-p2mKDWxFRV0SRl6diBhsWTrUg%3D%3D)

### Introduction

This is the first of three briefing papers to stimulate thinking prior to a series of stakeholder roundtables around public engagement and artificial intelligence (AI) hosted by Involve and DeepMind.

This briefing paper outlines the structure of the roundtables and some of the existing framings used by stakeholders when discussing AI and its ethical implications. It is not a systematic literature review of this topic but is designed as a prompt and starting point for discussions at the first roundtable.

The roundtables will focus on public engagement around automated-decision making in public services. The purpose and objectives for the roundtables are outlined below:

#### Purpose

* To investigate what meaningful public engagement looks like around 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.

As we will explore in the first roundtable, concerns about the impact of technology go right back to antiquity. However, recent advances in AI have highlighted potentially significant gaps between the public and decision-makers working in science and technology.

Within policy circles there have been growing concerns that public trust in governance and institutions is low and reducing. Indeed, as this briefing paper was being finalised Jim Al-Khalili, the incoming president of the British Science Association, [warned about a potential backlash](https://www.nature.com/articles/d41586-018-05773-y.epdf?author_access_token=zCLBurJLeXcAl6FFp9BSzdRgN0jAjWel9jnR3ZoTv0P3GZvi7nQK9vVPgInrkgE50-CnynbjPLK2DxxUpViq_Zh3a9r2CwHFKeFFlFO9Y2bIgfUBwV5r6HHx6CTc-p2mKDWxFRV0SRl6diBhsWTrUg%3D%3D) by the public against AI akin to the rejection of genetically modified crops. These roundtables are focused on resolving the challenge of how to engage the public effectively in the debate and in policy decisions around the use of AI.

There have already been several recent engagement processes focusing on AI and the ethical questions it poses. Some examples of these are: the 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) work; the ongoing [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 NESTA’s public dialogue on [AI and ethics](https://www.involve.org.uk/our-work/our-projects/practice/artificial-intelligence-what-do-public-really-think-about-its). We have also had the establishment of a dedicated government body, the Centre for Data Ethics and Innovation and the Nuffield Foundation’s Ada Lovelace Institute.

The discussions at these roundtables will build on this existing work and support interested stakeholders to identify a set of shared next steps to address the question of public engagement with AI and ethics.

### Structure & Attendees of the Roundtables

The idea for these roundtables was borne out of the above question of public engagement and the growing calls for the public to be involved in the numerous debates around AI and ethics. [Involve’s experience](https://www.involve.org.uk/our-work/our-projects/practice/how-can-public-inform-science-and-technology-policy) in leading dialogues and engaging the public on complex, technical and high-profile issues has shown us that a hugely diverse range of stakeholders will have very different opinions on how to engage the public.

These roundtables therefore intend to explore how a better understanding of engaging with public perspectives on the ethics of AI can be developed amongst these stakeholders.

Interest in the roundtables has been high and participants will be 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.

Each roundtable will have a different focus, building on the conclusions from the previous event. In summary the series of roundtables has the following structure:

1. Introductory roundtable mapping the different perspectives in the public debates on the research and deployment of AI and the impact it’s having on society.
2. Building on the insights generated at the first roundtable, we will explore how we might begin to interpret the outcomes from a range of public engagement processes.
3. Pulling together conclusions, we will support participating stakeholders to develop a shared understanding of the best ways to create a public debate which is more effective at shaping the use of AI in public service delivery.

All three roundtables will be structured to maximise the opportunities for participants to contribute and network amongst the group. There will be very short provocations invited from participants themselves. These will be followed by in-depth table and plenary discussions to test, challenge and begin to address the questions that each roundtable will pose.

### Existing Framings & Questions They Raise

Various attempts have been made to interpret the framings used by decision-makers to understand public reactions to AI. The diverse range of hopes and fears which form part of the debate about the impact of AI is evident in the briefest of scans of government policy documents, the latest technology news, science fiction and reviews on the cultural impact of AI on society.

For instance, Stephen Cave and Kanta Dihal’s recent piece “[The Automation Chronicles](https://www.nature.com/articles/d41586-018-05773-y.epdf?author_access_token=zCLBurJLeXcAl6FFp9BSzdRgN0jAjWel9jnR3ZoTv0P3GZvi7nQK9vVPgInrkgE50-CnynbjPLK2DxxUpViq_Zh3a9r2CwHFKeFFlFO9Y2bIgfUBwV5r6HHx6CTc-p2mKDWxFRV0SRl6diBhsWTrUg%3D%3D)” goes some way to unpicking the social and cultural influences on public opinion around automation. Cave and Dihal explain that we, the public, have “complex emotional responses to AI. *Understanding these and their deep history is crucial to making the most of life with intelligent machines.”*

Further examples, selected to highlight a few of the dominant narratives on this topic, explore some of these complex and wide-ranging perspectives. The UK Government’s Industrial Strategy “[Building a Britain fit for the future](https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future)” emphasises the huge potential benefits for the UK economy that AI can bring. The strategy outlines increased government investment in Research & Development (R & D), in part around AI, and explains the following potential impacts of this:

*“We could see a dramatic change in the use of R&D by industry, with our businesses creating the next generation of technologies to revolutionise productivity in all sectors from construction and agriculture to manufacturing and the creative industries. This will raise the standard of living and establish UK leadership in global markets.”*

In contrast to this a [BBC News article](https://www.bbc.com/news/business-45240758) from August 2018 highlights a Bank of England warning of: *“‘large swathes’ of people becoming ‘technologically unemployed’ as artificial intelligence makes many jobs obsolete.”*

There are, of course, a range of narratives beyond the policy and business framings. AI is presented in multiple ways in popular culture for example through Isaac Asimov’s *I, Robot* stories*;* Ann Leckie’s *Ancillary Justice s*eries; IainBanks’ *Culture Novels* and Alex Garland’s *Ex Machina*. Arguably it is through the framings created by this genre that the public has had most exposure to the world of AI and the ethical questions it brings.

In these brief examples, numerous narratives (that would be labelled as positive by some and negative by others) are raised such as the implications for the economy, employment, morality, power, ethics and basic human needs.

The purpose of these roundtables is not to create consensus around one narrative that should be used to frame AI and ethics. Instead we can start mapping different perspectives on the research and deployment of AI, and the impact it’s having on society. This will help us identify and better understand narratives that influence public opinion and stakeholder reaction.

Building on this we will challenge ourselves to consider further key questions:

* **What are the ethical tensions, dilemmas and debates that emerge from the different framings presented?**
* **What are the different aspirations and concerns of the public around AI that are raised by these narratives?**
* **How and where are we challenging and discussing the disconnect between public service delivery use of AI and public opinion/understanding of AI?**

The quote at the beginning of this paper highlights the polarised response to the development of AI and robotics. This series of roundtables aims to explore how we can start building a common understanding amongst diverse stakeholders on how to engage with public perspectives on the ethical questions of AI.

*Lizzie Adams & Simon Burall, September 2018.*