



**UCL**



# **AI for People & Planet**

*Policy Commentary: Equity*

## AI for People and Planet – Equity

The **Artificial Intelligence (AI) for Equity** policy roundtable discussion was organised as part of the UCL roundtable series '[AI for People and Planet](#)'. The aim of this roundtable was to convene a group of experts in AI and ethics to clarify thinking about the relation between equity and AI. The purpose of this document is:

- to understand AI's potential impacts on equity;
- to survey the range of threats and identify means of mitigation against each;
- and to develop a conceptualisation for how the considerations/issues/ideas/concepts that AI introduces can be integrated within existing ethical frameworks.

### Executive Summary

The key themes highlighted during the discussion included:

- To judge whether or not an AI model is 'good', we need clarity on the demonstrable difference to society its use could make.
- AI's ability to single out individuals is not just a threat to equity in making it possible to discriminate against them, but also a means of detecting and rectifying inequity.
- The principles of accountability and openness are particularly relevant to discussions of the interface between AI and equity. Moral accountability for the results of an algorithm should be ascribed to the person, or group of people (Government ministers, CEOs, etc.), who authorised the AI system.
- When a policy is based on the result of an AI model, the public, in theory, should be able to assess the model and the principles on which the decisions were based as far as possible.
- Being able to understand what variables and assumptions a model is based on and being able to compare similar models with each other is essential in helping to avoid biases. Conversely, it may not be necessary to know how a model works, but whether there is discrimination or bias.
- Full agency should not be given to the machine; accountability rests on the shoulders of humans.
- Having effective and advanced governance in the form of clear legislation, risk management structures, and monitoring and evaluation is crucial in accounting for ethical considerations and ensuring models are adaptable to different contexts.

These themes are discussed in greater detail below.

### Equity in the Context of AI

#### *What is AI?*

The term AI is used to refer to computational models that exceed in flexibility and expressivity conventional statistical models. Though often treated as categorially different, such models lie on a continuity of model complexity, and represent an evolution of computational modelling, even if the results of their application may be revolutionary.

Built into widely used software (e.g. Apple's Siri, Amazon's Alexa, Skype's Real-Time Translation tool), but seldom understood, AI is often debated as a tool for good versus tool for bad. To judge the moral status of AI, however, requires a deeper understanding of its nature and application than typically informs such debate.

*What is equity?*

It is important to distinguish between [equality and equity](#). Equality refers to the uniformity of a treatment or an outcome, whilst equity refers to the proportional sensitivity to circumstances of possibility and need. Equal treatment, iniquitously applied, may yield unequal outcomes; obtaining equal outcomes through unequal treatment may be iniquitous. The moral consequences of each aspect need to be examined case by case.

The World Health Organization (WHO) has published a report outlining [equity considerations](#) for the use of AI, which include:

- the digital divide (i.e. the gap between people at different socio-economic levels regarding their opportunities to access information and communication technologies);
- algorithmic bias and values;
- plurality of values across systems;
- and fair decision-making procedures.

The report argues that AI technologies should be designed and implemented such that they actively redress or eliminate inequities and promote greater equity, while also not sustaining or exacerbating existing inequities.

## **Principles of Public Life – Accountability and Openness**

The Committee on Standards in Public Life offers guidance on [The Seven Principles of Public Life](#), which apply to anyone who works as a public office-holder, for example, Government ministers, local government officials and civil servants. The principles include: selflessness, integrity, objectivity, accountability, openness and honesty. Accountability and openness are particularly relevant to discussions of AI and equity.

### *Accountability*

While emphasis has been placed on ensuring that humans can override a result, the issue of responsibility is rarely discussed. Automated decision-making is the process whereby algorithms make decisions without any involvement, for [example](#), in determining the decision of a loan application or an aptitude test. Responsibility cannot simply be ascribed to an algorithm, so who is accountable for the outcome of an automated decision?

It may be argued that there are at least [two forms of responsibility](#) – causal and moral, differentiated in *Table 1*.

Table 1 – Forms of responsibility

| <i>Form of responsibility</i> | <i>Description</i>  | <i>Example</i>  |
|-------------------------------|---|---|
| Causal                        | Purely causal responsibility arises when an agent causes an outcome, but without coherent purpose or intent. AI systems can have causal agency when, for example, their predictions guide policy decisions on welfare payments. | A child accidentally tripping and breaking her parents' vase is causally responsible for the damage.                        |
| Moral                         | Moral responsibility extends to causation whose agent can be meaningfully said to be responsive to reason and held to account by it. An AI system cannot be held morally responsible as it is not a rational agent.             | If the tripping child in fact intended to break the vase and did so on purpose, she would then also be morally responsible. |

If we agree that moral responsibility must rest with humans, on what grounds do we assign responsibility? Rather than placing the responsibility on the software developers, moral accountability should be ascribed to the person, or group of people (Government ministers, CEOs, etc.), who authorised use of the AI system and approved its characteristics.

#### *The A-level and GCSE results 'fiasco'*

As the [A-level and GCSE results 'fiasco'](#) unfolded, the Government's initial response was to ascribe responsibility to the algorithm itself. While Sally Collier, Head of Ofqual, England's exam regulator, did [resign from her post](#), many at the time called for Gavin Williamson, Secretary of State for Education, to also resign, as the one with ultimate accountability. The decision for the algorithm to base students' results on those of peers from similar schools and backgrounds runs counter to the very principles of equity.

#### *Openness*

The ability of AI to single out individuals (individuating model power) represents one of its greatest strengths, and conversely, one of its greatest risks. This ability to differentiate between individuals also makes it possible to discriminate between them. When a policy is based on the results of an AI model, the public, in theory, should be able to assess the model and the principles on which the decisions were based, as far as the necessary complexity of the model allows. Equally, when a decision has led to instances of inequity, openness and transparency surrounding how the decision was made is crucial.

There are technical solutions that can be used to reduce the risk of discrimination, such as limiting individuating model power, the complexity of the algorithm, and autonomous applications (software able to make decisions without the authorisation of human operators).

The assumptions behind these solutions are that non-individuating models (those that cannot identify individuals from datasets) are inherently fair, simpler models are better, and the scope of complex models should be limited. However, simple models should not always be preferred merely because they are easier to understand, as reality does not always map onto a simple model. Any bias against complexity is misguided, as what matters is the assessment of whether the model works. There is a trade-off between having a system that is adequately complex to perform sophisticated functions and being able to explain the intricacies of the system to society if something goes wrong.

## **The Role of Governance**

*How do we uphold administrative justice when the end result has negative consequences?*

Decision-makers need to proactively consider the consequences of the use of models and should be able to evidence that a particular model being considered is the most appropriate. This requires people in leadership positions to understand the variables that go into an AI system and use their professional expertise and experience, in conjunction with the results from the model, to come to a decision. Conversely, there is also the argument that it is not necessary to know how a model works; what is needed is knowing whether there is discrimination or bias.

A solution to ethical considerations is having effective and advanced governance – clear legislation, impact assessments, proper controls and risk management structures, setting responsibility for decision-making, monitoring and evaluation – that is adaptable to different contexts. It is the role of the person or group of people who commission the development of an algorithm to ensure that the appropriate guidance and checks are followed. Since ethical considerations are already entrenched in governance, the challenge now is how to enforce them.

## **Equity in AI and Public Policy**

*How do we convey the ‘truth’ in a way that is digestible within a policy framework?*

The lack of data to test models on subgroups of people from different racial, ethnic and socioeconomic backgrounds has made it difficult to ensure that equity and diversity are being adequately accounted for. Being able to understand what variables and assumptions a model is based on and being able to compare similar models with each other is essential.

Achieving a model that is accurate in its individuation without being biased is a challenge, as fairness and equity cannot be analysed as dichotomous variables (i.e. ‘equitable’ or ‘not equitable’). When people are exposed to risk, i.e. having their exam results predicted by an algorithm, it should be considered whether they would reasonably consent (or not) to this risk. When the model’s outcomes negatively affect people, an equity lens must be used to examine whether it was wrong across all population groups (thus fair across all groups), or whether the model worked disproportionately against those from protected characteristics and other underrepresented groups.

*What do you do if something goes wrong?*

Model transparency would help by revealing what variables were included in the model and how they were weighted. However, this is not always possible in a ‘black box’ AI system in which inputs and operations are not visible. AI models need to be assessed in a way that is workable for policy and, as far as possible, understandable by the public. Accountability should also be agreed on beforehand so that the appropriate structures are in place in the event that the model does make an error or the public disagrees with the results.

## Conclusion

The 'Equity' roundtable discussed AI's potential impacts and threats to equity and a conceptualisation for how the considerations that AI introduces can be integrated within existing ethical frameworks. The ability of to single out individuals (individuating model power) represents one of its greatest strengths, and conversely, one of the greatest threats to equity. To mitigate these risks, effective and advanced governance in the form of clear legislation, risk management structures, and monitoring and evaluation is critical. Principles of accountability and openness must be considered alongside equity frameworks, with the moral responsibility of results of models being attributed to those who authorised the AI system as fit for purpose. Being able to understand what variables and assumptions a model is based on and being able to compare similar models with each other is essential in helping to avoid biases. Conversely, it may not be necessary to know how a model works, but whether there is discrimination or bias. Ultimately, accountability rests on the shoulders of humans.

### Participants

#### UCL Participants

Professor Parashkev Nachev (Chair, UCL Queen Square Institute of Neurology)  
Dr Zeynep Engin (UCL Computer Science)

#### External Participants

Ms Amy Austin (Cabinet Office)  
Mr Parker Moss (Genomics England)  
Ms Zoë Porter (University of York)  
Dr Roland Tegeder (SW7 Group)  
Mr Paul Trenell (Government Equalities Office)

*This document was prepared under the Chatham House rule by [Ms Audrey Tan](#) and [Professor Parashkev Nachev](#). Please get in touch if you would like to contribute to the discussions outlined below.*

*The AI for People & Planet: Equity Roundtable is supported by:*

**UCL  
PUBLIC  
POLICY**

  
**UCL**  
INNOVATION  
& ENTERPRISE

 **GRAND  
CHALLENGES**