

Ethics and AI

What is AI: Issues with a definition

“We find that while AI researchers favor definitions of AI that emphasize technical functionality, policy-makers instead use definitions that compare systems to human thinking and behavior. We point out that definitions adhering closely to the functionality of AI systems are more inclusive of technologies in use today, whereas definitions that emphasize human-like capabilities are most applicable to hypothetical future technologies. As a result of this gap, ethical and regulatory efforts may overemphasize concern about future technologies at the expense of pressing issues with existing deployed technologies.”

A reasonable definition of AI

The Organisation for Economic Co-operation and Development (OECD) notes the following definitions pertaining to AI:

–*AI system*: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

AI programmes are made

Artificial intelligence is designed and made by humans for human purposes.

The design of AI is intentional and this process involves choices.

What is the impact of this?

Any choice which can have an impact in the world is subject to Law and Regulation.

Why Ethics in AI? Why Now?

AI systems are so embedded within technology that they are assumed and almost invisible. This also means that it's hard or impossible in many cases to say which ethical and other value issues are presented by AI per se, and which by other features of technology.

AI is already with us; many applications are fast developing and will be with us in the near to medium term. It is a disputed question when other even more advanced forms of AI including superintelligence will be here, if ever. Many consider that the emergence of superintelligence is an inevitability....

AI may involve robotics of varying complexity; it may involve the manipulation of vast amounts of data; it may involve machine learning. It may involve how we relate to others; it may involve how as individuals we think, remember and reason. It may have implications for the organisation of the labour market; it may involve how we trade; who has access to information; and how. It has implications for the economy, for politics, for culture. It has implications for those who directly use AI, such as those who use a robot butler; it has implications for those more remote, as for example when an algorithm developed by machine learning helps makes public policy decisions, or unemployment attributable to the use of AI means a worker can no longer afford to pay their child's music teacher.

AI is already in use

Machine translation
Robotics
Logistics planning
Spam fighting
Chat bots
Game playing
Scheduling and planning
Speech recognition
Facial recognition
Robotic vehicles

AI is increasingly being incorporated in autonomous roles

Two dilemmas

AI currently in use.

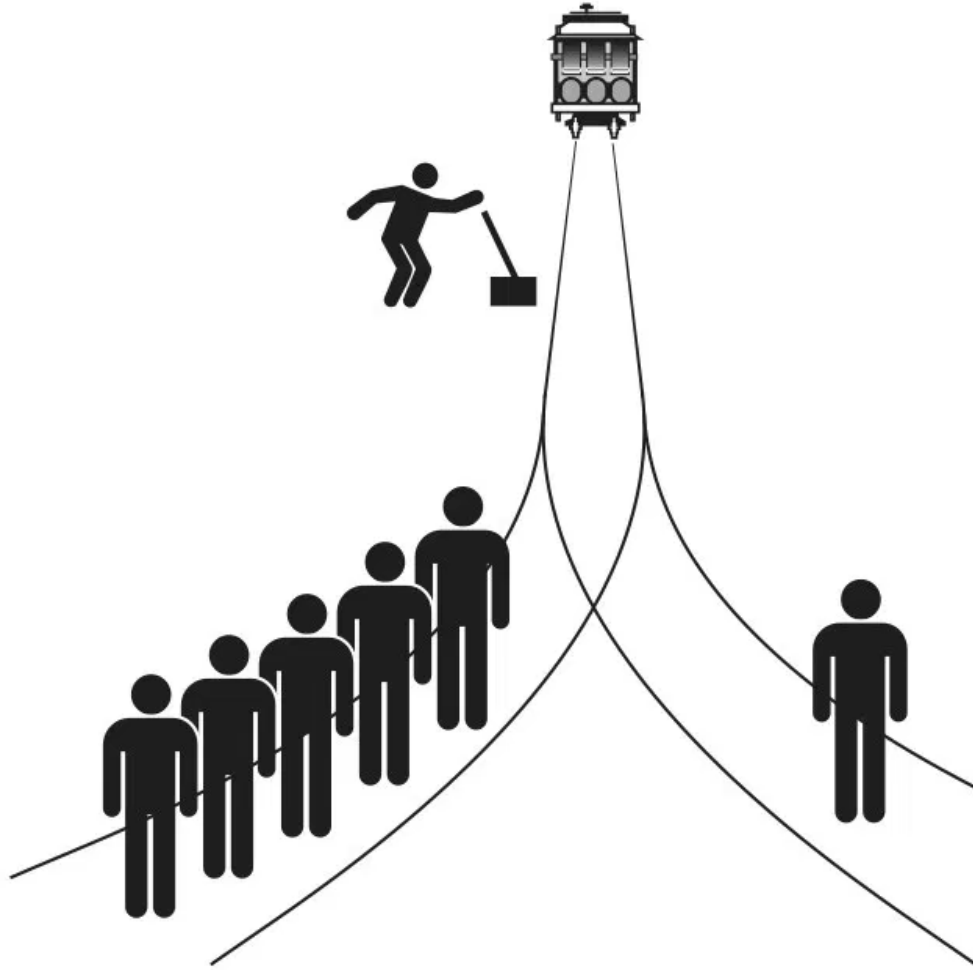
Should we be worried?

What should we be worried
about?

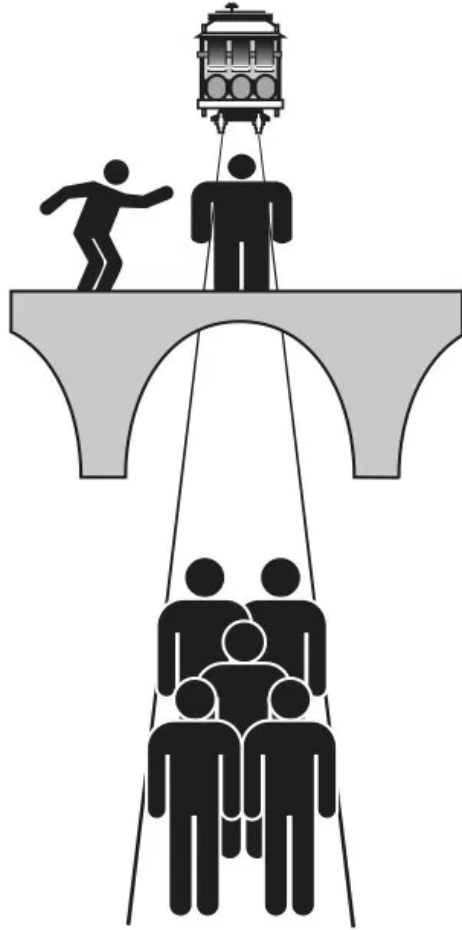
AI that may be made in the future

Should we be worried?

What should we be worried
about?



Trolley Problem



Driverless Cars

Are there any different scenarios to the Trolley Problem?

Designing Ethical AI

How would you design a code of ethics for the AI in:

- a driverless car?
- a robotic surgeon?
- a weaponised drone?
- a social media platform?

Useful Resources

For general discussion on Ethics: Stanford Encyclopedia of Philosophy.

For discussion on Ethics and AI: Talis Aspire Course Reading List has links to online books.

For example: P. Boddington, "Introduction: Artificial Intelligence and Ethics," in *Towards a Code of Ethics for Artificial Intelligence. Artificial Intelligence: Foundations, Theory, and Algorithms*, B. O'Sullivan and M. Wooldridge, Eds., Oxford, United Kingdom: Springer, Cham, 2017, pp. 1-5 [Online]. Available: https://doi.org/10.1007/978-3-319-60648-4_1

<https://intelligence.org/>

<https://openai.com/>

<https://ai100.stanford.edu/> (see 2020 report)

<http://lcfi.ac.uk/>

<https://www.partnershiponai.org/>

<https://www.ai-austin.org/>

<https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>

<https://www.stopkillerrobots.org/>

Videos you should watch:

<https://thoughtmaybe.com/cant-get-you-out-of-my-head/>

<https://www.youtube.com/watch?v=evBzPwCdeHI>

<https://www.youtube.com/watch?v=lzd2qOgOGQI>

<https://www.microsoft.com/en-us/research/video/future-ethics/>

https://www.youtube.com/watch?v=5dZ_lvDgevk

All of Black Mirror!