



The Law and Ethics of AI

How will the law respond to the rise and challenges of artificial intelligence?

with guest speaker
Professor Dan Hunter

Starting shortly





The Law and Ethics of AI

How will the law respond to the rise and challenges of artificial intelligence?

Introduced by
Professor Melinda Edwards

Starting shortly



Acknowledgement of Traditional Owners

QUT acknowledges the Turrbal and Yugara, as the First Nations owners of the lands where QUT now stands. We pay respect to their Elders, lores, customs and creation spirits. We recognise that these lands have always been places of teaching, research and learning.

QUT acknowledges the important role Aboriginal and Torres Strait Islander people play within the QUT community.



The machine learning age Challenges for law & ethics

QUTeX Real World Futures
February 2020

Dan Hunter





Overview

01

Introduction

In which we discuss the way that
Dustin Hoffman (of all people)
helps us understand AI

And also, in which we are introduced
to the Three Horsemen of the
AI Apocalypse.

02

What is AI, really?

How we can understand 64 years of
research and development in
approximately 6 slides
(not counting the digression)

03

Law & Ethics

International consensus, explained

04

6 problems

Not even remotely intended to be a
comprehensive description of the
kinds of issues we face.

(I don't even talk about deepfakes
and GANS, for example.)



Introduction

Machine learning



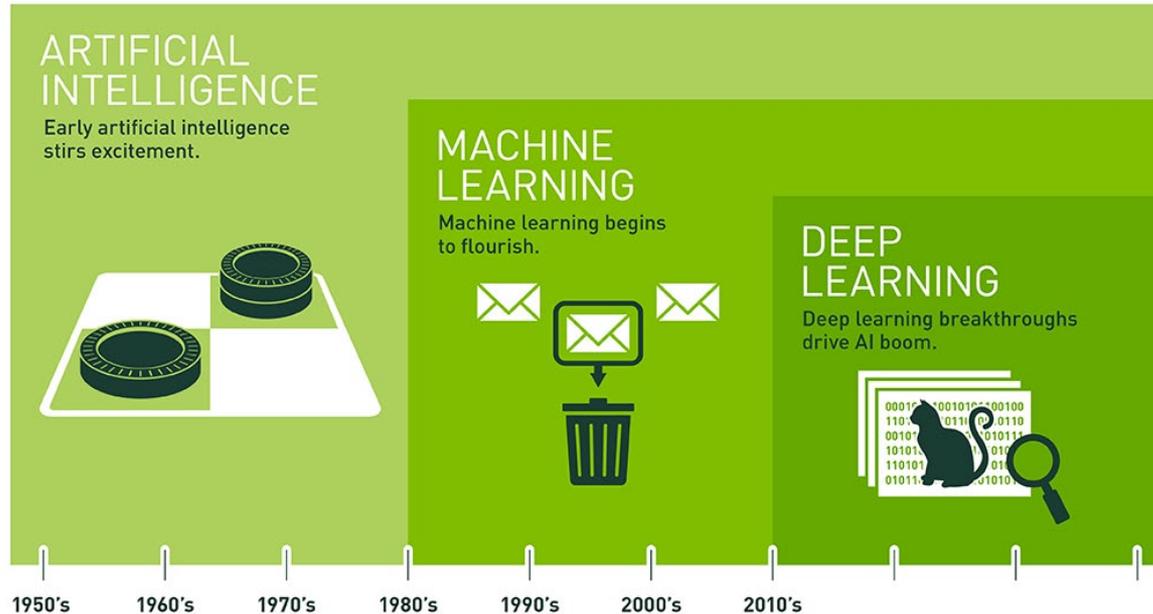
The Three Horsemen





What is AI, really?

Development of AI

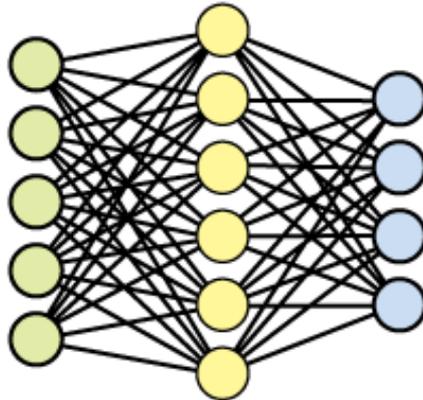


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Machine learning

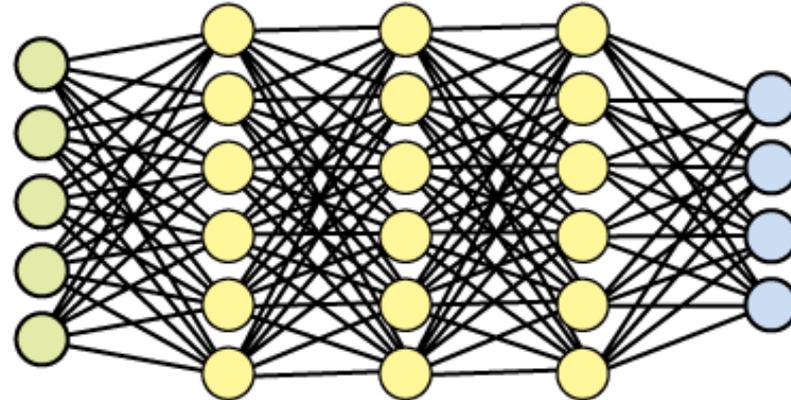
Neural network

Input Hidden Output



Deep neural network

Input Hidden Hidden Hidden Output



What is AI these days?

- When it's in crowded areas (**city centres**), it's a difficult situation because of the proximity of pedestrians ❶ and other vehicles without a real safety barrier between ❷ the two. We have to keep in mind that there will be people and vehicles approaching ❸ this car from all sides at different speeds
- Evolved **sensors** compared to the previous Google cars, which were based on the Prius
- Maximum speed is limited to 40km/h

LIDAR
This is the heart of the system, which has the task of creating a three-dimensional map of what's around the car. It has a 360° view and a range of up to 200 metres.

64 LASERS rotating at 600 rev/min
The Lidar is provided by a Silicon Valley company called Velodyne, and costs more than 50,000 Euros (₹40 lakhs). It has 64 laser emitters, which rotate at 600 revolutions per minute.

CAMERA AND RADAR
In addition to the sensor on the roof, the automatic driving mechanism uses radar in the front and back, plus front cameras.

BUTTONS
A touchscreen and very few buttons for setting a journey, starting, or stopping. No steering wheel to turn or pedals to depress.

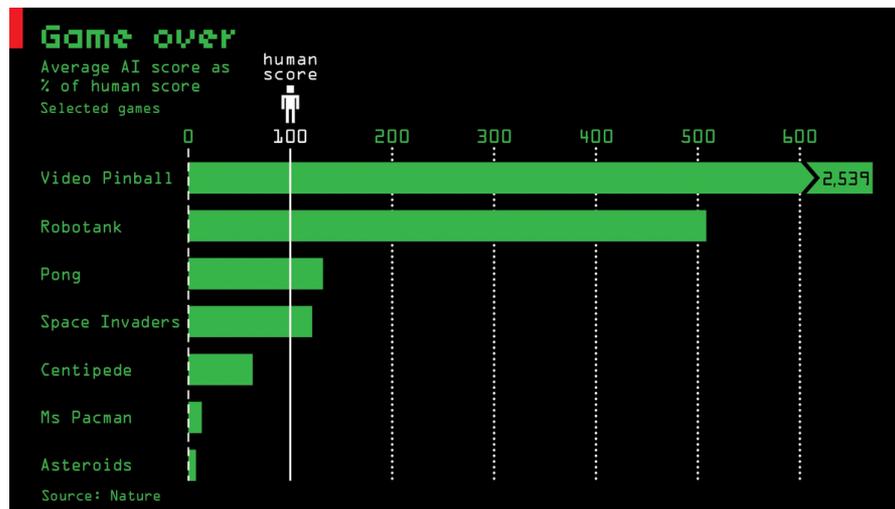
This is the first prototype. In the future, the production car will look different. Moreover, the car will not be produced directly by Google.

Self driving cars



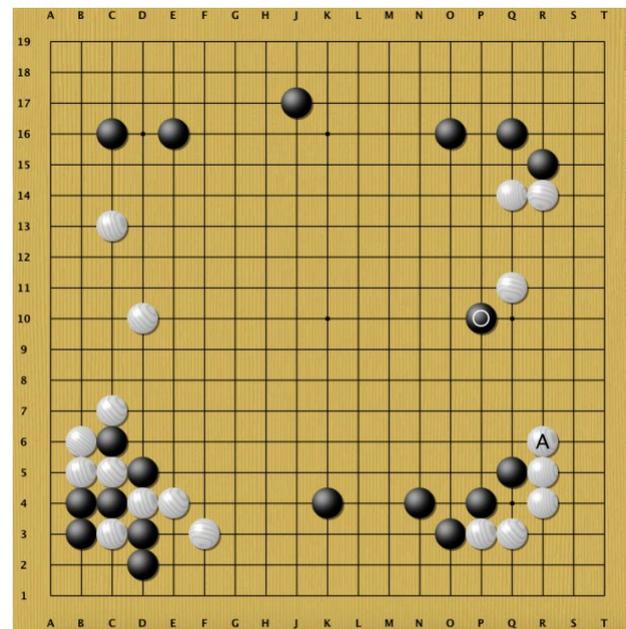
Siri

Fields: Games



Economist.com

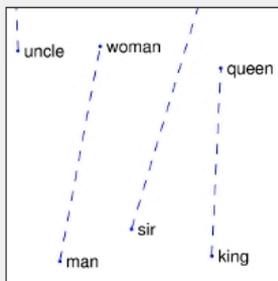
Moves 37



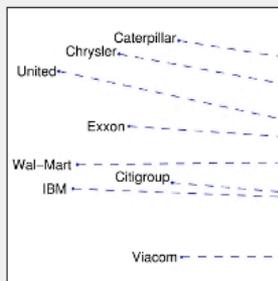
This figure has only one move: Black 37! This move proved so stunning that, when it appeared on the screen, many players thought the stone had been put down in the wrong place.

Fields: NLP/voice recognition

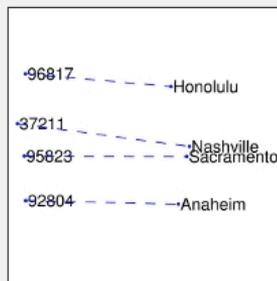
In order to capture in a quantitative way the nuance necessary to distinguish *man* from *woman*, it is necessary for a model to associate more than a single number to the word pair. A natural and simple candidate for an enlarged set of discriminative numbers is the vector difference between the two word vectors. GloVe is designed in order that such vector differences capture as much as possible the meaning specified by the juxtaposition of two words.



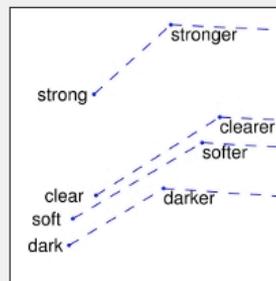
man - woman



company - ceo



city - zip code

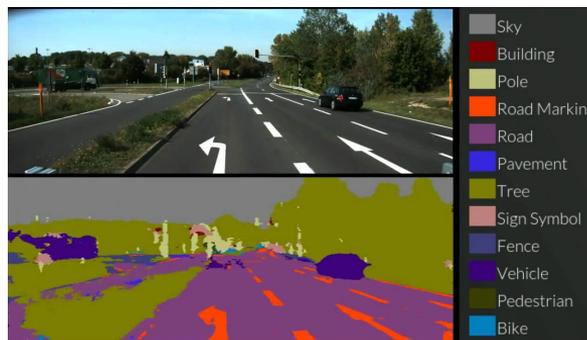
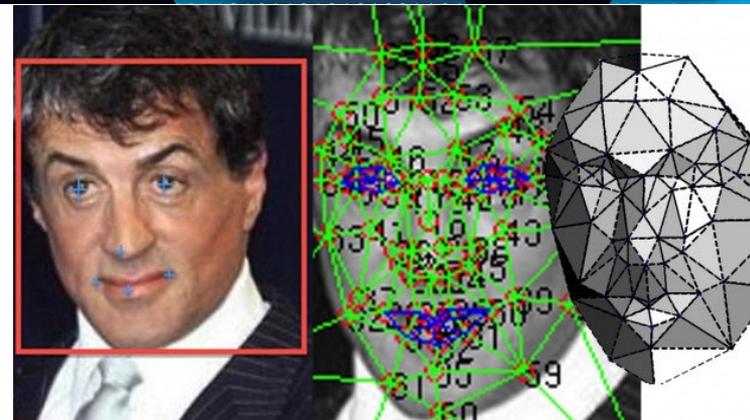
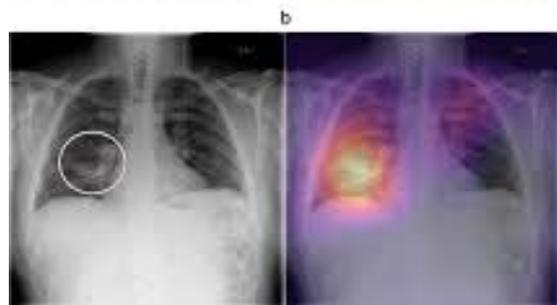
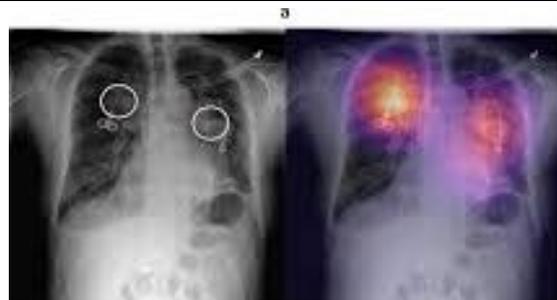


comparative - superlative

The underlying concept that distinguishes *man* from *woman*, i.e. sex or gender, may be equivalently specified by various other word pairs, such as *king* and *queen* or *brother* and *sister*. To state this observation mathematically, we might expect that the vector differences *man* - *woman*, *king* - *queen*, and *brother* - *sister* might all be roughly equal. This property and other interesting patterns can be observed in the above set of visualizations.



Fields: vision



...a digression





Law and Ethics

International approaches to AI law & ethics



OECD Legal Instruments

Login

[Home](#) [General information](#) [Full list](#) [Advanced search](#) [Adherences](#) [Key figures](#)

▸ FR



OECD/LEGAL/0449

📅 Adopted on: 22/05/2019

Recommendation of the Council on Artificial Intelligence

📄 In force [📄 Recommendation](#) [📄 Science and Technology](#)

Section 1: Principles for responsible stewardship of trustworthy AI

... **RECOMMENDS** that Members and non-Members adhering to this Recommendation (hereafter the “Adherents”) promote and implement the following principles for responsible stewardship of trustworthy AI, which are relevant to all stakeholders.

The screenshot shows the OECD Legal Instruments website. The header includes the OECD logo and the text 'OECD Legal Instruments' with a 'Login' link. A navigation menu contains 'Home', 'General information', 'Full list', 'Advanced search', 'Adherences', and 'Key figures'. A search bar with 'FR' and a magnifying glass icon is visible. The main content area displays 'OECD/LEGAL/0449' and 'Adopted on: 22/05/2019'. The title of the document is 'Recommendation of the Council on Artificial Intelligence'. Below the title, there are tags: 'In force', 'Recommendation', and 'Science and Technology'.



1.2. Human-centred values and fairness

- a) AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.
- b) To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.



The screenshot shows the OECD Legal Instruments website. The header includes the OECD logo and the text 'OECD Legal Instruments' with a 'Login' link. A navigation menu contains 'Home', 'General information', 'Full list', 'Advanced search', 'Adherences', and 'Key figures'. Below the menu, the document identifier 'OECD/LEGAL/0449' and the date 'Adopted on: 22/05/2019' are visible. The main title of the document is 'Recommendation of the Council on Artificial Intelligence'. At the bottom of the page, there are links for 'In force', 'Recommendation', and 'Science and Technology'.



1.4. Robustness, security and safety

a) AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.

b) To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.

c) AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.



The screenshot shows the OECD Legal Instruments website. The header includes the OECD logo and the text 'OECD Legal Instruments' with a 'Login' link. A navigation menu contains 'Home', 'General information', 'Full list', 'Advanced search', 'Adherences', and 'Key figures'. Below the menu, there is a search bar with 'FR' and a magnifying glass icon. The main content area displays 'OECD/LEGAL/0449' and 'Adopted on: 22/05/2019'. The title of the instrument is 'Recommendation of the Council on Artificial Intelligence'. At the bottom of the content area, there are tags: 'In force', 'Recommendation', and 'Science and Technology'.



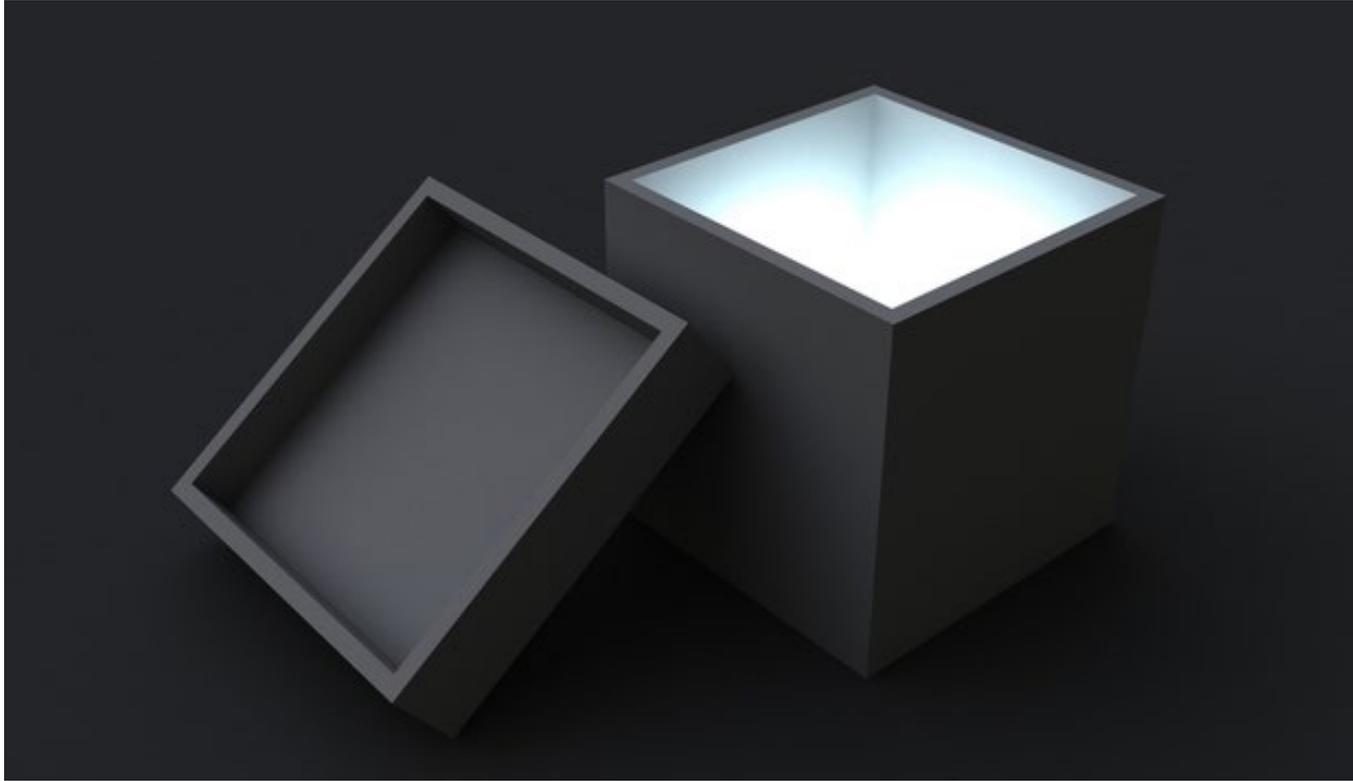
1.5. Accountability

AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.



6 problems

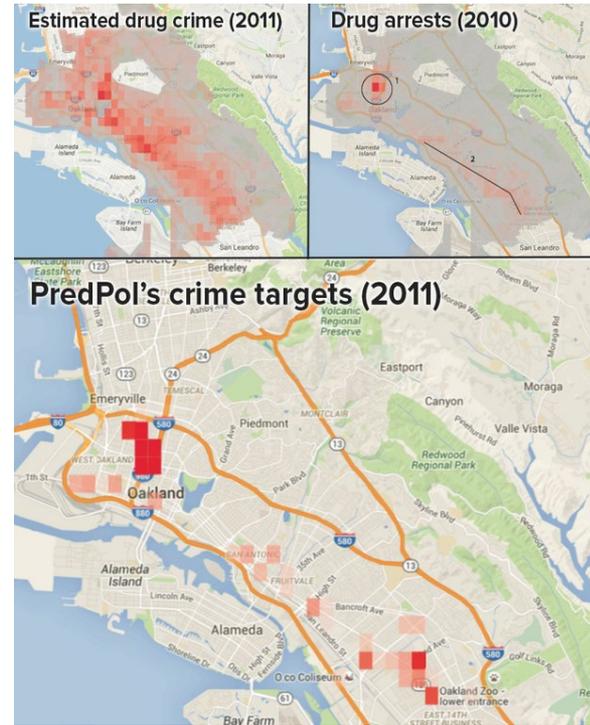
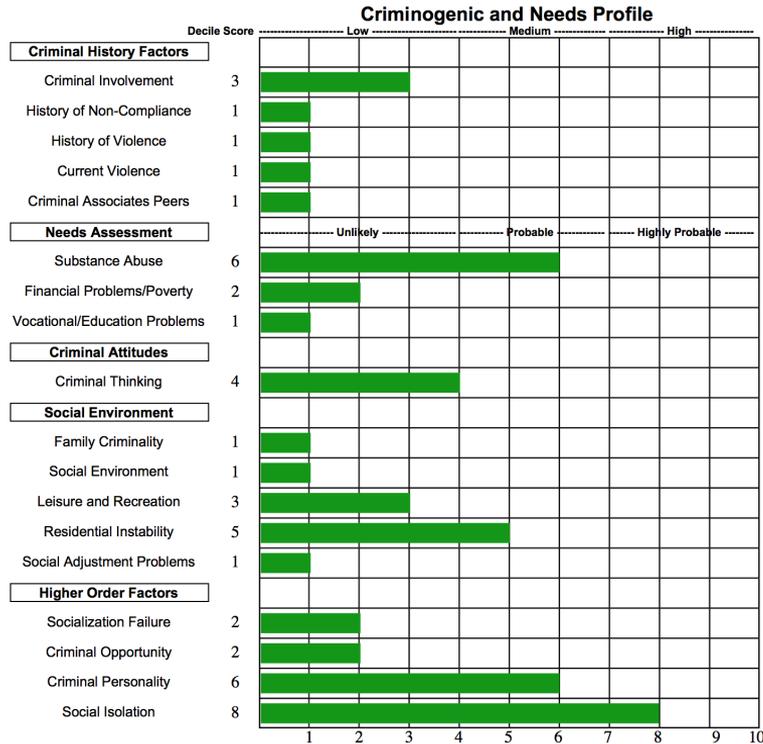
The XAI problem



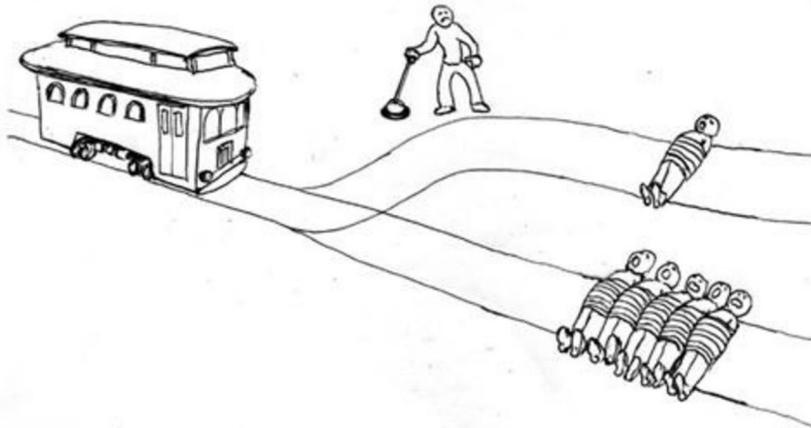
Facial recognition



Predpol & RAT

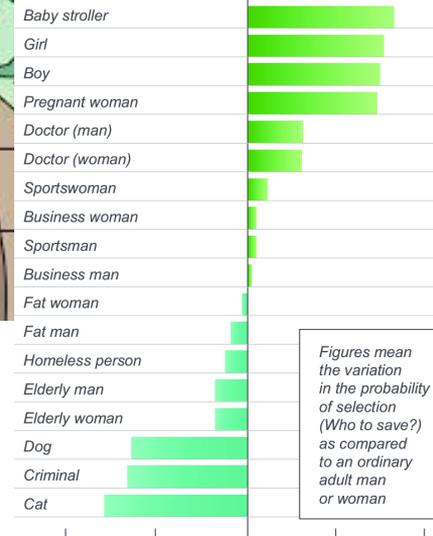
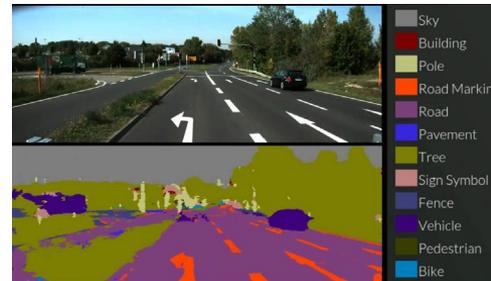


The trolley problem



You find yourself at a lever. A runaway trolley approaches five people who are tied to a set of tracks. Pulling the lever will divert the trolley to a different set of tracks, where only one person is tied down.

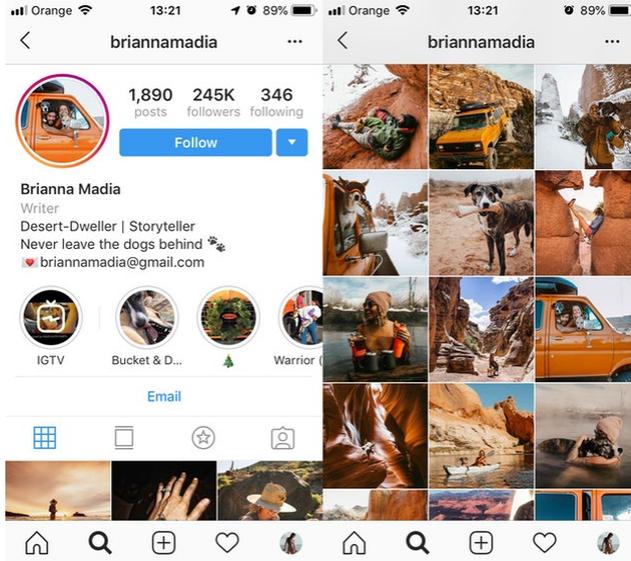
Do you pull the lever?



Killer robots



Real killer robots





Conclusion

What should you do?



1. Don't panic about the Singularity.
2. Expect traditional law & ethics to work fine about 53% of the time.
3. For the remaining 47% of the time, ask Australian federal and state governments to do something (or *anything*).



DAN HUNTER

EXECUTIVE DEAN, FACULTY OF LAW

CHIEF INVESTIGATOR, ARC COE FOR AUTOMATED DECISION-MAKING & SOCIETY

QUEENSLAND UNIVERSITY OF TECHNOLOGY

dan.hunter@qut.edu.au

END



Ethical Leadership in Practice

A series in creating and managing ethical leadership in organisations.





QUTeX Digital Capability Practice

A community for digital thought leadership,
life-long learning and partnership.



Let's navigate the future together.



QUTeX Digital Capability Practice

NEXT EVENT

Save the date

Thursday

14

May 2020



Technology is changing how we think, work and live.

QUTeX's Real World Futures program keeps track of what this means.

www.real-world-futures.qut.edu.au/



real world **futures**

#futureworking

#futurethinking

#futureliving