

## **Class 1 (Tuesday 13 February 2018) AI, ROBOTICS AND ETHICS**

### **BACKGROUND**

“Artificial Intelligence” means the design of machines which can do the kinds of things that make us humans seem intelligent. Initially this meant designing software programs capable of intellectual feats like reasoning, calculating, game-playing. Robotics added to this the hardware associated with *bodily* movement, sensing, and acting. It is sometimes called “embodied AI.” Recent developments in AI include the idea of ‘machine learning’ – a machine’s ability to identify patterns in the vast quantities of data that can now be fed into it.

Increasing sophistication with these developments have led to the idea of machines as “autonomous agents.” The claim is that machines can act autonomously when it comes to making decisions including ethical decisions. That is to say, they can do more than merely mimic humans’ reasoning processes and can actually make decisions for themselves without any human pre-programming. (Similar to Hal, the computer in *2001 Space Odyssey* which developed the ability to deceive its human controllers).

Philosophically, this raises questions about what it means to make an ethical judgement, for example, what should I do when faced with a choice between two equally unpalatable alternatives like killing a group of people in order to save the life of one child? Is morality just a matter of following pre-programmed rules? It also raises questions about the nature and locus of moral responsibility: who is to be held morally responsible for the actions of “autonomous agents” (e.g., the killing of civilians in warfare)? Can something non-human be held morally responsible and be accorded rights?

In this class we will begin by looking at some of the ethical issues raised by 3 different applications of AI: driverless cars; drone warfare; carer robots. We then turn to more general, and deeper questions about the nature of ethics, moral responsibility, and the broader context of these technological developments.

### **SPECIFIC QUESTIONS**

#### **1 Driverless Cars**

What are some of the ethical issues raised by driverless cars?

How useful is the doctrine of utilitarianism in programming a driverless car to make ethical judgements? Where does responsibility lie when things go wrong?

Is there more to morality than following a set of rules?

<https://aeon.co/essays/can-we-design-machines-to-make-ethical-decisions>

#### **2 Drone Warfare**

Can the very idea of a war fought by remotely controlled drones be judged to be unethical? Is drone warfare different from conventional warfare in an ethically relevant way?

Does utilitarianism have a satisfactory resolution for this issue?

Where does moral responsibility lie in the case of killing by drones?

Robo wars (P.W.Singer): <http://www.abc.net.au/radionational/programs/philosopherszone/the-morality-of-robo-wars---p.w.-singer/3701744>

### 3 Robot carers

What are the ethical issues raised by robotic carers?

Empathy can be defined as the capacity to see things from another person's point of view, and this is the basis for virtues such as compassion. Is empathy a necessary ingredient of morality?

What, if any, qualities should a robot exhibit if we are to consider granting it human rights? Citizenship?

Aged care robots (audio): <http://www.abc.net.au/radionational/programs/lifematters/2016-03-22/7256014>

Citizen robots: <http://www.hansonrobotics.com/robot/sophia/>

## GENERAL QUESTIONS

1 Can ethics be programmed? Can robots be ethical?

Ethical calculus: <https://theconversation.com/killer-robots-free-will-and-the-illusion-of-control-87460>

Can robots be ethical? (The Minefield):

<http://www.abc.net.au/radionational/programs/theminefield/can-robots-be-ethical/7300532>

2 Empathy can be defined as the capacity to see things from another person's point of view, and this is the basis for virtues such as compassion. Is empathy a necessary ingredient of morality?

<https://aeon.co/essays/creating-robots-capable-of-moral-reasoning-is-like-parenting>

(Video) robots and feeling pain: <https://aeon.co/videos/pain-leads-to-empathy-and-self-preservation-should-we-make-robots-feel-it>

3 Does AI lead to a debasement of human values? Does (genuine) ethical judgement require emotion and feeling (e.g., compassion, disgust)? Empathy? Trust?

**\*\* Citizen robots – is AI changing our understanding of what it means to be human? (a Minefield must): <http://www.abc.net.au/radionational/programs/theminefield/can-there-be-a-place-for-robots-within-the-human-community/9104338> NOTE: This is a key resource for the topic**

## FURTHER READING/LISTENING (OPTIONAL)

Driverless cars: <https://theconversation.com/at-last-the-worlds-first-ethical-guidelines-for-driverless-cars-83227>

Drone warfare: <http://www.abc.net.au/radionational/programs/allinthemind/do-you-read-me-hal-robot-wars-moral-machines-and/3070632#transcript>

(Robotic parenting) <https://theconversation.com/robot-babies-from-japan-raise-all-sorts-of-questions-about-how-parents-bond-with-ai-66815>

Human-like robots: <http://www.ethics.org.au/on-ethics/blog/december-2017/bladerunner-westworld-sexbot-suffering>

Robot love in Japan: <https://www.youtube.com/watch?v=YzzDLujpat4>

Programming ethics; hazards of bias:  
<http://www.abc.net.au/radionational/programs/futuretense/ai,-algorithmic-decision-making,-ethics-and-the-under-represent/9250270#transcript>

Examples of bias: <https://www.wired.com/story/artificial-intelligence-seeks-an-ethical-conscience/>

<https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/>

<http://www.abc.net.au/radionational/programs/futuretense/robots-and-ai/8804292#transcript>