

Can (Ro)bots Really Be Moral?

Moral Machine, Teaching Robot Right from Wrong – Wallach, Wendell Allen, Colin

AI-The ability of 'intelligent agents' in taking actions based on the information instructed by human to maximize the chances of success.

CHAPTER 4: Argument- current AIM of AMAs is not necessary to be compromised by the controversy of its limit

The discussion is useful if it suggest capacities that need to be engineered into AMAs.



Ethics of deterministic systems ???

- Free will ("unseen matter") – unable to formulate in AMAs
- Key features (Luciano Floridi and J.W. Sanders)
 1. Interactivity- Response to stimulate by change
 2. Autonomy- Ability to change state without stimulus
 3. Adaptability-ability to change the transition rules- learn by experience

increase the number of choices!!!

- Deterministic system as a moral agent is unanswerable

Understanding

'capacity to react appropriately and adaptively to social and physical environment'

- multimodel approaches & Theory of embodied cognition ignite the development of robot

It is a long way for robot to make ethical decision

BUT

- 1- Able to adapt in environment
- 2- reaction/responses without need for conscious thought

Consciousness

- Consciousness within artificial agents is being developed which covers 5 areas:

- 1. sense of self
- 2. imagination
- 3. focused attention
- 4. forward planning
- 5. emotion

Aim (AMA)- Should achieve Fuctional Equivalent Behavior

Stan Franklin- IDA- is archetecture and mechanism able to do many the same task as conscious human do

BUT

-Still long way off (Owen Hollen and Murray Shanahan)

Moral Turing Test

TURING TEST EXTRA CREDIT:
CONVINCE THE EXAMINER
THAT HE'S A COMPUTER.

YOU KNOW YOU MAKE
SOME REALLY GOOD POINTS.
/ I'M ... NOT EVEN SURE
WHO I AM ANYMORE.



- Possible to create?
- Necessary to define morality?
- Is positive outcome enough to label computer as AMA?
