Moral Machines

A brief summary

Susanne Bornelöv
Si Chen
Julia Paraskova
25-12-2010

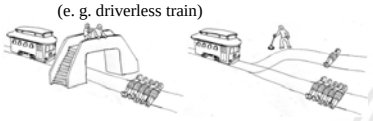
1

Why we need moral machines?

- Inevitability of autonomous systems
- People's lives affected by the decisions of these systems
- Capacity to make moral decisions

Why we need moral machines?

1. Trolley case:
   A trolley is running out of control down a track. What should IT do?
   (e. g. driverless train)



Picture from http://tijmz.wordpress.com/

2. Built-in ethical constraint for killing machines
   Noncombatants
   Enemy solders trying to surrender

W. Wallach and C. Allen, Moral machines: Teaching robots right from wrong. 3

Engineering morality

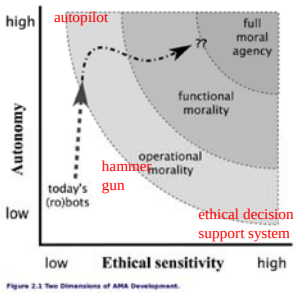


Figure 2.1 Two Dimensions of AMA Development.

W. Wallach and C. Allen, Moral machines: Teaching robots right from wrong.

Two independent dimensions: autonomy and sensitivity to values.

Operational morality: The moral significance is entirely in the hands of designers and users.

Functional morality: The machines themselves have the capacity for assessing and responding to moral challenges

4

Do we want computers making decisions for us?

- Fear of advanced technology (e. g. cloning, genetic engineering)
  - Things getting out of control
  - Fear of deeper understanding of human beings themselves
- Lack of human qualities, such as consciousness and emotions
- Effects on human culture
  - Delegating responsibility for decision making to machines
  - Becoming enslaved by machines

5

Design of AMAs

- Top-down approach (rule based)

Take an ethical theory,
A**nalyze** the informational and procedural requirements necessary to implement this theory in a computer system
Apply that analysis to the design of subsystems and the way they relate to each other in order to implement the theory

In top-down approach, the ethical principles are defined.

6

Design of AMAs

Bottom-up approach (revolution or learning based)
Create an environment where an agent explores courses of action and learns and is rewarded for behavior that is morally praiseworthy, like childhood development.

The ethical principles in bottom-up approach must be discovered or constructed.

7

Beyond reason

- Human-robot interaction is important for moral decision making
  (e.g. service robot interacting with people in a hospital or at home)
- Robots should
  have emotions
  be embodied in the world
  have social skills (e. g. the ability to read nonverbal cues and gestures)

8

Conclusions

Interest in AI is like interest in animals.

The discussion about AMAs raise the question about ethics itself.

Technology is here to stay, we need to deal with it.