AI Ethics

Groundwork in Moral Theory and its Implication in Technological Advancement

Contents of today‘s lecture

- Overarching question: Why does AI advancements perplex moral philosophers?
  - How is it challenging existing ideas and theories?

- 1. Trolley examples
- 2. (Moral) agency & (Moral) patiency
- 3. Instrumentalist Theory of Technology
- 4. Practical application of ethics in the context of AI

Agency

- The ability to do something that counts as an action (Himma 2009).
  - Basically doing something.
  - Actions as doings.

- So is simply exsisting enough?
  - Not quite.

- Requires a certain mental state.

Patiency

- Inanimate objects
- Something that is imposed upon by the will of an agent
  - A rock
  - A steering wheel
  - A trolley

Moral agency

- Beings whose behaviour is subject to moral requirements.
  - Moral obligations
  - Accountability for one‘s actions

- Agency is a prerequisite for moral agency
  - But not all agents are moral agents.

- What about autonomy?

Moral patiency

- Agents who do not meet requirement of moral agency
  - Someone who is owed at least one duty or obligation (Himma 2009)
    - Newborn infants
    - Animals

The Instrumentalist Theory of Technology (IT)

- Technology is a tool for humans. An extension of man.
  - The computer, the hammer, the lawnmower
- Any moral violations are clearly the responsibility of developers or end users.
- Eliminates the problem of the computer as a scapegoat.
- Sounds reasonable?

Problems for the IT

- Anthropocentric
  - Excluding

- Historically what counts as an agent or a moral agent has been prone to change.

Problems for the IT

- Does it hold that all technology is an extension of man?
  - Is it simply a tool that we use.
  - Under the IT, technology is thought of as having patiency.

- What about simple machines?
- Can we extend agency to technology?

Autonomy as Requirement for Moral Agency

- “In Kant‘s metaphysics, autonomy refers to the fundamental condition of free will – the capacity of the will to follow moral laws which it gives to itself” (Winner 1977)
- Kant contrasts this view with the concept of heteronomy – The rule of the will by external laws, or the deterministic laws of nature
- What do we think of “autonomous” technology when contrasted with this?
- It seems that when we say that technology is autonomous, it is nonheteronomous
  - Not governed by external law

Discussion

- Is there need to rethink moral agency?
  - Do we include AI as agents or moral agents?

- Does it matter, in the practical sence?
  - Debate on moral agency seems purely theoretical.
  - So what if we say that AI systems are moral agents?

- Plenty of debates to settle before we induct AI into moral agency
  - What do we mean when we say that a machine is autonomous?
  - Does it hold the same standard as human autonomy?

AI as autonomous moral agents?

- Although AI push the boundaries for the definition of the autonomous moral agent, assigning any form of responsibility or obligation to the entity itself is a perplexing and inconceivable endeavour.
  - Philosophers and legal scholars still struggle with this question.
  - Practical application of AI ethics seems to be taking a different stance
  - Rather than assigning moral responsibility to the AI system, responsibility resides with developers and users of the technology.
  - However, ethical challenges are recognised and addressed in the development and implementation process with ethical frameworks.

Questions of usefulness

- It is common that ethical challenges are addressed with ethical principles
  - Principles of beneficience, non-maleficence, explicability, well-being etc.

- Montréal Declaration for Responsible Development of Artificial Intelligence:
  - The well-being principle
    - 1. “AIS must help individuals improve their living conditions, their health, and their working conditions“
    - 2. “AIS must allow individuals to pursue their preferences, so long as they do not cause harm to other sentient beings“
    - 3. “AIS must allow people to exercise their mental and physical capacities“

Ticking ethical boxes

- Moral principles are all well and good up to a certain point.
  - Properly defined, they can give a description of some form of ideal.
  - Keep us focused on the ethical challenges we face.
- However, they are not enough an can be problematic.
  - Invite developers and users to look the other way by saying they‘ve fulfilled a certain principle.
- Ethics as guardrails to prevent a crash

A holistic approach

- AI as a socio-technical ecosystem
- As much about the development of the AI system as it is about the society in which this technology is being developed and deployed.
- “AI is not just about the automation of decisions and actions, the adaptability to learn from the changes affected in the environment, and the interactivity required to be sensitive to the actions and aims of other agents in that environment, and decide when to cooperate or to compete. It is mostly about the structures of power, participation and access to technology that determine who can influence which decisions or actions are being automated, which data, knowledge and resources are used to learn from, and how interactions between those that decide and those that are impacted are defined and maintained.” (Dignum 2022)