

## **AI for a better world?**

– Ethics and responsible development of new technology.

NTNU European Conference 2018, Brussels November 26

Siri Granum Carson

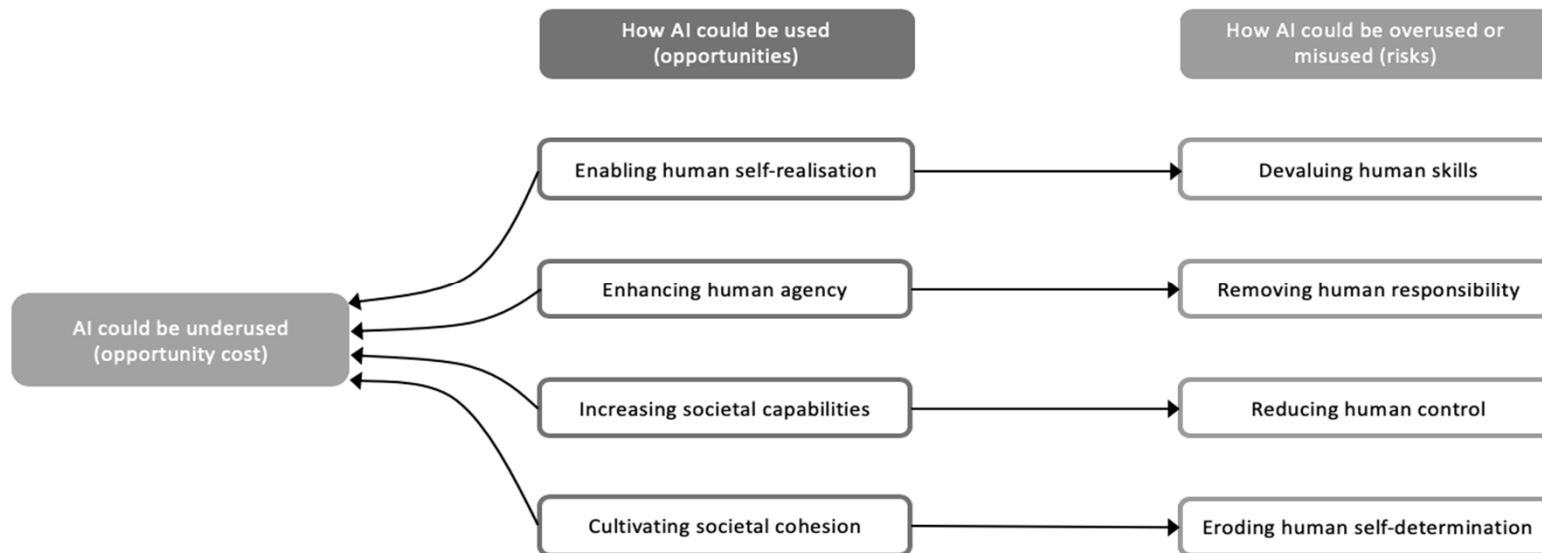
Programme for Applied Ethics, NTNU

# Basic questions in ethics

*What is right? What is a good life?*



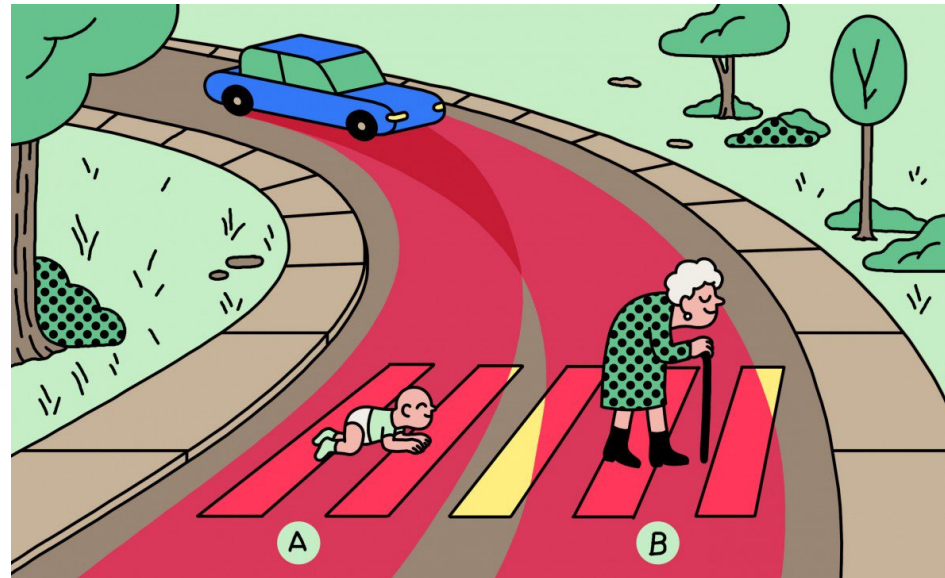
- How does AI affect our obligations or in other ways change our relationship with ourselves, others and the world?
- How can research, innovation and regulation of AI lead to a better world?



Ref.: Floridi et al. In *Minds and Machines*, 2018

# Whose ethics?

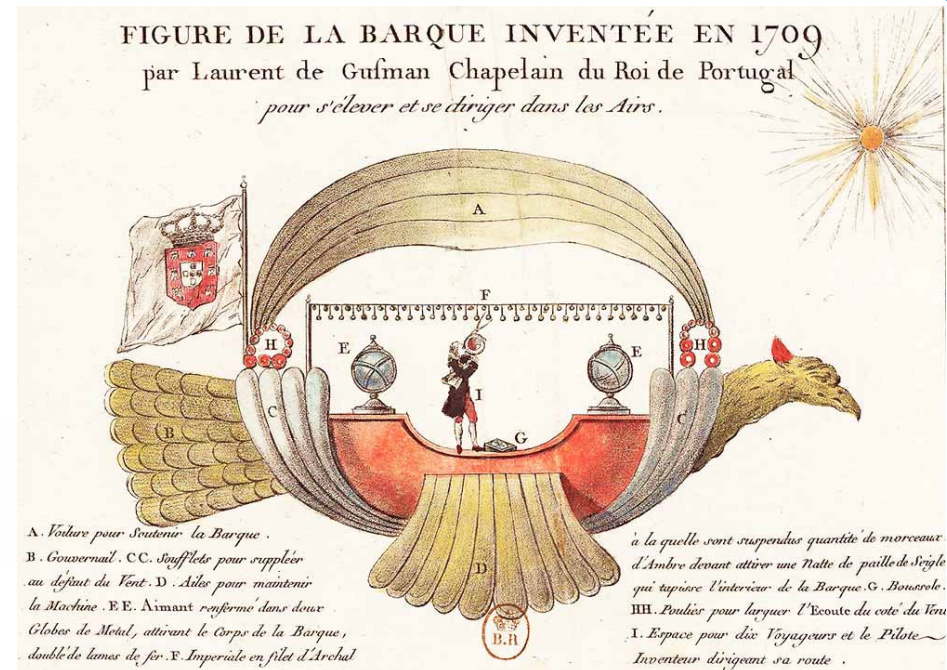
- The Trolley Problem.
- MIT's «Moral Machine» experiment illustrates how moral preferences vary, correlating with culture and economy.
- Implications for AI development?
- TechnoMoral change.



III: SIMON LANDREIN Ref.: MIT Technology Review, Oct. 2018

# From «ugly bird» to «hot robot»

- von Schomberg: «Old» vs. «new» understanding of responsible innovation.
- Collingridge dilemma: The information problem and the power problem.
- A solution with limitations: The precautionary principle.
- Beware the «ethics of hype»...



# Five core principles for ethical AI

*Cf. Beuchamp and Childress, 1979; Floridi et al., 2018.*



## I. Beneficence/«do good»

- Promoting human well-being, preserving dignity, sustaining the planet.

## II. Non-Maleficence/«do no harm»

- Securing privacy, security, «capability caution» (against overuse, misuse)

## III. Autonomy

- Balance decision-making power and delegation.

## IV. Justice

- Promoting shared benefit and prosperity; preserving solidarity

## V. Explicability

- Enabling I.-IV. by promoting intelligibility and accountability

# AI for a better world – the responsibility of NTNU

- Exposing researchers to research ethics and RRI training.
- Utilizing unique structures for cross-disciplinary research and close collaboration with industry.





**Thank you for the attention!**

