(De)constructing ethics for autonomous cars: A case study of Ethics Pen-Testing towards “AI for the Common Good”

By Bettina Berendt

Colton Asnes

> “*Many ethics codes…require that AI contribute to the* ***Common Good***

Common Good

- “*That* which *benefits society* as a whole” (Lee n.d.)
- That - natural environment, hospitals/schools, social institutions/practices (private property)
- Benefit - utility from an economic perspective or interests/values for an individual or group
- Society - country? Citizens or all human beings?
- Substantive, Procedural, Communal and Distributive Common Good

GOAL

Evaluate the Ethical Penetration Testing concept with four questions to respondents

The “Moral Machine” Experiment

Assume an autonomous vehicle that is about to crash and *cannot* save everyone

1\.

Machines make decisions that affect people.

“Machines are tasked not only to promote well-being and minimize harm, but also to distribute the well-being they create, and the harm they cannot eliminate” (Awad et al. 59)

2\.

Moral decisions should be based on social consensus

Survey method with majority voting

3\.

40 million decisions from 233 countries were analyzed

Relationship between demographics and ethics

Universality vs. Particularity of morality

Participants, Materials, and Procedure

Participants

Group GE

Humanities scholars at the KIAS AI, Ethics and Society Conference in Alberta, Canada

Group GP

Computer Science, AI/Law, and Philosophy experts at the ESME Conference in Pisa, Italy

Group GG

Experts in data science from a variety of fields at the Trust in Data Science Summer School in Ghent, Belgium

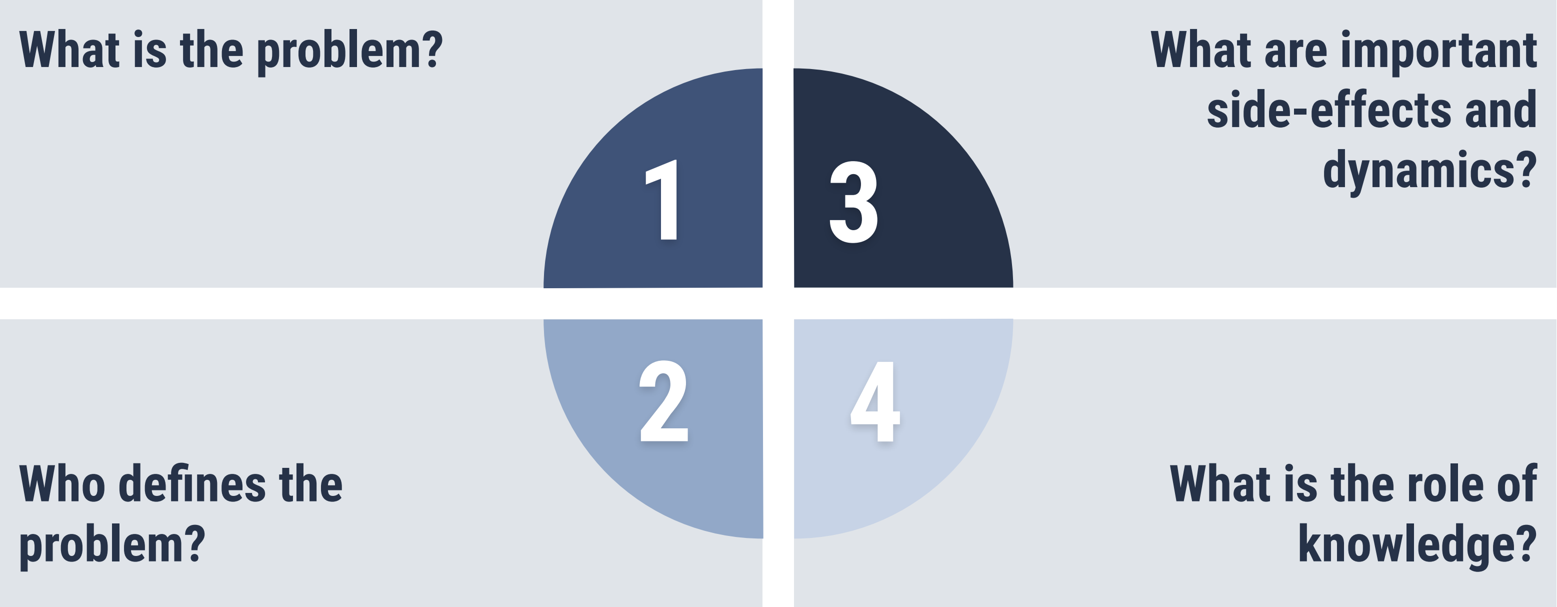
Group GB

Graduate students and faculty at the VeriLearn research project

Group GM

Authors of the original paper

Questions



1. What is the problem?
2. Who defines the problem?
3. What are important side-effects and dynamics?
4. What is the role of knowledge?

Results

1. What is the problem?

Responses:

Research Priorities

“Wrong priority: concentrate on making normal behavior as safe as possible” (GB)

“Top-level question should be how to prevent accidents” (GG)

Socio-Technical Systems

Separate roads for AVs, why not public transport? (GG, GE)

Domain knowledge about EV’s

“Can’t the cars just stop” (GP)

“Stop and self-destruct” (GB)

Machine Ethics

“We don’t trust people to make these moral decisions, why should we trust machines” (GE)

Ethics and Democracy

Results

2. Who defines the problem?

Responses:

“The main stakeholder is a paying customer” (GG)

“This is philosophical BS” (GE)

“People other than engineers should think about this” (GB)

Results

3. What are important side-effects and dynamics?

Responses:

“The assumption is made that the victims just stand there” (GE)

“The experimental setup assumes no uncertainties in the life-and-death outcomes” (GM)

‘“Confirmation bias: People will focus on this kind of scenario and not about alternatives” (GG)

Results

4. What is the role of knowledge?

Responses:

“The experimental setup assumes no uncertainties in the recognition of personal characteristics of the victims” (GM)

“The assumption is made that all properties of the ‘victims’ can be observed/assessed by the AI” (GE)

“‘Autonomous cars have killed people because they mistook them for a plastic bag. We assume that the decision to drive over an inanimate object in an emergency is ethically unproblematic, but the concrete value that a probability threshold in a recognition function has, can lead to such mistakes. Should we make programmers aware of this potential effect of their setting a probability threshold?” (GB)

Conclusion

The remarks made by participants in four parallelized EPT sessions lent themselves well to being structured following the four lead questions.

The discussions also showed how important thorough technical domain knowledge is in order to have a meaningful discussion about the ethical challenges and options

The results also demonstrated the importance of framing and education. It is much easier to capture engineers' attention to and willingness to participate in ethical decisions when focusing on statistical trolley problems and AIs, when compared to asking them to think “directly” about ethics and enticing them to think of their machines making conscious ethical decisions.

Acceptance/Rejection

- Broad Questions
- Unstructured

Discussion

- The paper referenced the “human dignity” importance in the German constitution, where sacrificing human lives should not be considered. What are your thoughts on this?
- https://www.moralmachine.net/