Novel perspectives on Ethics and Autonomous Vehicles

Andreia Martins Martinho
15-11-2018
0. Introduction

1. Should AI systems be furnished with moral capacities?

2. How can AI systems be furnished with moral capacities?

Autonomous Vehicles
1. Ethics of AVs
1.1 Anatomy of literature on the Ethics of AVs

2. Ethical Issues
3. Social Dilemma
   Assumption: AVs will be faced with moral dilemmas.

1. Social Benefits
4. Solution
1.2. How to equip AVs with moral values?

- **Traditional**
  Translating moral theories into algorithms for AVs agency.

- **Data-Driven**
  Using Big Data to understand global social preferences OR to assist AVs through learning algorithms.

- **Heterogeneous**
  Combination of different moral theories OR moral theories with data OR Humanlike AVs.

**Costs & Constraints**
Describing ethical principles either as costs or constraints (rules) in Optimal Control Theory (Gerdes & Thornton, 2016)
1.3. In a nutshell: Costs & Constraints Ethical Approach

**PHILOSOPHY**
- Consequentialism
- Deontology

**OPTIMAL CONTROL THEORY**
- Cost Function
- Constraints

**OPTIMIZATION**

**ACTIONS**
- Restricted

**Dilemma**

**Soft Constraints**
1.4. The Moral Dilemma

- Collision is imminent and unavoidable;
- The agent is able to distribute the harms that arise as a result of the collision;
- The decision situation is one of certain.
Do trolley cases really pose a central challenge in the ethics of AVs? Recent research hints otherwise.

- Outliers
- Unrealistic
- Inconsistent
- Limitations in Design
- Evidential Value
The Moral Machine experiment

Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon & Iyad Rahwan

Nature 563, 59–64 (2018) | Download Citation

Abstract

With the rapid development of artificial intelligence have come concerns about how machines will make moral decisions, and the major challenge of quantifying societal expectations about the ethical principles that should guide machine behaviour. To address this challenge, we deployed the Moral Machine, an online experimental platform designed to explore the moral dilemmas faced by autonomous vehicles. This platform gathered 40 million decisions in ten languages from millions of people in 233 countries and territories. Here we describe the results of this experiment. First, we summarize global moral preferences. Second, we document individual variations in preferences, based on respondents’ demographics. Third, we report cross-cultural ethical variation, and uncover three major clusters of countries. Fourth, we show that these differences correlate with modern institutions and deep cultural traits. We discuss how these

“car manufacturers and policymakers are currently struggling with these moral dilemmas (…)"

→ no study is used as a reference to support such claims.
2. Beyond the Trolley

2.1. Ethical Issues reported in the literature

- Mundane traffic situations
- Cost/benefit issues
2.2 Qualitative Research Study

- **Methodology:** Semi-structured interviews held at Delft University of Technology (TUD)
- **Selection of Subjects:** TUD website + personal references
- **Recruitment of the Subjects:** email
- **Subjects:** $n = 6$ [5 in the Faculty of TUD 1 Senior Scientist of TUD/TNO. Departments: Cognitive Robotics; Delft Center for Systems and Control; Delft Haptics Laboratory]
- **Limitations:** small sample
- **Findings:**

1. Trolley problems are **rare** or **unrealistic**.
2. Contexts where AVs will need to make moral decisions:
   - Selection of passengers;
   - Collisions;
   - Sharing of data;
3. Only one subject was somewhat aware of the ethical discussions about equipping AVs with moral values.
3. Wrapping Up

1. Should AI systems be furnished with moral capacities?

**Assumption:** AVs will be faced with moral dilemmas.

2. How can AI systems be furnished with moral capacities?

**Cost & Constraints Approach.**
References

- Wolkenstein, A. (2018). "What has the Trolley Dilemma ever done for us (and what will it do in the future)? On some recent debates about the ethics of self-driving cars." Ethics and Information Technology 1:11.
- pp. 169-173.
Thank you.