

# Introduction to Machine-Ethics

Philip Wälde, M.A  
Freie Universität Berlin, Institute of Computer Science  
Seminar: Normative Reasoning and Machine Ethics  
Winter Term 2019/20

## Outline<sup>1</sup>

### I. Konzeptuelle Grundlagen

Die Maschinenethik beschäftigt sich mit der Möglichkeit Maschinen mit der Fähigkeit ethische Entscheidungen bzw. Handlungen auszuführen. Zentrale Dabei wird bereits auf die Problematik definitorisch nicht abschließbarer Begriffe eingegangen und es werden Definitionsvorschläge für den Kontext des Vortrags explizit angegeben. Wir stellen uns folgende Frage:

Können Maschinen moralisch handeln? Wenn Ja, lässt sich eine moralische Entscheidung modellieren und letztlich programmieren?

Aus dieser Fragestellung lassen sich bereits einige grundlegenden Begriffe ableiten:

1. Moral/Ethik
  - 1.1 normative Ethik: Utilitarismus und Pflichtethik
2. Subjekte/Akteure
3. Handlung/Entscheidung
4. Freiheit/Autonomie
  - 4.1 Handlungs-/Willensfreiheit
5. Verantwortung

Folgende Kriterien ergänzen die Einführung und dienen als Vorschlag zur Bewertung der ethischer Fragestellungen: 1. Normen und Werte als Maßstab 2. Allgemeingültigkeit 3. Universalisierbarkeit 4. Unbedingtheit 5. Vorrangigkeit. Die Begriffe stehen in vielfältigen Beziehungen zueinander. Exemplarisch werden solche Beziehungen durch die Argumente im nächsten Schritt verdeutlicht. Bereits nach dem einleitenden Teil sollte deutlich sein, dass die Lösung ethischer Problem mehrere Antworten und/oder Dilemmata zulässt, moralische Begriffe vage sein können und verschiedene Interpretationen zulassen.<sup>2</sup>

---

<sup>1</sup> An english Version will follow shortly! I want this to be well written.

<sup>2</sup> Leibniz und Kant werden als Primärquelle für zentrale Definitionen und Begriffe angeführt.

### **I. a) Argumente für eine Maschinenethik**

- Wer ist verantwortlich für das Handeln von Maschinen?
  - Stichwort: Ratgebersysteme und eigenständige Maschinen
  - Rationales Handeln und Transparenz
  - Kantische Pseudoethik: geboten, verboten und erlaubt als Grundlage deontischer Logik

### **I. b) Argumente gegen eine Maschinenethik**

- Starke Begriffe und Phänomene von moralischer Person, Willensfreiheit und Verantwortung als Grundlage der Rechtsordnung lassen sich durch algorithmische Systeme nicht repräsentieren.
  - Eine Pseudoethik wird von diesem Einwand nicht getroffen und begnügt sich mit schwächeren Begriffen.

### **Literatur:**

Janina Loh. Roboterethik - Eine Einführung. Suhrkamp 2019.

Catrin Misselhorn. Grundfragen der Maschinenethik Reclam 2018.

Ludger Heidbrink, Claus Langbehn, Janina Loh (Hrsg.). Handbuch Verantwortung. Springer 2017.

Luís Moniz Pereira, Ari Saptawijaya. Programming Machine Ethics. Springer 2016.

Wendell Wallach, Colin Allen. Moral Machines: Teaching Robots Right from Wrong. Oxford University Press 2008.

McNamara, Paul, "Deontic Logic", The Stanford Encyclopedia of Philosophy (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/logic-deontic/>.