

---

## **Autonome Autos können ethisch-moralisches Verhalten lernen**

Von Hans-Robert Richarz

Maschinen können zukünftig sehr wahrscheinlich menschlich-ethische Entscheidungen fällen. Das besagt eine Studie, die das Institut für Kognitionswissenschaft der Universität Osnabrück jetzt veröffentlicht hat. Damit wäre die häufig erhobene Kritik am vollautomatischen Fahrzeug vom Tisch, die besagt, dass ein solches Auto nie in der Lage wäre, zu entscheiden, was im Fall des Falles moralisch vertretbar sei und was nicht: nämlich nicht mit einem Menschen sondern statt dessen beispielsweise mit einem Tier zu kollidieren und es somit zu töten.

Die fortschreitende Entwicklung autonomer Autos wirft Fragen auf. Zum Beispiel danach, wie seine Systeme entscheiden sollen, wenn es um Leben und Tod geht. Sollten selbstfahrende Fahrzeuge zum Beispiel ihre Insassen um jeden Preis schützen, oder sollten sie die Menschen im Inneren opfern, um eine größere Gruppe von Menschen draußen zu retten? Ein Forscherteam der Uni Osnabrück glaubt, dass Moral bald eine Rolle bei den Entscheidungen selbstfahrender Autos spielen könnte.

Erst vor wenigen Wochen hatte die von Bundesverkehrsminister Alexander Dobrindt eingesetzte Ethik-Kommission zum automatisierten Fahren ihren Bericht vorgelegt. Darin hieß es unter anderem: "Die Technik muss nach ihrem jeweiligen Stand so ausgelegt sein, dass kritische Situationen gar nicht erst entstehen. Dazu gehören auch Dilemma-Situationen, also eine Lage, in der ein automatisiertes Fahrzeug vor der Entscheidung steht, eines von zwei nicht abwägungsfähigen Übeln notwendig verwirklichen zu müssen." Weiter hieß es: "Technische Systeme sind nicht eindeutig normierbar und auch nicht ethisch zweifelsfrei programmierbar. Sie müssen auf Unfallvermeidung ausgelegt werden, sind aber auf eine komplexe oder intuitive Unfallfolgenabschätzung nicht so normierbar, dass sie die Entscheidung eines sittlich urteilsfähigen, verantwortlichen Fahrzeugführers ersetzen oder vorwegnehmen könnten."

Das sind sie doch, antwortet Doktorand Leon Sütfeld, der Hauptautor der Osnabrücker Studie und erklärt das so: „Das menschliche moralische Verhalten lässt sich durch den Vergleich von einer Wertigkeit des Lebens, das mit jedem Menschen, jedem Tier oder jedem Objekt assoziiert ist, erklären beziehungsweise mit beachtlicher Präzision vorhersagen. Das zeigt, dass menschliche moralische Entscheidungen prinzipiell mit Regeln beschrieben werden können und dass diese Regeln als Konsequenz auch von Maschinen genutzt werden könnten.“

Prof. Dr. Gordon Pipa, einer der leitenden Wissenschaftler der Studie, fügt hinzu: „Um Regeln oder Empfehlungen definieren zu können, sind zwei Schritte notwendig. Als Erstes muss man menschliche moralische Entscheidungen in kritischen Situationen analysieren und verstehen. Als zweiten Schritt muss man das menschliche Verhalten statistisch beschreiben, um Regeln ableiten zu können, die dann in Maschinen genutzt werden können.“

Um beide Schritte zu realisieren, nutzten die Autoren eine virtuelle Realität, um das Verhalten von Versuchspersonen in simulierten Verkehrssituationen zu beobachten. Etwas mehr als 100 Probanden setzten sich eine spezielle Brille auf, die ihnen das Gefühl gab, tatsächlich Auto zu fahren. Das Dilemma: Die Probanden mussten entscheiden, ob sie Erwachsene, Kinder, Tiere oder Gegenstände wie Heuballen oder Mülltonnen

---

überfahren. Ein Ausweichen war unmöglich. In einigen Fällen ging es um die Entscheidung zwischen Leben und Tod. „Definitiv haben wir herausgefunden, dass Menschen mehr wert sind als Tiere. Tiere aber auch wieder mehr wert sind als Gegenstände. Außerdem hat die Studie ergeben, dass ein Mensch mit Tier mehr wert ist, als ein Mensch allein“, sagt Sütfeld.

Mit diesen Ergebnissen der Versuche fütterten die Wissenschaftler dann einen Computer. Allerdings bekam die Maschine nicht alle Informationen. Etwa zehn Prozent der moralischen Entscheidungen, die die Menschen getroffen hatten, enthielten die Forscher der Maschine vor. Stattdessen musste der Computer die Versuche selbst machen. Und: Er hatte gelernt und entschied auch in den Fällen, die er noch nicht kannte, fast immer so, wie es auch die Menschen in den Versuchen gemacht hatten. Damit war der Beweis erbracht: Der Computer kann moralische Entscheidungen von Menschen lernen und nachahmen.

„Nun, da wir jetzt wissen, wie wir moralische Entscheidungen in die Maschinen implementieren können, bleiben uns trotzdem noch zwei moralische Dilemmata“, sagt Prof. Dr. Peter König, ein weiterer Autor dieser Veröffentlichung, und fügt hinzu: „Erstens müssen wir uns über den Einfluss von moralischen Werten auf die Richtlinien für maschinelles Verhalten entscheiden. Zweitens müssen wir uns überlegen, ob wir es wollen, dass Maschinen sich (nur) menschlich verhalten sollen.“ Die genauen Ergebnisse der Studie sind auf Englisch im Internet <http://journal.frontiersin.org/article/10.3389/fnbeh.2017.00122/full> nachzulesen. (ampnet/hrr)

---

## Bilder zum Artikel



Moral und virtuelle Realität im Dienst des autonomen Fahrens.

Foto: Auto-Medienportal.Net/Universität Osnabrück



Moral und virtuelle Realität im Dienst des autonomen Fahrens.

Foto: Auto-Medienportal.Net/Universität Osnabrück