17-07-17

Maschinen können bald moralisches Verhalten von Menschen imitieren - Neue Studie vorgelegt

Autonome selbstfahrende Autos sind die erste Generation von Robotern, die den alltäglichen Lebensraum mit uns teilen. Deshalb ist es unabdingbar, Regeln und Erwartungen an autonome Systeme zu erarbeiten, die definieren, wie sich solche Systeme in kritischen Situationen verhalten sollen. Das Institut für Kognitionswissenschaft der Universität Osnabrück hat nun eine Studie in Frontiers in Behavioral Neuroscience (http://journal.frontiersin.org/article/10.3389/fnbeh.2017.00122/full) veröffentlicht, die zeigt, dass menschlich-ethische Entscheidungen in Maschinen implementiert werden können und autonome Fahrzeuge bald moralische Dilemmata im Straßenverkehr bewältigen.

Probanden saßen am Steuer eines virtuellen PKWs, der in einem Vorstadt-Setting auf Hindernisse zufuhr. Eine Kollision war unausweichlich, es konnte lediglich die Spur ausgewählt werden (Grafik: Universität Osnabrück)

Politisch wird die Debatte zur Modellierbarkeit von moralischen Entscheidungen durch eine Initiative des Bundesministeriums für Transport und Digitale Infrastruktur (BMVI) begleitet, welche 20 ethische Prinzipien formuliert hat. Die Osnabrücker Studie liefert dazu erste empirische wissenschaftliche Daten.

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 , erklärt Prof. Dr. Gordon Pipa, einer der leitenden Wissenschaftler der Studie.

Um beide Schritte zu realisieren, nutzten die Autoren eine virtuelle Realität, um das Verhalten von Versuchspersonen in simulierten Verkehrssituationen zu beobachten. Die Teilnehmer der Studie fuhren dazu an einem nebeligen Tag durch die Straßen eines typischen Vorortes. Im Verlauf der Experimente kam es dabei zu unvermeidlichen und unerwarteten Dilemma-Situationen, bei denen Menschen, Tiere oder Objekte als Hindernisse auf den Fahrspuren standen. Um den Hindernissen auf einer der beiden Spuren ausweichen zu können, war deshalb eine moralische Abwägung notwendig. Die beobachteten Entscheidungen wurden dann durch eine statistische Analyse ausgewertet und in Regeln übersetzt. Die Ergebnisse weisen darauf hin, dass im Rahmen dieser unvermeidbaren Unfälle moralisches Verhalten durch eine einfache Wertigkeit des Lebens erklärt werden kann, für jeden Menschen, jedes Tier und jedes Objekt.

Leon Sütfeld, der Hauptautor der Studie, erklärt dies 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 bzw. 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.

Diese neuen Osnabrücker Erkenntnisse stehen im Widerspruch zu dem achten Prinzip des BMVI-Berichtes, das auf der Annahme gründet, dass moralische Entscheidungen nicht modellierbar sind.

Wie kann dieser grundlegende Unterschied erklärt werden? Algorithmen können entweder durch Regeln beschrieben werden oder durch statistische Modelle, die mehrere Faktoren miteinander in Bezug setzen können. Gesetze, zum Beispiel, sind regelbasiert. Menschliches Verhalten und moderne künstliche intelligente Systeme nutzen dazu im Gegensatz eher komplexes statistisches Abwägen. Dieses Abwägen erlaubt es beiden - dem Menschen und den modernen künstlichen Intelligenzen - auch neue Situationen bewerten zu können, denen diese bisher nicht ausgesetzt waren. In der wissenschaftlichen Arbeit von Sütfeld wurde nun eine solche dem menschlichen Verhalten ähnliche Methodik zur Beschreibung der Daten genutzt. Deshalb müssen die Regeln nicht abstrakt am Schreibtisch durch einen Menschen formuliert, sondern aus dem menschlichen Verhalten abgeleitet und gelernt werden. So stellt sich die Frage, ob man diese nun gelernten und konzeptualisierten Regeln nicht auch als moralischen Aspekt in Maschinen nutzen sollte , so Sütfeld.

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, 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 Ergebnisse der Studie Using Virtual Reality to Assess Ethical Decisions in Road Traffic Scenarios: Applicability of Value-of-Life-Based Models and Influences of Time Pressure sind erschienen in Frontiers in Behavioral Neuroscience (http://journal.frontiersin.org/article/10.3389/fnbeh.2017.00122/full)

https://idw-online.de/de/news677927