Künstliche Moral im dynamischen Risikomanagement

Autonome Systeme können ohne menschliche Steuerung oder detaillierte Programmierung ein vorgegebenes Ziel selbstständig und an die Situation angepasst erreichen. Dabei ist es oft auch eine ethische Frage was der beste Weg ist um das vorgegebene Ziel zu erreichen. Es ist beispielsweise eine ethische Frage welches Verhalten in einer Situation noch sicher ist. Bei klassischen Systemen wird die Antwort auf diese Frage explizit einprogrammiert oder die Frage kommt gar nicht auf, weil das System nur den Anweisungen des Menschen folgt. Bei autonomen Systemen ist es aber oft nicht mehr möglich jede Situation einzeln zu betrachten und zu entscheiden was noch sicher genug ist. Mittels dynamischem Risikomanagement [1] können Systeme aber in die Lage versetzt werden Risiken zu erkennen, zu bewerten und zu kontrollieren. Beispielsweise können Einflussfaktoren auf das Risiko identifiziert und in einem regelbasierten Expertensystem oder einem Bayesschen Netz [2] so miteinander verknüpft werden, dass die Ausgabe möglichst gut das Risiko bezüglich eines Personenschadens, Sachschadens oder Umweltschadens quantifiziert. Unter Berücksichtigung des quantifizierten Risikos und den dafür ursächlichen Einflussfaktoren kann das autonome System dann Maßnahmen zur Risikoreduktion ergreifen. Die Art und Weise wie ein autonomes System Maßnahmen auswählt um Risiken zu reduzieren, um Risiken auszubalancieren und um ein akzeptables Verhältnis von Risiken und Nutzen zu erreichen ist Teil seiner künstlichen Moral [3]. Diese Art von künstlicher Moral ist essentiell für das Engineering autonomer Systeme und unterscheidet sich deutlich von dem Umgang mit Dilemma-Situationen [4]. Es müssen Methoden erforscht werden wie man diese künstliche Moral so engineered, dass Sie im Einklang mit individuellen und gesellschaftlichen Werten steht.
Der Vortrag zeigt aktuelle Herausforderungen beim Engineering dieser Art der künstlichen Moral auf, diskutiert grundsätzliche Lösungsansätze und gibt einen Ausblick im Hinblick auf die Evolution von heutigen Systemen bis hin zu komplexen HCPS (Human-Cyber-Physical Systems, safetrans Roadmap).

1. https://www.iese.fraunhofer.de/en/seminare_training/edcc-workshop.html#1207475977
2. J. Reich and M. Trapp, "SINADRA: Towards a Framework for Assurable Situation-Aware Dynamic Risk Assessment of Autonomous Vehicles," *2020 16th European Dependable Computing Conference (EDCC)*, 2020, pp. 47-50, doi: 10.1109/EDCC51268.2020.00017
3. Misselhorn, Catrin: Artificial Intelligence and Ethics, #INFORMATIK2018. https://doi.org/10.5446/40420
4. Autonomous Driving Ethics: from Trolley Problem to Ethics of Risk, M. Geisslinger, F. Poszler, J. Betz, C. Luetge and M. Lienkamp, Philosophy & Technology 2021 doi: 10.1007/s13347-021-00449-4