TU Clausthal

Abschlussvortrag Bachelorarbeit Tobias Marcel Kühnel

„Extracting and Formalizing Ethical Properties for AI Applications“

With the increasing spread of artificial intelligence (AI) applications comes a rising demand for assurances and safety guarantees, as well as ethical behaviour. While there exist ethical guidelines for AI systems, they are often too broad and hardly applicable. The high complexity of ethics necessitates precise domain and application-specific rules. To this end, this bachelor thesis surveys how ethical AI behaviour can be formalised and aims to provide a first step towards theoretical morals and practically applicable ethical properties for AI applications. Simultaneously, it seeks to explore the feasibility of automatically verifying these properties by analysing when and how this is possible using neural network (NN) formal verification. Therefore, we conduct a selective literature review identifying existing domain-specific ideas on desirable ethical properties. From these implicit requirements, we extract explicit ethical rules and formalise them in First-Order Logic (FOL). Our work focuses on four properties drawn from the domains of autonomous vehicles, natural language processing (NLP), and recommender systems, respectively, serving as illustrative examples of how to formulate ethical behaviour mathematically.

Betreuer der Arbeit:	Prof. Dr. Rüdiger Ehlers, Prof. Dr. Benjamin Leiding
Datum:	Dienstag, 14. November 2023, 10:30 Uhr
Ort:	Institut für Software and Systems Engineering Besprechungsraum 120 Arnold-Sommerfeld-Straße 1 38678 Clausthal-Zellerfeld Webkonferenz: https://webconf.tu-clausthal.de/b/sim-uc9-rvy