Karlsruher Institut für Technologie

Institut für Mess- und Regelungstechnik

Stellenausschreibung

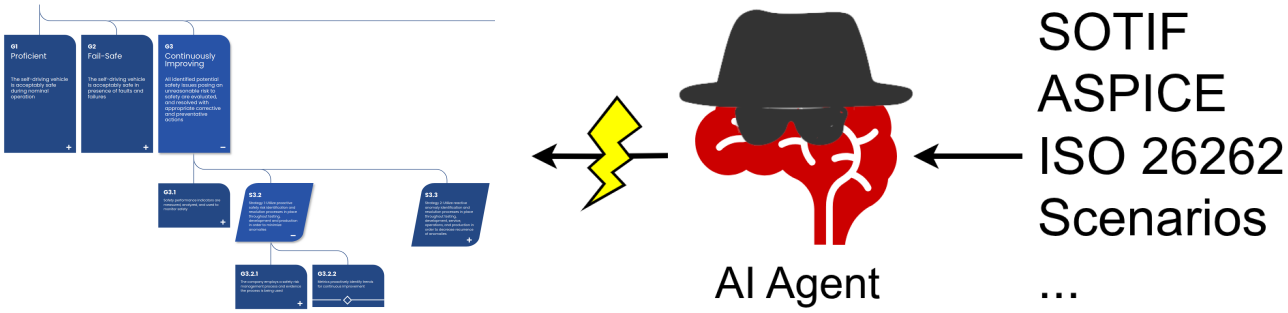
Akademische/r Mitarbeiter/in (w/m/d)

KI-basierte Freigabeargumentation mittels Defeater Agents

Das Institut für Mess- und Regelungstechnik (MRT) hat seinen Forschungsschwerpunkt seit über einem Jahrzehnt im Bereich Automatisiertes Fahren. Alleinstellungsmerkmal sind hier die Versuchsträger mit Sensorik, Aktorik und einem vollständigen Softwaresystem zum automatisierten, urbanen Fahren.
Die vorhandene Ausstattung und Erfahrung ermöglichen nicht nur kontinuierliche Spitzenforschung in allen Teilen des Forschungsfeldes, sondern auch Teilnahme an Wettbewerben zum autonomen Fahren und vielfältige Kooperationen mit führenden Industriepartnern. Darüber veröffentlicht das MRT regelmäßig Datensätze wie die KITTI Vision Benchmark Suite oder den INTERACTION-Datensatz, die das Forschungsfeld oft maßgeblich prägen.

Aufgaben

Um automatisierte Fahrfunktionen sicher zuzulassen wird eine lückenlose Freigabeargumentation benötigt, die je nach Standard (bspw. Automotive SPICE oder SOTIF) und Fahrfunktion variiert. Die spezifische und detaillierte Erstellung sowie die – in Zeiten von Over-the-Air-Updates und Software-Defined Vehicles normal gewordene – ständige Anpassung von solchen Argumentationen ist umständlich, bislang kaum automatisiert und damit eine der großen Hürden auf dem Weg zu hochautomatisierten Fahrzeugen.
In diesem Projekt, das in enger Zusammenarbeit mit einer deutschen Premium-Automarke erfolgt, soll das sog. Defeater-Konzept auf KI-Agenten angewendet werden um Freigabeargumentationen zu prüfen und zu verbessern. Hierbei wird das ungemeine Allgemeinwissen großer Sprachmodelle ausgenutzt um Lücken in der Argumentation aufzudecken sowie fehlende Aspekte oder Gefahren, menschliche Fehler oder gar logische Fehlschlüsse zu erkennen. Weiterhin sollen die KI-Agenten durch sog. Self-Play gegeneinander antreten um am Ende sich selbst oder die Argumentationen zu verbessern.



Der Aurora Safety Case[1] als beispielhafte Freigabeargumentation, die von einem AI-Defeater Agent angegriffen wird.

Wir bieten

+ Eine wissenschaftlich anspruchsvolle Aufgabe, die Publikationen auf führenden Konferenzen in den Bereichen automatisiertes Fahren, Robotik und künstlicher Intelligenz (NeurIPS, ICLR, IV, ICRA) ermöglicht,
+ Zusammenarbeit in einem leistungsfähigen, international ausgewiesenen Team am MRT,
+ Einbindung in Doktorandencluster beim Industriepartner,
+ Möglichkeit zur Promotion zum Dr.-Ing. am Karlsruher Institut für Technologie (KIT) sowie
+ Vergütung nach einer vollen TV-L E13 Stelle (sofern fachliche und persönliche Voraussetzungen erfüllt sind).

Wir suchen

Engagierte und innovationsfreudige Kollegen (m/w/d) mit
+ einem sehr gut abgeschlossenen Masterstudium in den Bereichen Informatik, Maschinenbau, Elektrotechnik o.ä.,
+ Erfahrung mit der Benutzung und dem Training von tiefen neuronalen Netzen und maschinellen Lernverfahren,
+ Erfahrung im Bereich Softwareentwicklung (Python),
+ Freude am selbständigen, wissenschaftlichen Arbeiten,
+ Teamgeist, der aktiv gelebt wird,
+ Reisebereitschaft zum Industriepartner sowie
+ fließenden Deutsch- und Englischkenntnissen.

Die Tätigkeit verbindet sich mit dem Ziel der Promotion.

Bewerbung

Bei Interesse freuen wir uns auf Ihre Bewerbung mit Lebenslauf, Zeugnissen bzw. Notenauszug sowie optional Codebeispielen oder einer Veröffentlichungsliste an: **Prof. Dr.-Ing. Christoph Stiller** (stiller@kit.edu). Fachliche Auskünfte erteilt Ihnen gerne Herr Jan-Hendrik Pauls (0721 608-43599, pauls@kit.edu).

1: Öffentlich verfügbar unter https://safetycaseframework.aurora.tech/gsn.