

Themenvorschlag für SafeTRANS Industrial Day 2024

## **Quantitative Restrisikobestimmung von KI-basierten Systemen – ist es möglich?**

Im Bereich der hochautomatisierten Systeme spielt das Thema des unvollständigen Wissens eine immer größere werdende Rolle. Durch den Einsatz der Systeme in offenen Betriebskontexten, dem Auftreten von Messungenauigkeiten, dem Einsatz neuer Technologien und der damit zusammenhängenden zunehmenden Komplexität, entstehen Unsicherheiten hinsichtlich des vollständigen Verständnisses des Gefährdungsrisikos und der Effektivität der gewählten Mitigationsmaßnahmen. Der Einsatz von KI zur Realisierung solcher hochautomatisierten Systeme fügt hierbei durch ihr unerklärbares Verhalten eine weitere Dimension der Unsicherheit hinzu.

Standards wie ISO 21448 (Safety of the Intended Functionality) oder auch entstehende Standards wie ISO/AWI PAS 8800 (Safety and Artificial Intelligence) adressieren diese Probleme, lassen den Lösungsraum hinsichtlich der tatsächlich anzuwendenden Methoden aber entsprechend vage. Dieser Spielraum führt nicht selten zu Bedenken hinsichtlich der ausreichenden Auswahl an entsprechenden Methoden zur Erreichung eines akzeptablen Restrisikos. Im akademischen Kontext existieren diverse Lösungsansätze zur Absicherung solcher Systeme, bei denen der Reifegrad jedoch schwer ersichtlich ist. Somit ist es oft nicht einfach zu beurteilen, ob ein bestimmtes Absicherungsverfahren ein noch nicht ausreichend validiertes akademisches Konzept darstellt oder bereits für den industriellen Einsatz ausgereift ist.

Konkret wird im diesem Vortrag ein mögliches Verfahren für die Verifikation und Validierung von KI-basierten Systemen im Kontext von ISO 21448 und der damit einhergehenden quantitative Restrisikobestimmung vorgestellt. Besonderer Fokus wird hierbei auf die statistische Belastbarkeit der Ergebnisse gelegt. Zudem werden für einige interpretationsbedürftige Anforderungen des Standards, wie z.B. der Auswahl eines hinreichenden statistischen Konfidenzniveaus, die Bedeutung von „Performance Targets“ im Kontext von KI oder die Reduzierbarkeit von „Validation Targets“ Lösungen erarbeitet. Darüber hinaus soll die Diskussion angeregt werden für welche Art von System ein solches Güteniveau hinsichtlich der statistischen Belastbarkeit der Restrisikobestimmung überhaupt praktisch realisierbar ist.

Sprecher: Philipp Schleiß forscht seit 2013 bei Fraunhofer an industriell-anwendbaren Methoden zur Erhöhung der Zuverlässigkeit von eingebetteten und autonomen Systemen und leitete dabei erfolgreich eine Vielzahl von Kundenprojekten im Automobilbereich. Im Jahr 2019 begann er mit dem Aufbau einer Abteilung im Bereich „Systems Safety Engineerings“ am Fraunhofer IKS (Institut für Kognitive Systeme), wodurch der Stand der Technik auf dem Gebiet Absicherung von unsicherheitsbehafteten Systemen erheblich vorangetrieben werden konnte. Seine derzeitigen Forschungsinteressen konzentrieren sich auf die werkzeuggestützte Designautomatisierung, die kontinuierliche Absicherung und die Sicherstellung des Echtzeitverhalten für das automatisierte Fahren.