Absicherungsprozess und -toolbox für KI-basierte Komponenten

Mirko Knaak, Christian Kruschel, IAV GmbH

Die Absicherung einer KI-basierten Komponente hat besondere Herausforderungen, weil sie als lernendes System die Sicherheitsanforderungen nicht über spezifische Regeln, sondern über Daten abdecken muss. Darum müssen bisherige Absicherungstechniken sowie Standards für die funktionale Sicherheit auf um die besonderen Entwicklungszyklen beim Maschinellen Lernen erweitert werden.

Obgleich das Thema Zertifizierung von KI-Komponenten in den letzten Jahren eine große Aufmerksamkeit sowohl in der Forschung als auch in der Standardisierung genossen hat, verbleiben zahlreiche offene Punkte. Zu diesen gehört eine mangelnde Verknüpfung zwischen konkreten technische prüfbaren Kriterien auf der einen Seite und den abstrakten Anforderungen auf der anderen Seite. Auf der abstrakten Ebene gibt es bereits viele Vorschläge und potentielle Anforderungen an eine sichere KI. Beispiele sind der Leitfaden zur Gestaltung von vertrauenswürdiger KI oder verschiedene Spezifikationen für Normen wie DIN SPEC 92001. Für konkrete Prüfkriterien gibt es ebenfalls gute Forschungsergebnisse und erste Toolboxen für den Nachweis, dass auf der konkreten Detailebene bestimmte Ziele der KI erreicht werden.

Was jedoch fehlt, ist ein strukturiertes Vorgehen, wie man aus diesen Einzelnachweisen eine Argumentation erstellt, dass die auf der oberen Ebene erstellten Zielvorgaben tatsächlich erreicht werden. Diese Zusammenstellung ist aktuell noch manuell und einzelfallbasiert.

Im Vortrag wird darum ein strukturiertes Vorgehen entlang des Lebenszyklus einer KI vorgeschlagen. Darauf basierend wird eine Toolbox für die partielle Automatisierung dieses Prozesses und der finalen Abprüfung vorgestellt. Den Abschluss bildet eine Anwendung auf ein reales Beispiel für eine KI auf Zeitreihenbasis.