Springer Vieweg

ATZ extra

BIG DATA I KI

KI in sicherheitskritischen Automobilanwendungen

itk ENGINEERING

KI in sicherheitskritischen Automobilanwendungen

Künstliche Intelligenz (KI) hebt die Entwicklung von Fahrzeugen und ihren Funktionen auf ein neues Level. ITK Engineering widmet sich der Frage, mit welchen Analysemethoden und systematischen Ansätzen die notwendigen Sicherheitsaspekte von Systemen maschinellen Lernens zur Einhaltung der ISO PAS 8800 gewährleistet werden können.

Die rasanten Fortschritte im Bereich der künstlichen Intelligenz eröffnen einerseits weitreichende Möglichkeiten zur Entwicklung innovativer Fahrzeugfunktionen. Andererseits stellt sich vor dem Hintergrund des Innovationsdrucks in der Fahrzeugbranche, von Unfällen im Zusammenhang mit automatisierten Fahrzeugen [1] und der voranschreitenden Regulierung [2] die Frage, wie ein standardisiertes Vorgehen zur Entwicklung sicherheitsrelevanter KI-Systeme im Fahrzeug aussehen kann. Die Spezifikation ISO PAS 8800 wird einen solchen Rahmen für die Entwicklung dieser Systeme im Fahrzeug vorgeben und damit Orientierung für alle Hersteller und deren Zulieferer bieten, die KI-Systeme für Fahrzeugfunktionen entwickeln [3].

HERAUSFORDERUNGEN SICHERER KI-SYSTEME

KI-Systeme kommen meist zur Lösung sehr herausfordernder Problemstellungen zum Einsatz. So finden diese im Fahrzeug allen voran im Bereich der automatisierten Fahrfunktionen Anwendung. Von besonderer Bedeutung sind dabei Deep-Learning-Methoden, die es ermöglichen, komplexe Funktionen – etwa zur Objekterkennung – unter Nutzung großer Datenmengen algorithmisch zu erlernen. Durch den dabei vorliegenden Open-World-Kontext muss ein Vorgehen gewählt werden, das eine kontinuierliche Entwicklung und Sicherheitsbewertung von KI-Systemen im Fahrzeug zulässt. Genauso bringen die speziellen Eigenschaften von KI-Methoden weitreichende Herausforderungen mit sich. Beispielsweise schränken zum einen die Nicht-Linearität und die hohe Komplexität von Deep Neural Networks (DNNs) deren Nachvollziehbarkeit stark ein, und zum anderen limitieren Grenzen in der Generalisierungsfähigkeit und Robustheit von DNNs deren Funktion. Da das Verhalten von KI-Modellen auf Basis von maschinellem Lernen (ML) implizit durch die

VERFASST VON

Dr. Stefan Held
ist Lead Engineer bei ITK Engineering in Holzkirchen bei München.

Andreas Bossert
ist Fachreferent Verifikation und Validierung bei ITK Engineering in Holzkirchen bei München.

Dr. Frank Lenzen
ist Expert Engineer Funktionale Sicherheit bei ITK Engineering in Rülzheim.

Dr. Ulrich Sutter
ist Senior Manager Funktionale Sicherheit bei ITK Engineering in Rülzheim.

verwendeten Daten spezifiziert und bestimmt wird, müssen datenbezogenen Arbeitsschritte (Datensammlung, Annotation und Datensatz-Aufbau) eine besondere Betrachtung finden.

SICHERHEITSARGUMENTATION FÜR KI-SYSTEME

Kernprozess jeder Entwicklung gemäß einer Sicherheitsnorm ist die Erarbeitung einer Argumentation, die eine ausreichende Sicherheit des Produkts belegen soll (im Folgenden: Assurance Argument). Sicherheit bedeutet in diesem Zusammenhang, dass das Risiko eines Personenschadens durch einen Fehler im System beziehungsweise durch einen vorhersehbaren Fehlgebrauch durch Nutzer auf ein akzeptables Maß reduziert wurde. Dabei soll das Assurance Argument einer systematischen Struktur folgen [4] und kann beispielsweise mit einer Zielstrukturierungs-Notation (Goal Structuring Notation, GSN) schematisch dargestellt werden.

Speziell für KI-Systeme muss das Assurance Argument Bezug auf die besonderen Herausforderungen von KI nehmen. Ausgangspunkt sind dabei unter anderem die abgeleiteten Sicherheitsanforderungen an das KI-System, der Kontext, in dem die KI zum Gesamtsystem steht, und der Eingaberaum. Letzterer beschreibt die möglichen Eingabewerte des KI-Systems und steht in Zusammenhang mit der Definition eines exakten Einsatzbereichs (Operational Design Domain, ODD). Erst eine ausreichend detaillierte ODD-Spezifikation erlaubt es, ein KI-System auf mögliche funktionale Unzulänglichkeiten innerhalb des Open-World-Kontexts zu untersuchen [5]. Insbesondere muss dabei eine ausreichende Leistungsfähigkeit des KI-Systems auf dem Eingaberaum belegt werden, sowohl qualitativ als auch quantitativ. Neben systematischen Analysen sind auch statistische Betrachtungen notwendig, wie zum Beispiel das Abschätzen der Auftretenswahrscheinlichkeit. Je größer die Dimensionalität des Eingaberaums ist, desto größer ist die Herausforderung einer ausreichenden Abdeckung. Hier muss projektspezifisch die richtige Herangehensweise gewählt werden. Es ist zu zeigen, dass sowohl die Vielzahl von Standardsituationen als auch das Auftreten seltener, aber kritischer Situationen abgedeckt sind, und dass die Wahrscheinlichkeit von unbekannten Auslösebedingungen, sogenannter Triggering Conditions (siehe ISO 21448), für funktionale Unzulänglichkeiten ausreichend klein ist.

KI-SICHERHEITSLEBENSZYKLUS

Um sicherheitsrelevante Aktivitäten einer KI-Systementwicklung im Kontext eines Gesamtsystems abzubilden, wird nach [3] ein Sicherheitslebenszyklus definiert. **BILD 1** zeigt alle Aktivitäten eines beispielhaften Sicherheitslebenszyklus, die zur kontinuierlichen Erbringung des Assurance Arguments beitragen sollen. Allgemein werden von den übergeordneten Systemanforderungen entsprechende Sicherheitsanforderungen an das KI-System abgeleitet. Diese Anforderungen bilden die Eingabe für eine Entwicklung und Absicherung des KI-Systems. So werden Datensätze zur Entwicklung ML-basierter KI-Systeme typischerweise kontinuierlich und begleitend zur iterativen Umsetzung der KI-Komponente aufgebaut. Hinzu kommt, dass häufig erst im Verlauf der Entwicklung evident wird, welcher algorithmische Ansatz und welche Methoden die gegebenen KI-Anforderungen erfüllen können. Daher müssen auch die Sicherheitsanalysen kontinuierlich und iterativ als Teil der KI-Systementwicklung erfolgen. Meist kann aufgrund der zeitveränderlichen, hochkomplexen ODD und der spezifischen Eigenschaften der KI-Systeme nicht garantiert werden, dass KI-Systeme sich im Betrieb dauerhaft fehlerfrei verhalten. Deshalb muss das KI-System mit entsprechenden Monitoring-Systemen fortlaufend überwacht und geprüft werden, ob das Assurance Argument für das KI-System weiterhin

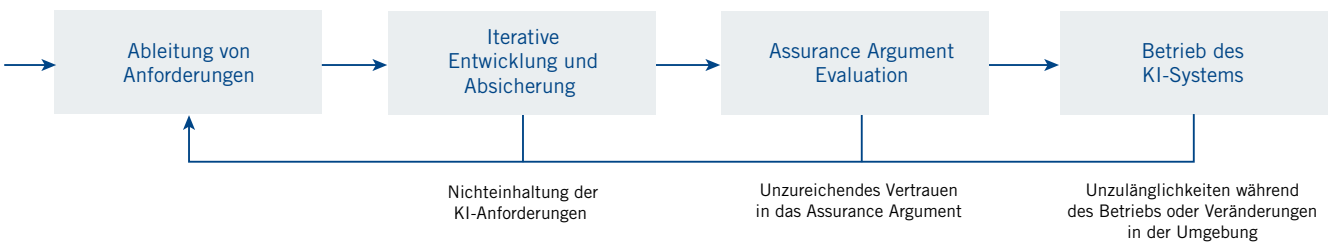

BILD 1 Beispielhafter AI-Datenlebenszyklus – geänderte Darstellung nach [3] (© ITK Engineering GmbH)

Gültigkeit besitzt (Continuous Assurance). Ist dies nicht der Fall, müssen das KI-System angepasst, die Anforderungen an das KI-System geändert oder entsprechende Maßnahmen auf übergeordneter Systemebene getroffen werden, um die Gültigkeit des Assurance Arguments herzustellen.

DATENLEBENSZYKLUS

Das Verhalten von KI-Systemen auf Basis von ML wird maßgeblich von den Eigenschaften der zur Entwicklung verwendeten Daten bestimmt. Diese ändern sich während der Entwicklung und über die Produktlebensdauer aufgrund neuer Erkenntnisse und einer sich verändernden Umwelt. Daher ist es sinnvoll, Anforderungen an einen Datenlebenszyklus [3] zu stellen, um zu gewährleisten, dass für die KI-Entwicklung aktuelle und konsistente Daten verwendet werden. Damit bildet der Datenlebenszyklus die Grundlage zum Aufbau eines validen Assurance Arguments.

Der in **BILD 2** dargestellte Datenlebenszyklus beginnt typischerweise mit der Datensicherheitsanalyse. Diese zielt darauf ab, potenziell sicherheitsrelevante Defizite zu identifizieren, geeignete Gegenmaßnahmen zu entwickeln und Messgrößen zur Bewertung der Vermeidung zu definieren. Die ermittelten Unzulänglichkeiten, Ursachen und Auswirkungen dienen als Grundlage für die Phasen Datensatzanforderungsentwicklung, dem Datensatzdesign und der Datensatzimplementierung.

In der Datensatzanforderungsentwicklung erfolgt die Formulierung der Anforderungen an die Datensätze und die Ableitung der Anforderungen an die Qualitätssicherung. Dies geschieht unter der Annahme, dass die in ISO 26262-3 für die Item Definition und die in ISO 21448 Abschnitt 7 für die Triggering Conditions angegebene Methode befolgt wurde. Das Item ist dabei ein System oder eine Kombination von Systemen.

Die Datensatzanforderungen stellen die Grundlage für die Designphase dar, die sich mit der Art und Weise der Erstellung beschäftigt. Dazu gehören Aspekte wie die Zusammenstellung der Daten aus physikalisch, synthetisch [6] oder augmentierten Daten, nötige Vorverarbeitungsschritte und Metadaten. Vorbereitung, die Definition der Prozesse und Methoden zum Annotieren und das eigentliche Annotieren werden in der Phase der Datensatzimplementierung durchgeführt. Die Konsistenz und Korrektheit in den Datensätzen und die Einhaltung der Anforderungen an die Datensätze betrifft die Datensatzverifikation. In der Phase der Datensatzvalidierung wird die Anforderungskonformität geprüft, was bedeutet, dass die abgeleiteten Anforderungen an die Datensätze den Erwartungen entsprechen. Die Wartung der Datensätze stellt sicher, dass mit aktuellen und konformen Datensätzen gearbeitet wird.

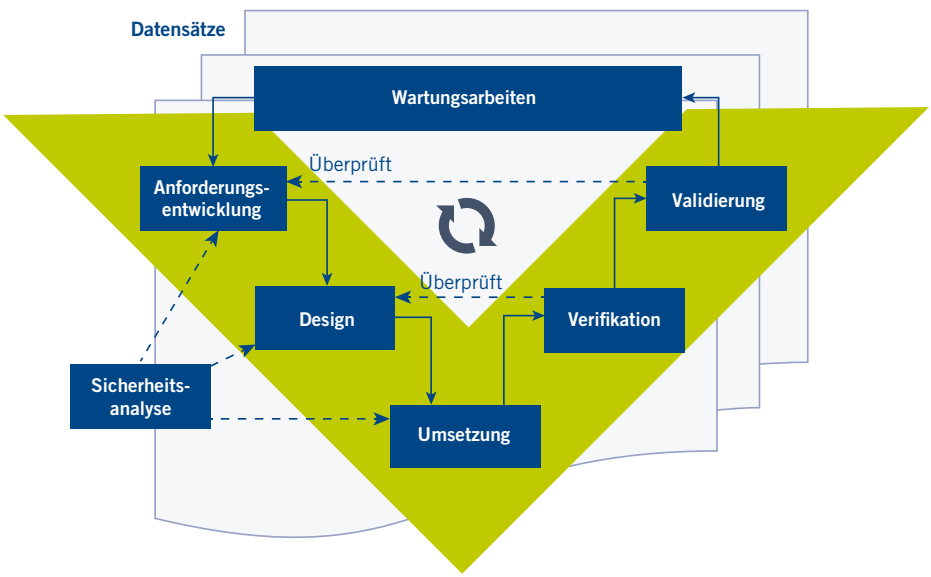

BILD 2 Phasen des Datenlebenszyklus (© ITK Engineering GmbH)

ABLEITUNG VON KI-ANFORDERUNGEN

Die Anforderungen, die vom Gesamtsystem für das KI-System hergeleitet werden, sind zumeist nicht spezifisch genug, um aus deren Nichteinhaltung direkte Maßnahmen zu definieren. Um Risiken während des KI-Entwicklungsprozesses zu minimieren, ist es sinnvoll, Einflussfaktoren zu definieren, die die Formulierung qualitativer Anforderungen ermöglichen. Ein möglicher Ansatzpunkt sind die in [7] genannten Sicherheitsbelange. Ein Teil dieser Sicherheitsbelange sind messbare Eigenschaften des trainierten KI-Modells, der Daten oder Entwicklungsprozesse, wie zum Beispiel Robustheit, Generalisierungsfähigkeit und Erklärbarkeit, und können so einen quantitativen Ansatz zur Verfeinerung von KI-Sicherheitsanforderungen bieten.

Um KI-spezifische Sicherheitsanforderungen zu verfeinern, werden mittels Sicherheitsanalysen jene Eigenschaften identifiziert und bewertet, die maximal mit der Verletzung einer KI-Anforderung korrelieren. Viele dieser Eigenschaften sind entweder direkt oder durch korrelierte Messungen bestimmbar. Je nach Anwendung und Modell gibt es hierfür verschiedene Methoden, deren Eignung für den Anwendungsfall bewertet und die implementiert werden müssen. Auf Grundlage der erfolgten Bewertung können konkrete Maßnahmen abgeleitet werden, die entweder die Entwicklung oder die Architektur des KI-Systems betreffen.

FAZIT UND AUSBLICK

Neue Normen und Spezifikationen wie die ISO PAS 8800 werden in der Automotivebranche hinsichtlich Sicherheit und KI auf den neuen Stand der Technik wirken. Dabei sind spezielle Anforderungen und Prozesse zu erfüllen (zum Beispiel für den KI- und den Datenlebenszyklus), die Änderungen der Sicherheitsanforderungen über die Zeit und nach dem Produktionsbeginn vorsehen. Da Daten zum Trainieren von ML-Systemen unerlässlich sind und somit Spezifikationen und Anforderungen in diesem Bereich teilweise ersetzen, wird künftig ein systematischer Ansatz in Form eines Datenlebenszyklus notwendig sein. Des Weiteren werden zusätzliche Sicherheitsanalyse-Methoden benötigt (zum Beispiel GSN und STPA), die eine holistische Betrachtung

BILD 3 Sicherheits- und KI-spezifische Herausforderungen (© ITK Engineering GmbH)

der Sicherheitsaspekte ermöglichen sowie den speziellen Eigenschaften von KI-Methoden Rechnung tragen sollen.

Künftige Prüfungen und Bewertungen bezüglich Sicherheit und KI werden eine große Herausforderung darstellen, da das Kompetenzprofil von Sicherheitsexperten hinsichtlich KI erweitert oder KI-Expertise unterstützend hinzugezogen werden muss. Zusätzlich stellen kontinuierliche Entwicklung und kontinuierliche Gewährleistung aus Sicherheitssicht einen Paradigmenwechsel dar, für den bisher wenig Erfahrungswerte vorliegen.

Insofern ist KI für sicherheitsrelevante Automobilanwendungen Fluch und Segen zugleich. Zum einen können mit KI leistungsfähige Lösungen für bestehende Problemstellungen, wie etwa die Umfelderkennung, umgesetzt werden. Zum anderen bringt die Entwicklung und der Betrieb sicherer KI-Systeme große Herausforderungen mit sich.

Eine tiefgehende Expertise von Fachspezialisten in den Bereichen Sicherheit, KI sowie Verifizierung und Validierung ist die Grundlage, um sicherheits- und KI-spezifische Herausforderungen zu lösen, **BILD 3**.

LITERATURHINWEISE

[1] National Transportation Safety Board: Automated Vehicles – Investigations. Online: https://www.ntsb.gov/Advocacy/safety-topics/Pages/automated-vehicles-investigations.aspx, aufgerufen: 15. Mai 2024
[2] EUR-Lex: Document 52021PC0206. Online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206, aufgerufen: 15. Mai 2024
[3] Burton, S.: Standardisation of safe, data-driven AI Development & Tooling. KI Data Tooling Final Event. Online: https://www.ki-datatooling.de/fileadmin/KI_DataTooling/Downloads/Final_Results/FE_Presentations/20231205_KI_Data_Burton.pdf, aufgerufen: 15. Mai 2024
[4] Klaes, M. et al.: Using Complementary Risk Acceptance Criteria to Structure Assurance Cases for Safety-Critical AI Components. Proceedings of the Workshop on Artificial Intelligence Safety, 2021, Vol-2916, Online: https://ceur-ws.org/Vol-2916/paper_9.pdf, aufgerufen: 15. Mai 2024
[5] ISO 34503: Road Vehicles – Test scenarios for automated driving systems – Specification for operational design domain
[6] ITK Engineering GmbH: Individual Virtual Environment and Sensor Simulation (iVESS). Online: https://www.itk-engineering.de/stories/individual-virtual-environment-and-sensor-simulation-ivess/, aufgerufen: 15. Mai 2024
[7] Abrecht, S. et al.: Deep Learning Safety Concerns in Automated Driving Perception. Online: https://arxiv.org/pdf/2309.03774v1, aufgerufen: 15. Mai 2024

DANKE

Die Autoren bedanken sich bei Philipp Leopold, KI-Experte bei ITK Engineering, für die Unterstützung bei der Erstellung dieses Beitrags.

IMPRESSUM
Sonderausgabe 2024 in Kooperation mit ITK Engineering GmbH, Bergfeldstraße 2, 83607 Holzkirchen; Springer Fachmedien Wiesbaden GmbH, Postfach 1546, 65173 Wiesbaden, Amtsgericht Wiesbaden, HRB 9754, USt-IdNr. DE81148419

GESCHÄFTSFÜHRER:
Stefanie Burgmaier I Andreas Funk I Joachim Krieger
PROJEKTMANAGEMENT: Anja Trabusch

TITELBILD: © istockphoto I metamorworks

ITK Engineering

Seit der Firmengründung 1994 stehen wir für Stabilität, Sicherheit und Methodenexpertise. Damals wie heute bildet branchenübergreifendes Spezialwissen insbesondere im Bereich der Regelungstechnik und der modellbasierten Entwicklung die Basis, um unsere Kunden von der Idee bis zur Serienproduktion durchgängig und partnerschaftlich zu begleiten.

Unsere Kompetenzen umfassen u.a.:

- Softwareentwicklung
- Hardwareentwicklung
- Elektrik/Elektronik
- Systemintegration
- Software als Produkt
- Komplettlösungen
- Auftragsentwicklung
- Technische Beratung
- Schulungen
- Qualitätssicherung

Die Zufriedenheit all unserer Partner und ein respektvolles Miteinander prägen unsere Unternehmensphilosophie, in der vier Werte fest verankert sind: Lesen Sie gerne mehr darüber im Web.

V.1.0.0_d_2021

ITK Engineering GmbH
Hauptsitz Rülzheim
Im Speyerer Tal 6
76761 Rülzheim
Tel.: +49 (0)7272 7703-0
Fax: +49 (0)7272 7703-100
info@itk-engineering.de

www.itk-engineering.de

Gegründet 1994 –
heute hat ITK deutschlandweit Niederlassungen und ist international vertreten.

Folgen Sie uns auch auf: