| | |
|---|---|
| Projekttitel: Prüfverfahren für KI-basierte Komponenten im Eisenbahnbetrieb (KI-LOK)<br>Autoren: Michael Leuschel, Jan Gruteser, Jan Roßbach<br>Version: 1.0<br>Datum: 17.12.2024 | hhu Heinrich Heine Universität Düsseldorf |

# KI-LOK

## Prüfverfahren für KI-basierte Komponenten im Eisenbahnbetrieb

**Abschlussbericht**
**Heinrich-Heine-Universität Düsseldorf**
**17. Dezember 2024**

**Prof. Dr. Michael Leuschel, Jan Gruteser, Jan Roßbach**
**Technischer Bericht HHU-STUPS 2024-12-1**

Projektpartner

- ITPower Solutions GmbH
- neurocat GmbH
- Hitachi Rail (GTS Deutschland GmbH) - ehemals Thales Deutschland GmbH
- Fraunhofer Institut für Offene Kommunikationssysteme FOKUS
- Heinrich-Heine-Universität Düsseldorf

Projektlaufzeit: 01.04.2021 – 30.09.2024

Finanziert von der Europäischen Union
NextGenerationEU

Gefördert durch:

Bundesministerium für Wirtschaft und Klimaschutz

aufgrund eines Beschlusses des Deutschen Bundestages

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Wirtschaft und Klimaschutz unter den Förderkennzeichen 19I21007A gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Projektpartnern.

# Inhalt

## I.1 Aufgabenstellung

Im KI-LOK-Projekt sollen neue Testverfahren und Methoden zur Absicherung und Zertifizierung von KI-gestützten Technologien für sicherheitskritische Anwendungen in der Bahntechnik entwickelt werden.

## I.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Der Entwurf und Betrieb innovativer Fahrzeuge im schienengebundenen Verkehr fordert immer stärker den Einsatz neuartiger KI-basierter, autonomer, lernender Systeme. Das Verhalten KI-basierter Systeme ist jedoch, im Gegensatz zu bisherigen Systemen, nicht deterministisch nachvollziehbar. Dies erfordert gegenüber dem heutigen Stand der Technik adäquate, angepasste und intelligente Test- und Prüfsysteme, um die dynamischen Wechselwirkungen mit der realen Umwelt abzubilden und die Systeme auf ihre Intelligenz, Verlässlichkeit und Robustheit zu prüfen.
Im Projekt wurden daher Techniken für die Validierung und Verifikation KI-basierter Systeme der Bahntechnik sowie modellbasierte Methoden zum Test von autonomen Zugsystemen erforscht. Als Fallstudien dienen dabei ein Lokführerassistenzsystem zur Objekterkennung im Lichtraumprofil des Fahrweges, und ein visuelles Odometrie und Positionssystem. Auf Basis dieser Ergebnisse sollen Prozesse und Werkzeugketten definiert werden, mit denen in Zusammenarbeit mit den nationalen und europäischen Genehmigungsbehörden eine Zulassung KI-basierter Zugsysteme zum Betrieb erreicht werden kann.

## I.3 Planung und Ablauf des Vorhabens

Im Projekt werden schwerpunktmäßig vier F&E-Aspekte betrachtet, die jeweils eines der oben genannten Teilziele unterstützen: Definition der Anforderungen aus der Bahntechnik (AP1), Verifikation und Validierung von KI (AP2), Entwicklung modellbasierter Techniken für den Test KI-basierter Bahntechnikkomponenten (AP3) und Methodik der Absicherung (AP4). So gliedert sich das Projekt in die folgenden fünf Hauptarbeitspakete:

- *(AP 1)* Das HAP1 umfasst die Definition der Fallstudien, die Erhebung ihrer Anforderung an die Techniken und Methoden, die im Projekt erarbeitet werden (sowohl in wissenschaftlicher als auch in technischer Hinsicht), die Integration der Projektergebnisse und dessen Evaluation entlang der Fallstudien.

- *(AP 2)* Das HAP2 beschäftigt sich mit dem Schwerpunkt (S1), d.h. der Entwicklung von Validierungstechniken für ML, und erarbeitet Techniken, Methoden und Werkzeuge zur Prüfung von Modellqualität, Datenqualität sowie Testtechniken für ML.

- *(AP 3)* Gegenstand der Arbeiten von HAP3 ist Schwerpunkt (S2), d.h. die Erforschung von Validierungstechniken für AS, die ML-Komponenten enthalten können. Arbeitsschwerpunkte sind Testtechniken, die selber wiederum KI-Technologie bzw. andere such- und datengetriebene Optimierungsverfahren

verwenden sowie risikobasierte Testverfahren mit dem Ziel, Abnahme, Zulassung und Zertifizierung von AS abzudecken.

- *(AP 4)* Das HAP4 integriert die Ergebnisse des AP2 und AP3 durch die Bereitstellung einer umfassenden Methodik für die Verifikation und Validierung sowie einer Experimentierplattform, die der Integration der in AP2 und AP3 entwickelten Werkzeuge dient, und zu Lehrzwecken mit entsprechenden Schulungsangeboten angereichert wird. Damit wird Schwerpunkt (S3) adressiert.

- *(AP 5)* Das HAP5 beschäftigt sich mit den projektübergreifenden Aktivitäten. Hierzu zählt die die Koordination des Forschungsvorhabens sowie die Verbreitung der Projektergebnisse und ihre Verwertung. Es deckt dabei die wissenschaftlichen, technischen Resultate und die Ergebnisse der Evaluation ab und organisiert den Transfer der Projekterkenntnisse in Standardisierungs- und Normungsgremien.

## I.4 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

### Bekannte Konstruktionen, Verfahren und Schutzrechte, die für die Durchführung des Vorhabens benutzt wurden

Für das Projekt wurden formale Modelle für Bahnsysteme in B und Event-B entwickelt. Diese Verfahren sind im Bahnbereich bekannt, wurden hier aber für KI-Komponenten weiterentwickelt. Viele Arbeiten der HHU basieren auch auf dem ProB Validierungswerkzeug, welches für viele industrielle Anwendungen im Bahnbereich eingesetzt wird. Das Werkzeug wurde im Projekt um relevante Fähigkeiten erweitert, zum Beispiel die Fähigkeit KI-Komponenten mit einem formalen Modell zu verknüpfen. Formale Methoden können aber auch zur Systemanalyse verwendet werden. Hier wird nicht nur eine Softwarekomponente formal spezifiziert, sondern ein gesamtes System samt Umgebungsmodell. Die von Abrial entwickelte Event-B Methode[1] ist eine beliebte und ausdrucksstarke Möglichkeit diese Art der Modellierung zu betreiben, mit der mehrere Fallstudien im Fahrzeugbereich durchgeführt wurden. Innerhalb von Shift2Rail wurde ein existierendes vollautomatisches Zugsystem modelliert und damit ein fundiertes Sicherheitskonzept abgeleitet[2]. Im Rahmen eines Forschungsprojektes haben die HHU und Thales Aspekte des von Thales entwickelten RBC (Radio Block Centre) und der neuen ETCS Hybrid Level 3 Spezifikation modelliert[3]. Diese Arbeiten führten zur Aufdeckung von über 40 Fehlern in der HL3 Spezifikation der ERTMS Users Group und zu mehreren Demonstratoren, bei denen das formale mathematische Modell in Echtzeit ausgeführt wurde. Ein Demonstrator wurde bei der InnoTrans 2018

[1] J.-R. Abrial. Modeling in Event-B: System and Software Engineering. Cambridge University Press, 2010.
[2] Mathieu Comptier, Michael Leuschel, Luis-Fernando Mejia, Julien Molinero Perez, Mareike Mutz: Property-Based Modelling and Validation of a CBTC Zone Controller in Event-B. RSSRail 2019: 202-212
[3] Dominik Hansen, Michael Leuschel, David Schneider, Sebastian Krings, Philipp Körner, Thomas Naulin, Nader Nayeri, Frank Skowron: Using a Formal B Model at Runtime in a Demonstration of the ETCS Hybrid Level 3 Concept with Real Trains. ABZ 2018: 292-306

vorgeführt, ein anderer im DB Living Lab[4]. Die Event-B Methode wird von Werkzeugen wie Rodin, Atelier-B und ProB unterstützt. Diese ermöglichen eine formale Beweisführung, aber auch die Animation und ausgiebige Simulation des Systems.
Im KI-LOK Projekt wurde an diese Erfahrung angeknüpft und mehrere Ablauf- und Umgebungsmodelle in Event-B entwickelt. Mit den technologischen Neuentwicklungen von KI-LOK kann mit diesen Modellen zum Beispiel die Fehlerwahrscheinlichkeiten von KI-basierten Systemen abzuschätzen.

**Verwendete Fachliteratur sowie benutzte Informations- und Dokumentationsdienste**

Ein Großteil der Literatur wurde über die Zugänge der Universität bezogen. Die wichtigsten Herausgeber waren dabei: Springer-Verlag (insbesondere LNCS Online), ACM, Elsevier und IEEE.

## I.5 Zusammenarbeit mit anderen Stellen

Die HHU hat die geplanten Arbeitspakete durch Einsatz eigener Kapazitäten durchgeführt, es sind keine Unteraufträge vergeben worden. Die mittel-neutrale Zusammenarbeit mit Dritten (Behörden, Standardisierungsgremien, wissenschaftlichen Organisationen) wie zum Beispiel dem DZSF (Deutsches Zentrum für Schienenverkehrsforschung) ist im Kapitel 4 (Verwertungsplan) beschrieben.

[4] https://www.youtube.com/watch?v=FjKnugbmrP4

## II.1 Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen, mit Gegenüberstellung der vorgegebenen Ziele

Die Zuwendung wurde verwendet, um die geplanten Arbeiten an diesen Arbeitspaketen durchzuführen:

- AP1: Fallstudien
- AP2: Entwicklung von Validierungstechniken für ML, und Erarbeitung Techniken, Methoden und Werkzeuge zur Prüfung von Modellqualität, Datenqualität sowie Testtechniken für ML
- AP3: Validierungstechniken für AS, die ML-Komponenten enthalten können
- AP4: Methodik für die Verifikation und Validierung sowie einer Experimentierplattform
- AP5: Projektübergreifende Aktivitäten: Verbreitung der Projektergebnisse und ihre Verwertung

Hierbei hat HHU das AP4 geleitet.

Im AP1 hat die HHU Anforderungen für die Fallstudien identifiziert, damit die späteren Aufgaben des Projekts (Erstellung von formalen Systemmodellen, Datengenerierung mit Simulation, formale Absicherungsmethodik) durchgeführt werden können.
In AP2 hat die HHU formale Modelle zur Überwachung des Lernens von ML-Systemen und Verifikationsstrategien erstellt. Hierbei wurden verschiedene Ansätze eingesetzt: „Certifying Control“ und von Sicherheitsbedingungen abgeleitete „Shields”, um das Lernen zu verbessern oder KI-Fehler zur Laufzeit aufzudecken. Die HHU hat domänenspezifische Visualisierungen erstellt, um KI-Entscheidungen für Domänenexperten einleuchtend darzustellen.
In AP3 erstellte die HHU formale Umgebungsmodelle in der B Sprache für die Fallstudien. Bei der Erstellung der Modelle konnte auf die Erfahrung mit Hybrid-Level 3 zurückgegriffen werden. Mit der Simulation dieser Modelle wurden Testszenarien erstellt. Die Modelle wurden zur automatisierten Ausführung von Tests mit simulierter oder echter KI verwendet. Die Umgebungsmodelle wurden in AP4 mit den Formalisierungen der Sicherheitsanforderungen verknüpft. In AP4 erfolgte auch die Identifikation und Formalisierung von Gefährdungen sowie eine formale Analyse der B Systemmodelle. In AP4 wurde eine Absicherungsmethodik basierend auf klassischen formalen Methoden und neuartigen Techniken aus AP2 und AP3 entwickelt (und veröffentlicht). Schlussendlich, erfolgte in AP4 die Entwicklung von Teilen der Werkzeugkette und Integration der von der HHU entwickelten Werkzeuge. Ein Demonstrator für die Absicherungsmethodik und Werkzeugkette für die Fallstudien wurde erstellt und auf der InnoTrans 2024 Messe vorgeführt. Der Demonstrator beinhaltete die formalen B Modelle mit ProB als Ausführungsmotor inklusive domänenspezifischer Visualisierungen. In AP5 hat die HHU wissenschaftliche Artikel veröffentlicht und Vorträge gehalten, um die wissenschaftlichen Errungenschaften des Projekts zu verbreiten.

Nach Abschluss des KI-LOK Projekts liegt zusätzlich zum Demonstrator jetzt eine werkzeuggestützte Methode zur Simulation und Validierung von formalen Systemmodellen mit KI-Komponenten vor. In diesem Rahmen wurde das

Validierungswerkzeugs *ProB* a) zur Erstellung von KI-Trainingsdaten anhand formaler Systemmodelle und b) zur Simulation und Validierung von Systemmodellen mit KI-Komponenten erweitert.

Die Zuwendung wurde zur Präsentation der Ergebnisse auf zahlreichen internationalen Konferenzen verwendet:

- *Reliability, Safety and Security of Railway Systems (RSSRail) 2023*
  Das ist für das KI-LOK Projekt eine wichtige Konferenz; sie vereint Forscher und Industriepersonen im Bereich von Sicherheit und Informatik im Bahnbereich. Der vorgestellte KI-LOK Artikel wurde jetzt schon mehrfach zitiert (obwohl im Oktober 2023 erschienen) und wir haben im Anschluss den Ansatz mehreren Firmen in Einzelgesprächen vorgestellt.
- *Formal Methods for Autonomous Systems (FMAS) 2023/2024*
  Dies ist eine neue Reihe zu formalen Methoden und autonomen System, gleichzeitig mit der "integrated Formal Methods" Konferenz. Hier wurden unsere KI-LOK Ergebnisse zu Certified Control (2023) und zur KI-gesteuerten Simulation des formalen Modells (2024) vorgetragen.
  Der Beitrag bei der FMAS verbreitet die KI-LOK Forschungsergebnisse auch in anderen Industriesektoren (autonome Fahrzeuge). Auf Grundlage des Vortrages zu Certified Control wurden wir auch für eine "Special Issue" der internationalen Zeitschrift "Science of Computer Programming" eingeladen. Dort haben wir eine erweiterte Version der KI-LOK-Forschungsergebnisse zu Certified Control eingereicht.
- *Formal Methods for Industrial Critical Systems (FMICS) 2022*
  FMICS ist eine Konferenz, die sich auf die Industrieanwendungen der formalen Methoden konzentriert. Sie ist nicht auf den Bahnsektor eingeschränkt. Hier wurden auch KI-LOK-Forschungsergebnisse, diesmal auf Ebene der grundlegenden Methodik zur Visualisierung von formalen Modellen für Domänenexperten, zusammen mit neuen Toolentwicklungen vorgestellt.
  Wie bei FMAS wurden wir dank des Beitrags zu einer "Special Issue" eingeladen. Dieser erweiterte Zeitschriftenartikel wurde eingereicht und in der Zeitschrift "International Journal on Software and Tools for Technology Transfer" veröffentlicht (Int. J. Softw. Tools Technol. Transf. 26(2): 147-168 (2024)). Ohne die Reise wäre die Dissemination der Projektergebnisse und deren zusätzliche Verbreitung (in Form eines Zeitschriftenartikels) nicht möglich gewesen.
- *International Conference on Engineering of Complex Computer Systems (ICECCS) 2024*
  Präsentation des Artikels „Validation of railML using ProB“ (flexible Eisenbahn-Topologiedaten für formale Modelle). Wie bei FMICS wurden hier auch grundlegende Forschungsergebnisse für eine breitere Öffentlichkeit dargestellt. railML ist ein internationaler Standard zum Datenaustausch im Bahnsektor. Diese Aktivitäten sind wichtig für den KI-LOK Demonstrator und der geplanten Vorstellung bei Innotrans 2024. Wir sind diesbezüglich auch in Kontakt mit der internationalen Organisation "railML.org", um das ProB Werkzeug und die railML Anbindung zertifizieren zu lassen.

- *16th NASA Formal Methods Symposium (NFM) 2024*
  Präsentation des Artikels "Validation of Reinforcement Learning Agents and Safety Shields with ProB". Hier wurden Arbeiten vorgestellt, die es ermöglichen eine "echte" KI mit dem Validierungswerkzeug ProB zu kombinieren. Auch dieser Vortrag hat Interesse erweckt. Ähnlich wie bei FMAS dient dieser Vortrag der Dissemination über den Bahnsektor hinaus, um auf globaler Ebene die KI-LOK Arbeiten sichtbar zu machen.

Andere wichtige Ergebnisse der HHU im KI-LOK Projekt:
• Experimentierplattform und Demonstrator auf der internationalen Fachmesse Innotrans 2024 (siehe unten)
• formales B Modell zur Einbindung von KI-Komponenten mit deterministischen Steuersystemen (siehe unten)

## Detaillierte Beschreibung der Projektergebnisse

### Anforderungsanalysen

Aus Sicht der HHU sind die Anforderungsanalysen in UAP 1.1, UAP 2.1, 3.1 und 4.1 verwoben. Wir präsentieren hier die Ergebnisse gebündelt. Die HHU hat sich mit den Grenzen der Formalisierung und der Erstellung von formalen Modellen beschäftigt, besonders im Hinblick auf die Absicherungsmethodik (AP4).

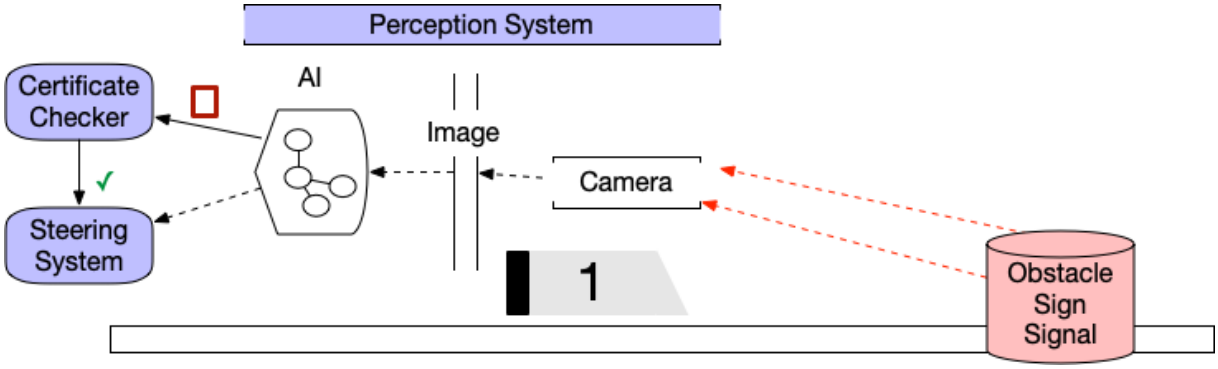



*Abb. 1: Fallstudie 1, Hinderniserkennung*

Für den Kern der ersten Fallstudien von AP1 (Hinderniserkennung, vgl. Abb. 1) ist die Anwendung klassischer formaler Methoden nicht vielversprechend. Zum einen besteht die Schwierigkeit die Korrektheit und „Ground Truth“ formal zu fassen, zum anderen die Schwierigkeit nicht triviale Eigenschaften von komplexen neuronalen Netzen formal abzuleiten. Die Korrektheit von Hinderniserkennung lässt sich formal mathematisch nicht beschreiben.

Schlussfolgerung der Analysen war, dass die HHU ein formales Modell nur für die äußere Schale des Gesamtsystems inklusive nachgelagerter Bremsentscheidung entwickeln wird. Man kann damit einen Teil der möglichen Fehlerquellen der Fallstudie

bändigen: Wir konzentrieren uns im Nachweis auf das Analyse-, Decision- und Steering-System (und führen eine betriebliche Gefährdungsanalyse durch). Die Zertifizierung der KI-Kernkomponente wird dann ohne formalen Beweis und nur auf Basis von Tests und Robustheit durchgeführt. Die Interaktion zwischen probabilistischem KI-Kern und deterministischem System ist unter Umständen komplexer als gedacht (z.B. wird vom deterministischen System ein Bahnsteig erwartet). Die HHU hat in diesem Zusammenhang eine Klassifikation der möglichen Fehlerquellen des KI-Kerns entwickelt (siehe Abbildung 2).

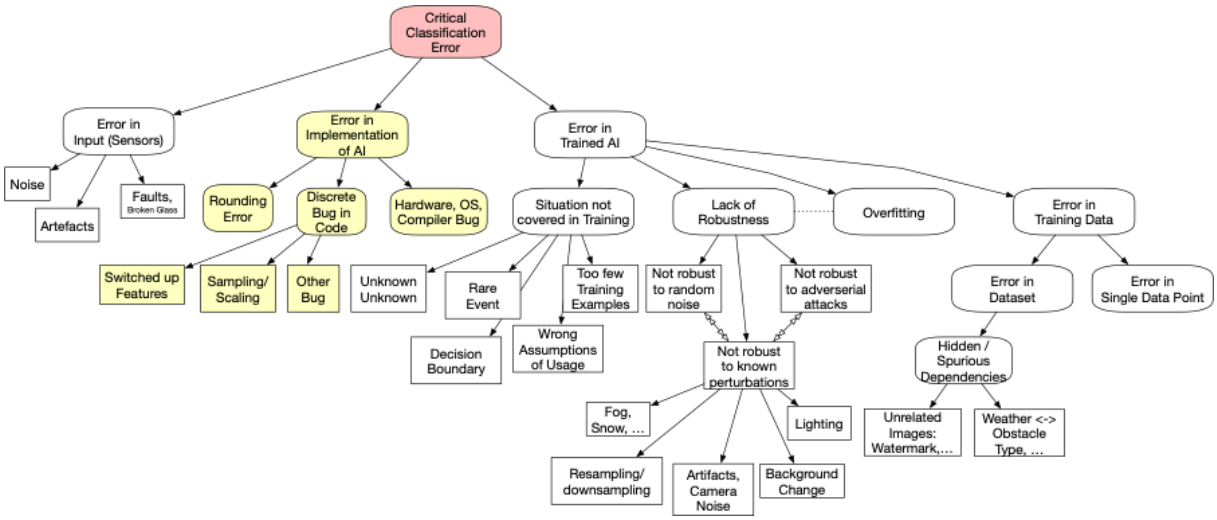



*Abb. 2: Klassifikation der möglichen Fehlerquellen*

**Normung und Richtlinien für die KI-Absicherung (Unterarbeitspaket 1.1b)**
Gemeinsam mit den Projektpartnern erfolgte eine Recherche zu existierenden Normen und Richtlinien für die KI-Absicherung in den Bereichen Bahntechnik, Automotive und Medizintechnik. Die relevanten Dokumente wurden festgestellt und im Detail analysiert. Im Ergebnis der Analyse wurden Anforderungen an die KI-Absicherungsmethoden im Kontext des Projekts formuliert.

**Ziele und Strategien zum automatisierten Testen von KI-basierten Objekterkennungssystemen in der Bahntechnik (Unterarbeitspakete 2.1 und 3.1)**
Was muss alles zu einer Zertifizierung eines KI-Systems nachgewiesen werden? Welche Ziele sind zu erreichen und wie kommen wir dahin? In einem gemeinsamen Dokument *Objectives and strategies for automated testing of AI-based perception systems in railroad engineering* (Ziele und Strategien zum automatisierten Testen von KI-basierten Objekterkennungssystemen in der Bahntechnik) wurden von den verschiedenen Projektpartnern dazu Ziele und Strategien aus ihren Kompetenzbereichen abgeleitet und festgehalten.

**Formales Ablaufmodell (AP 3.2, AP 4.1)**

Für die vom Projektpartner Hitachi (vorm. Thales) bereitgestellten Fallstudien (AP1) wurde ein formales Ablaufmodell in der formalen B-Methode entwickelt. Dieses deckt Rangierfahrten mit der zu zertifizierenden KI-Komponente und der Umgebung (Weichen, Signale, Hindernisse) ab. Die Formalisierung ist von früheren B-Modellen (ETCS Hybrid Level 3 und Moving Block) inspiriert. Zusätzlich berücksichtigt das Modell die unterschiedlichen in der Fallstudie beschriebenen Anforderungen wie das Lichtraumprofil oder unterschiedliche Annahmen für das Steuerungssystem des Zuges, z.B. vorab bekannte Signalpositionen.

Wir modellieren die Umgebung (Environment) und haben das eigentlich zu verifizierende System in seine einzelnen Komponenten aufgeteilt, um die KI und das Steuerungssystem im formalen Modell getrennt zu behandeln (vgl. Abb. 3). Die KI wird hierbei als „Black Box“ betrachtet, das heißt, wir betrachten ausschließlich den Rückgabewert der KI (Vision). Anschließend werden diese Werte an das deterministische Steuerungssystem (Control) übergeben und verarbeitet. Die Verarbeitung erfolgt direkt im formalen Modell, weshalb dieses später auch als Safety Shield eingesetzt werden kann. Das Modell kann mit unserem Werkzeug ProB simuliert und visualisiert werden.

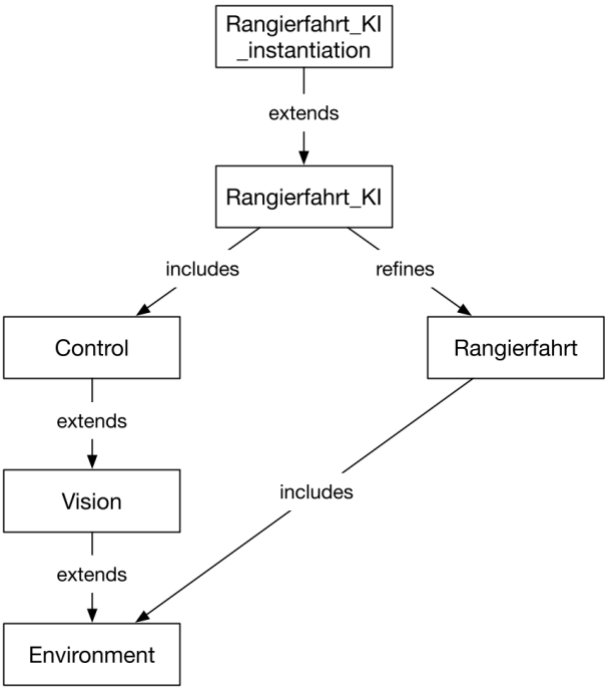



*Abb. 3: Aufteilung der System-komponenten für die formale Modellierung*

Ein wissenschaftlicher Artikel zu dem Ablaufmodell mit dem Titel „A Formal Model of Train Control with AI-based Obstacle Detection“ wurde bei der internationalen Konferenz **RSSRail 2023** veröffentlicht und im Oktober 2023 vorgestellt. In diesem haben wir das Verhalten für verschiedene Szenarien mit dem in ProB integrierten Komponente SimB simuliert. SimB ermöglicht probabilistische Simulationen formaler Modelle. Hierzu müssen Wahrscheinlichkeiten und Timing-Eigenschaften für die KI und dessen Umgebung von einem Modellierer manuell enkodiert werden. Die Herausforderung besteht darin, dass die Wahrscheinlichkeiten und das Timingverhalten auch tatsächlich das Verhalten der KI erfassen.

In unserer Simulation haben wir Wahrscheinlichkeiten für die korrekte, falsche oder fehlende Erkennung von Signalen und Weichen formuliert (vgl. Tab. 1). Diese hängen von der Distanz zu den jeweiligen Feldobjekten ab. Timing-Eigenschaften haben wir für die Bewegung der Lokomotive, sowie für das Ändern der Weichen, Signale und Gleissperren formuliert. Dadurch ist es möglich (1) die Geschwindigkeit der Lokomotive zu definieren sowie (2) das Zeitintervall für den Zustand der Weichenpositionen, der Signale und Gleissperren zu definieren.

*Tabelle 1: Übersicht aller Wahrscheinlichkeiten für das Perzeptionssystem der KI; CD: Korrekte Erkennung, WD: Falsche Erkennung, I: Ignorieren*

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Signal | Distance | 0-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71-80 | 81-90 | 91-100 | 101-110 |
| | CD | 99.9% | 99.9% | 64.9% | 49.9% | 39.9% | 29.9% | 19.9% | 14.9% | 9.9% | 4.9% | 0.0% |
| | WD | 0.01% | 0.01% | 3.51% | 5.01% | 6.01% | 7.01% | 8.01% | 8.51% | 9.01% | 9.51% | 0.0% |
| | I | 0.09% | 0.09% | 31.59% | 45.09% | 54.09% | 63.09% | 72.09% | 76.59% | 81.09% | 85.59% | 100.00% |
| Point Positioning | Distance | 0-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71-80 | 81-90 | 91-100 | 101-110 |
| | CD | 99.9% | 99.9% | 54.9% | 34.9% | 19.9% | 9.9% | 4.9% | 0.0% | 0.0% | 0.0% | 0.0% |
| | WD | 0.01% | 0.01% | 4.51% | 6.51% | 8.01% | 9.01% | 9.41% | 0.0% | 0.0% | 0.0% | 0.0% |
| | I | 0.09% | 0.09% | 40.59% | 58.59% | 72.09% | 81.09% | 85.59% | 100.0% | 100.0% | 100.0% | 100.0% |
| Derailer | Distance | 0-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71-80 | 81-90 | 91-100 | 101-110 |
| | CD | 99.9% | 99.9% | 64.9% | 49.9% | 39.9% | 29.9% | 19.9% | 14.9% | 9.9% | 4.9% | 0.0% |
| | WD | 0.01% | 0.01% | 3.51% | 5.01% | 6.01% | 7.01% | 8.01% | 8.51% | 9.01% | 9.51% | 0.0% |
| | I | 0.09% | 0.09% | 31.59% | 45.09% | 54.09% | 63.09% | 72.09% | 76.59% | 81.09% | 85.59% | 100.00% |
| Wagon | Distance | 0-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71-80 | 81-90 | 91-100 | 101-110 |
| | CD | 99.9% | 99.9% | 64.9% | 49.9% | 39.9% | 29.9% | 24.9% | 19.9% | 14.9% | 9.9% | 4.9% |
| | WD | - | - | - | - | - | - | - | - | - | - | - |
| | I | 0.1% | 0.1% | 35.1% | 50.1% | 60.1% | 70.1% | 75.1% | 80.1% | 85.1% | 90.1% | 95.1% |
| Person | Distance | 0-10 | 11-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71-80 | 81-90 | 91-100 | 101-110 |
| | CD | 99.9% | 99.9% | 64.9% | 49.9% | 39.9% | 29.9% | 24.9% | 19.9% | 14.9% | 9.9% | 4.9% |
| | WD | - | - | - | - | - | - | - | - | - | - | - |
| | I | 0.1% | 0.1% | 35.1% | 50.1% | 60.1% | 70.1% | 75.1% | 80.1% | 85.1% | 90.1% | 95.1% |

Basierend darauf können wir entweder eine einzelne Simulation in Echtzeit oder Monte-Carlo Simulationen ausführen. Während sich eine Echtzeitsimulation dazu eignet, ein einzelnes Szenario in Echtzeit zu betrachten, können basierend auf Monte-Carlo Simulationen statistische Validierungstechniken angewendet werden. Diese haben wir in unserem Ansatz verwendet, um die Wahrscheinlichkeit einer gefährlichen Situation, und für das Erreichen des Ziels (die „Mission Order“) abzuschätzen. Um das Verhalten der KI und dessen Umgebung präzise zu erfassen, haben wir ausgehend von diesen Ergebnissen daran gearbeitet, die Simulation direkt durch die KI ausführen zu lassen (s. Abschnitt „KI-gesteuerte Simulation des formalen Ablaufmodells“).

Zusätzlich zur Formalisierung wurde eine domänenspezifische Visualisierung für die Visualisierungskomponente VisB (Teil des ProB Werkzeuges) erstellt (vgl. Abb. 4, 5). Es ist möglich, diese nach der Validierung als einzelne HTML-Datei zu exportieren, welche Domänenexperten das Abspielen des validierten Szenarios ohne Kenntnisse von ProB in einem Browser ermöglicht. Eine Simulation kann auch zusammen mit der Visualisierung betrachtet werden:

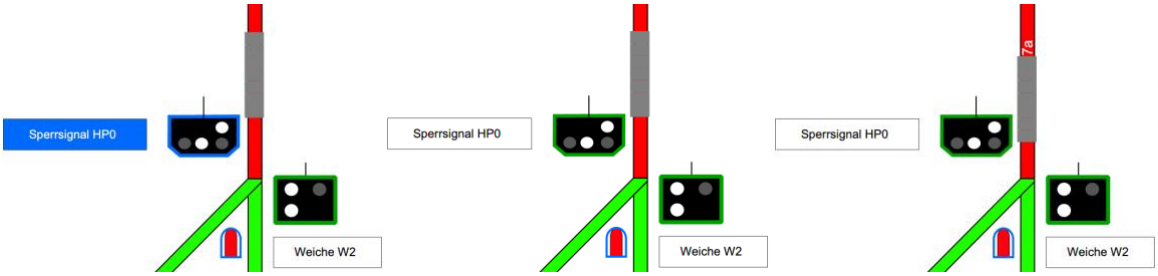



*Abb. 4: Zug nähert sich Sh1-Signal, erkennt dieses korrekt und setzt die Fahrt fort*

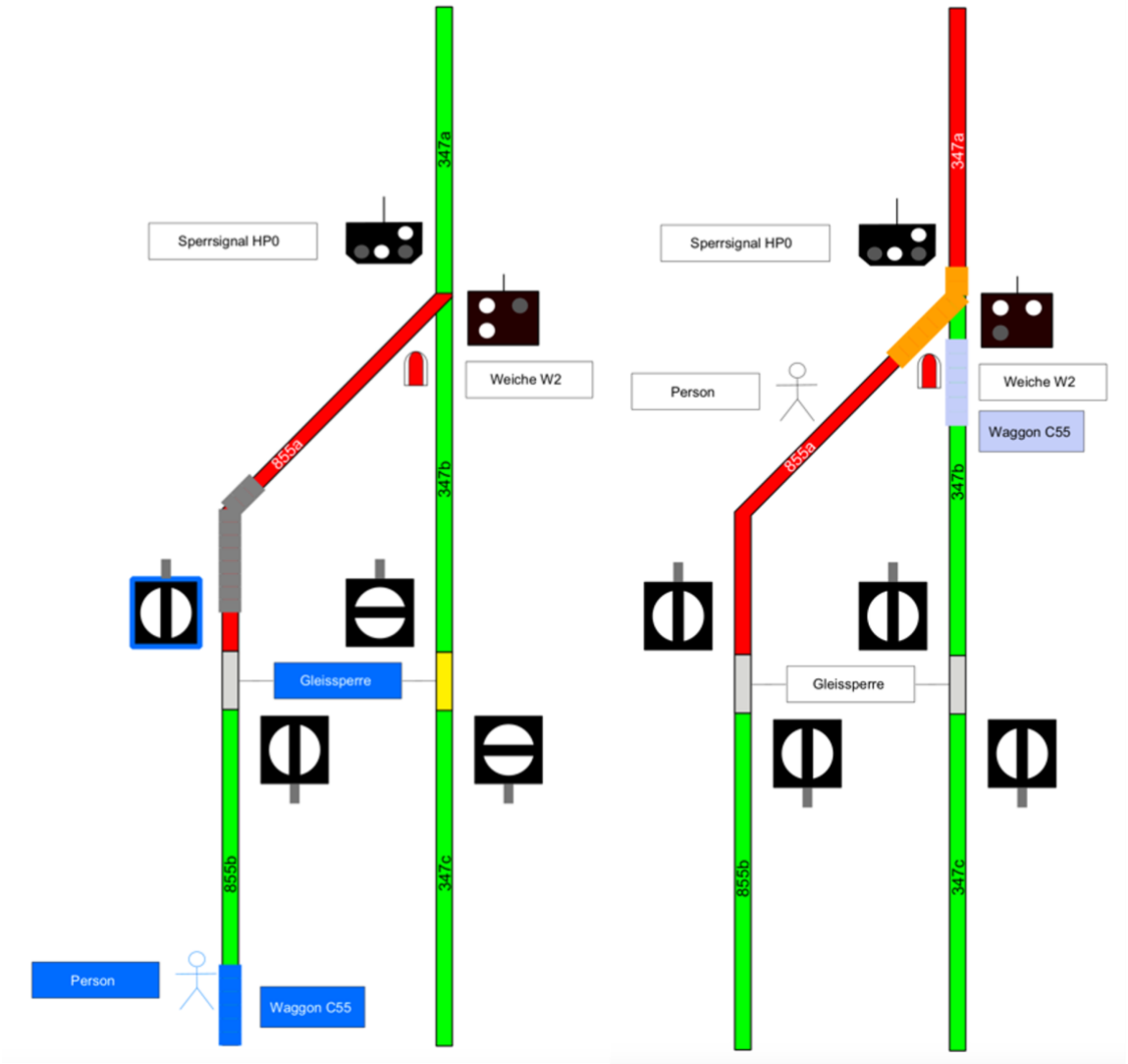



*Abb. 5: Domänenspezifische VisB-Visualisierung für zwei verschiedene Szenarien aus AP1*

Neben der Simulation haben wir auch weitere Validierungstechniken wie *Lineare Temporale Logik (LTL)* und *Model-Checking* angewandt. Eine der LTL-Eigenschaften, die wir validiert haben, entspricht der folgenden Formel:

***G*** *({"Zug fährt vorwärts" ⇒*
***Y*** *("Steuerungseinheit trifft Entscheidung, Zug vorwärts zu fahren" ∧*
*"Zug hat alle Signale korrekt erkannt" ∧*
*"Zug hat alle Weichen korrekt erkannt" ∧*
*"Zug hat alle Hindernisse korrekt erkannt" ∧*
*"Zug hat Gleisverlauf korrekt erkannt"})*
*⇒* ***G*** *({"Zug erreicht keine sicherheitskritische Situation"}).*

Wir konnten diese Eigenschaft für reduzierte Teile des formalen Modells, z.B. unter der Annahme bekannter Signalpositionen, erfolgreich verifizieren.

*Tabelle 2: Model-Checking Ergebnisse für ausgewählte reduzierte Modelle*

| Model | Operations | Variables/ Constants | States | Transitions | Time (min) | Memory (GB) |
|---|---|---|---|---|---|---|
| CD | 13 | 34 | 269 153 | 2 240 046 | 6.8 | 1.3 |
| + WS | 14 | 34 | 480 409 | 5 403 158 | 12.3 | 2.6 |
| + WP | 15 | 34 | 807 001 | 10 733 462 | 23.4 | 4.8 |
| + WP_DT | 15 | 34 | >16 785 959 | >185 250 252 | >530 | >80 |
| complete | 22 | 46 | n/a | n/a | n/a | n/a |

Für diese reduzierten Modelle haben wir anschließend Model-Checking durchgeführt, was zu den Ergebnissen in Tabelle 2 geführt hat. Dies hat gezeigt, dass der Zustandsraum für das vollständige Modell mit Model-Checking realistischerweise nicht verifizierbar ist, weshalb wir uns auf den Forschungsansatz mit Validierung durch Simulation fokussiert haben (s.a. „KI-gesteuerte Simulation des formalen Ablaufmodells“).

Für die flexible Konfiguration der Szenarien wurde am Import von railML-Topologiedaten nach ProB gearbeitet. railML ist ein XML-basiertes Austauschformat für diverse Eisenbahndaten (Infrastruktur, Stellwerk, Rollmaterial, Fahrplan). Hierbei ist die Motivation, verschiedene Szenarien automatisch in das formale Modell laden zu können, ohne die Umgebung für jedes Szenario manuell kodieren zu müssen. Als ein Nebenergebnis ist es möglich, railML-Spezifikationen mit ProB zu visualisieren, zu validieren und zu animieren. Die im Rahmen des Projektes weiterentwickelten Simulationsmöglichkeiten mit SimB kommen hierbei ebenso zum Einsatz. In diesem Kontext wurde im Juni 2024 auf der Konferenz **„International Conference on Engineering of Complex Computer Systems“** der Artikel „Validation of railML Using ProB“ vorgestellt. Weiterhin wurden die Ergebnisse im November 2024 internationalen Industriepartnern aus dem Eisenbahnbereich im Rahmen der 46. railML Konferenz vorgestellt. Eine Stärkung der Zusammenarbeit ist beabsichtigt.

## Datengenerierung (AP 2.2)

Für die Durchführung von aussagekräftigen Machine-Learning-Experimenten zur Schilderkennung benötigte die HHU einen sehr großen Korpus an annotierten Bildern aus der Perspektive einer Lokomotive.

Das Projekt KI-LOK hat sich aufgrund des Mangels an öffentlich verfügbaren Datensätzen in diesem Feld für die Generierung von synthetischen Testdaten entschieden.

Für die Experimente der HHU wurde ein Datensatz für die Segmentierung generiert, basierend auf Videos der Bahnstrecken Bremen-Oldenburg[5] und Freiburg-Seebrugg[6].

[5] https://www.youtube.com/watch?v=7ycjaWRnETU
[6] https://www.youtube.com/watch?v=axEpTqB3NZw

Um die Bilder brauchbar für die Bilderkennung und Klassifizierung zu machen, müssen Schilder hinzugefügt werden und die entsprechenden Bilder mit der Schildklasse annotiert werden. Hilfsweise haben wir uns dazu entschieden, Bilder von Schildern in die Bahnbilder einzusetzen. Vom Projektpartner Thales haben wir eine Liste der für Rangierfahrten relevanten Schildern erhalten und uns auf die folgenden neun verschiedenen Schildklassen konzentriert:

- EL6, Ra10, Ra12, Sh1, Wn7, Hp0, Ra11, Sh0, Sh2 (siehe Unterlagen zu AP1).

Da der obige Ansatz einige Probleme hatte, wurde er durch die folgenden Maßnahmen überarbeitet. Aus YouTube-Videos und Google Image-Bildern der Rangierschilder wurde ein neuer, kleinerer Datensatz erstellt, in dem die gesuchten Schilder-Klassen vorkommen.

Diese Daten wurden anschließend manuell gelabelt. Die Menge der Daten, die selbst gelabelt werden können, reicht zwar nicht für das Trainieren einer neuen KI, aber durchaus für ein erfolgreiches Fine-Tuning eines schon vor-trainierten Modells.

Außerdem wurde ein größerer Datensatz generiert, indem verschiedene Basisbildern der relevanten Schilder durch eine Reihe von Permutationen angepasst wurden (Resizing, Gaussian Noise, Brightness ...; s. Abb. 6) und in Streckenbilder aus der Lok-Perspektive eingefügt wurden. Dieser so erstellte Datensatz wurde nur zum Testen der KI und der Certifying Control Software benutzt, nicht zum Training der KI.

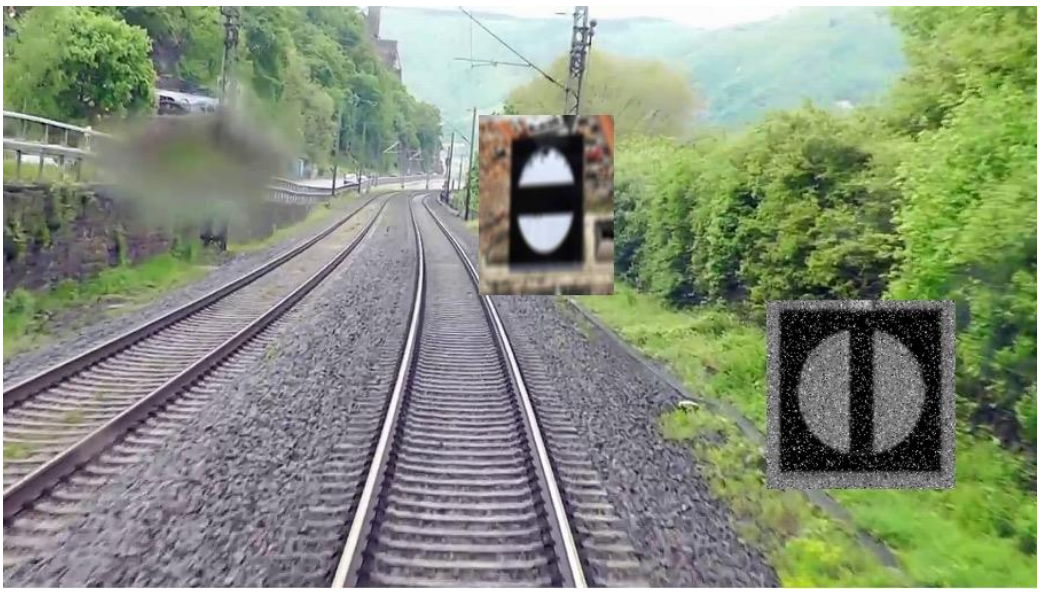


*Abb. 6: Beispiel für ein generiertes Bild mit Noise und Resizing*

## Machine Learning Experimente (AP 4.2 und 2.3)

Für die Experimente zu den vorliegenden Fallstudien von AP1 hat die HHU Deep Learning Modelle trainiert, die die zu validierende KI repräsentieren. Insbesondere wurde das Open-Source YOLOv8 Modell von Ultralytics (https://github.com/ultralytics/ultralytics) auf den oben aufgeführen Daten trainiert, die Rangierschlider im Datensatz zu lokalisieren.

Das Training wurde mit folgenden Ergebnissen abgeschlossen.

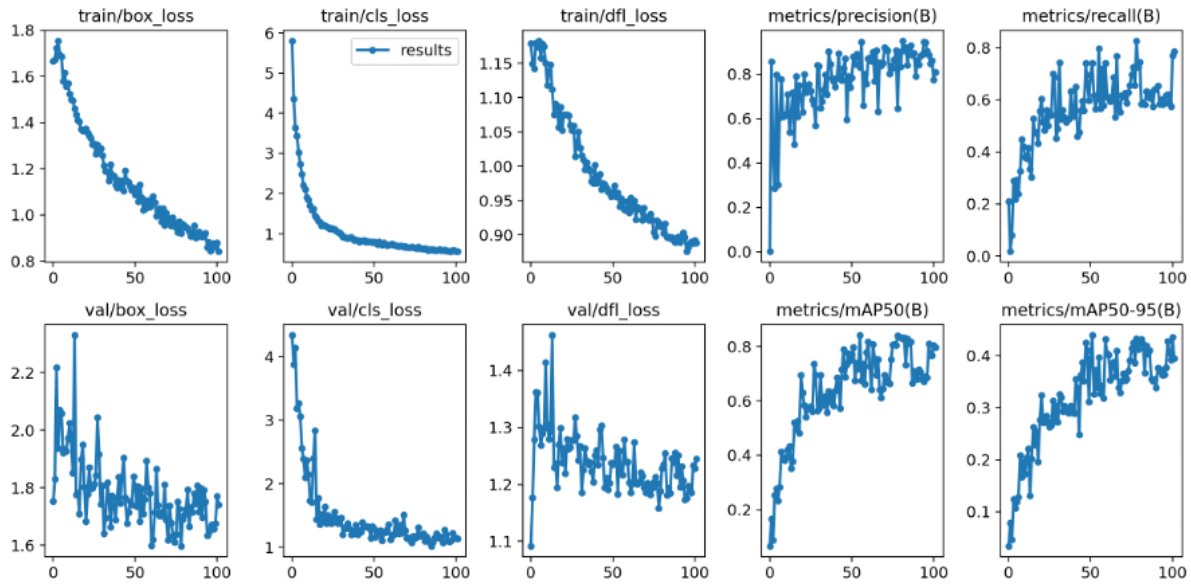



*Abb. 7: Ergebnisse nach dem Nachtrainieren des YOLOv8-Modells*

Die Ergebnisse (Abb. 7) zeigen gute Resultate in Bezug auf den Verlust auf den Validierungsdaten, was eine potenzielle Überanpassung (Overfitting) ausschließt. Die mAP50 und mAP50-95 sind für das gegebene Problem auch gute Werte. Dies zeigt, dass eine YOLO KI auf kleiner Datengrundlage für solche Aufgaben nach-trainiert werden kann.
Bei unserem Training stand allerdings nicht im Vordergrund, dass die KI sehr gute Resultate erzielt. Vielmehr soll diese auch Negativbeispiele liefern, um unsere Validierungstechniken damit adäquat erforschen zu können.

## Studium von Certified Control (AP 2.1 und 2.3)

Im Rahmen des Projektes hat die HHU auf den akademischen Arbeiten vom MIT bezüglich „Certifying Control“ aufgebaut. Hier wurde mit der formalen Sprache Alloy ein kleiner Durchbruch für den Sicherheitsnachweis von KI in autonomen Fahrzeugen erreicht:

- Daniel Jackson et al. Certified Control: An Architecture for Verifiable Safety of Autonomous Vehicles. Preprint: https://arxiv.org/abs/2104.06178

Mit Hilfe von SMT-Solvern und formaler Modellierung wurde ein Certificate Checker für Objekterkennung mit Lidar bei Schnee entwickelt. Die Anwendung ist deutlich eingeschränkter als die KI-LOK-Fallstudien. Objekte werden nicht klassifiziert; die KI soll „nur“ bestimmen, ob große Objekte in Fahrtrichtung vorhanden sind und ob eine Vollbremsung eingeleitet werden soll.

Die HHU hat diese Arbeiten studiert und daraus verschiedene aussichtsreiche Ansätze für die KI-LOK Fallstudie identifiziert.

Wir haben uns mit der Erkennung von Rangierschildern auf ein eingeschränktes Unterproblem konzentriert. Im Vergleich zur kompletten Hinderniserkennung sind hier die Objekte von einfacherer geometrischer Natur und lassen sich eventuell sogar formal beschreiben. Alternativ kann man durch klassische Computer-Vision-Algorithmen die Erkennung eines Schildes prüfen.

Für dieses Unterproblem untersuchten wir folgende Punkte:

- Was sind die Grenzen der Präzision von Deep Learning bei perfekter, idealisierter Datenlage und reiner Klassifikation?
- Können mit Ensembletechniken Fehler reduziert oder fast ausgeschlossen werden?
- Kann man für dieses Unterproblem den Ansatz von „Certifying Control“ einsetzen? Mit Deep Learning werden die Bilder segmentiert und Computer-Vision-Algorithmen werden als Certificate Checker eingesetzt.

Mit Hilfe von herkömmlichen Computer-Vision Algorithmen, wie dem Konturerkennungsalgorithmus der OpenCV Bibliothek, ist es gelungen, Sh0, Sh1 und Wn7 Rangierschilder anhand der Winkel der vorhandenen Halbkreise zuverlässig (in 100% der Fälle in allen Tests), unter Voraussetzung einer ausreichenden Bildqualität, zu erkennen und zu unterscheiden.

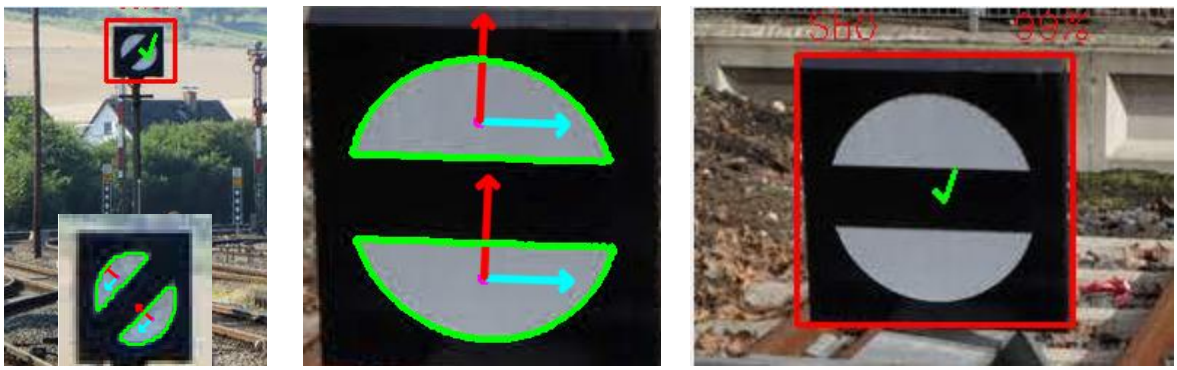


*Abb. 8: Durch die KI erkannte, mittels Certificate Checker zertifizierte Bounding Boxes*


*Abb. 9: Beispiel für eine falsch zertifizierte und eine falsch abgelehnte Bounding Box*

Dies wurde erfolgreich benutzt, um die Bounding Boxes aus der KI-Objekterkennung im Sinne von Certifying Control zu zertifizieren, das heißt, es können von der KI falsch erkannte Schilder verhindert werden. In Abbildung 8 sind die korrekte Zertifizierung von einem Sh0- und einem Sh1-Schild illustriert. Der Nachteil ist eine Ablehnung einiger von der KI eigentlich richtig erkannten Schilder, die im Bild zu klein sind oder wo der Bildbereich zu verschoben ist, um von der Software erkannt werden zu können (vgl. Abb. 9). Ebenso können Falscherkennungen fälschlicherweise als korrekt zertifiziert werden (vgl. Abb. 9).

Diese Technik haben wir für verschiedene YOLO-Modelle auf den synthetisch generierten Daten ausgewertet, was die folgenden Ergebnisse ergab.

| Model | Detected | TP | FP |
|---|---|---|---|
| n | 30111 | 25514 | 4597 |
| s | 30335 | 26790 | 3545 |
| m | 28672 | 22728 | 5944 |

(a) Results without Monitor

| Model | Detected | TP | FP |
|---|---|---|---|
| n | 21716 | 21714 | 2 |
| s | 22834 | 22831 | 3 |
| m | 20460 | 20460 | 0 |

(b) Results with Monitor

Table 1: Raw numbers for Models on Generated Data

| Model | Precision | Recall | $F_1$ score |
|---|---|---|---|
| n | 0.85 | 0.58 | 0.69 |
| s | 0.88 | 0.61 | 0.72 |
| m | 0.79 | 0.52 | 0.63 |

(a) Results without Monitor

| Model | Precision | Recall | $F_1$ score |
|---|---|---|---|
| n | 1.00 | 0.50 | 0.67 |
| s | 1.00 | 0.52 | 0.68 |
| m | 1.00 | 0.47 | 0.64 |

(b) Results with Monitor

*Tabelle 3: Modellmetriken für generierte Daten (Werte auf zwei Dezimalstellen gerundet)*

Wie man in der Tabelle 3 sehen kann, ist es nach einigen Experimenten und Verbesserungen am Algorithmus gelungen, die Reduktion des Recall auf ein Minimum zu beschränken während **falsch positive Erkennungen weiterhin effektiv verhindert werden**. Dies zeigt das Potential der Methode in dieser und weiteren Anwendungen.

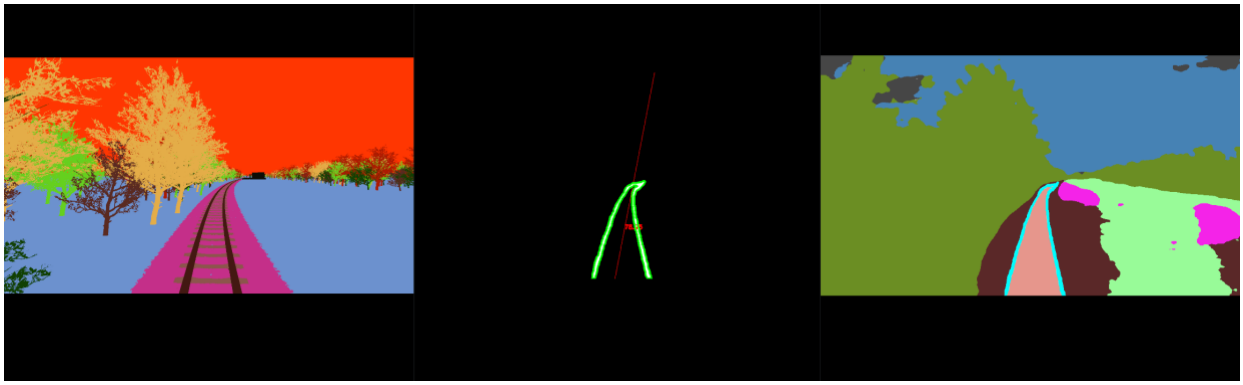
So können potentiell selbst KI-Komponenten, die natürlicherweise eine niedrige Präzision haben, zum Beispiel wegen der Schwierigkeit der Aufgabe, mit minimalem Overhead eingesetzt werden und auf die notwendigen Erfolgsmetriken in der Praxis kommen.

Eine Verallgemeinerung des Konzeptes auf andere Schilder- und Objektklassen scheint allerdings weiterhin wenig erfolgversprechend.

Diese Arbeiten wurden auf der FMAS 2023 vorgestellt und veröffentlicht.

- Jan Roßbach & Michael Leuschel. Certified Control for Train Sign Classification: https://arxiv.org/abs/2104.06178

Eine Journal Version des Artikels ist aktuell noch in Review.



*a) Segmentierungslabels* *b) Monitor Visualisierung* *c) SUT Predictions*

*Abb. 10: Schienensegmentierung mit Certified Control*

Das Konzept Certified Control wurde auch auf die zweite von Thales bereitgestellte Fallstudie zur Schienensegmentierung angewendet. Hierbei sollten die Segmentierungsmasken des System under Test (siehe Abb. 10c), validiert werden und insbesondere Fehler in der vorhergesagten Richtung der Kurve identifiziert werden. Dafür hat die HHU die Segmentierungsmasken analysiert und mit Hilfe von OpenCV die Schienen isoliert, auf denen sich die eigene Lok gerade befindet und den Kurvenwinkel berechnet (siehe Abb. 10b). Dies kann im Sinne von Certified Control benutzt werden, um die für die Zuglokalisierung wichtigen Daten im Perceptionsystem abzugleichen, die von anderen Modellen erkannt werden sollen.

## KI-gesteuerte Simulation des formalen Ablaufmodells mit Certificate Checker (AP 3.2, 4.1)

Für die ersten Simulationen von Szenarien mithilfe des formalen B-Ablaufmodells haben wir Wahrscheinlichkeiten für die Events per Hand kodiert. In weiteren Arbeiten haben wir an einem Ansatz geforscht, mit welchem die Events direkt von der KI ausgeführt werden. Da die Simulation direkt durch die KI erfolgt, muss man keine Simulation mehr manuell enkodieren und wir erreichen dadurch, dass Wahrscheinlichkeiten und Timing-Eigenschaften für das Verhalten der Simulation den realen Eigenschaften der KI entsprechen. Die zugrundeliegende Technik wurde im

Rahmen einer parallelen Arbeit implementiert, die sich mit der Validierung von Reinforcement Learning Agenten befasst. In diesem Zusammenhang wurde im Juni 2024 der Artikel „Validation of Reinforcement Learning Agents and Safety Shields with ProB“ bei der internationalen Konferenz **„NASA Formal Methods“ 2024** vorgestellt.

In diesem Artikel arbeiten wir ausführlich mit einer „Highway“ KI, welche mit Reinforcement Learning trainiert wurde (https://github.com/Farama-Foundation/HighwayEnv). Anhand von weiteren Fallstudien haben wir gezeigt, dass dieser Ansatz auch für weitere Reinforcement Learning Agenten wie z.B. Autonomous Drone Swarm (siehe https://github.com/hhu-stups/reinforcement-learning-b-models) skaliert. In dem Artikel haben wir ebenso die Effizienz der Safety Shield Technik demonstriert. Die Validierungstechniken konnten außerdem ohne weitere Modifikationen direkt angewendet werden. Genauso wie bei der KI-LOK Fallstudie muss hier ein formales Modell enkodiert werden, welches das Verhalten der KI und dessen Umgebung erfasst. Zudem haben die Validierungstechniken es ermöglicht, Schwächen in der KI zu erkennen, und diese anschließend in der Reward Funktion und dem Safety Shield zu verbessern. Insbesondere haben wir verschiedene Eigenschaften, u.a. die Unfallquote mit und ohne Safety Shield evaluiert. Hierbei konnten wir bestätigen, dass ein Safety Shield in der Lage ist, die Anzahl gefährlicher Situation stark zu verringern.

Diese Technik konnten wir erfolgreich in das KI-LOK Projekt übertragen, um die Simulation direkt durch die KI ausführen zu lassen und dadurch Schwachstellen der KI aufzudecken. Zusätzlich integrieren wir Certified Control in die Simulation des formalen Modells. Hierzu wurden das formale Modell sowie die Schnittstellen zwischen Python und SimB (ermöglicht Simulationen von formalen Modellen mit ProB) überarbeitet und angepasst. Abbildung 11 zeigt die Funktionsweise dieser Integration des formalen Modells mit einer „echten“ KI und dem „echten“ Certificate Checker.

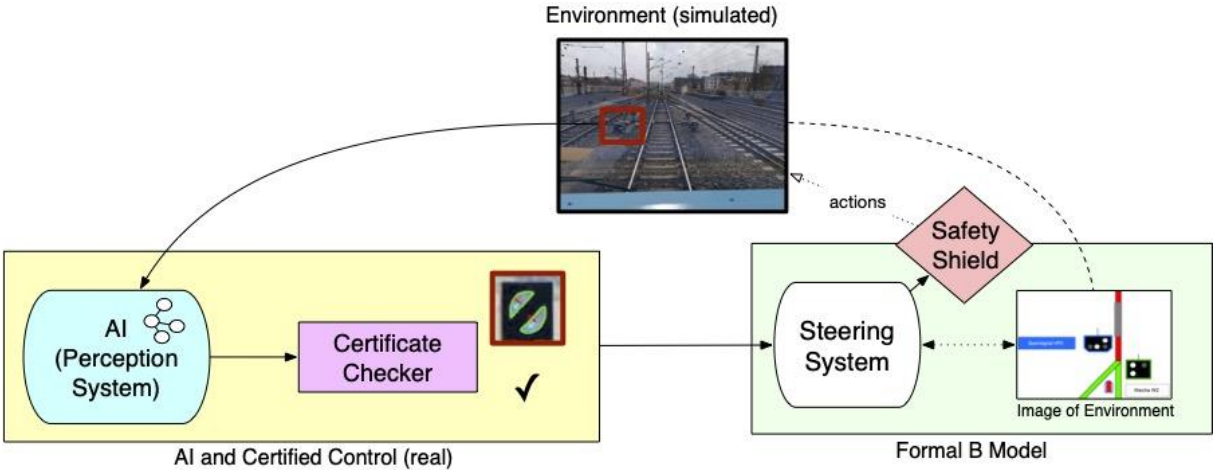



*Abb. 11: Konzept der KI-gesteuerten Simulation des formalen Modells mit Certified Control*

Der Zustand des gesamten Systems wird durch das formale B-Modell verwaltet und überwacht. Auf Grundlage der Zugposition und der dort zu erwartenden Signal wird

ein Bild aus der Umgebung gewählt und an das Perzeptionssystem (die KI-Komponente) zu Auswertung übergeben. Das Ergebnis der KI wird durch den Certificate Checker geprüft. Dieses Ergebnis wiederum wird zurück an das formale Modell geschickt, welches basierend darauf ein die Erkennung repräsentierendes Event auslöst und dem Steering System so mitteilt, wo ein Signal erkannt wurde. Basierend darauf steuert das deterministische Steering System, welches zuvor unter der Annahme einer korrekt funktionierenden KI formal validiert wurde, die Bewegungen des Zuges. Zusätzlich haben wir die Möglichkeit, Safety Shields im formalen B-Modell einzusetzen, um potenziell gefährliche Operationen nicht auszuführen.

Basierend auf diesem Kreislauf können wir wie zuvor entweder Echtzeitsimulationen (vgl. Abb. 12) oder Monte-Carlo Simulationen in Verbindung mit statistischen Validierungstechniken durchführen. Mithilfe der Fallstudie für KI-LOK können wir mit unseren Experimenten bereits zeigen, dass sich mit Anwendung von Certified Control und einem einfachen Safety Shield, z.B. dass der Lok die zu erwartenden Signalpositionen bekannt sind, die sicher zurückgelegte Strecke der Lok deutlich verbessern lässt. Während ohne Anwendung beider Techniken eine Weiterfahrt aufgrund vieler falsch positiver Haltsignalerkennungen nahezu unmöglich ist, filtert Certified Control diese Erkennungen fast immer zuverlässig. Das Safety Shield hilft dabei, die vom Certificate Checker fälschlicherweise abgelehnten korrekten Erkennungen abzuschwächen, in dem vor einem nicht erkannten Signal mit bekannter Position sicherheitshalber angehalten werden kann.

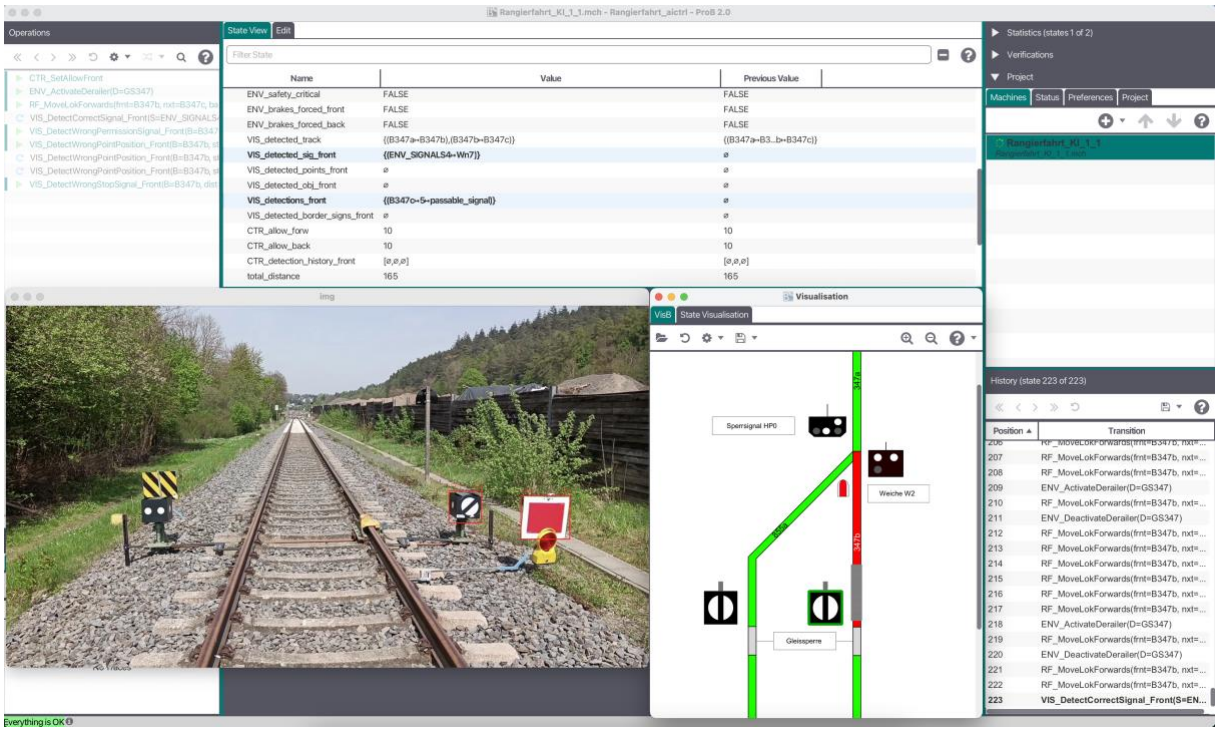



*Abb. 12: Echtzeitsimulation mit SimB im Werkzeug ProB2-UI, korrekte Erkennung des Sh1 durch die KI und Visualisierung des aktuellen Zustands im formalen Modell*

Für die durch die KI ausgewerteten Bilder verwenden wir positions- und zustandsabhängige Bilder, die zufällig aus Videos von Abfahrten auf Rangierbahnhöfen entnommen werden. Perspektivisch können auch die mithilfe der Simulationsumgebung des Projektpartners Fraunhofer FOKUS generierten Bilder verwendet werden.

Dieser Ansatz und erste Resultate wurden im wissenschaftlichen Artikel „Using Formal Models, Safety Shields and Certified Control to Validate AI-Based Train Systems“ bei dem internationalen Workshop „Formal Methods for Autonomous Systems“ im November 2024 präsentiert. Außerdem können auf diese Weise untersuchte Szenarien im KI-LOK Demonstrator inspiziert und vorgeführt werden (siehe auch III.5).

## Absicherungsmethodik (AP 4.2)

In Zusammenarbeit mit Neurocat hat die HHU an einer wissenschaftlichen Publikation gearbeitet, die den abschließenden Stand der Absicherungsmethodik (**Safety Case**) ausarbeitet. Es sollen die Methoden der Projektpartner zusammengefasst und integriert werden zu einer übergreifenden Sicherheitsevaluationsmethodik, die unabhängig vom Anwendungsfall auf KI-Komponenten im automatisierten Bahnbetrieb angewendet werden kann. Die Methodik wird im Artikel ausführlich erklärt, mit den relevanten Zertifizierungsstandards verglichen und an einem Beispiel in Goal Structuring Notation (GSN) illustriert.

- Jan Roßbach, Oliver De Candido, Ahmed Hammam & Michael Leuschel. Evaluating AI-Based Components in Autonomous Railway Systems: https://link.springer.com/chapter/10.1007/978-3-031-70893-0_14

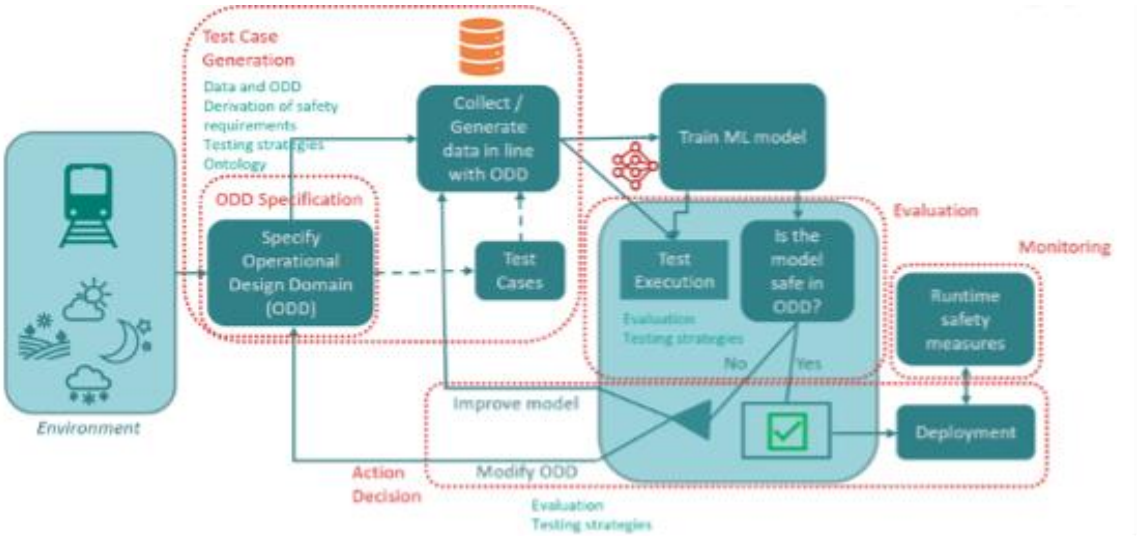



*Abb. 13: Übersicht über die KI-LOK Methodik*

Die Methodik (Abb. 13/14) beschreibt wie eine KI-Komponente mit den Projektmethoden zertifiziert werden kann. Dabei wird wie folgt vorgegangen:

1. Ontologie Spezifikation: Erstellung einer formalen Ontologie zur Definition der Operational Design Domain (ODD). Die Ontologie dient als Basis für systematische Tests und Datenanalyse, erweitert durch probabilistische Informationen zur realistischen Modellierung von Umweltszenarien.

2. Testfall-Generierung: Nutzung einer 3D-Simulationsumgebung zur Erstellung von realistischen, diversifizierten Testdaten. Diese Simulation berücksichtigt Faktoren wie Streckengeometrie, Wetterbedingungen und Objekte in der Umgebung. Ziel ist die Abbildung möglichst vieler relevanter Szenarien.

3. Evaluation: Bewertung der KI-Komponenten sowohl einzeln als auch in ihrer Integration ins Gesamtsystem. Hierbei werden robuste Testverfahren angewendet, einschließlich der Analyse von adversarialen Angriffen und der Suche nach Schwachstellen in den Eingabedaten. Zudem erfolgt eine formale Modellierung des Gesamtsystems, um Sicherheitsgarantien zu gewährleisten.

4. Monitoring: Einsatz von Laufzeitüberwachung, um sicherheitskritische Eigenschaften dynamisch zu verifizieren. Dabei werden beispielsweise Zertifikate zur Bewertung der Korrektheit von KI-Ausgaben genutzt, ohne dass die Modelle selbst formal verifiziert werden müssen (Certified Control).

5. Bei unzureichender Sicherheit wird entweder die ODD angepasst oder die KI-Komponente durch zusätzliche Daten verbessert.

Am Ende kann man durch die genannte Methodik die Häufigkeit feststellen, mit der in einer realen Situation, unter der Bedingung, dass die Testdaten repräsentativ sind, im Einsatz Fehler der KI zu erwarten sind. Diese Daten können dann in anderen Argumenten im Safety-Case benutzt werden, um ausreichende Sicherheit nachzuweisen.

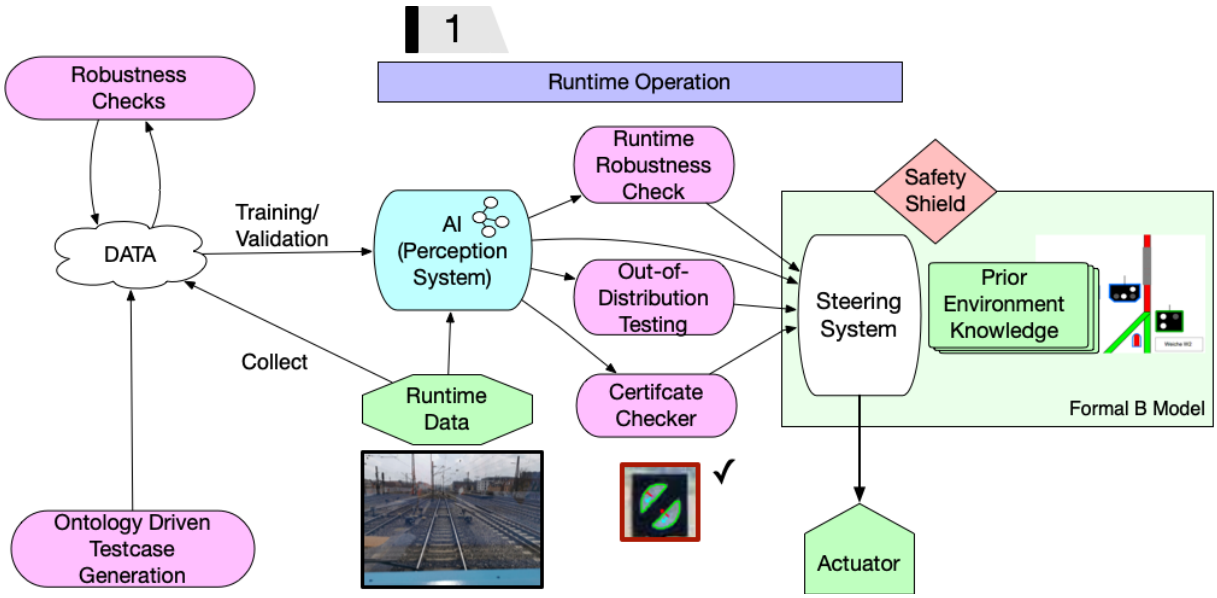



*Abb. 14: Einordnung der HHU-Arbeiten in den Gesamtüberblick*

### Demonstration der integrierten Werkzeugkette (AP 4.3)

Zur Demonstration des Ablaufes der im Projekt entwickelten Absicherungsmethodik (integrierte Werkzeugkette) wird das Programm Rerun benutzt. Dieses ermöglicht die einfache Aufzeichnung und wiederholte Darstellung von verschiedenen Durchläufen der integrierten Werkzeugkette in einer grafischen Benutzeroberfäche (GUI). Hierzu stellte jeder Projektpartner für die entwickelten Fallstudien (AP 1) die Daten für den betreffenden Teil der Werkzeugkette bereit.

Die HHU hat hierzu mit einer Visualisierung des Certified Control Prozesses beigetragen. Einerseits wird die Signalerkennung nach Auswertung der erkannten Bounding Boxes durch den Certificate Checker dargestellt (vgl. Abb. 15 und 16). Hierdurch lassen sich die erkannten Konturen analysieren und eventuelle Fehler feststellen. Weiterhin werden die Schienensegmentierung und die hierbei erkannten Eigenschaften durch den Certificate Checker visualisiert (vgl. Abb. 15).

Ebenso integriert wurde eine Visualisierung der Interaktion mit dem formalen Modell über die im Rahmen des Projektes entwickelte Schnittstelle mit SimB. Die entwickelte VisB-Visualisierung wurde für die Darstellung des aktuellen Zustands aus Sicht des formalen Modells integriert (vgl. Abb. 16).

Die integrierte Demonstration mit den Ergebnissen aller Projektpartner wurde auf der Messe „InnoTrans 2024“ in Berlin einem internationalen Fachpublikum vorgeführt.

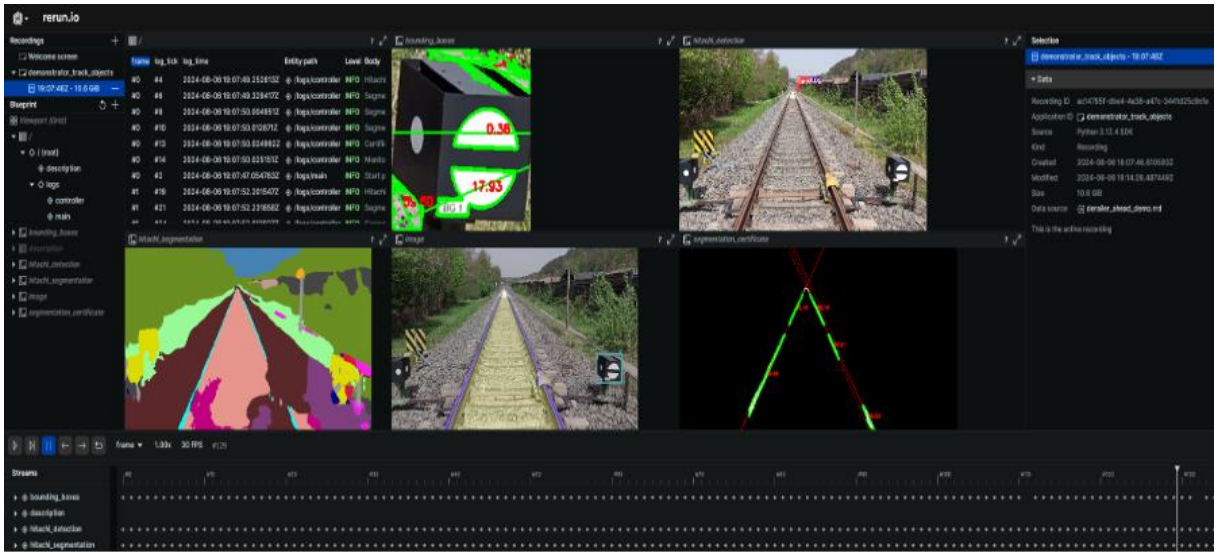



*Abb. 15: Nutzung von Rerun für Demonstration von Certified Control*

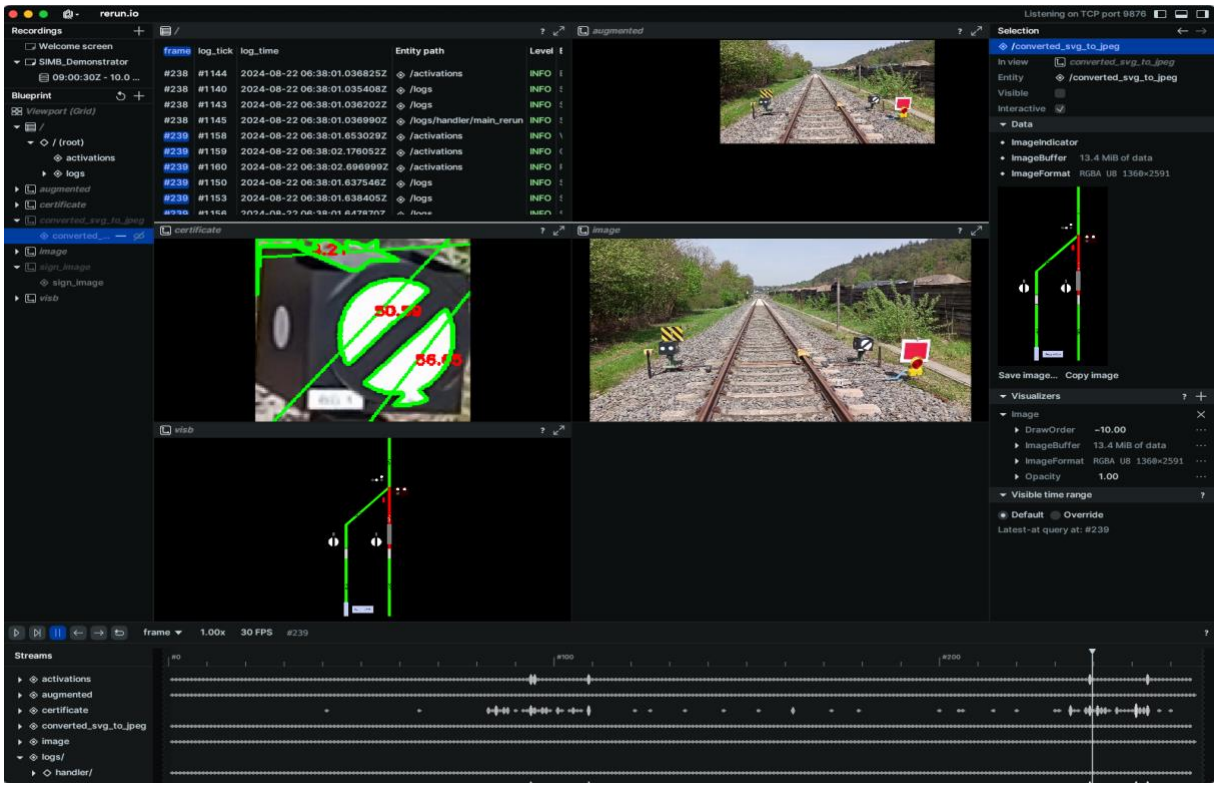



*Abb. 16: Nutzung von Rerun für Demonstration des formalen B-Modells*

## II.2 Wichtigste Positionen des zahlenmäßigen Nachweises

Der überwiegende Anteil (94,1 %) der Förderung ist für Personalausgaben verwendet worden. Die Forschungsarbeiten waren nicht nur vom Umfang her aufwändig, sondern erforderten auch eine breite Expertise (Softwareentwicklung, Künstliche Intelligenz, Schienenverkehr, ...) und somit den Einsatz von unterschiedlichen Mitarbeitern.

Die restlichen Mittel sind hauptsächlich für Reisen zu Projekttreffen, Konferenzen und Workshops verwendet worden. In der Tat sind internationale Konferenzen und Workshops das maßgebliche Mittel, um Forschungsergebnisse in der Informatik zu verbreiten (siehe Punkt II.3).

## II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Das Projekt KI-LOK hat erfolgreich als Wegbereiter für nachhaltige Partnerschaften zwischen Unternehmen und Organisationen zur Erforschung und Entwicklung von sicherheitskritischen Systemen mit KI-Komponenten fungiert. Die innovativen Forschungs- und Entwicklungsergebnisse des Projekts stärken den Standort Deutschland, indem sie Test- und Prüflösungen bereitstellen, die effektive Maßnahmen zur Qualitätssteigerung von KI-Komponenten ermöglichen. Die durchgeführten Arbeiten sowie die eingesetzten Ressourcen waren erforderlich und angemessen, da sie der im Projektantrag detaillierten Planung entsprachen und alle im Arbeitsplan festgelegten Aufgaben erfolgreich bearbeitet und mit innovativen Ergebnissen abgeschlossen wurden.

**Personal**

- *KI-Expertise (Machine-Learning Modelle erstellen, trainieren, validieren):* Jan Roßbach
- *Bahn Expertise (realistische formale Modelle für Bahnanwendungen erstellen, Import von realistischen Bahndaten (railML)):* Jan Gruteser
- *Werkzeugentwicklung*
  - *Expertise in der Softwareentwicklung und Weiterentwicklung von ProB, u.a. für die Visualisierung für Bahnexperten, Simulation von formalen Modellen mit KI:* Fabian Vu
  - *ProB für Testfallgenerierung für ODDs, grundlegende Erweiterungen von ProB:* David Geleßus

**Konferenz-/Workshopteilnahmen**

In unserem Fachbereich sind Konferenzen das wichtigste Mittel Forschungsergebnisse zu vermitteln und somit sicherzustellen, dass die Projektergebnisse auf Dauer einen Impakt für Forschung und Entwicklung haben. Diese Konferenzpublikationen sind immer an eine Reise gebunden (die Konferenzorganisatoren verlangen eine Einschreibung, bevor die Artikel veröffentlicht werden). In diesem Sinne sind die Mittel mit dem Zweck verbunden, sicherzustellen, dass die Projektergebnisse nachhaltig von einem breiten Publikum genutzt werden können.

In der Tat sind unsere Arbeiten aus 2023 schon mehrfach zitiert worden und wir wurden von mehreren Firmen im Nachgang an die Konferenzen gebeten, unsere Projektideen und Ergebnisse zu erörtern. Die Reisen sind auch wichtig, damit die jungen Forscher Feedback, Lösungsansätze und neue Ideen zu ihren Forschungsaufgaben bekommen und Ihnen somit ermöglicht wird, die Projektziele besser zu erreichen.

## II.4 Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

Die HHU möchte in den kommenden Jahren ihre Aktivitäten im Bereich der künstlichen Intelligenz und der Datenwissenschaften weiter erhöhen. Das *Heine Center for Artificial Intelligence and Data Science* (https://www.heicad.hhu.de) koordiniert die Aktivitäten in diesem Bereich. Anhand der Fallstudien von KI-LOK hat die HHU ihre Expertise im Rahmen der formalen Systemmodellierung erweitert und ist in der Lage, komplexe sicherheitskritische Systeme mit KI-Komponenten auch in anderen Bereichen zu simulieren, analysieren und validieren. Die neu entwickelten Techniken, Werkzeuge und die gewonnene Expertise wird es der HHU ermöglichen, wissenschaftliche und wirtschaftliche Drittmittelprojekte für die Verifikation sicherheitskritischer Anwendungen der Künstlichen Intelligenz im Rahmen dieses Zentrums einzuwerben, und dies nicht nur im Bereich der Fahrzeugautomatisierung, sondern auch in anderen Bereichen, wie zum Beispiel der Entwicklung von Medizinprodukten. Ein Projekt im medizinischen Bereich ist unter Auflagen bewilligt. Ein Forschungsantrag im Automobil- und Bahnbereich beim BMWK ist bis Jahresende geplant.

## II.5 Während der Durchführung des Vorhabens dem ZE bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Durchführung des Vorhabens sind neue Ergebnisse anderer Projekte bekannt geworden. Dazu gehören die Erstellung und Veröffentlichung eines Multi Sensor Datensatzes für den Schienenbereich (https://data.fid-move.de/dataset/osdar23/resource/fcbf8856-fc46-43d1-ac1a-147c8ab87ff4) vom Deutschen Zentrum für Schienenverkehrsforschung (DZSF).
Der Datensatz beinhaltet gelabelte Daten für mehrere verschiedene Sensoren (Kamera, LIDAR, Radar etc.) in unterschiedlichen Sicherheitskritischen Szenarien. Dies ermöglicht wichtige zukünftige Forschung im Bereich Sensorfusion, konnte aber für dieses Projekt nicht zur Anwendung kommen, da wichtige Klassen im Bereich Schilderkennung für die Fallstudien fehlen.

Andere Ergebnisse beinhalten die Abschätzung von Anforderungen an einzelne KI-Komponenten für eine erfolgreiche KI-Verifizierung durch Projekte wie ATO-Risk. Dort wurde eine Abschätzung vorgenommen, wie gut eine KI in verschiedenen Aufgaben sein muss, um besser als ein vergleichbarer menschlicher Lokführer zu sein.

Ein weiteres Projekt, das während der Durchführung des Vorhabens Ergebnisse präsentiert hat, ist Projekt HiDyVe, deren technischer Report

(https://arxiv.org/pdf/2401.06156) eine Möglichkeit die Fehlerwahrscheinlichkeit eines integrierten KI-basierten Systems abzuschätzen zeigt, indem mehrere KIs für dieselbe Aufgabe benutzt werden. Wenn die Fehlerwahrscheinlichkeiten der einzelnen KIs stochastisch unabhängig voneinander sein sind, kann gezeigt werden, dass die Gesamtfehlerwahrscheinlichkeit kleiner ist als die vom Standard vorgeschriebene Toleranzgrenze.

## II.6 Erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr.11

Georg Hemzal, Timo Strobel, Jürgen Großmann, Bernd-Holger Schlingloff, Michael Leuschel, Sadegh Sadeghipour, Jörg Firnkorn
KI-LOK – Ein Verbundprojekt über Prüfverfahren für KI-basierte Komponenten im Eisenbahnbetrieb.
In Signal+Draht, 10/2021, 6--15, https://eurailpress-archiv.de/SingleView.aspx?show=2911184

Fabian Vu, Christopher Happe, Michael Leuschel
Generating Domain-Specific Interactive Validation Documents.
In Proceedings FMICS, LNCS, 13487, Springer, 32--49, 2022.
https://doi.org/10.1007/978-3-031-15008-1_4

Georg Hemzal, Timo Strobel, Michael Leuschel, Jürgen Großmann, Dorian Knoblauch, Mariia Kucheiko, Nicolas Grube, Roman Krajewski
KI-LOK – Ein Verbundprojekt über Prüfverfahren für KI-basierte Komponenten im Eisenbahnbetrieb.
In Signal+Draht, 04/2023, 37--45, https://eurailpress-archiv.de/SingleView.aspx?show=4975840

Jan Gruteser, David Geleßus, Michael Leuschel, Jan Roßbach, Fabian Vu
A Formal Model of Train Control with AI-Based Obstacle Detection.
In Proceedings RSSRail 2023, LNCS, 14198, Springer, 128--145, 2023.
https://doi.org/10.1007/978-3-031-43366-5_8

Jan Roßbach, Michael Leuschel
Certified Control for Train Sign Classification.
In Proceedings FMAS 2023, EPTCS, 395, 69--76, 2023.
https://doi.org/10.4204/EPTCS.395.5

Fabian Vu, Christopher Happe, Michael Leuschel
Generating interactive documents for domain-specific validation of formal models.
In International Journal on Software Tools for Technology Transfer, 26, Springer, 147--168, 2024.
https://doi.org/10.1007/s10009-024-00739-0

Fabian Vu, Jannik Dunkelau, Michael Leuschel
Validation of Reinforcement Learning Agents and Safety Shields with ProB.
In Proceedings NFM 2024, LNCS, 14627, Springer, 279--297, 2024.

https://doi.org/10.1007/978-3-031-60698-4_16

Jan Gruteser, Michael Leuschel
Validation of RailML Using ProB.
In Proceedings ICECCS 2024, LNCS, 14784, Springer, 245--256, 2024.
https://doi.org/10.1007/978-3-031-66456-4_13

Jan Roßbach, Oliver De Candido, Ahmed Hammam, Michael Leuschel
Evaluating AI-Based Components in Autonomous Railway Systems.
In Proceedings KI 2024, LNAI, 14992, Springer, 190--203, 2024.
https://doi.org/10.1007/978-3-031-70893-0_14

Jan Gruteser, Jan Roßbach, Fabian Vu, Michael Leuschel
Using Formal Models, Safety Shields and Certified Control to Validate AI-Based Train Systems.
In Proceedings FMAS 2024, EPTCS, 411, 151--159, 2024.
https://doi.org/10.4204/EPTCS.411.10

*Geplante Veröffentlichungen:*
KI-LOK – Ein Verbundprojekt über Prüfverfahren für KI-basierte Komponenten im Eisenbahnbetrieb.
In Signal+Draht 1/2, 2025.

Jan Roßbach, Michael Leuschel
Certified Control for Train Sign Classification.
AFMAS 2023.