Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P2933-cd

Safety Argumentation for ML-Enabled Perception Systems for Autonomous Trains State of the Discussion and Perspectives

Carsten Thomas
Computer Engineering, HTW Berlin, Germany. E-mail: carsten.thomas@htw-berlin.de

Mirko Conrad
samoconsult GmbH, Berlin, Germany. E-mail: mirko.conrad@samoconsult.de

Railway is constantly gaining importance as a sustainable mode of transportation. Highly automated train operation is a means to increase the utilization of existing networks. Technical solutions for driverless operation (Grade of Automation GoA3 and higher) typically rely on Machine Learning (ML) components to evaluate sensor data and establish situational awareness. Due to the inherent complexity and black-box nature of ML components, traditional approaches for safety argumentation are not directly applicable to ML-enabled perception systems.
In our paper we present the state of the discussion on this subject and sketch-out potential approaches for technical solutions and the associated safety argumentation. First we discuss safety goals and objectives of perception systems for autonomous trains. We then highlight problem areas that prevent the use of traditional methods for arguing the safety of ML components and propose possible technical solutions and methods at ML component level and at the level of ML-enabled systems as a whole, taking into account the status of work in ongoing major funded projects. Finally we discuss strategies for safety argumentation and the role of safety argumentation as a driver for development decisions.

Keywords: Safety Case, Safety Argumentation, Automatic Train Operation, Artificial Intelligence, ML-Enabled Systems, Artificial Neural Networks

1. Introduction

Railways are becoming increasingly important as a sustainable means of transporting people and goods (cf. SRU (2017)). Nevertheless, growth in the sector through the construction of new railway lines is limited due to the substantial investments and prolonged planning periods involved. Digitalization and *automatic train operation (ATO)* are thus moving into focus as approaches to achieving improvements within the existing rail network through faster train sequences, more flexible utilization of existing railroad lines and lower operating costs (cf. Büker et al. (2024)) .

The transition to driverless or unattended train operation (GoA3 or higher, cf. IEC 62290-1 (2014)) marks a significant milestone to achieve the aforementioned objectives. In this domain, technical solutions frequently employ machine learning (ML) approaches, leveraging methods such as artificial neural networks to evaluate sensor data. Examples include on-board perception systems that allow to identify the train's location and route and to detect potential dangerous obstacles on the train's track.

Safety is paramount in the deployment of ML-enabled ATO solutions. The complexity and unpredictability of real-world environments necessitate a robust safety argumentation framework to demonstrate that these systems can operate safely under diverse permissible conditions. Due to the inherent complexity and black-box nature of ML components, traditional approaches for safety argumentation are applicable to these ML-enabled solutions only to a very limited extent.

This paper aims to explore safety objectives, technical solutions and safety argumentation for ML-enabled systems in autonomous trains, and to highlight important areas for future research. It identifies safety goals and objectives for ML-enabled perception systems (section 3.1), investigates related technical challenges (section 3.2) and discusses potential technical solutions (section 4) and the associated safety argumentation

(section 5).

2. Related Work

In this chapter, we highlight work that focuses on safety objectives for ML-enabled systems for autonomous trains and the approaches to safety argumentation in this context, and provide a brief overview of related standards and guidance from the railway, automotive, and aviation sectors.

2.1. *ML-enabled Systems for Autonomous Trains*

Highly automated driving on mainline railway routes is the core target of current research in Europe and elsewhere. Whilst initially, development of the required technologies, e.g., for ML-enabled perception systems, was the core focus (see, e.g., Ristic-Durrant et al. (2021)), recently projects also focus on investigating the related assurance approaches.

Safety objectives for perception systems for autonomous trains have been the subject of various analyses in the context of autonomous train projects and standardization activities, e.g., Rangra et al. (2018); Braband (2021); Braband et al. (2023); Lahneche et al. (2024). The results of these analyses are partially diverging, which highlights the need for further work, as these safety objectives set the frame for the safety argumentation of ML-enabled systems.

Within the framework of the *Autonomous Train* program at *Railenium*, Tonk et al. (2023) have developed a structured safety assurance methodology for autonomous trains that differentiates three different levels: (1) overall (train) system level, (2) level of the ML-enabled systems including the perception system, (3) level of the ML software component. Various safety-related engineering activities and technical approaches are described according to these levels.

In the *safe.trAIn* project, a safety-enabling architecture, metrics for the performance assessment of AI-enabled systems, and assurance approaches have been developed (Zeller et al. (2023)). The project's safety argumentation strategy is built on five pillars: (1) processes tailored to the perception specifics, (2) non-conventional redundancies in the safety-enabling architecture, (3) sufficient understanding of the causalities of functional behavior, (4) testing with real and simulated data, and (5) safety monitoring during operation, employing methods like out-of-distribution detection. The safety argumentation strategy is guided by a "landscape of AI concerns" developed in the project (Schnitzer et al. (2024)), which is also used to derive verifiable requirements and performance measures used to demonstrate the adequacy of ML-based solution elements.

2.2. *Standards and Guidelines*

Standards and guidelines also provide valuable insight on the challenges in utilizing ML-enabled systems for highly automated vehicle operation as well as guidance on technical solutions and safety argumentation.

Railway: In the railway sector, the certification framework is still strongly focused on "classical", procedural forms of computation (cf. Jenn et al. (2020)). Some standards such as EN 50716 (2023) do mention Artificial Intelligence technologies, but do not yet provide guidance on the preconditions and limits of their application nor on their implementation and certification.

Other domains, such as automotive and aviation, are more advanced regarding the formalization of safety assurance and certification aspects. Adaptation of these approaches into the safety culture of the railway domain is to be expected in the coming years.

Automotive: *ISO/PAS 8800 (2024) 'Road vehicles — Safety and artificial intelligence'* addresses the risk of unintended safety-related behavior at vehicle level due to output insufficiencies, systematic errors, and random hardware errors of AI elements, thereby adding additional aspects complementing the existing automotive safety standards ISO 26262 (2018) and ISO 21448 (2022) for electric/electronic in-vehicle systems.

The most recent version of the *ASPICE* framework (Automotive SPICE 4.0 (2023)) adds a new group of *machine learning engineering (MLE) processes* to augment the pre-existing system engineering, software engineering, and hardware engineering process groups.

Aviation: In the aviation domain, the *ED-324 (202x) 'Process Standard for Development and Certification/Approval of Aeronautical Safety-Related Products Implementing AI'* was planned to be published end of 2024 as *ED-324 / SAE ARP 6983*[a], but has not been released yet.

As an alternative to certification based on fulfilling the detailed objectives of a prescriptive standard, the overarching properties framework Holloway (2019) has been proposed. Under this framework, a set of safety arguments proves that the system possesses the three overarching properties (or meta objectives) 'intent', 'correctness', and 'innocuity'.

The analysis of related work indicates that, while the railway domain yet lacks comprehensive guidance regarding the safety argumentation for ML-based systems, required elements and partial approaches have been developed and demonstrated, and that guidance is available from other domains which may potentially be transferred and adapted for application in the railway sector.

3. Application Context: Perception

To achieve the goal of Grade of Automation (GoA) 3 or higher, all driver tasks must be taken over by on-board automation systems. One of the core functions to be realized by these systems is to monitor the track for obstacles and to derive appropriate action depending on the type of obstacle. The detection and classification tasks are difficult to formalize and thus lend themselves to the use of artificial intelligence – primarily machine learning (ML) – technologies.

In this paper, we focus on such ML-enabled perception systems and the technical and assurance approaches guaranteeing and demonstrating their safety.

3.1. *Safety Goals and Safety Objectives*

Typical *safety goals* for ML-enabled systems for autonomous trains include (cf. Braband et al. (2023)):

- Potentially dangerous objects in the train's standard clearance envelope shall be recognized.
- Persons who unintentionally appear in the in the train's standard clearance envelope or in the suction area of the train shall be recognized.
- Unjustified emergency braking shall be avoided.

For these safety goals, risk acceptance criteria must be identified according to CSM (EU (2013)), following three primary methods:

Use of code of practice: This method focuses on the application of universally accepted rule sets, such as safety norms. Currently, there are no universal rule sets available for safety-critical ML-enabled systems, neither in the railway domain nor in other industries. However, some selected rules from the railway domain could be applicable in combination with the "explicit risk estimation" method.

Explicit risk estimation: In this method, for each of the basic system functions, qualitative and quantitative safety criteria are derived in a detailed analysis (Braband (2021)), also using information on the severity and frequency of events. Braband et al. (2023) have established detailed safety integrity objectives for selected automated driving functions, ranging from SIL1 to SIL3 in exceptional cases. Based on additional considerations extending this work, the *safe.trAIn* project considers 1% probability of failure on demand as an appropriate safety objective.

Use of similar systems as a reference: This method uses similar systems, which are proven in use in a comparable operating environment, as a reference, assuming that a system fulfilling the same safety objectives would be acceptably safe. Unfortunately, existing driverless metro systems are not sufficiently comparable since they operate in a protected environment, such that only selected safety objectives can be taken over.

Using a human train driver as a reference system is an approach currently discussed in industry, but the quantification of train driver performance is difficult and subject of ongoing research.

[a] cf. https://eurocae.net/about-us/working-groups/ (accessed 2025-01-05)

Simplistic approaches lead to the assumption of 10^{-3}/hr human error probability, which – applied to perception systems – would lead to comparatively high safety objectives. On the contrary, recent studies have shown that train drivers have only limited chances for timely obstacles detection and appropriate reaction, such that even in good visibility conditions only 30% of collisions can be avoided (Lahneche et al. (2024)). This strongly suggests the need to revisit the approach considering human driver performance as reference system. One potential input for such reconsideration could be the work currently performed temporary standards working group in Germany on DIN SPEC 91516 *'Human performance with regard to the dynamic driving task for the specification of AI for ATO'*.

3.2. *Technical Challenges*

Using ML-enabled systems in safety-critical applications poses a multitude of challenges regarding suitable technical solutions and the demonstration of their adequacy (Pereira and Thomas (2020); Jenn et al. (2020); Perez-Cerrolaza et al. (2024)).

One group of challenges arises from the fact that AI-enabled systems are primarily applied for "non-algorithmic" computational tasks. Such tasks are often defined in an implicit way, e.g., by ML training data, potentially leading to ambiguity, bias, and lack of completeness in the definition of the task.

Another group of challenges arises from the black-box nature of ML components. Their behavior results from a combination of algorithm and parametrization, where most of the behavior specifics are the result of the parametrization. Lack of explainability and robustness are two challenges often mentioned in this context.

Comparing with traditional approaches for the development and the certification of software-based systems, Dmitriev et al. (2021) differentiate the following issues when using ML technology for safety-critical applications:

Traceability Issue: Conventional software is implemented as human-readable source code, which can be traced back to the specific requirements it implements. In comparison, the functional behavior of an ML component is primarily characterized by a multitude of model parameters, which are calibrated during the training of the model. Due to the fact that an ML model is generally not directly comprehensible by humans, it is practically impossible to trace the values of these model parameters to specific low-level requirements or functions implemented by the ML model (Dmitriev et al. (2021)). As a result, the related traceability objectives of the safety standards cannot be achieved.

Coverage Issue: To assess completeness of software testing, software-related safety standards such as EN 50716 (2023), ISO 26262 (2018) or DO-178C (2011) usually require an assessment of (1) requirements coverage and (2) structural coverage of the source code achieved by the available tests. The first criterion evaluates the extent to which the created tests exercise the intended behavior stipulated by the software requirements. Due to the traceability issue explained above, test coverage of the requirements implicitly expressed by the ML model cannot be demonstrated.

The second criterion assists in identifying source code elements that are not covered by requirements-based tests. These elements may indicate unintended functionality in the source code or deficiencies in the requirements or tests (Dmitriev et al. (2021)). The established metrics for structural code coverage are not well suited for a typical ML model implementation, which has a very simple control flow. Here, a single randomly selected test input can achieve a high level of code coverage, thereby rendering this criterion ineffective for ML models. As a result, additional activities will be necessary to detect unintended functionality in the source code and to achieve confidence in the correctness and completeness of the requirements-based tests (Dmitriev et al. (2021)).

Verification Issue: If an ML model expresses software requirements, the verification objectives of the safety standard for such requirements apply. As an ML model is generally not directly comprehensible by humans, traditional human reviews to assess aspects such as accuracy, consistency,

and traceability are very restricted or impossible. However, testing of the ML model with test data sets can be claimed to support some of the verification objectives (Dmitriev et al. (2021)). Consequently, the verification objectives of traditional software safety standards are at least partially achievable.

4. Solution Approaches

4.1. *ML Component Level*

Addressing the traceability, coverage, and verification issues identified by Dmitriev et al. (2021) for individual ML components is one step towards ensuring and demonstrating the safety of ML-enabled systems.

Traceability issue: One part of the high-level requirements defines rather generic functional and safety requirements for the ML component. These requirements translate into low-level requirements for the computing platform and for the algorithmic realization of the ML component, and can be traced and verified with methods compliant to traditional railway safety standards such as EN 50716 (2023).

The majority of high-level requirements describe the perception tasks, with the definition of the operational design domain (ODD) as an important element. These perception-specific requirements are not refined into traditional low-level requirements, but rather implicitly documented in the data sets used to train and verify the ML component.

Schleiss et al. (2022) have introduced the concept of micro operational design domains (μODD). Defining the ODD as a collection of μODD in a very detailed way enables a refinement of perception-related high-level requirements into a set of intermediate-level requirements which can be individually addressed and verified. Verification includes confirming that the individual μODD are properly represented in the training data, and that the solution performs properly for the respective μODD. Introducing intermediate-level requirements for the individual μODDs partially bridges the semantic gap between the perception-related high-level requirements and the ML model expressing the software requirements. As a result, the scope of the traceability issue will be reduced. In addition, μODDs also target the (requirements) coverage issue, and the verification issue (see below).

Driving this concept one step further, the required behavior for each μODD could be specified explicitly, and could be illustrated with representative examples of input-out data for the ML component. Applying XAI methods, those elements of the ML component can be identified that actively contribute to the fulfillment of the requirement. In this way, traceability from low-level requirements to elements of the ML component could be established, thereby at least partially solving the traceability issue for low-level perception related requirements.

Coverage issue: The concept would also enable to establish sets of test cases required to demonstrate correct behavior of the ML component for each of the μODD. This would adequately address the requirements coverage criterion of the coverage issue.

However, the structural coverage criterion would not be addressable in this way, since currently there is no universally accepted approach to assess the structural coverage of an ML component as a means for quality assurance (Dong et al. (2020); Wang et al. (2024)).

Verification issue: The concept of low-level requirements associated to μODD would also allow to perform detailed tests, thereby establishing confidence that the ML component is adequately performing for these μODD.

These measures addressing traceability, coverage and verification at ML component level applied in combination with additional assurance measures safeguarding, e.g., the quality of training data with respect to representativity and absence of bias, have the potential to ensure and demonstrate a certain level of safety integrity for ML components. Depending on the chosen measures and the level of rigor applied, this level might be similar to ”quality managed (QM)” (ISO 26262 (2018)) or ”basic integrity” (EN 50716 (2023)), or might already achieve an initial level of safety integrity.

4.2. *Architecture Level*

For most safety-critical applications, the lower levels of safety integrity that are achievable for individual ML components are not sufficient. To address these applications, ML-enabled systems must additionally use architectural measures to handle capability insufficiency of individual ML components and to compensate residual errors. Such measures include among others:

- Combination of ML components and conventional components (doer-checker pairs or groups of complementary ML and algorithmic components),
- Parallel dissimilar ML components and voting and/or result merging,
- ML component monitoring during operation (e.g., out-of-distribution checking),
- Plausibility checking with dissimilar information (maps, GNSS, odometry).
- Extensive virtual testing based on synthetic fuzzing (Miller et al. (1990)) of existing test scenarios.

Relevant standards such as ISO/PAS 8800 (2024) provide generic guidance for the development of architectures for ML-enabled systems in safety-critical applications. Yet, these architectures must be developed in light of the specific challenges of the concrete application, taking into account hazards arising from external sources and failure modes of the components forming the ML-enabled system.

Systematic exploration of the possible architecture design space, including possible architectural variability patterns and variability aspects of system components, is key to eventually define an adequate architectural solution for the specific safety-critical application. Product-line engineering methods combined with methods for assessing the safety characteristics of the defined architecture variants help to perform this potentially large exploration task in an efficient manner (Thomas and Jaß (2024)).

5. Safety Argumentation

The solution approaches discussed in section 4 are building blocks for ensuring the safety of ML-enabled systems for highly automated driving in railway. Each of the approaches influences aspects relevant for the safety argumentation at overall system level. In order to eventually define an optimal system solution it is advisable to develop the safety case in parallel and to make development decisions using the safety case as guiding information.

However, most safety-related aspects of a potential ML-enabled system are associated with uncertainty. Many arguments in the safety case are not black-and-white statements, but should rather be considered taking into account the related uncertainties. Approaches to structure and understand, quantify and explicitly manage uncertainties have been described, e.g., by Burton and Herd (2023) and Idmessaoud et al. (2024).

The explicit management of uncertainties related to the safety argumentation follows two objectives: (1) Uncertainty management helps to establish and weight the confidence that can be placed in the combined safety argumentation for the ML-enabled system, thereby supporting the judgment if the realized system behaves adequately safe in its intended usage context. (2) Uncertainty management may be used to drive development decisions at various levels, including product-oriented choices like the selection of optimal architectural variants, component characteristics and algorithms, but also process-oriented choices like the selection of testing approaches and decisions about additional process assurance measures.

6. Summary

In this paper we summarize the state of discussion regarding the safety argumentation for ML-enabled systems for autonomous trains and highlight topics that are essential for further development of the field.

For highly automated driving in railway, the safety objectives to be fulfilled are the driving element that define the technical approaches and the development methods to be applied. Yet, the discussion is continuing regarding the use of human train drivers (and their perception and reaction capabilities) as reference model and the appropriate

consideration of obstacle frequency data. These aspects might lead - if sufficiently substantiated by appropriate data and accepted by authorities and public - to substantially reduced safety objectives for perception systems, making it more viable to provide adequate technical solutions.

Regarding the solutions for ML-enabled perception systems, we differentiate between the core ML components and ML-enabled perception systems as a whole. For ML components, the traceability, coverage, and verification issues highlighted by Dmitriev et al. (2021) must be solved. This is at least partially possible by applying the μODD approach as a means to define, trace and verify low-level requirements for the perception task. Remaining functional deficiencies and residual errors of individual ML components must be mastered through appropriate measures at architectural level. This requires to systematically explore the architectural design space and to evaluate architecture variants regarding their compliance with the defined safety objectives.

The intended safety argumentation should be a driving factor during development, influencing development decisions and thereby making sure that the final solution is optimally fulfilling the safety objectives. To achieve this, the safety case must be created in parallel to the development activities. Assurance uncertainties related to individual arguments should be explicitly managed and should be taken into account when making development decisions based on the intended safety argumentation.

Acknowledgement

This work was partially supported by the project 'CertML – Certifiable machine learning based controls for safety-critical applications' (2022 - 2025), funded by the German Federal Ministry of Education and Research (grant 01IS22029A).

References

Automotive SPICE 4.0 (2023). Automotive SPICE – Process Reference Model, Process Assessment Model. Technical report, VDA QMC.

Braband, J. (2021, Oct). Approach for a risk analysis for automated train operation. *SIGNAL + DRAHT* (113), 25–34.

Braband, J., M. Kinas, C. Klotz, and B. Milius (2023, Jun). Risk analysis for automatic train operation. In A. Oetting and B. Milius (Eds.), *Scientific Railway Signalling Symposium 2023*, pp. 7–24.

Burton, S. and B. Herd (2023). Addressing uncertainty in the safety assurance of machine-learning. *Frontiers in Computer Science 5*.

Büker, T., S. Heller, and E. H. et al. (2024). Zum verkehrlichen Nutzen der Digitalen Schiene Deutschland. In *Der Eisenbahningenieur Feb. 2024:47-52*.

Dmitriev, K., J. Schumann, and F. Holzapfel (2021). Toward certification of machine-learning systems for low criticality airborne applications. In *2021 IEEE/AIAA 40st Digital Avionics Systems Conference (DASC)*.

DO-178C (2011). RTCA DO-178C:2011 Software Considerations in Airborne Systems and Equipment Certification. Publicly available specification, RTCA.

Dong, Y., P. Zhang, J. Wang, S. Liu, J. Sun, J. Hao, X. Wang, L. Wang, J. Dong, and T. Dai (2020). An empirical study on correlation between coverage and robustness for deep neural networks. In *2020 25th International Conference on Engineering of Complex Computer Systems (ICECCS)*, pp. 73–82.

ED-324 (202x). Process Standard for Development and Certification Approval of Aeronautical Products Implementing AI. Technical report, EUROCAE.

EN 50716 (2023). EN 50716:2023 Railway Applications - Requirements for software development. Technical report, CENELEC.

EU (2013). Commission Implementing Regulation (EU) No 402/2013 on the common safety method for risk evaluation and assessment.

Holloway, C. M. (2019). Understanding the overarching properties. Technical report, NASA.

Idmessaoud, Y., J.-L. Farges, E. Jenn, V. Mussot, A. Fernandes Pires, F. Chenevier, and R. Conejo Laguna (2024). Uncertainty in assurance case pattern for machine learning. In *Embedded Real Time Systems (ERTS)*, Toulouse, France.

IEC 62290-1 (2014). IEC 62290-1 Railway appli-

cations - Urban guided transport management and command/control systems - Part 1: System principles and fundamental concepts. Technical report, IEC.

ISO 21448 (2022). ISO 21448:2022 Road vehicles – Safety of the intended functionality. International standard, International Organization for Standardization, Geneva, CH.

ISO 26262 (2018). ISO 26262:2018 Road vehicles – Functional safety. International standard, International Organization for Standardization, Geneva, CH.

ISO/PAS 8800 (2024). ISO/PAS 8800:2024 Road vehicles – Safety and artificial intelligence. Publicly available specification, International Organization for Standardization, Geneva, CH.

Jenn, E., A. Albore, F. Mamalet, G. Flandin, C. Gabreau, H. Delseny, A. Gauffriau, H. Bonnin, L. Alecu, J. Pirard, et al. (2020). Identifying challenges to the certification of machine learning for safety critical systems. In *European congress on embedded real time systems (ERTS 2020)*.

Lahneche, O., A. Haag, P. Dendorfer, V. Aravantinos, M. Guilbert, and M. Sallak (2024). Analysing railway accidents: A statistical approach to evaluating human performance in obstacle detection. In J. Pombo (Ed.), *Proceedings of the Sixth International Conference on Railway Technology: Research, Development and Maintenance*, Volume 7 of *Civil-Comp Conferences*, Prague, Czech Republic.

Miller, B., L. Fredriksen, and B. So (1990). An empirical study of the reliability of unix utilities. In *Commun ACM. 1990 Dec 1; 33 (12). p. 32–44*.

Pereira, A. and C. Thomas (2020). Challenges of machine learning applied to safety-critical cyber-physical systems. *Machine Learning and Knowledge Extraction 2*(4), 579–602.

Perez-Cerrolaza, J., J. Abella, M. Borg, C. Donzella, J. Cerquides, F. J. Cazorla, C. Englund, M. Tauber, G. Nikolakopoulos, and J. L. Flores (2024, apr). Artificial intelligence for safety-critical systems in industrial and transportation domains: A survey. *ACM Comput. Surv. 56*(7).

Rangra, S., M. Sallak, W. Schön, and F. Belmonte (2018). Risk and safety analysis of main line autonomous train operation: Context, challenges and solutions. In *Conference: Congrès Lambda Mu 21 de Maîtrise des Risques et de Sûreté de Fonctionnement*.

Ristic-Durrant, D., M. Franke, and K. Michels (2021). A review of vision-based on-board obstacle detection and distance estimation in railways. *Sensors 21*(10), 3452.

Schleiss, P., Y. Hagiwara, I. Kurzidem, and F. Carella (2022). Towards the quantitative verification of deep learning for safe perception. In *2022 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, pp. 208–215.

Schnitzer, R., A. Hapfelmeier, S. Gaube, and S. Zillner (2024). *AI Hazard Management: A Framework for the Systematic Management of Root Causes for AI Risks*, pp. 359–375. Springer Nature Singapore.

SRU (2017). Time to take a turn: Climate action in the transport sector. Technical report, German Advisory Council on the Environment (SRU).

Thomas, C. and P. Jaß (2024). Product line engineering applied to perception system architectures for autonomous trains. In *2024 IEEE International Conference on Recent Advances in Systems Science and Engineering*, pp. 1–9.

Tonk, A., M. Chelouati, A. Boussif, J. Beugin, and M. E. Koursi (2023). A safety assurance methodology for autonomous trains. *Transportation Research Procedia 72*, 3016–3023. TRA Lisbon 2022 Conference Proceedings Transport Research Arena (TRA Lisbon 2022),14th-17th November 2022, Lisboa, Portugal.

Wang, Z., S. Xu, L. Fan, X. Cai, L. Li, and Z. Liu (2024). Can coverage criteria guide failure discovery for image classifiers? an empirical study. *ACM Trans. Softw. Eng. Methodol. 33*(7).

Zeller, M., T. Waschulzik, M. Rothfelder, and C. Klein (2023). Safety assurance of a driverless regional train – Insight in the safe.trAIn project. In *2023 IEEE 34th International Symposium on Software Reliability Engineering Workshops (ISSREW)*, pp. 41–42.