# Abstract Traffic Scenario Specifications for Improving Safety and Trustworthiness of AI-based Mobility Systems

Ishan Saxena[0000-0003-0575-4402], Dominik Grundt[0000-0002-8233-7429],
Eike Möhlmann[0000-0003-3815-6353], and Bernd Westphal[0000-0002-6824-0567]

German Aerospace Center (DLR),
Institute of Systems Engineering for Future Mobility,
Oldenburg, Germany

## Extended Abstract

### Introduction

Autonomous vehicles increasingly depend on AI technologies for automated driving. There is a large semantic gap between sensor data processing and fusion, perception chains, and planning in automated driving and high-level safety requirements. Both, traditional requirements engineering methods and traditional methods to verify and validate that automated driving systems meet their specified requirements face new challenges. One approach to guaranteeing acceptable safety of automated driving is the concept of the operational design domain (ODD) [18], the *safety of intended functionality* (SOTIF) [12] and scenario-based verification [20]. An ODD description specifies conditions for which safe operation of a particular system is guaranteed. For road traffic this can be weather and road conditions, but can also consider geographic constraints such as limitations to certain quarters of a city. Checking the ODD conditions at runtime is then one building block in the safety concept of automated driving in that a violation of the conditions should trigger safeguarding operations such as minimum risk manoeuvres. The ODD concept can be the basis for a gradual introduction and continuous improvement of the driving function based on the analysis of data from operation time when some ODD conditions have been violated.

If automated driving is brought to public roads or targets passenger transportation, there is the additional demand for trustworthiness of the AI-based driving system. There are proposals [4, 8] to complement the verification and validation of AI-based driving systems with monitoring and shielding mechanisms that detect traffic situations and scenarios that need attention of the operators or developers, e.g., based on world knowledge such as physical plausibility of data provided by the perception chain. The approach of scenario-based verification and validation of driving systems [14] has led to the development of specific formalisms to abstractly capture and specify driving behaviour, e.g., to formalise traffic rules and regulations related to driving. One such formalism is Traffic Sequence Charts (TSC), originally developed for specifying spatio-temporal properties in the road-transportation domain [6, 7, 5], and recently adapted to the maritime domain [1].

In this Extended Abstract, we show how abstract scenario specification formalisms from the discipline of Requirements Engineering can be used to devise procedures for continuous safeguarding and improvement of autonomous driving systems thereby contributing towards their increased reliability, safety, and trustworthiness. In particular, we propose how traffic behaviour can be classified at system runtime using scenario-based Verification and Validation techniques, namely TSC runtime monitoring. Here, we extend the concept of novel traffic behaviour from [15] to define out-of-distribution, anomalous and implausible traffic behaviour. We further present a framework that uses these concepts and TSC runtime monitoring to continuously check traffic behaviour at vehicle runtime and triggers hand-over actions and scenario data recording upon detection of unseen or erroneous behaviour. The framework also utilises the work [10] by allowing passengers to be shown abstract scenario-based explanations of future vehicle behaviour, even when an unseen or erroneous behaviour is encountered.

## Terminology

The literature on AI-technology in safety critical systems considers numerous approaches [13] on how to achieve safe operation despite the opacity of the involved AI models. One branch proposes to complement the AI models with knowledge from the corresponding domain and to trigger safeguarding mechanisms in case of, e.g., implausibilities (shielding [4]).

Here, we focus on three properties that can be assigned to situations and scenarios in order to assess the performance of driving systems wrt. them, namely *out-of-distribution (OOD)*, *anomalies*, and *plausibility* (see [21] for a taxonomy of generalised OOD), and in addition the concept of *novelty*. A novelty is data and its processing outcome (images, sensor data, object lists, etc.) that has (as such or conceptionally) not been considered during development of the AI component. An example in automated driving could be a new kind of vehicle or traffic participant, or a new driving behaviour of surrounding traffic caused by manual traffic management from a traffic police officer.

Out-of-distribution (OOD) data is data that has not been used to train the AI component. Note that OOD data is not necessarily novel (because it may have intentionally not been used), but a novelty is OOD. An anomaly is input or output data of the AI component that is not normal wrt. a given definition of normality. An example in automated driving is the observation of own or other participants' behaviour that is against a traffic law that is usually obeyed in the region of operation. A similar property is plausibility. Input or output data can be assessed for a given definition of plausibility which is based on world knowledge [3]. An example for an implausible observation would be another car hovering 5 m above the ground or the proposal to jump over cars to overtake. Note that an implausible observation is an anomaly, but an anomaly may well be plausible according to the considered knowledge.

The concepts named above are mainly considered for improving the safety of AI-based systems and shielding mechanisms that rely on these concepts (such as anomaly detection at runtime) can contribute to trustworthiness through the safety argumentation. Another branch of work towards trustworthiness of AI-based systems considers explanations (explainable AI), of which there are works [17] that study notions of explanations of the inner workings of AI components (mostly for developers and trained personnel). In contrast, the previous work [10] addresses users or passengers of AI-based systems in automated driving by providing explanations on the abstraction level of traffic situations and scenarios. The procedure shows abstract descriptions of *expected* manoeuvres that are provided for trigger specifications corresponding to the current traffic situation.

### Towards Trustworthy Automated Driving Systems

Scenario-based approaches in the development of automated driving systems [20] distinguish *(traffic) situations* and *(traffic) scenarios*. A concrete (traffic) situation (in a given scope) is a model of the traffic participants and environment in the scope (e.g., geographical area, or perception horizon) according to a given data model (possibly including spatial or conceptual relations between traffic participants) at one point in time. This can be attributes like speed, position, heading, or relations like the nearest car to 'ego' on the opposite lane. A concrete *(traffic) scenario* is a temporal evolution of concrete situations over a period of time. An abstract (traffic) situation specification is a formal description of a set of concrete situations. In the TSC formalism, so-called spatial views serve as abstract situation specifications over an object-oriented data model. Given a spatial view $\mathcal{SV}$ and a concrete situation $\sigma$, there is a satisfaction relation $\sigma \models \mathcal{SV}$. An alternative on slightly different level of abstraction would be Abstract Scene Graphs [16]. Analogously, an abstract scenario specification is a formal description of a set of concrete scenarios, in case of TSC this is the purpose of charts $\mathcal{T}$ over spatial views. TSCs have been used to formalise requirements on driving behaviour [19] and to capture traffic rules from the maritime domain [11, 2]. In [15], we have shown how to capture novel situations with a collection (or catalogue) of spatial views. For a finite set $\mathcal{C}$ of spatial views, we write $\sigma \in \mathcal{C}$ if and only if there is a spatial view $\mathcal{SV} \in \mathcal{C}$ s.t. $\sigma \models \mathcal{SV}$, and analogously for TSC charts.

Given sets of abstract situation specifications that capture *nominal behaviour* $\mathcal{C}_{nom}$, *plausible behaviour* $\mathcal{C}_{pls}$, *in-distribution behaviour* $\mathcal{C}_{ind}$ and *known behaviour* $\mathcal{C}_{knn}$, Algorithm 1 shows how traffic situations encountered by an autonomous vehicle during its operation can be checked continuously using, e.g., the TSC monitoring approach of [19]. For situations $\sigma$, that are known and normal, plausible, and in-distribution, the future behaviour of the vehicle is explained using a visual representation of expected and unexpected manoeuvres, as shown in [10], to the passengers and driving continues. Any other case needs attention. E.g. encountering a wrong-way driver on a highway would violate the check for nominal behaviour. The cases where an anomalous, implausible, or

**Algorithm 1** Using abstract situation specifications for safety and trustworthiness of AI-based driving functions.

```
if $\sigma \in \mathcal{C}_{nom} \cap \mathcal{C}_{pls} \cap \mathcal{C}_{ind} \cap \mathcal{C}_{knn}$ then
    provide explanation for manoeuvre
else
    if $\sigma \in \mathcal{C}_{knn}$ then
        provide explanation for hand-over due to $\mathcal{C}_{nom}$, $\mathcal{C}_{pls}$, or $\mathcal{C}_{ind}$ as applicable
    else
        provide explananation for hand-over due to $\mathcal{C}_{knn}$ & start scenario data recording
    end if
end if
```

out-of-distribution behaviour are detected may need a (remote) driver to take over or the vehicle to perform a minimum risk manoeuvre. In this case, the algorithm can give the information for which (conceptual) reason the safety procedure is triggered. If the situation is unknown wrt. $\mathcal{C}_{knn}$ though, a novel situation is detected and a recording of the current concrete driving is started for offline analysis. This selective and labelled recording of sensor data helps reduce on-vehicle data storage needs and further saves developer-time and -resources during offline analysis of the data, which is required for continuous improvement of the driving system.

Following the same procedure as [15] for novel situations, an analysis of the concrete scenario recording should give rise to an extended or improved catalogue $\mathcal{C}'_{knn}$ that covers in particular the just analysed recording. The new catalogue $\mathcal{C}'_{knn}$ can be obtained by generalising existing spatial views (or abstract scene graphs), or by adding new ones. The benefit of abstract situation (and scenario) specifications is that they cover a whole set of concrete situations at once based on an understanding of the developers. The newly recorded scenario should then also be considered in $\mathcal{C}'_{nom}$, $\mathcal{C}'_{pls}$, $\mathcal{C}'_{ind}$ so that nominal, plausible, and in-distribution situations are in particular not novel according to $\mathcal{C}'_{knn}$.

## Conclusion

AI-based driving systems are not perfect, because in the open world context we can not assume that we have trained models for all situations. Hence, there is a possibility of encountering unseen or erroneous behaviour in traffic. We first define the concepts of out-of-distribution, anomalous, implausible and novel traffic behaviour. We then present a framework that uses the requirements engineering formalism of TSCs and runtime monitoring to classify the observed traffic behaviour and trigger safeguarding mechanisms or sensor data recording as needed. The framework helps AI developers by 1) providing a mechanism that allows only recording data where something unseen/erroneous happened (saving of data storage space and developer time), and 2) providing additional information about why the data was captured, which further saves developer time and aids in understanding of data. We also help vehicle passengers by providing explanations about the future actions of the automated driving system and thereby helping increase the trustworthiness of the system. We further show how techniques usually used for monitoring systems wrt. system requirements can simultaneously help in this case.

Future work includes implementing and testing the framework on simulation and real-world datasets and testing the real-time capabilities of the runtime monitors on automotive grade hardware. In addition, tool support is required for scenario catalogue coverage argumentations. Finally, efforts are also required for expanding the framework to include uncertainty-aware runtime monitoring techniques [9].

**Acknowledgments.** The research leading to these results is funded by the German Federal Ministry of Education and Research under grant agreement No. 16MEE044 (EdgeAI-Trust), by the Chips Joint Undertaking under grant agreement No. 101139892 (EdgeAI-Trust) and by the German Federal Ministry for Economic Affairs and Energy within the project "NXT GEN AI METHODS - Generative Methoden für Perzeption, Prädiktion und Planung".

## References

1. Austel, A., Hake, G., Wetzig, N.: A visual formalism for the specification of maritime traffic scenarios. In: Guizzardi, R., Araújo, J. (eds.) Requirements Engineering: Foundation for Software Quality. pp. 75–90. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-21423-2_6
2. Austel, A., Wetzig, N., Westphal, B.: Formal specification of colreg steering and sailing rules (2026), accepted for publication at SMRC x MTEC/ICMASS 2026
3. Borchers, P., Hagemann, W., Grundt, D., Werner, T., Müller, J.: Using traffic sequence charts for knowledge formalization and ai-application. In: Arai, K. (ed.) Intelligent Systems and Applications. pp. 198–220. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-66428-1_12
4. Brorholt, A.H., Jensen, P.G., Larsen, K.G., Lorber, F., Schilling, C.: Shielded reinforcement learning for hybrid systems. In: Steffen, B. (ed.) Bridging the Gap Between AI and Reality. pp. 33–54. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-46002-9_3
5. Damm, W., Kemper, S., Möhlmann, E., Peikenkamp, T., Rakow, A.: Using traffic sequence charts for the development of havs. In: ERTS. pp. 1–11. SEE and 3AF associations, Toulouse, France (2018), https://hal.science/hal-01714060
6. Damm, W., Kemper, S., Möhlmann, E., Peikenkamp, T., Rakow, A.: Traffic sequence charts - from visualization to semantics. Tech. rep. (10 2017). https://doi.org/10.13140/RG.2.2.15190.42563
7. Damm, W., Möhlmann, E., Peikenkamp, T., Rakow, A.: A Formal Semantics for Traffic Sequence Charts, pp. 182–205. Springer (2018). https://doi.org/10.1007/978-3-319-95246-8_11
8. Ferreira, R.S., Guérin, J., Delmas, K., Guiochet, J., Waeselynck, H.: Safety monitoring of machine learning perception functions: A survey. Computational Intelligence **41**(2), e70032 (2025). https://doi.org/10.1111/coin.70032

9. Finkbeiner, B., Fränzle, M., Kohn, F., Kröger, P.: Stream-based monitoring under measurement noise. In: Ábrahám, E., Abbas, H. (eds.) Runtime Verification. pp. 22–39. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-74234-7_2
10. Grundt, D., Saxena, I., Petersen, M., Westphal, B., Möhlmann, E.: Context-aware, ante-hoc explanations of driving behaviour. In: Luckcuck, M., Schwammberger, M., Xu, M. (eds.) FMAS. EPTCS, vol. 436, pp. 114–135 (Nov 2025). https://doi.org/10.4204/EPTCS.436.11
11. Hake, G., Austel, A., Becker, J.S., Putze, L., Wetzig, N.: Safety assessment of maritime autonomous surface ships: A scenario-based approach. Journal of Physics: Conference Series **3123**(1), 012022 (oct 2025). https://doi.org/10.1088/1742-6596/3123/1/012022
12. International Organization for Standardization: Road vehicles – safety of the intended functionality. ISO 21448:2022 (2022), iSO Standard
13. Könighofer, B., Bloem, R., Ehlers, R., Pek, C.: Correct-by-construction runtime enforcement in ai – a survey. In: Raskin, J.F., Chatterjee, K., Doyen, L., Majumdar, R. (eds.) Principles of Systems Design: Essays Dedicated to Thomas A. Henzinger on the Occasion of His 60th Birthday. pp. 650–663. Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-22337-2_31
14. Riedmaier, S., Ponn, T., Ludwig, D., Schick, B., Diermeyer, F.: Survey on scenario-based safety assessment of automated vehicles. IEEE Access **8**, 87456–87477 (2020). https://doi.org/10.1109/ACCESS.2020.2993730
15. Saxena, I., Grundt, D., Möhlmann, E., Westphal, B.: Towards runtime detection of novel traffic situations. In: OVERLAY. vol. 4142. CEUR Workshop Proceedings (Dec 2025). https://doi.org/10.5281/zenodo.18298437
16. Saxena, I., Westphal, B., Fränzle, M.: Abstract scene graphs: Formalizing and monitoring spatial properties of automated driving functions. In: Luckcuck, M., Schwammberger, M., Xu, M. (eds.) FMAS. EPTCS, vol. 436, pp. 3–14 (Nov 2025). https://doi.org/10.4204/EPTCS.436.3
17. Schwammberger, M., Mirandola, R., Wenninghoff, N.: Explainability engineering challenges: Connecting explainability levels to run-time explainability. In: Longo, L., Lapuschkin, S., Seifert, C. (eds.) Explainable Artificial Intelligence. pp. 205–218. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-63803-9_11
18. Shakeri, A.: Formalization of operational domain and operational design domain for automated vehicles. In: QRS-C. IEEE (April 2024). https://doi.org/10.1109/QRS-C63300.2024.00131
19. Stemmer, R., Saxena, I., Panneke, L., Grundt, D., Austel, A., Möhlmann, E., Westphal, B.: Runtime monitoring of complex scenario-based requirements for autonomous driving functions. SCP **244**, 103301 (2025). https://doi.org/10.1016/j.scico.2025.103301
20. Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., Maurer, M.: Defining and substantiating the terms scene, situation, and scenario for automated driving. In: ITSC. pp. 982–988 (2015). https://doi.org/10.1109/ITSC.2015.164
21. Yang, J., Zhou, K., Li, Y., Liu, Z.: Generalized out-of-distribution detection: A survey. Int. J. Comput. Vis. **132**(12), 5635–5662 (2024). https://doi.org/10.1007/S11263-024-02117-4