Safe AI in the Automotive Domain
Track at AISoLA 2023

Falk Howar[1] and Hardi Hungar[2]

[1] Dortmund University of Technology and Fraunhofer ISST, Dortmund, Germany
falk.howar@tu-dortmund.de
[2] German Aerospace Center, Braunschweig, Germany
hardi.hungar@dlr.de

Abstract. Today, the most prominent application of AI technology in the automotive domain is in the realm of environment perception. The diversity of the traffic environment and the complexity of sensor readings make it impossible to specify and implement perception functionality manually. Deep learning technology, on the other hand, has proven itself capable of solving the task very well. However, it is important to note that effectiveness alone does not guarantee a comprehensive solution, and the issue of validation currently remains unsatisfactorily resolved. The track provided different perspectives on the challenges pertaining to the use of AI/ML technology in highly automated driving functions, including considerations on safety verification and validation techniques for AI-based autonomous vehicles, formal methods and their application in assuring the safety of AI-based autonomous systems robustness and resilience of AI algorithms in uncertain and open environments, system architectures for AI-based autonomous vehicles, data-driven approaches for safety assurance and risk analysis in autonomous driving, safety standards, regulations, and certification processes for AI-based autonomous vehicles, as well as testing, simulation, and validation methodologies for autonomous vehicle systems.

Track Description

Autonomous vehicles and highly automated driving functions have been a focus of research and development for roughly two decades since the 2004 DARPA grand challenge [5]. Most major carmakers and tech companies as well as a bunch of startups have engaged in a fierce competition for the first truly driverless vehicle.

Today (2023), fleets of driverless taxis are being tested in San Francisco (California), Phoenix (Arizona), and in several other cities across the world [8]. The company Cruise, e.g., operates driverless taxis that do not have a safety driver in the vehicle in San Francisco. Tesla delivered the biggest fleet of vehicles with a advanced driver assistance systems: Autopilot / FSD is a SAE Level 2 assistance systems that controls lateral and longitudinal movement of a vehicle under constant supervision of its driver. As of early 2023, more than 360,000 vehicles

outfitted with this system are being driven on public roads by their owners [9]. Mercedes Benz is the first company to announce the release of a SAE Level 3 automated driving function. The so-called "drive pilot" will take over driving completely on highways in certain conditions. The human driver/operator does not have to constantly supervise the systems but must be able to take over driving when requested by the system within a reasonable amount of time. The system has been certified in the U.S. state of Nevada in January of 2023 and is announced to be released in the U.S. in 2024 [3].

From this description of the current state of development, one could incorrectly infer that autonomous driving is a solved problem. This, however, is not true. Several major challenges remain unsolved, e.g., pertaining to system design and development, safety assurance, as well as legal and economic aspects (which we will not further discuss here):

System Design and Development

One of the greatest challenges for automated driving lies in the complexity and heterogeneity of the environment in which the vehicle must operate. Above all, the automation has to form a sufficiently comprehensive and accurate picture of the environment. All relevant elements must be seen, recognized, and assessed as far as possible. And where uncertainties remain, these must be taken into consideration. AI/ML technology today provides the essential building blocks for practical solutions (e.g., semantic classification for analyzing camera images).

Models for these tasks, however, are trained, which relies on training data, and when the training data does not include certain situations, the vehicle will not recognize these situations. Reported examples of misclassified or missed objects include chlid-like objects [6]. Then, these models do not encode common sense: they e.g., have been reported to recognize traffic lights transported in the back of a truck as floating on the freeway. Mechanisms for dealing with features of the environment that were not present in the training data and for validating the perceived environment remain to be developed.

Safety Assurance

To this day, it is not clear if and how we can document the safety of automated driving systems. The used AI/ML components are black-boxes. Verification techniques for such components are still in their infancy and mostly focus on robustness. It remains an open question, if and how their intended behavior can be specified beyond the set labelled training data. At the system level, it is not clear how safety can be assured in complex open environments. One current strategy is breaking down automated driving into many specific operational design domains (ODD)s and then specifying, developing, testing, and releasing functions for domains of increasing complexity and increasingly associated with risk over time – starting with highway driving, to urban driving, to mixed environments.

A number of industry standards is being developed that e.g., mandate ODD specification (ISO 34503 [7]), scenario-based testing of the intended functionality (ISO 21448 [1]), and safety of automotive AI components (ISO 8800 [2]). However, finding a balance between pace of deployment and safety is not trivial: Autonomous taxis have caused disruptions and outright dangerous situations in their urban environments. Reported incidents include blocking lanes when autonomous driving fails, not following police instructions, blocking ambulances, interfering with fire fighters, and creating a multiple vehicle roadblock close to a bigger event (triggered by overload in the mobile network) [4].

Contributions

The track on Safe AI in the Automotive Domain at AISoLA 2023 aimed at bringing together researchers, practitioners, and experts from formal methods, AI/ML communities, and the automotive domain to discuss the sketched challenges in the broader context of "bridging the gap between AI and reality". The contributions to this track study different aspects of the role of AI technology in autonomous systems. They span the spectrum from requirements to implementation to verification of AI in perception, and the approaches partly employ AI themselves.

1. Starting with the requirements for the systems, today there is no established approach how to formalize them. First of all, a language is needed that precisely grasps the phenomena of the traffic world. The presentation on *"Situation Recognition in Complex Operational Domains using Temporal and Description Logics - A Motivation from the Automotive Domain"* by Westhofen, Neurohr, Neider and Jung proposes to use Description Logics (ontologies) for this purpose. These provide a basis for expressing the traffic objects, their states and static relationships to each other. On top of that, they use temporal logic operators, with which durations and sequences of conditions can be described. This results in a description language that is powerful enough to express the relevant sequences, such as the way in which a particular critical situation arises. On the other hand, the constructors of the language are chosen such that it remains computable when a description applies to an observed sequence of events. The authors announce that they will develop an evaluation routine for their language: *"Mission-Time Linear Temporal Logic over Conjunctive Queries"*.
2. When considering the need to prove the safety of the system for homologation, the standards ISO 26262 (functional safety) and ISO 21448 (safety of the intended functionality, SOTIF) must be observed. A SOTIF analysis entails the study of triggering conditions leading to unintended functionality. How to identify, analyze and test such condition are questions only sparsely covered by research so far. *"Identifying and Testing SOTIF Triggering Condition for the Safety Verification and Validation of Automated Driving Systems"* by Zhu and Howar addresses this topic systematically.

The contribution provides a formalization of three main types of triggering conditions. Then, it introduces a knowledge-driven method for systematically identifying such conditions. With a data-driven method, scenarios relevant to a condition can be extracted from real-life test data. And based on that, a strategy is developed by which triggering conditions can be incorporated into a testing process complying to the requirements of the ISO 21448.

3. Concerning the functionalities of perception, their AI based implementation makes verification and validation very difficult. In his presentation on *"A Note on Confidence Awareness in Automotive Perception"*, Hungar makes a point of making confidence in perception a first-order citizen in the argumentation. Results of verification and validation would be much better if the quality information of the perception output was computed compositionally over the system architecture, starting with adequately captured current quality of sensor readings. The AI components interpreting the sensor readings would be one stage of the system particularly important to be characterized in their contribution to potential inaccuracies or mistakes. Also, the sensor models used in simulating an automated driving system would have to produce the additional quality information.

4. Another aspect of compositionality concerns not the verification, but the construction of the perception itself. The current focus of applying AI technology is on single AI models, i.e., it is model-centric, disregarding the challenges of engineering systems with multiple components that need to interact to realize complex functionality. Applications of machine learning (ML) should be able to support architectures that can integrate and chain ML components. This requires systems-centric methods and tools. This is discussed in *"Towards ML-Integration and Training Patterns for AI-Enabled Systems"* by Peldszus, Knopp, Sens, and Berger. They analyze the limitations of currently applied training processes when engineering multi-ML systems, and discuss possible patterns for training and integration to facilitate the effective and efficient development, maintenance, and evolution of such complex systems.

5. Even if today's AI technology, both in multi-level systems as well as in the form of single models, does not yet provide perception results of the desired level of confidence, in the way one makes use of the results one may attain the desired level of safety. This is shown by Fränzle in *"Maximizing Confidence in Safety-Critical Decisions of Automated Vehicles that are Grounded in ML-based Environmental Perception — Rendering AVs 'Safer than Perception'"*. The approach starts from the observation that critical maneuvers are generally safeguarded by complex spatio-temporal conditions that combine multiple percepts. Evaluation of these conditions, and thereby drawing safety-relevant decisions, exposes all kinds of masking effects between individual misperceptions. The masking can be influenced by rewriting the conditions. This implies that these conditions can be analyzed and modified for their error propagation, optimizing them to limit the negative safety-

related effect of ML-induced misperceptions to a societally acceptable level.

6. Finally, AI techniques can be used not only in the construction of automated driving systems, but also in their verification, as Hungar presents in the contribution *"Using AI in the Verification and Validation of Automated Driving Systems"*. This includes the analysis of data from the real world and from simulation, recognition of maneuvers in real-world data, detection and evaluation of criticality, construction of scenarios, compilation of scenario catalogs, different kinds of simulations, in particular the exploration of scenario spaces, and assignment of real-world data to scenarios.

Moreover, the track included the demonstration of the DevOps-inspired process implemented by TU Dortmund university's formula student team in the development of their autonomous driving system.

References

1. ISO 21448:2022. Road vehicles – Safety of the intended functionality. Standard, International Organization for Standardization, Geneva, CH, June 2022.
2. ISO/CD PAS 8800. Road Vehicles -— Safety and artificial intelligence. Standard, International Organization for Standardization, Geneva, CH, (under development).
3. Andrew J. Hawkins (The Verge). Mercedes-Benz is the first to bring Level 3 automated driving to the US. `https://www.theverge.com/2023/1/27/23572942/mercedes-drive-pilot-level-3-approved-nevada`, 2023. Accessed September 2023.
4. Public Utilities Comission Of The State Of California. Resolution approving authorization for cruise llc's expanded service in autonomous vehicle passenger service phase i driverless deployment program. `https://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M516/K812/516812218.PDF`, 2023. Accessed September 2023.
5. Defense Advanced Research Projects Agency. Grand Challenge 2004: Final Report. `https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/DARPA/15-F-0059_GC_2004_FINAL_RPT_7-30-2004.pdf`, 2004. Accessed September 2023.
6. Edward Helmore (The Guardian). Tesla's self-driving technology fails to detect children in the road, group claims. `https://www.theguardian.com/technology/2022/aug/09/tesla-self-driving-technology-safety-children`, 2023. Accessed September 2023.
7. Road Vehicles — Test scenarios for automated driving systems — Specification for operational design domain. Standard, International Organization for Standardization, Geneva, CH, August 2023.
8. Joann Muller (Axios). Robotaxis hit the accelerator in growing list of cities nationwide. `https://www.axios.com/2023/08/29/cities-testing-self-driving-driverless-taxis-robotaxi-waymo`, 2023. Accessed September 2023.
9. National Highway Traffic Safety Administration. Full Self-Driving Software May Cause Crash. `https://static.nhtsa.gov/odi/rcl/2023/RCAK-23V085-2525.pdf`, 2023. Accessed September 2023.