

Safety Assurance for Autonomous Mobility

Jyotirmoy Deshmukh^{*1}, Bettina Könighofer^{*2}, Dejan Ničković^{*3}, and Filip Cano^{†4}

1 USC – Los Angeles, US. jyotirmoy.deshmukh@usc.edu

2 TU Graz, AT. bettina.koenighofer@iaik.tugraz.at

3 AIT – Austrian Institute of Technology – Wien, AT. dejan.nickovic@ait.ac.at

4 TU Graz, AT. filip.cano@iaik.tugraz.at

Abstract

This report documents the program and the outcomes of the Dagstuhl Seminar “Safety Assurance for Autonomous Mobility” (24071). The seminar brought together an interdisciplinary group of researchers and practitioners from the fields of formal methods, cyber-physical systems, and artificial intelligence, with a common interest in autonomous mobility. Through a series of talks, working groups, and open problem discussions, participants explored the challenges and opportunities associated with ensuring the safety of autonomous systems in various domains, including industrial automation, automotive, railways, and aerospace. Key topics addressed included the need for industrial-grade autonomous products to operate reliably in safety-critical environments, highlighting the lack of standardized procedures for obtaining safety certifications for AI-based systems. Recent advancements in the verification and validation (V&V) of autonomous mobility systems were presented, focusing on requirements verification, testing, certification, and correct-by-design approaches. Overall, the seminar provided a comprehensive overview of the current state and future directions in safe autonomous mobility, emphasizing the need for interdisciplinary collaboration and innovation to address the complex challenges in this rapidly evolving field.

Seminar February 11–16, 2024 – <https://www.dagstuhl.de/24071>

2012 ACM Subject Classification Applied computing → Transportation; Computer systems organization → Embedded and cyber-physical systems; Hardware → Robustness; Software and its engineering → Software verification and validation

Keywords and phrases aerospace, automotive, autonomy, formal methods, railway

Digital Object Identifier 10.4230/DagRep.14.2.95

1 Executive Summary

Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT)

Jyotirmoy Deshmukh (USC – Los Angeles, US)

Bettina Könighofer (TU Graz, AT)

Filip Cano (TU Graz, AT)

License © Creative Commons BY 4.0 International license

© Dejan Ničković, Jyotirmoy Deshmukh, Bettina Könighofer, and Filip Cano

As autonomous mobility systems gain traction worldwide, ensuring their safety, robustness, and dependability has become a paramount concern for their implementation at scale. The Dagstuhl Seminar on “Safety Assurance for Autonomous Mobility” gathered experts from academia, and industry to address the critical challenges and opportunities posed by the rapid

* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Safety Assurance for Autonomous Mobility, *Dagstuhl Reports*, Vol. 14, Issue 2, pp. 95–119

Editors: Jyotirmoy Deshmukh, Bettina Könighofer, and Dejan Ničković



Dagstuhl Reports

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

growth of autonomous technologies across various mobility domains, including automotive, aerospace, robotics, and railways. This seminar provided a much-needed platform for researchers and practitioners to exchange insights, collaborate on emerging ideas, and build a shared understanding of safety assurance in this rapidly evolving field.

Seminar Context and Structure

The seminar convened a diverse and multidisciplinary group of participants, each bringing specialized expertise in formal methods, software verification, embedded systems, and transportation safety. We believe that autonomous mobility is a field where progress cannot be just limited to academic research. The ideas and methods developed in theoretical settings build more robust applications, and the challenges faced in industrial settings guide theoretical research towards productive solutions. To reflect this drive, the group of participants was chosen to strike a balance between academia and industry, with many participants having experience in both domains. The seminar was structured around discussions in small working groups, different each day. Each group had a topic or problem to tackle, and the key challenges and state of the art solutions were shared at the end of each day to all participants of the seminar. One of the key objectives of this seminar was to bring together ideas and researchers from academic and industrial backgrounds. To this end, a sessions on Wednesday were focused on the current state of the practice being used in industrial applications, with each industrial partner sharing knowledge about their respective application fields.

Key Themes and Discussions

- Participants explored various formal methods and verification techniques designed to enhance the reliability of autonomous systems. Discussions highlighted the need to advance state-of-the-art formal verification approaches to accommodate the complexity of modern autonomous systems.
- The seminar also emphasized the importance of building resilient systems capable of functioning reliably in dynamic environments. Discussions tackled challenges related to the various modules (e.g. perception, motion planning, etc.) in autonomous mobile cyber-physical systems, considering how to incorporate robustness into the design phase and beyond.
- The specific challenges unique to each transportation sector were discussed, emphasizing tailored strategies for addressing safety assurance in automotive, aerospace, and railway systems. The cross-sectoral dialogue shed light on shared challenges and provided new perspectives that will inform future efforts.

Outcomes and Future Directions

The seminar generated a consensus on the urgent need for more research and collaboration across sectors. Participants emphasized the importance of combining expertise from different domains to address the interdisciplinary nature of safety assurance in autonomous mobility.

Moreover, the discussions underscored the potential for ongoing interdisciplinary seminars and follow-up workshops that would ensure continuous engagement among stakeholders. These future events would also provide venues for updating each other on progress, refining safety standards, and accelerating technological advancements in this field.

Overall, the seminar succeeded in creating a collaborative environment that not only identified existing challenges but also laid the groundwork for innovative solutions in safety assurance for autonomous mobility.

2 Table of Contents

Executive Summary

Dejan Ničković, Jyotirmoy Deshmukh, Bettina Könighofer, and Filip Cano 95

Overview of Talks

Industrial-grade Autonomous Products – Challenge and Applied Approaches for Safety <i>Christof Budnik</i>	99
Industrial Challenges in Assuring Autonomy <i>Mauricio Castillo-Effen</i>	99
Challenges in Autonomous Vehicle Development <i>Patricia Derler</i>	100
Future Mobility: Challenges and Recent Developments in V&V <i>Bardh Hoxha</i>	100
Clearsy: Safety Critical Systems And Forthcoming Autonomy In the Railways <i>Thierry Lecomte</i>	101
Introduction to Conformal Prediction <i>Lars Lindemann</i>	102
Denso: Critical Scenario Identification <i>Selma Music</i>	103
AVL framework for close loop testing <i>Darko Stern</i>	103
Safety Assurance of Automated Driving Systems – Selected Concepts, Standards, and Approaches <i>Dirk Ziegenbein</i>	104

Working groups

What Do We Need to Provide Safety Assurance? <i>Houssam Abbas, Ezio Bartocci, Chih-Hong Cheng, Ichiro Hasuo, Panagiotis Katsaros, Assaf Marron, Stefan Pranger, and Alessandro Zanardi</i>	104
Methods for Safety Assurance <i>Dejan Ničković, Christof Budnik, Mauricio Castillo-Effen, Jyotirmoy Deshmukh, Marie Farrell, Rong Gu, Thierry Lecomte, Lars Lindemann, Selma Music, Necmiye Ozay, Giulia Pedrielli, Doron A. Peled, and Darko Stern</i>	105
How To Design for Assurance <i>Bettina Könighofer, Rayna Dimitrova, Michael Fisher, Martin Fränzle, Mahsa Ghasemi, Radu Grosu, Bardh Hoxha, Sayan Mitra, and Andoni Rodríguez</i>	106
Application Domain: Automotive – Challenges <i>Patricia Derler, Ezio Bartocci, Radu Grosu, Ichiro Hasuo, Panagiotis Katsaros, Assaf Marron, Selma Music, Dejan Ničković, Darko Stern, Alessandro Zanardi, and Dirk Ziegenbein</i>	107

Application Domain: Automotive – Solutions <i>Patricia Derler, Ezio Bartocci, Radu Grosu, Ichiro Hasuo, Panagiotis Katsaros, Assaf Marron, Selma Music, Dejan Ničković, Darko Stern, Alessandro Zanardi, and Dirk Ziegenbein</i>	108
Application Domain: Robotics – Challenges <i>Stefan Pranger, Rayna Dimitrova, Michael Fisher, Mahsa Ghasemi, Rong Gu, Bardh Hoxha, and Bettina Könighofer</i>	109
Application Domain: Robotics – Solutions <i>Stefan Pranger, Filip Cano, Rayna Dimitrova, Michael Fisher, Mahsa Ghasemi, Rong Gu, Bardh Hoxha, Bettina Könighofer, and Selma Music</i>	110
Application Domain: Railway – Challenges <i>Chih-Hong Cheng, Christof Budnik, Martin Fränzle, Thierry Lecomte, Doron A. Peled, and Andoni Rodríguez</i>	111
Application Domain: Railway – Solutions <i>Christof Budnik, Chih-Hong Cheng, Martin Fränzle, Thierry Lecomte, Doron A. Peled, and Andoni Rodríguez</i>	111
Application Domain: Aerospace – Challenges & Solutions <i>Lars Lindemann, Houssam Abbas, Mauricio Castillo-Effen, Jyotirmoy Deshmukh, Marie Farrell, Taylor T. Johnson, Panagiotis Katsaros, Necmiye Ozay, and Giulia Pedrielli</i>	113
Techniques for Safety Assurance: : Low-level Control <i>Taylor T. Johnson, Mauricio Castillo-Effen, Patricia Derler, Jyotirmoy Deshmukh, Panagiotis Katsaros, Lars Lindemann, Selma Music, and Necmiye Ozay</i>	114
Techniques for Safety Assurance: : Perception <i>Dejan Ničković, Christof Budnik, Martin Fränzle, Rong Gu, Thierry Lecomte, Stefan Pranger, Andoni Rodríguez, Darko Stern, and Dirk Ziegenbein</i>	115
Techniques for Safety Assurance: : Machine Learning in Planning <i>Alessandro Zanardi, Ezio Bartocci, Filip Cano, Rayna Dimitrova, Marie Farrell, Mahsa Ghasemi, Radu Grosu, Ichiro Hasuo, Bettina Könighofer, Assaf Marron, and Doron A. Peled</i>	116
Techniques for Safety Assurance: Large Language Models <i>Marie Farrell, Christof Budnik, Filip Cano, Mauricio Castillo-Effen, Jyotirmoy Deshmukh, Rayna Dimitrova, Martin Fränzle, Rong Gu, Bardh Hoxha, Taylor T. Johnson, Panagiotis Katsaros, Selma Music, Dejan Ničković, Necmiye Ozay, Andoni Rodríguez, and Darko Stern</i>	117
Participants	119

3 Overview of Talks

3.1 Industrial-grade Autonomous Products – Challenge and Applied Approaches for Safety

Christof Budnik (Siemens – Princeton, US)

License  Creative Commons BY 4.0 International license
© Christof Budnik

Many businesses recognize the vast potential of artificial intelligence (AI) to enhance efficiency, productivity, and quality in autonomous industrial production. While numerous prototypes have emerged, only a limited number of products exhibit the capability to operate reliably in safety-critical environments. Notably, the absence of a secure person locator in manufacturing poses a significant challenge, hindering worker collaboration with autonomously operating machines in open environments. The elevated risk of AI failures causing harm to workers underscores the urgent need for solutions to mitigate such risks and avoid adverse consequences for companies. In this context, gaining a competitive edge in verifying and measuring AI safety becomes imperative for quality-focused enterprises. This imperative is further underscored by the establishment of standards and industrial regulations. The presented discussion outlined prevailing challenges and barriers associated with implementing safe AI for autonomous robots in the manufacturing floor. Specifically, it provided insights into how autonomous systems are further shaping the future in mobility and smart cities, presenting challenges that resonate within industry practices. One prominent challenge discussed pertains to the absence of standardized procedures for obtaining safety certifications for AI-based systems. Addressing this issue, the presentation introduced three distinct approaches: the development of a comprehensive test infrastructure supporting the entire DevOps lifecycle, leveraging AI to generate test cases, and ensuring end-to-end assurance of autonomous machines from an architectural perspective. By exploring these strategies, businesses can actively navigate the complexities of AI safety certification and enhance their ability to deploy reliable and secure autonomous systems in industrial settings.

3.2 Industrial Challenges in Assuring Autonomy

Mauricio Castillo-Effen (Lockheed Systems – Arlington, US)

License  Creative Commons BY 4.0 International license
© Mauricio Castillo-Effen

In this talk, I introduced three ideas:

- The necessity and feasibility of agility and adaptability,
- The importance of context, and
- The challenges in modeling system evolution.

I demonstrated that current Systems Engineering practice has shortcomings because it assumes systems must be developed with all foreseeable conditions in mind. This is impractical and significantly burdens design assurance, limiting timely deployment opportunities and hindering learning from operations. This issue can be mitigated by focusing on targeted conditions, reusing assurance artifacts, and enabling continuous improvement, specifically by gradually expanding the operational domain under the protection of runtime assurance and safe learning mechanisms.

In the second part of the talk, Operational Design Domains (ODDs) were presented as a framework for modeling context. Seven research challenges were proposed to advance our understanding of ODD similarity, coverage, and scenarios as ODD samples.

The final section revealed a conceptual diagram illustrating the evolution of systems through the interplay of four categories of artifacts: specifications, design, implementation, and assurance. Systems evolve over time and through a spectrum of variants. It is necessary to develop formalisms and algorithmic reasoning to address system evolution effectively, focusing on assurance to create safe configuration baselines and variants offering guarantees.

3.3 Challenges in Autonomous Vehicle Development

Patricia Derler (Zoox Inc. – Foster City, US)

License  Creative Commons BY 4.0 International license
© Patricia Derler


Automotive autonomous vehicles (AVs) development faces many challenges stemming from (a) the need to express the desired behavior, (b) the need to (formally) capture this desired behavior, and (c) the need to verify whether the AV exhibited the desired behavior. This talk discusses some of the challenges surrounding specification of requirements from traffic rules as written in the rules of the road (e.g. how should one formalize the rule in the Austrian road traffic act stating that one “is not allowed to drive so fast that he dirty other road users or things on the road” – § 20 in [1], the assumptions that one needs to make about other road users (should one always assume that the pedestrian on the side walk could jump in front of the ego vehicle as it is passing?), the timing and latency requirements (e.g. what is the latency requirement from first sensing a previously occluded road participant to braking?), and the lack of good, useful, formal models at various levels of abstractions that lend themselves to formal verification.

References

- 1 Austria, Straßenverkehrsordnung 1960, BGBl. Nr. 159/1960. (Austrian Road Traffic Act of 1960)

3.4 Future Mobility: Challenges and Recent Developments in V&V

Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US)

License  Creative Commons BY 4.0 International license
© Bardh Hoxha

Joint work of Bardh Hoxha, Danil Prokhorov, Dejan Ničković, Georgios Fainekos, Hideki Okamoto, Hoang-Dung Tran, Jacob Anderson, Jyotirmoy Deshmukh, Navid Hashemi, Sungwoo Choi, Tomoya Yamaguchi, Xiaodong Yang

The talk delves into the emerging challenges and recent advancements in the Verification and Validation (V&V) of autonomous mobility systems. Focused on enhancing safety assurance, the discussion encompasses a broad range of cyber-physical systems, including smart cities, medical devices, and autonomous driving systems (ADS). The research emphasizes the critical role of requirements verification, validation, testing, certification, and correct-by-design approaches. The talk provides a number of methods for safety verification of machine learning-enabled CPS, safe planning and control of heterogeneous multi-agent systems, and

human-robot interaction. The team aims to provide formal guarantees on system functionality and performance under uncertainty. The presentation showcases tools and case studies, highlighting the importance of rigorous V&V processes in realizing the future of autonomous mobility safely and efficiently.

References

- 1 Jacob Anderson, Georgios Fainekos, Bardh Hoxha, Hideki Okamoto, Danil Prokhorov, *Pattern Matching for Perception Streams*, Runtime Verification (RV), 2023
- 2 Tomoya Yamaguchi, Bardh Hoxha, Dejan Ničković, *RTAMT – Runtime Robustness Monitors with Application to CPS and Robotics*, International Journal on Software Tools for Technology Transfer (STTT), 2023
- 3 Hoang-Dung Tran, Sungwoo Choi, Hideki Okamoto, Bardh Hoxha, Georgios Fainekos, Danil Prokhorov, *Quantitative Verification for Neural Networks using ProbStars*, Hybrid Systems: Computation and Control (HSCC), 2023
- 4 Hoang-Dung Tran, Sungwoo Choi, Xiaodong Yang, Tomoya Yamaguchi, Bardh Hoxha, Danil Prokhorov, *Verification of Recurrent Neural Networks with Star Reachability*, Hybrid Systems: Computation and Control (HSCC), 2023
- 5 Navid Hashemi, Bardh Hoxha, Tomoya Yamaguchi, Danil Prokhorov, Georgios Fainekos, Jyotirmoy Deshmukh, *A Neurosymbolic Approach to the Verification of Temporal Logic Properties of Learning enabled Control Systems*, ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS), 2023

3.5 ClearSY: Safety Critical Systems And Forthcoming Autonomy In the Railways

Thierry Lecomte (*CLEARSY – Aix-en-Provence, FR*)

License © Creative Commons BY 4.0 International license
© Thierry Lecomte

Joint work of Jan Peleska, Anne E. Haxthausen, Thierry Lecomte

Main reference Jan Peleska, Anne E. Haxthausen, Thierry Lecomte: “Standardisation Considerations for Autonomous Train Control”, in Proc. of the Leveraging Applications of Formal Methods, Verification and Validation. Practice – 11th International Symposium, ISoLA 2022, Rhodes, Greece, October 22-30, 2022, Proceedings, Part IV, Lecture Notes in Computer Science, Vol. 13704, pp. 286–307, Springer, 2022.

URL https://doi.org/10.1007/978-3-031-19762-8_22

This presentation demonstrates how safety critical systems are designed, developed, and certified in the railways. Safety is about failing systems. Failing parts to consider are exposed: wrong specification, wrong program, wrong binary, wrong execution, bad hardware (hardware caonatains other functions/interface that described in the datasheet), failing hardware (entropy leading to dysfunctional gates, drifting clock, etc.), wrong environment specification, and wrong exploitation procedure. Failure, either systematic or random, have an impact on the behaviour of the system. Safety is about keeping the probability of catastrophic failure below a treshold. Safety demonstration has to convince an independent expert that the feared events are not going to happen more frequently than expected. CLEARSY is using formal methods and related tools at different levels (software, data, system) to complete this safety demonstration, in accordance with the safety standards. The introduction of ML-based technologies to enable autonomous driving is raising safety issues. AI is at the moment not recommended to develop the most critical function in the railways, the main argument geing the lack of explainability of the ML black box function. UIC (<http://uic.org>) has launched a project titled “New methods for safety demonstration” that is aimed at

helping human certifiers to envisage the certification of functions developed with AI or using IoT/Cybersecurity. Several projects of autonomous trains have been initiated in France, some are related to existing lines (freight, passengers, high-speed) while others are targeting low traffic, regional lines with specific/adapted infrastructure. These projects are expected to contribute to the standards, based on the results obtained.

3.6 Introduction to Conformal Prediction

Lars Lindemann (USC – Los Angeles, US)

License  Creative Commons BY 4.0 International license
 Lars Lindemann

Learning-enabled systems promise to enable many future technologies such as autonomous driving, intelligent transportation, and robotics. Accelerated by the computational advances in machine learning and AI, there has been tremendous success in the design of learning-enabled systems. At the same time, however, new fundamental challenges arise regarding the safety and reliability of these increasingly complex systems that operate in unknown dynamic environments. In this tutorial, I will provide new insights and discuss exciting opportunities to address these challenges by using conformal prediction (CP), a statistical tool for uncertainty quantification. I will advocate for the use of CP in systems theory due to its simplicity, generality, and efficiency as opposed to existing model-based techniques that are either conservative or have scalability issues.

References

- 1 Shafer, Glenn, and Vladimir Vovk. *A tutorial on conformal prediction*, Journal of Machine Learning Research (JMLR) 2008.
- 2 Angelopoulos, Anastasios N., and Stephen Bates. *Conformal prediction: A gentle introduction*, Foundations and Trends in Machine Learning 2023.
- 3 Romano, Yaniv, Evan Patterson, and Emmanuel Candes. *Conformalized quantile regression*, Conference on Neural Information Processing Systems (NeurIPS) 2019.
- 4 Ryan J. Tibshirani, Rina Foygel Barber, Emmanuel Candes, Aaditya Ramdas. *Conformal prediction under covariate shift*, Conference on Neural Information Processing Systems (NeurIPS) 2019.
- 5 Lars Lindemann, Matthew Cleaveland, Gihyun Shim, George J. Pappas. *Safe planning in dynamic environments using conformal prediction*, IEEE Robotics and Automation Letters (RA-L) 2023.
- 6 Matthew Cleaveland, Insup Lee, George J. Pappas, Lars Lindemann. *Conformal Prediction for Time Series using Linear Complementarity Programming*, AAAI Conference on Artificial Intelligence (AAAI) 2024.
- 7 Lars Lindemann, Xin Qin, Jyotirmoy V. Deshmukh, George J. Pappas. *Conformal prediction for STL runtime verification*, ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS) 2023.
- 8 Yiqi Zhao, Bardh Hoxha, Georgios Fainekos, Jyotirmoy V. Deshmukh, Lars Lindemann. *Robust Conformal Prediction for STL Runtime Verification under Distribution Shift*, ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS) 2024.
- 9 Yiqi Zhao, Xinyi Yu, Jyotirmoy V. Deshmukh, Lars Lindemann. *Conformal Predictive Programming for Chance Constrained Optimization* arXiv 2024.

3.7 Denso: Critical Scenario Identification

Selma Music (Denso Automotive – Echting, DE)

License © Creative Commons BY 4.0 International license
© Selma Music

Proving the safe operation of autonomous driving is one of the biggest challenges in the automotive domain. Conventional test methods are not sufficient due to the large amount of test kilometers that are needed to argue about system safety guarantees. Therefore, scenario-based testing is a promising and widely studied method to address this challenge in simulation. In this talk, I will discuss critical scenario identification (CSI), one of the ways to perform scenario-based testing. We will refer to requirements from the safety standard ISO 21448 (SOTIF) and focus on discovering scenario “unknown unknowns” using CSI methods. We will mainly focus on intelligent testing methods within CSI, using optimization-based testing, and we will discuss open challenges, e.g., the selection of appropriate input parameters and the appropriate design of the scenario. The talk will contain illustrative examples and results to demonstrate the effectiveness of the CSI approach for the safety assurance of autonomous driving.

3.8 AVL framework for close loop testing


Darko Stern (AVL – Graz, AT)

License © Creative Commons BY 4.0 International license
© Darko Stern

Benefits to automated vehicles are greater road safety, cost savings, more productivity and reduced use of fuels, thus contributing to a green future. End users are ready to invest in autonomous vehicles, however, 64% demand increased safety for autonomous vehicles. Customers expect that autonomous systems cause no dangerous situations and do not influence the traffic flow, however, most systems currently on the market cannot handle complex situations. So, how to make sure that the automated vehicle behaves correctly in EVERY situation? In my talk, I presented a framework for close-loop testing and verification of automated vehicles that are under development in the AVL List. I presented ways how to avoid expensive full factorial testing, by using the Active DoE approach, followed by a comprehensive overview of the co-simulation environment for its execution. The reliability of the test depends on the fidelity of the 3D environment, which needs to support close-loop testing and the possibility of creating rare events. I finished my talk with the progress AVL made in modelling sensors and pedestrian behaviour.

3.9 Safety Assurance of Automated Driving Systems – Selected Concepts, Standards, and Approaches

Dirk Ziegenbein (Robert Bosch GmbH – Renningen, DE)


License  Creative Commons BY 4.0 International license
© Dirk Ziegenbein

The talk covered several topics in the area of safety assurance for automated driving systems. Based on a model to reason about the interaction of required, specified and implemented behaviors as well as the definition of the Operational Design Domain, the purposes and scopes of two most prevalent safety standards for road vehicles have been discussed. Furthermore, the Pegasus-VVM framework for scenario-based safety assurance as well as a specific method for open context domain analysis have been introduced.

4 Working groups

4.1 What Do We Need to Provide Safety Assurance?

Houssam Abbas (Oregon State University – Corvallis, US), Ezio Bartocci (TU Wien, AT), Chih-Hong Cheng (Universität Hildesheim, DE), Ichiro Hasuo (National Institute of Informatics – Tokyo, JP), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Assaf Marron (Weizmann Institute – Rehovot, IL), Stefan Pranger (TU Graz, AT), and Alessandro Zanardi (ETH Zürich, CH)

License  Creative Commons BY 4.0 International license
© Houssam Abbas, Ezio Bartocci, Chih-Hong Cheng, Ichiro Hasuo, Panagiotis Katsaros, Assaf Marron, Stefan Pranger, and Alessandro Zanardi

This working group discussed what artifacts are need to build a safety assurance case for an autonomous vehicle or system of autonomous vehicles. We started by considering what things we think are currently working or should be part of an eventual reliable solution. Answers included:

- 1) Natural language specifications (unstructured natural language vs constrained natural language)
- 2) Tools that ensure “horizontal” traceability from natural language requirement to formal requirements to test cases to results (with intermediate steps in the middle)
- 3) Syntactic generators of scenarios: while still in infancy in this domain, they are showing usefulness within industry. Declarative/program-like specifications

We did an overview of types of specs (functional, performance, security, architectural design, ODD etc.) and some attendees suggested that rather than adhere strictly to the use of unambiguous specifications, perhaps we should allow multiple interpretations with some margins of flexibility to account for genuine uncertainty (borne out of the designers’ uncertainty, as opposed to wrong interpretations.) Another things that works really well is the existence of a near-universal system description language (like Verilog for hardware) which enables industry-wide tool development and agreement on pre-competitive technologies.

In answer to “Where do you see the biggest gap?” the attendees offered:

- 1) No ways to efficiently capture the different interpretations of certain requirements, or which ones are relevant.

- 2) How to account efficiently for new factors in the assurance case? (E.g. we discover a new factor that affects a given outcome). Measures of coverage?
- 3) How to capture expert (implicit) knowledge and intuition?
- 4) How to create vertical traceability, from application-specific assurance outcomes (e.g. no running over pedestrians) to assurance requirements one level below, and one level below that, etc.

4.2 Methods for Safety Assurance

Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Christof Budnik (Siemens – Princeton, US), Mauricio Castillo-Effen (Lockheed Systems – Arlington, US), Jyotirmoy Deshmukh (USC – Los Angeles, US), Marie Farrell (University of Manchester, GB), Rong Gu (Mälardalen University – Västerås, SE), Thierry Lecomte (CLEARSY – Aix-en-Provence, FR), Lars Lindemann (USC – Los Angeles, US), Selma Music (Denso Automotive – Eching, DE), Necmiye Ozay (University of Michigan – Ann Arbor, US), Giulia Pedrielli (Arizona State University – Tempe, US), Doron A. Peled (Bar-Ilan University – Ramat Gan, IL), and Darko Stern (AVL – Graz, AT)

License © Creative Commons BY 4.0 International license
 © Dejan Ničković, Christof Budnik, Mauricio Castillo-Effen, Jyotirmoy Deshmukh, Marie Farrell, Rong Gu, Thierry Lecomte, Lars Lindemann, Selma Music, Necmiye Ozay, Giulia Pedrielli, Doron A. Peled, and Darko Stern


The discussion was centered around the prerequisites needed in to provide safety assurance to autonomous mobile systems in terms of requirements, specifications, assumptions, descriptions of the Operational Design Domain (ODD), architecture models, digital twins and simulators. This rather broad list of topics was narrowed down during the discussion and the participants identified the most critical aspects with respect to the safety assurance.

The first observation was that the requirements are still predominantly given in the form of natural language with possibly some guidelines regarding the structure of the text. In the past, there have been works on constrained natural language templates for specifications, but they had limited success of practical adoption. On the other hand, formal specifications are needed to reduce ambiguities, facilitate exchange of requirements between teams, and automate certain design and verification activities. The process of formalizing requirements is not straightforward as certain natural language statements can sometimes have (even on purpose) multiple interpretations, and it is hard to capture different interpretations with formal specification. Formal specifications typically require providing one precise interpretation of the natural language sentence. Another problem is that there are no efficient ways to capture implicit knowledge and intuitions with formal specifications, which are often part of any design. Formal specifications can take many different forms, from declarative (e.g. temporal logics) to operational (e.g. program-like specifications). Specifications can also address many different aspects of the design such as its functional properties, performance, security, architecture design, ODDs, etc. Finally, the participants identified the need for tools that ensure horizontal traceability throughout the design cycle (from natural language requirements to formal specifications to test cases with everything in between). More specifically, there is the question of how to create vertical traceability, from application-specific assurance outcomes (e.g. no running over pedestrians) to assurance requirements one level below, down to the requirements of individual components. In order to facilitate the safety assurance reasoning, the participants identified a framework for reusability of

artefacts as part of a potential solution. Such a framework would permit to keep track of specifications, historical evidence on prior arguments, scenarios and domain models, raw datasets and assurance patterns.

4.3 How To Design for Assurance

Bettina Könighofer (TU Graz, AT), Rayna Dimitrova (CISPA – Saarbrücken, DE), Michael Fisher (University of Manchester, GB), Martin Fränzle (Universität Oldenburg, DE), Mahsa Ghasemi (Purdue University – West Lafayette, US), Radu Grosu (TU Wien, AT), Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US), Sayan Mitra (University of Illinois – Urbana Champaign, US), and Andoni Rodríguez (IMDEA Software Institute – Madrid, ES)

License  Creative Commons BY 4.0 International license

© Bettina Könighofer, Rayna Dimitrova, Michael Fisher, Martin Fränzle, Mahsa Ghasemi, Radu Grosu, Bardh Hoxha, Sayan Mitra, and Andoni Rodríguez

A pivotal area of discussion in our group was how to present evidence for safety assurance. We discussed the importance of constructing and showcasing worst-case scenarios from each contract. However, we quickly realized that these scenarios might contradict each other. This led us to the consensus that analyzing combined worst-case behaviors, instead of individual ones, would be more effective, especially to tackle dependence problems and ensure the system’s safety through invariance properties.

When it came to integrating specifications, contracts, and tests, we saw much potential in mining contracts from tests and generating tests to cover scenarios not addressed by the contracts. This approach ensures a comprehensive coverage of potential failures and strengthens the safety argument.

The group recognized the challenges in verifying large codebases and model-based design. We acknowledged that tools are not scalable enough yet but concluded that focusing verification on critical properties and having good layered architectural designs could be a workaround. Similarly, for model-based testing, the difficulty lies in obtaining models, so we suggested advancing techniques like abstract interpretation, which automatically generates models from code.

We also tackled the issue of information overflow and how to pinpoint the relevant properties that need our attention. Integrating different kinds of evidence into a consistent assurance argument emerged as a significant challenge. Here, we saw a need for heterogeneous contracts that can seamlessly combine probabilistic and Boolean elements, despite the current lack of extensive research in this area.

Specific attention went to the verification of perception systems. We discussed managing various uncertainties by defining precise contracts that set tolerated error ranges and ensure consistency across data sequences. This method aims to mitigate risks associated with perception errors.

The synthesis of systems brought up concerns over explainability. We all agreed that guided synthesis, imposing constraints on architectural design, could make the synthesis process quicker and the systems more understandable.

Lastly, we discussed existing solutions that we think have high potential. Layered and error-aware architectures are great because they allow for error tracing and include mechanisms for recovery. They are built to make safe decisions based on the current knowledge and have plans in place for potential deviations. Contracts that mediate trust between different layers of an architecture are also part of the solution.

In sum, our discussions underscored a multifaceted approach to designing safety into autonomous mobility products. We believe that a combination of worst-case analysis, contract-based design, targeted verification, and sophisticated handling of uncertainties will pave the way for safer autonomous systems.

4.4 Application Domain: Automotive – Challenges

Patricia Derler (Zoox Inc. – Foster City, US), Ezio Bartocci (TU Wien, AT), Radu Grosu (TU Wien, AT), Ichiro Hasuo (National Institute of Informatics – Tokyo, JP), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Assaf Marron (Weizmann Institute – Rehovot, IL), Selma Music (Denso Automotive – Eching, DE), Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Darko Stern (AVL – Graz, AT), Alessandro Zanardi (ETH Zürich, CH), and Dirk Ziegenbein (Robert Bosch GmbH – Renningen, DE)

License © Creative Commons BY 4.0 International license

© Patricia Derler, Ezio Bartocci, Radu Grosu, Ichiro Hasuo, Panagiotis Katsaros, Assaf Marron, Selma Music, Dejan Ničković, Darko Stern, Alessandro Zanardi, and Dirk Ziegenbein

This working group session discussed important and hard problems in the area of automotive autonomous mobility. The problems were roughly divided into the areas of

- (a) requirements and specification engineering,
- (b) the development of automotive autonomous systems, and
- (c) verification and validation (V&V) activities.

For (a), participants highlighted the difficulty in formalizing requirements, especially given the diverse range of requirement types such as collision avoidance, comfort, adherence to traffic and social rules, among others. Moreover, concerns arose regarding the verification of requirements' completeness across different levels, spanning from the vehicle to component levels. Complex Operational Design Domains (ODDs) further compounded these challenges, prompting discussions on how to formalize, refine, and extend them effectively while balancing usefulness and safety. Defining safety requirements proved intricate, with passive safety measures like stopping deemed insufficient for freeways. Additionally, challenges in formulating requirements that help gain trust in autonomous systems were discussed. Such requirements necessitate interpreting social behaviors and developing robust models for the environment, AV, as well as components and sub-components, reflecting the multifaceted nature of autonomous automotive systems.

As for problems surrounding the development (b), the complexities of handling and certifying machine learning (ML) components were front and center. Debates arose regarding the autonomy levels and the suitability of supervised versus unsupervised approaches. There were concerns regarding the ambiguous nature of certification processes and the need to address performance issues, including the adequacy of data for training and the challenges of generalizing it effectively. Architecture flaws, particularly the lack of redundancy, emerged as a focal point, emphasizing the necessity for robust system designs. Operational Design Domains posed significant hurdles, raising questions about developing and detecting specific ODDs and the system's behavior beyond these domains. The workshop also shed light on the myriad tooling and engineering issues inherent in designing and deploying autonomous systems, underscoring the multifaceted nature of the development process.

Key problems identified in (c) are the balance between simulation and testing, questioning what could effectively be simulated versus what necessitated physical testing for robust validation. Concerns arose regarding the fidelity of simulation environments and determining

the appropriate level of simulation, whether at the holistic AV and environment level or at the component level. Closed-loop testing presented difficulties, along with debates surrounding the utility of abstractions in models and simulations, particularly in mapping them to real-world scenarios. Coverage of Operational Design Domains in simulation and testing remained a significant challenge, as did the processing and extraction of information from data to uncover special cases effectively.

4.5 Application Domain: Automotive – Solutions

Patricia Derler (Zoox Inc. – Foster City, US), Ezio Bartocci (TU Wien, AT), Radu Grosu (TU Wien, AT), Ichiro Hasuo (National Institute of Informatics – Tokyo, JP), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Assaf Marron (Weizmann Institute – Rehovot, IL), Selma Music (Denso Automotive – Eching, DE), Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Darko Stern (AVL – Graz, AT), Alessandro Zanardi (ETH Zürich, CH), and Dirk Ziegenbein (Robert Bosch GmbH – Renningen, DE)

License © Creative Commons BY 4.0 International license

© Patricia Derler, Ezio Bartocci, Radu Grosu, Ichiro Hasuo, Panagiotis Katsaros, Assaf Marron, Selma Music, Dejan Ničković, Darko Stern, Alessandro Zanardi, and Dirk Ziegenbein

Following the working group discussions on problems for autonomous automotive systems, this session focused on solution approaches and existing solutions. International safety standards were recognized as pivotal, although some standards posed additional challenges or lacked actionable directives, underscoring the importance of concerted efforts in this area. Advances in simulation, exemplified by initiatives like VISTA 2.0 [1], offered promising avenues for enhancing capabilities in multimodal sensing and policy learning for autonomous vehicles. Bridging natural language and formal methods emerged as a key strategy, with solutions integrating language and logic for formalizing Operational Design Domains (ODDs) and requirements to provide clearer and more actionable specifications. Moreover, leveraging probabilistic modeling and verification techniques showed promise in enhancing the robustness of autonomous systems, particularly in interpreting social behaviors and demonstrating intent. Digital twin frameworks were identified as valuable tools for gaining insights into real-world scenarios and validating autonomous systems. Open source reference implementations were also highlighted for fostering collaboration, transparency, and innovation within the autonomous automotive community. Specific solutions discussed included improving AI/ML explainability through redundancy, independence assumptions, and model-based approaches like neuro-symbolic ML/AI. The importance of explainable AI was emphasized, stressing the need for comprehensive understanding and enabling replayable errors for effective analysis and improvement. Safety frameworks, such as the Safety Case by Waymo [2], were recognized for providing structured methodologies to ensure the safety and reliability of autonomous systems. Formalization efforts, such as Responsibility-Sensitive Safety (RSS) [5], the Pegasus Project [3], and IEEE 2846 [4], were deemed essential for establishing clear guidelines and standards for system behavior and domain coverage. Communication protocols like Collaborative Awareness Messages [6] and Collective Perception Messages [7] were seen as critical for facilitating effective communication between autonomous vehicles and infrastructure, thereby enhancing overall system efficiency and safety.

References

- 1 Alexander Amini, Tsun-Hsuan Wang, Igor Gilitschenski, Wilko Schwarting, Zhijian Liu, Song Han, Sertac Karaman, Daniela Rus. *Vista 2.0: An open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles*, 2022 International Conference on Robotics and Automation (ICRA), 2022.
- 2 The Waymo Team, *A Blueprint for AV Safety: Waymo's Toolkit For Building a Credible Safety Case*, 2023.
- 3 Hermann Winner, Karsten Lemmer, Thomas Form, Jens Mazzega. *PEGASUS—First steps for the safe introduction of automated driving*. Road Vehicle Automation 5. Springer International Publishing, 2019.
- 4 IEEE, *IEEE Standard for Assumptions in Safety-Related Models for Automated Driving Systems*. 2022.
- 5 Shalev-Shwartz, Shai, Shaked Shammah, and Amnon Shashua. *On a formal model of safe and scalable self-driving cars*. arXiv preprint arXiv:1708.06374 (2017).
- 6 F. Raviglione, S. Zocca, A. Minetto, M. Malinverno, C. Casetti, C. F. Chiasserini, and F. Dovis, *From collaborative awareness to collaborative information enhancement in vehicular networks*, Vehicular Communications, vol. 36, p. 100497, 2022.
- 7 Mohammad Raashid Ansari, Jean-Philippe Monteuis, Jonathan Petit, Cong Chen. *V2x misbehavior and collective perception service: Considerations for standardization*, IEEE Conference on Standards for Communications and Networking (CSCN) 2021.

4.6 Application Domain: Robotics – Challenges

Stefan Pranger (TU Graz, AT), Rayna Dimitrova (CISPA – Saarbrücken, DE), Michael Fisher (University of Manchester, GB), Mahsa Ghasemi (Purdue University – West Lafayette, US), Rong Gu (Mälardalen University – Västerås, SE), Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US), and Bettina Könighofer (TU Graz, AT)

License © Creative Commons BY 4.0 International license
© Stefan Pranger, Rayna Dimitrova, Michael Fisher, Mahsa Ghasemi, Rong Gu, Bardh Hoxha, and Bettina Könighofer

This working group discussed important problems in the field of autonomous robotics. The topics discussed can roughly be divided into problems that arise during design time and problems faced after deployment of an autonomous actor.

The discussion in this working group started off with problems that occur during design time. The first hard problems that have been discussed concerned the design and definition of concrete requirements and specifications, as well as assumptions that can be taken at design time. Clear definitions of safety and performance are often highly dependent on the specific application area. The group discussed several examples from industry, an example to highlight the intricacy of the problem are autonomous delivery robots in a Japanese hospital. The robots are used to pick up and deliver e.g. blood samples and are therefore allowed to use elevators. Since the hospital is in an area prone to earthquakes it is of special interest that robots do not hinder the fast evacuation of patients and staff in case of an earthquake. The group put a special emphasis on the arising complexity w.r.t. the different application areas of autonomous robots. In contrast to the problems faced in autonomous mobility, an area in which the legislative body has already compiled a manifest that autonomous systems need to adhere to. The group discussed a second issue that arises during the design time with respect to simulation software, especially the robot operating system (ROS). The issue raised regarding ROS is its dependency on the scheduling of the underlying operating system.

The simulation might not be faithful as it cannot ensure real-time execution due to this dependency. This issue has been discussed in the seminar group, we refer the reader to the abstract regarding discussed solutions.

The discussion continued with topics related to the deployment of autonomous robots. The discussed problems heavily relate to the problem mentioned above regarding the different application areas, especially under the presence of humans. The problem mentioned here is twofold: It is an open problem, how a robot should behave to be the least harmful towards humans in safety critical situations, and it is a formidable problem to incorporate arbitrary human behaviour, such that autonomous robots can behave appropriately. Additionally, the problem of reliable perception has been labeled as very interesting for industry partners.

Lastly, the group wanted to include security as a hard problem. The topic has initially been mentioned, but has not been discussed.

4.7 Application Domain: Robotics – Solutions

Stefan Pranger (TU Graz, AT), Filip Cano (TU Graz, AT), Rayna Dimitrova (CISPA – Saarbrücken, DE), Michael Fisher (University of Manchester, GB), Mahsa Ghasemi (Purdue University – West Lafayette, US), Rong Gu (Mälardalen University – Västerås, SE), Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US), Bettina Könighofer (TU Graz, AT), and Selma Music (Denso Automotive – Echting, DE)

License © Creative Commons BY 4.0 International license
© Stefan Pranger, Filip Cano, Rayna Dimitrova, Michael Fisher, Mahsa Ghasemi, Rong Gu, Bardh Hoxha, Bettina Könighofer, and Selma Music


The group followed up on the discussion about hard problems for autonomous robotics by discussing potential solution methods. The complexity of generalizing specifications regarding safety and utility in the domain of autonomous robotics has been identified as a key problem. This stems from the wide range of application areas. Therefore, the need for formal definitions of Operational Design Domains (ODDs) per applications has been identified as key driver during the discussion. The group highlighted the different aspects that are to be taken into account, specifically with regards to human interaction. In order to be able to allow safe behavior there is the need for dynamic modelling of human behaviour. The group mentioned data-driven approaches as a solution. Apart from a model of potential human behaviour, specifications of the types of intended interaction with an autonomous robot needs to be part of an ODD.

The group continued by discussing methods that enhance the explainability through counterfactual analysis.

Regarding open problems related to the faithfulness of simulations, the group discussed how parts of the ROS can be formally verified, and the capabilities of ROS to allow real-time simulations.

4.8 Application Domain: Railway – Challenges

Chih-Hong Cheng (Universität Hildesheim, DE), Christof Budnik (Siemens – Princeton, US), Martin Fränzle (Universität Oldenburg, DE), Thierry Lecomte (CLEARSY – Aix-en-Provence, FR), Doron A. Peled (Bar-Ilan University – Ramat Gan, IL), and Andoni Rodríguez (IMDEA Software Institute – Madrid, ES)

License  Creative Commons BY 4.0 International license
 © Chih-Hong Cheng, Christof Budnik, Martin Fränzle, Thierry Lecomte, Doron A. Peled, and Andoni Rodríguez


The development of machine learning has created great promises, not only for autonomous driving but also for other domains. Within the railway domain, projects such as safe.trAIIn aim to uncover the potential of ML-based perception systems for increasing the Grade of Automation (GoA) via enabling driverless regional trains.

When aligning railway applications with urban autonomous driving, one can find some challenges sitting on the common ground. The first is the positive contribution of ML in safety arguments. The second is the introduction of security attacks. The third is the offering of guarantees to cover sufficiently the environmental conditions. The fourth is the ability to explain why ML fails by generalizing a single prediction error.

Nevertheless, there are also some railway-specific challenges. The first challenge comes with the problem where trains are operated on high-speed and have a long stopping distance. Thus, precisely characterizing acceptable behavior by considering aspects such as safety margin is a concern. The long stopping distance also implies that the decision making requires (1) a different sensor suite that allows look ahead way further than autonomous driving, and (2) considering high uncertainties associated with prediction due to detected objects being very far. The final challenge is related to the use of synthetic data. In contrast to autonomous driving where obtaining data is relatively straightforward, for railway applications such as defect inspection or rail track covered by muds, they occur very rarely, implying a must to use synthetic data. Arguing whether the degree of simulation fidelity is “enough” remains a challenging issue.

4.9 Application Domain: Railway – Solutions

Christof Budnik (Siemens – Princeton, US), Chih-Hong Cheng (Universität Hildesheim, DE), Martin Fränzle (Universität Oldenburg, DE), Thierry Lecomte (CLEARSY – Aix-en-Provence, FR), Doron A. Peled (Bar-Ilan University – Ramat Gan, IL), and Andoni Rodríguez (IMDEA Software Institute – Madrid, ES)

License  Creative Commons BY 4.0 International license
 © Christof Budnik, Chih-Hong Cheng, Martin Fränzle, Thierry Lecomte, Doron A. Peled, and Andoni Rodríguez

In the intricate world of autonomous train safety, a blend of traditional and innovative solutions have been discussed which can ensure a secure railway environment.

Traditional safety measures like guard-rails, isolation protocols, and radio signals act as the backbone, guiding trains safely on their tracks and preventing collisions. These time-tested methods establish a reliable foundation for safe railway operations. Safety is further fortified through sensor diverse redundancy, where an array of safety-focused sensors act as vigilant guardians. Redundancy ensures that even if one sensor encounters an issue, others are ready to step in, contributing to fail-safe detection capabilities. This

principle of sensor redundancy has been extensively studied and discussed in sensor fusion for autonomous car engineering utilizing different sensor technologies and their detection ranges. The introduction of telescope cameras for instance adds a futuristic dimension to safety measures, allowing us to “see the future” of train travel. This innovation poses interesting challenges, particularly in the context of Railway Vehicles (RV), and has been a subject of exploration in the field of railway safety research. In alignment with the philosophy of designing for verifiability verification processes and approaches can be streamlined such as when tracks are only designed as straight-line tracks. This design approach simplifies safety checks and addresses railway infrastructure design. Furthermore, if full specifiability can be reached then there is no need for complex Deep Neural Networks (DNNs).

Looking towards promising and innovative solutions online runtime monitoring has been identified as a beacon for ensuring safety in real-time operations. The emphasis on making these monitors as safe as possible is a foundational principle, underscored in safety studies on autonomous systems. This real-time vigilance serves as a safeguard against potential risks, ensuring the continuous safety of autonomous trains throughout their journeys. However, the advent of the autonomous train era brings forth new challenges, notably the domain’s distribution shift. Placing monitors in the safety loop requires thoughtful consideration, as an excessive presence might overwhelm the system. This delicate balance between effective monitoring and system efficiency is a topic to be further explored on autonomous system architectures. Striking the right equilibrium becomes pivotal in managing the distribution shift effectively without compromising safety.

Another aspect discussed is the adapting infrastructure as a forward-thinking strategy, heralding a promising future for autonomous trains. Rather than solely replacing the driver, reshaping the environment in which trains operate is a paradigm shift discussed on autonomous transportation infrastructure. This approach ensures a holistic enhancement of safety measures while leveraging existing infrastructure investments. The characteristic of the rarity of failures in autonomous train systems sparked the idea of injecting artificial failures to address the synthetic data gap. This proactive approach to failure simulation aims to bridge the gap in rare failure occurrences, ensuring comprehensive safety testing. Rigorous envelope protection, with an emphasis on minimum braking capabilities, establishes another robust standard for safety, ensuring autonomous trains possess the necessary safeguards to handle unforeseen circumstances. Finally, the integration of neuro-symbolic approaches, incorporating temporal stability, consistency across modalities, and physics-assisted techniques, is seen as cutting-edge methodology to filter potential safety issues in autonomous train operations. Neuro-symbolic approaches leverage a combination of symbolic reasoning and neural networks, offering a sophisticated methodology to enhance safety in autonomous train operations.

4.10 Application Domain: Aerospace – Challenges & Solutions

Lars Lindemann (USC – Los Angeles, US), Houssam Abbas (Oregon State University – Corvallis, US), Mauricio Castillo-Effen (Lockheed Systems – Arlington, US), Jyotirmoy Deshmukh (USC – Los Angeles, US), Marie Farrell (University of Manchester, GB), Taylor T. Johnson (Vanderbilt University – Nashville, US), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Necmiye Ozay (University of Michigan – Ann Arbor, US), and Giulia Pedrielli (Arizona State University – Tempe, US)

License © Creative Commons BY 4.0 International license

© Lars Lindemann, Houssam Abbas, Mauricio Castillo-Effen, Jyotirmoy Deshmukh, Marie Farrell, Taylor T. Johnson, Panagiotis Katsaros, Necmiye Ozay, and Giulia Pedrielli

For aerospace system design, the discussion initially centered around the critical balance between safety and utility, the quantification of risk, and the pursuit of more modular and principled design methodologies. The discussion further centered around the multifaceted challenge of reconciling the inherent trade-offs between ensuring safety and achieving optimal performance and utility. Central to these discussions is the notion of risk, traditionally quantified as the product of severity and probability. This classical framework, however, prompts a reevaluation in the context of both offline design and runtime adaption of aerospace systems. The discourse underscored the necessity for nuanced conceptions of risk (such as the conditional value of risk) that are responsive to the dynamic operational environments and the uncertainties that pervade these phases. The discussion also focused on the complexities of flight certification, with risk-friendly and risk-averse approaches, and challenges due to the sim2real gap. Specifically, models in aerospace may be highly nontrivial and not correct due to lack of data. Another point of discussion was on academic strategic repositioning towards emphasizing the economic benefits of safety verification (“Safety does not sell”) in accelerating system development and deployment. Marie brought up another point of discussion, that of the regulatory landscapes of NASA, ESA, and private entities like SpaceX, the latter being in absence of regulatory frameworks. In the end, the conversation touched upon educational and research methodologies, discussing the value of integrating courses on assurance cases and system safety into the aerospace (or more broadly any engineering) curriculum. This is to equip future engineers with the skills necessary to navigate the complex interplay of safety, risk, and performance in aerospace design and operation and to prepare them best for real world challenges. Lastly, the discussion centered on the development of case studies tailored to the aerospace and space sectors that link academia and industry more closely (further bridging the gap between theoretical safety tools and their application in real-world scenarios).

4.11 Techniques for Safety Assurance: : Low-level Control

Taylor T. Johnson (Vanderbilt University – Nashville, US), Mauricio Castillo-Effen (Lockheed Systems – Arlington, US), Patricia Derler (Zoox Inc. – Foster City, US), Jyotirmoy Deshmukh (USC – Los Angeles, US), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Lars Lindemann (USC – Los Angeles, US), Selma Music (Denso Automotive – Eching, DE), and Necmiye Ozay (University of Michigan – Ann Arbor, US)

License © Creative Commons BY 4.0 International license
© Taylor T. Johnson, Mauricio Castillo-Effen, Patricia Derler, Jyotirmoy Deshmukh, Panagiotis Katsaros, Lars Lindemann, Selma Music, and Necmiye Ozay

In the working group dedicated to low-level control, we discussed the integration of various autonomous mobility components, such as perception, planning, and low-level control, and the types of guarantees that can be provided for safe execution. A key focus was on the architectures used in system design, including the distinction between functional and physical architectures and how structure is defined through interfaces. We explored the use of languages and standards such as AADL, SysML, and Autosar for specifying system architecture from various perspectives, including functionality, safety, behavior, timing, and enabling different types of analyses. Contract-based verification was highlighted as a critical method for ensuring system reliability. Challenges included integrating systems composed of many components, particularly when dealing with legacy systems and proprietary models, especially interfaces to low-level control through application programmer interfaces (APIs) that often may necessitate crossing cross-layer boundaries between high-level and low-level control, similar to cross-layer optimizations in network protocols. Overall, the discussion covered the need for a more flexible and adaptive approach to system design, one that can accommodate changing requirements and negotiations between stakeholders. Various techniques for software development and verification were discussed, including structure, composition, certification, schedulability analysis, modeling, and verification.

The discussion also touched upon the aerospace industry’s adoption of architectures defined by the customer, similar to Autosar, underlining the importance of system architecture in the design and verification process. The challenge of low-level control not being fully accessible to the control designer was discussed, emphasizing the gap created by hidden system models and the necessity for system identification, and the use of lookup tables (LUTs) to manage nonlinearities, as well as approximations of large LUTs with other methods, such as via neural networks. More broadly, there was discussion on verification of these types of systems, ranging from the classical designs in low-level control like LUTs, to the usage of artificial intelligence (AI) and machine learning surrogates for these and broader tasks, where robustness of these components are critical.

We delved into the complexities of managing systems with time-varying parameters, such as battery degradation and the effects of aging, and discussed the potential for systems to self-identify and monitor changes. The conversation also covered the concept of conducting verification not only a priori but online, through the generation of conditional evidence and dynamic assurances, so that overall system-level and component-level specifications are monitored for assurance during operation.

The ability to refine high-level architectures through methods such as optimization modulo theories was considered crucial for overcoming system integration challenges, particularly the issue of abstraction leading to loss of detail and mismatches at the low level. We debated the importance of establishing low-level metrics that aid in reasoning about system-level integration and verification, including considerations for common mode failures, logical sensor fusion, and the balance between safety and availability.

Risk modulation and data-driven approaches were discussed for measuring risk either fleet-wise or at the individual vehicle level, with considerations on scoring individual artifacts, Safety Integrity Levels (SIL), and blame analysis. Questions were raised about the systematic approach to decision-making, such as selecting driving routes based on safety assessments.

Finally, the session touched upon the role of falsification in continuous integration and continuous delivery (CI/CD) processes, underscoring the evolving nature of safety and verification in the context of autonomous mobility. The discussion illuminated the multifaceted challenges and innovative strategies involved in ensuring the safe execution of autonomous robots, cars, planes, and trains, with a particular emphasis on the crucial role of low-level control in the broader context of autonomous system safety, all of which will require collaboration in teams of varied expertise to address.

4.12 Techniques for Safety Assurance: : Perception

Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Christof Budnik (Siemens – Princeton, US), Martin Fränzle (Universität Oldenburg, DE), Rong Gu (Mälardalen University – Västerås, SE), Thierry Lecomte (CLEARSY – Aix-en-Provence, FR), Stefan Pranger (TU Graz, AT), Andoni Rodríguez (IMDEA Software Institute – Madrid, ES), Darko Stern (AVL – Graz, AT), and Dirk Ziegenbein (Robert Bosch GmbH – Renningen, DE)


License © Creative Commons BY 4.0 International license
 © Dejan Ničković, Christof Budnik, Martin Fränzle, Rong Gu, Thierry Lecomte, Stefan Pranger, Andoni Rodríguez, Darko Stern, and Dirk Ziegenbein

The main topic of the discussion was the perception module and its role in the safety of smart mobility applications. The perception module is a complex system that typically relies on multiple sensors (e.g. camera, lidar and radar in the automotive domain), which are fused together, sometimes even using additional information extracted from other sources such as maps. The type and configuration of sensors used for the perception is domain-specific. For instance, the railways domain relies more on the infrastructure sensors rather than the train sensors, due to the large breaking time for trains. Modern perception systems must also have predictive capabilities (e.g. predict the next action of the detected pedestrian). Finally, perception has to deal with multiple sources of uncertainty. The participants identified three sources of uncertainty: state uncertainty, classification of uncertainty and existential uncertainty. Perception contracts were proposed as a potential solution towards dealing with uncertainty and building safe perception with error bounds on the state estimation.

The participants noted that the problem of testing perception is quite different from operational perception. Testing perception is notoriously hard and opens many challenging problems – from the synthetic generation of realistic inputs to sensors to the integration of perception in closed-loop testing with the ability to catch rare scenarios. There is also the problem of storage when testing perception. For instance, recording one hour of raw sensor data in a car requires several TBs of storage. There is hence a need to pre-process and filter this data on the edge, before sending it to the cloud. When looking at the perception module in isolation, it is not clear what is the right set of properties to verify. The participants noted that testing perceptions shall be done on a system level, since the important part is the functional impact of a misperception on the overall system. Perception contracts were proposed as a potential solution for building safe perception with error bounds on the state estimation.

4.13 Techniques for Safety Assurance: : Machine Learning in Planning

Alessandro Zanardi (ETH Zürich, CH), Ezio Bartocci (TU Wien, AT), Filip Cano (TU Graz, AT), Rayna Dimitrova (CISPA – Saarbrücken, DE), Marie Farrell (University of Manchester, GB), Mahsa Ghasemi (Purdue University – West Lafayette, US), Radu Grosu (TU Wien, AT), Ichiro Hasuo (National Institute of Informatics – Tokyo, JP), Bettina Könighofer (TU Graz, AT), Assaf Marron (Weizmann Institute – Rehovot, IL), and Doron A. Peled (Bar-Ilan University – Ramat Gan, IL)

License  Creative Commons BY 4.0 International license

© Alessandro Zanardi, Ezio Bartocci, Filip Cano, Rayna Dimitrova, Marie Farrell, Mahsa Ghasemi, Radu Grosu, Ichiro Hasuo, Bettina Könighofer, Assaf Marron, and Doron A. Peled

The discussion concentrated primarily on the crucial role of Machine Learning (ML) in effective planning, emphasizing its indispensable nature in trajectory planning. ML’s ability to consider the conduct of a multitude of actors in scenarios such as switching lanes in traffic jams is indeed invaluable. One of the core discussion points was the significance of re-evaluating our predictions about the environment. This point underscores the necessity for dynamically adapting to changes that were previously unforeseen.

In terms of alternatives and security, the group outlined the importance of having a Plan B to fall back on when assumptions are not met while planning. They also presented the idea of having logical safeguards in place, alongside a Simplex architecture, to ensure that the planning process is robust and able to handle unexpected changes or challenges. A substantial portion of the discussion was devoted to explainable planning, highlighting the need for ML and symbolic planning to work in parallel. The symbolic planner, they detailed, would be capable of elucidating decisions when inquired, an essential feature for understanding and improving the plan’s operation.

The group also shed light on how planning often involves juggling various objectives that might be conflicting. Addressing this challenge, the concept of minimal-violation planning was introduced, which emphasized having different property levels, treating safety and performance at distinct levels.

The discussion then directed toward handling risks and uncertainties, recognizing the crucial role of conformal prediction in this context. Motion Planning, as deterministic as it may appear, also contains a slew of uncertainties, often arising due to perception nuance. Therefore, developing strategies to evaluate and handle these risks plays a part in successful planning. The involvement of symbolic and sub-symbolic elements in end-to-end modular learning, which encompasses perception, planning, and control, was clarified. The group also elaborated on the utilization of Belief-Desire-Intention (BDI) agents to increase explainability and transparency.

Practical aspects like the costs to monitor, the timing, and the demands towards optimization engines were explored, with a particular focus on time budgets. The meeting acknowledged that there are numerous different solutions, and the challenge lies in how to combine them effectively. Emphasis was laid upon verification aspects and how to propagate results between modules—an essential part of improving and refining the planning process.

Lastly, the discussion proposed possible solutions, advocating for the use of all the ML resources, including black-box models, yet enclosing them within a safety net. This safety net could comprise Hoare logic-based controllers, safety filters, as well as rulebooks.

4.14 Techniques for Safety Assurance: Large Language Models

Marie Farrell (University of Manchester, GB), Christof Budnik (Siemens – Princeton, US), Filip Cano (TU Graz, AT), Mauricio Castillo-Effen (Lockheed Systems – Arlington, US), Jyotirmoy Deshmukh (USC – Los Angeles, US), Rayna Dimitrova (CISPA – Saarbrücken, DE), Martin Fränzle (Universität Oldenburg, DE), Rong Gu (Mälardalen University – Västerås, SE), Bardh Hoxha (Toyota Research Institute North America- Ann Arbor, US), Taylor T. Johnson (Vanderbilt University – Nashville, US), Panagiotis Katsaros (Aristotle University of Thessaloniki, GR), Selma Music (Denso Automotive – Eching, DE), Dejan Ničković (AIT – Austrian Institute of Technology – Wien, AT), Necmiye Ozay (University of Michigan – Ann Arbor, US), Andoni Rodríguez (IMDEA Software Institute – Madrid, ES), and Darko Stern (AVL – Graz, AT)

License © Creative Commons BY 4.0 International license

© Marie Farrell, Christof Budnik, Filip Cano, Mauricio Castillo-Effen, Jyotirmoy Deshmukh, Rayna Dimitrova, Martin Fränzle, Rong Gu, Bardh Hoxha, Taylor T. Johnson, Panagiotis Katsaros, Selma Music, Dejan Ničković, Necmiye Ozay, Andoni Rodríguez, and Darko Stern

The discussion centered on the use of LLMs in the development of safety-critical autonomous mobile vehicles. During this discussion, we used ChatGPT to guide our discussion. We asked questions including: What are the hot topics regarding the use of LLMs in the development of safe autonomous systems (for mobility), what are specific applications of LLMs in the development of safety-critical autonomous mobility systems, how can LLMs be used in the development process of safety-critical autonomous mobility systems, what makes the use of LLMs in this domain unreliable and untrustworthy, what role can LLMs have in the development process of planning for autonomous mobility systems, and what are the dangers of using LLMs in this domain?

It is clear that LLMs bring both benefit and potentially create undesirable outcomes when used in the development of safety-critical autonomous mobility systems. Some of the positive ways that LLMs can be used are as follows. LLMs can improve the natural-language understanding of autonomous vehicles, potentially improving the safety of the interactions that the system has with the user. For example, responding accurately to verbal commands. They can also be used to provide clear instructions to passengers in an emergency. LLMs can be used to recognise and interpret traffic signs, signals and road markings. They can potentially even be used in adaptive navigation by assisting in dynamic route planning. One particular strength of LLMs in this domain is to provide fast incident reporting and documentation which can be used in insurance claims, compliance, and post-incident analysis. During the development process itself, LLMs can be used to generate requirement specifications and associated documentation. Other uses include the provision of a natural language interface for design and configuration. This would facilitate intuitive communication with system designers and engineers during the design phase. LLMs can be used to summarise regulatory compliance and standards documentation to help the developers to demonstrate compliance. Along this vein, LLMs can be used to carry out risk/hazard analyses by processing relevant pre-existing documentation. They can also provide useful ways of generating user manuals and training materials as well as improve communication with stakeholders. Each of these positive perspectives can contribute to the overall safety and reliability of these systems.

However, this all comes with some major caveats. LLMs can struggle to gain a deep understanding of context in real-world, dynamic situations and current documentation that LLMs are trained on is likely incomplete. Since LLMs are sensitive to ambiguities in natural language, misinterpretation of cues or context can lead to incorrect, potentially dangerous situations. The age-old adage of “garbage in, garbage out” holds true and LLMs are likely to

inherit biases in training data that could inspire inequitable decision-making which would raise ethical concerns. They are highly dependent on the quality and diversity of their training data but these models may not accurately reflect real-world conditions, leading to unsafe responses. LLMs are vulnerable to adversarial attacks that could be exploited by malicious agents and compromise the safety of the system. The environments that mobile autonomous systems operate within are rapidly changing and LLMs may struggle to adapt to dynamic scenarios such as road closures. Although LLMs are great at producing explanations, they do not possess a deep understanding of the operation at hand, merely they produce reasonable explanations based on data they have seen. However, understanding natural-language and its nuances is not trivial and it is possible that LLMs may misinterpret instructions that are given by the user. Though LLMs produce explanations, the LLM itself is not transparent and does not explain how it made the decisions that it made, this lack of transparency and explainability can make it difficult to trust LLMs and the output that they generate.

Participants

- Houssam Abbas
Oregon State University –
Corvallis, US
- Ezio Bartocci
TU Wien, AT
- Christof Budnik
Siemens – Princeton, US
- Filip Cano
TU Graz, AT
- Mauricio Castillo-Effen
Lockheed Systems –
Arlington, US
- Chih-Hong Cheng
Universität Hildesheim, DE
- Patricia Derler
Zoox Inc. – Foster City, US
- Jyotirmoy Deshmukh
USC – Los Angeles, US
- Rayna Dimitrova
CISPA – Saarbrücken, DE
- Marie Farrell
University of Manchester, GB
- Michael Fisher
University of Manchester, GB
- Martin Fränzle
Universität Oldenburg, DE
- Mahsa Ghasemi
Purdue University –
West Lafayette, US
- Radu Grosu
TU Wien, AT
- Rong Gu
Mälardalen University –
Västerås, SE
- Ichiro Hasuo
National Institute of Informatics –
Tokyo, JP
- Bardh Hoxha
Toyota Research Institute North
America- Ann Arbor, US
- Taylor T. Johnson
Vanderbilt University –
Nashville, US
- Panagiotis Katsaros
Aristotle University of
Thessaloniki, GR
- Bettina Könighofer
TU Graz, AT
- Thierry Lecomte
CLEARSY –
Aix-en-Provence, FR
- Lars Lindemann
USC – Los Angeles, US
- Assaf Marron
Weizmann Institute –
Rehovot, IL
- Sayan Mitra
University of Illinois –
Urbana Champaign, US
- Selma Music
Denso Automotive – Eching, DE
- Dejan Nickovic
AIT – Austrian Institute of
Technology – Wien, AT
- Necmiye Ozay
University of Michigan –
Ann Arbor, US
- Giulia Pedrielli
Arizona State University –
Tempe, US
- Doron A. Peled
Bar-Ilan University –
Ramat Gan, IL
- Stefan Pranger
TU Graz, AT
- Andoni Rodríguez
IMDEA Software Institute –
Madrid, ES
- Darko Stern
AVL – Graz, AT
- Alessandro Zanardi
ETH Zürich, CH
- Dirk Ziegenbein
Robert Bosch GmbH –
Renningen, DE

