Notes from the Safety-Case world-café table

At this table, a wide array of topics related to creating safety cases for autonomous systems were discussed. A specific question that all participants discussed was if safety cases for autonomous systems must become rigorous artifacts in the sense of proving claims with complete certainty or a least quantifying the level of confidence in a claim.

Main discussion points:

- *Currently, significant increase of the level of rigor is unlikely*. Almost unanimously, the participants agreed that increasing the level of rigor to the level of proofs, or explicitly quantifying the confidence in safety-case-claims is nearly impossible. The reasons for this are multiple and some of them are: i) high cost of such analyses, ii) high sensitivity of estimated confidence levels to slightest modifications of used evidence and argument structure, iii) inability to guarantee the completeness of hazard analysis, iv) difficulties in assessing reliability of machine-learning-based components based on testing and simulation results, v) difficulties with applicability and scalability of formal methods for verification of systems properties etc.
- *Standardizing required verification activities is needed*. The overall complexity of autonomous systems, and particularly the introduction of machine-learning-based components, implies that more verification tasks should be performed, but as of now it is not clear which amount and what type of verification is sufficient to claim that an autonomous system is developed to the specified integrity levels. This challenge could be alleviated by developing standardized verification tasks for different types of autonomous systems and different integrity levels.
- *Automotive domain can learn from other domains.* Automotive domain can learn from experiences and best practices in domains like aerospace or railway who have been developing and certifying highly automated, safety critical systems for a long period of time. For example, the concept of autopilot is long present in the aerospace domain, and vehicle to infrastructure communication is long present in the railway domain.
- *Positioning with respect to the quickly evolving state-of-the-art is challenging*. In cases when systems are involved in accidents and accident investigation leads to procedures in front of a court, safety cases can be used to argue about that the system has been developed and verified using appropriate methods. It is customary to also show that the used methods were aligned with the state-of-the-art. In the case of autonomous systems, positioning with respect to the state-of-the-art would be challenging because of the increasing number of methods needed for autonomous systems development and their constant and fast evolution.
- *Addressing confirmation bias is increasingly important*. Confirmation bias is known problem when developing safety cases. Because of the increased criticality of functions performed by autonomous systems, the importance of assuring absence of confirmation bias is even more important. This can be achieved by: i) constructing a single part of a safety cases using several different methodologies and then comparing the results, ii) constructing a “non-safety” case where argumentation argues against the safety case claims etc.