講演番号	250

Considerations regarding the safety assurance of AI-based Automated Driving Systems

Olaf Op den Camp [1)] **Jan-Pieter Paardekooper** [1) 2)]

1) TNO Integrated Vehicle Safety,
Automotive Campus 30, 5708 JZ Helmond, the Netherlands (E-mail: olaf.opdencamp@tno.nl)
2) Radboud University, Donders Institute for Brain, Cognition and Behavior,
P.O. Box 9010, 6500 GL, Nijmegen, the Netherlands

ABSTRACT: Automated Driving System manufacturers and AV-developers make more and more use of AI-based systems in the development of their products. In some cases, an end-to-end (E2E) AI approach is followed in which no longer a distinction is made between perception, path planning and actuation in the ADS of the vehicle. The paper presents considerations regarding the safety assurance of AI-based systems. The vulnerabilities of AI-based systems and how these vulnerabilities have a negative influence on safety assurance will be discussed. It will be shown how the design of AI-based systems can be improved to allow for proper safety assurance.

KEY WORDS: **Automated Driving, Safety Assurance, AI, Machine Learning**

The paper proposes an approach that enables a proper safety assessment for Automated Driving System (ADS) configurations that are based on e2e trained path planning. In this approach a *guard rail* is added in parallel to the AI-trained ADS pipeline. The guard rail consists of a perception and reasoning mechanism on a symbolic level that is human-interpretable and that provides a safe fallback trajectory for the vehicle in case of unexpected and undesired behavior of the AI path planner. The guard rail is completed by a Meta-Cognitive Component (MCC) that is capable to decide between the trajectory provided by the AI planner and a safe fallback trajectory that is based on the context provided by the reasoning mechanism.

The proposed approach allows the use of an e2e-trained ADS that, if developed correctly, provides comfortable human-like driving behavior that is also recognizable and interpretable for other human road users. While it is not possible to guarantee safe behavior of standalone e2e-trained ADSs under all circumstances and scenarios, the addition of the guard rail enables proper safety assurance for the complete system. This comes with the cost of development of an additional perception and reasoning mechanism and an MCC, and the additional cost of safety assessment of the additional systems.

Future research directions include the efficiency and scalability of the computation of the guard rails and how and if to make use of the perception output of the e2e network, if available. In addition, further research into the explainability aspect of the MCC component is needed, including the trade-off between optimal functionality of the ADS and the ability of the guard rails to provide enough safety assurance. Depending on the choice of configuration, different options to generate a safe context-aware fallback trajectory will be studied.

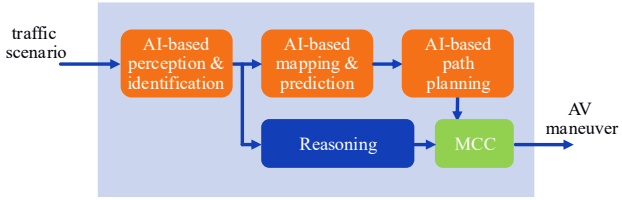

Fig. 1 Guard rail architecture for traditionally trained AI modules in ADS.

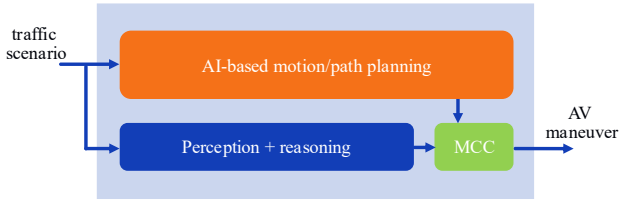

Fig. 2 Guard rail architecture for 'vanilla' e2e trained AI-based path planning in ADS.

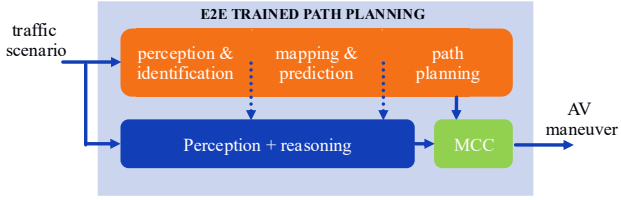

Fig. 3 Guard rail architecture for e2e trained path planning in a modular design.