

## ABSTRACT

Self-Driving cars are one of the hottest topic of the decade. With the rise of *machine learning*, it seems like we are going to see cars driven by artificial intelligences on the roads very soon. The complexity of these systems require extra care due to the complexity of the environment in which they provide their service. The architectural features of these systems and the ultra-high dependability required, makes them fall in the category of *Safety-Critical Cyber-Physical System of Systems*. At the same time, the techniques used to teach to a computer how to drive a car are still being studied, but little attention is given to the *emergent behaviour* that comes from the combination of artificial intelligences and error checkers. In this work we explored this topic, putting our focus on the system in its entirety, aiming at developing a monitoring activity capable of giving an overview of the system's performances, starting from common measures of interest well known in the systems engineering community. Moreover, due to the asymmetric nature of these systems, we also focused on the long-term effectiveness of these error checkers when the neural network becomes an "*expert driver*", wondering if these AIs can get so expert that the use of an error-checker actually becomes *detrimental* for the system in providing its service.

To conduct this study, an *Object-Detection Module* was developed using non-AI techniques, to be used as an error-checker. A *DDPG Agent* was at first trained with a *default strategy* and then retrained using different strategies, to observe how these could have an impact on both the agent's performances and the error-checker's performances.

This thesis presents an overview of the current state-of-art in the *dependability assessment and monitoring* for safety-critical systems and the issues that prevents us to deploy self-driving cars worldwide. An experimental methodology was developed, respecting all the requirements that a monitoring activity must have to be acceptable, and used to monitor these systems in a simulated environment.

The methodology proved to be an effective mean for the evaluation of the whole system in a simulated environment, by combining few metrics together we had a better understanding of its behaviour. The results collected seems to suggest that there is a correlation between *how good* the neural network is in driving a car, and the accuracy of the error-checker. This work is nothing but a starting point for further studies, aiming at understanding the complexity of these systems.