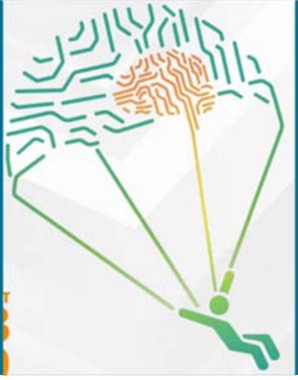
Fraunhofer
IESE

FRAUNHOFER INSTITUTE FOR EXPERIMENTAL SOFTWARE ENGINEERING IESE



SAFE AI – SAFE SOLUTIONS WITH AI INSIDE

Autonomy of technical systems is one of the current megatrends in industry. Artificial Intelligence (AI) is one key enabler for its realization. Machine learning (ML) and in particular, Deep Learning (DL) is used for implementing functionality for autonomy Success cases show an excellent performance for selected tasks, such as the classification of objects in live images. Being part of a larger ecosystem, failure of an autonomous system can cause economic or even personal damage. Therefore, special measures to ensure the functional safety of systems and in particular those with AI inside are required. Existing methods prescribed by standards such as IEC 61508 are only partly applicable, or do not cover the full scope of the issue of functional safety of autonomous systems.

ML also goes along with a shift in the development paradigm: While developing a classical system, we typically specify, implement, and test it, considering safety requirements. If we use ML for certain, typically complex tasks, an algorithm generates a model that is used in the final system from data (or try and error). One could say the algorithms produce the „code". However, the way how those models are created is known to be (often) not deterministic and even worst, models' quality depend on, among others on the data, the process, the algorithms applied (product), and developers' awareness (people). To make it even worse, current verification and validation methods are impacted by a lack of specification and non-interpretability of ML-based solutions.

Fraunhofer-Institut für Experimentelles Software Engineering IESE

Fraunhofer-Platz 1
67663 Kaiserslautern

Kontakt
Dr. Andreas Jedlitschka
Tel. +49 631 6800-2260
andreas.jedlitschka@iese.fraunhofer.de
www.iese.fraunhofer.de

In a recent project, we developed a catalog of methods and identified best practices to address common challenges and to increase the safety of solutions with AI inside.

A development team shall be aware of the following challenges, among others:

- related to the process: inherently different development paradigm, inability to obtain a representative and comprehensive data set, verification and validation, assurance of safety-related properties
- for the model: generalization and convergence of models, interpretability of the resulting model,
- during operation: how to deal with noisy data, run-time safety, robustness of the model against adversarial attacks.

Best Practices such as simulation, automated data generation, data augmentation, or parametric controllers can be used to address the data challenge.

Rule extraction, equivalent class-based testing, and approaches as DeepXplore are examples for addressing the verification and validation challenge.

The selection and combination of the most appropriate techniques are solution specific, however, within this tutorial, we will go through the challenges and characterize the best practices how to address them aiming at a safe solution.

Within this seminar, we present the state-of-the-art in safety and artificial intelligence. We discuss the challenges resulting from the use of AI approaches in safety-critical solutions and present how to address them by specific steps in the development process, by exploiting AI products' (approaches) characteristics, and by raising people awareness.

Language: English or German

Target group: Safety Engineers and Data-Scientist from companies that use or develop AI-Components in/for safety-critical solutions.

Content: Motivation, State-of-practice in safety and AI, Challenges and Best Practices (Process, Product, People), and outlook.