SAFEXPLAIN: SAFE AND EXPLAINABLE CRITICAL EMBEDDED SYSTEMS BASED ON AI

Description

Deep Learning (DL) techniques are key for most future advanced software functions in Critical Autonomous AI-based Systems (CAIS) in cars, trains and satellites. Hence, those CAIS industries depend on their ability to design, implement, qualify, and certify DL-basedsoftware products under bounded effort/cost.There is a fundamental gap between Functional Safety (FUSA) requirements of CAIS and the nature of DL solutions needed to satisfythose requirements. The lack of transparency (mainly explainability and traceability), and the data-dependent and stochastic natureof DL software clash against the need for deterministic, verifiable and pass/fail test-based software solutions for CAIS.

SAFEXPLAIN tackles this challenge by providing a novel and flexible approach to allow the certification hence adoption of DLbasedsolutions in CAIS by

- architecting transparent DL solutions that allow explaining why they satisfy FUSA requirements, withend-to-end traceability, with specific approaches to explain whether predictions can be trusted, and with strategies to reach (andprove) correct operation, in accordance with certification standards.
- devise alternative and increasinglycomplex FUSA design safety patterns for different DL usage levels (i.e. with varying safety requirements) that will allow using DL inany CAIS functionality, for varying levels of criticality and fault tolerance.

SAFEXPLAIN brings together a highly skilled and complementary consortium to successfully tackle this endeavor including 3 research centers, RISE (AI expertise), IKR (FUSA expertise), and BSC (platform expertise); and 3 CAIS case studies, automotive (NAV), space(AIKO), and railway (IKR). SAFEXPLAIN DL-based solutions are assessed in an industrial toolset (EXI). Finally, to prove that transparency levels are fully compliant with FUSA, solutions are reviewed by internal certification experts (EXI), and external ones subcontracted for an independent assessment.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación