OC - (26348) - WHAT WE KNOW WE DON'T KNOW: SYSTEMATIC RISK IDENTIFICATION FROM CONTRADICTORY SAFETY DOCUMENTATION

Ebrahimi, Masoud (Sweden)[1]; Hänninen, Kaj (Sweden)[1]; Lundqvist, Kristina (Sweden)[1]; Sirjani, Marjan (Sweden)[1]

1 - Mälardalen University

Ensuring accountability and responsibility in human-centered safety-critical systems, such as autonomous vehicles and medical devices, is essential. These systems must assure safety, comply with normative constraints, ethical principles, and regulatory standards while navigating complex human-machine interactions. Existing methods often fall short in formally verifying normative properties and attributing responsibility for many reasons, insufficient knowledge of the unknown risks that remain in systems being one of them. In this paper we propose an approach to bridge the gap between the Open World Assumptions (OWA), produced e.g. by ontological knowledge graphs, and the Closed World Assumptions (CWA) analysis, using Kripke structure construction and formal analysis. The approach presented can increase knowledge of the unidentified risks in the set of the closed world assumptions of the total residual risk. Our methodology addresses the challenge that knowledge graph extraction from informal documentation is not formulated concisely and completely enough to be fully validated or verified, preventing automation and correctness guarantees. We systematically construct Kripke structures where each state represents a world model satisfying specific knowledge graph configurations derived from seemingly contradictory documentation. Additionally, the framework can integrate expert feedback, ensuring alignment with domain knowledge and regulatory expectations. As such, it serves as a complement to existing analysis methods, and thereby also verification and validation processes of safety-critical systems. A case study on an automotive seatbelt system demonstrates the framework's ability to construct systematic Kripke models, use it to identify contradictory specifications, and enable rigorous safety analysis. By bridging the gap between informal documentation and machine-checkable models, our framework offers a rigorous and transparent solution for trustworthy human-centered systems.

Palavras-chave : Unknown risks identification, Accountability and responsibility, Safety-critical systems, Open-World Assumption, Known Unknown