

LESS Guidance: Using LLMs for Requirement Formalization and Test Generation in Safety-Critical Engineering

Abhishek Shrestha, Holger Schlingloff, and Jürgen Großmann

Speaker: Abhishek Shrestha (abhishek.shrestha@fokus.fraunhofer.de), Fraunhofer FOKUS

Abstract:

Requirement formalization and test case generation, whether through manual effort or script-based automation, remains a major challenge in systems engineering, often requiring significant time and resources. While Large Language Models (LLMs) offer a promising alternative for automation, generating accurate and structured outputs remains difficult due to the ambiguity of natural language and limitations such as hallucinations and syntactic errors.

In this presentation, we discuss an LLM-based workflow for transforming natural-language safety and security requirements into a controlled natural language representation called LESS (Language for Embedded Safety and Security). LESS is a domain-specific language with a formally specified grammar. It is designed to express safety and security requirements in a structured, precise, and machine-interpretable form while remaining close to natural language.

More specifically, we explore structured prompt engineering as a lightweight method for guiding LLMs in generating LESS-compliant requirements. While existing approaches such as Grammar-Constrained Decoding (GCD) enforce syntactic validity, they require high implementation overhead. We illustrate that grammar-compliant generation can be achieved through prompt design with Few-Shot Learning (FSL), making integration into existing workflows faster and more flexible. The generated LESS requirements are then validated for syntactic correctness and semantic consistency before being used as an intermediate representation (IR) for LLM-supported test case generation.

The workflow is evaluated in the context of safety and security requirements from two industrial critical-system use cases: a Ventricular Assist Device case study and the E-GAS monitoring concept. We benchmark LLMs with different capabilities and compare LLM-generated test cases produced directly from natural-language requirements with those generated from LESS-formatted requirements.

The work was carried out as part of the ongoing ITEA GENIUS project on Generative AI for the Software Development Life Cycle¹. Our results indicate that well-designed prompts can improve the accuracy and consistency of generated requirements. Furthermore, using LESS as an IR reduces ambiguity during test generation, resulting in fewer hallucinations and fewer incorrect details. At the same time, the use of LLMs in this context introduces important risks and drawbacks. These include hallucinated interpretations, syntactically correct but semantically wrong formalizations, the need for developers to understand the grammar structure in order to formulate well-engineered prompts. Therefore, the talk will focus on both the opportunities and open questions around AI-supported requirements engineering and test generation, including reliability and validation efforts.

¹ <https://itea4.org/project/genius.html>