

Evaluating LLMs for Automated Requirement and Test Case Generation in Railway Signaling Systems

Friday 19 September 2025 14:15 (15 minutes)

Large Language Models (LLMs) have shown potential in supporting requirements engineering through automation, especially in regulated and safety-critical domains. This paper evaluates the capabilities of 3 well-known LLMs (GPT-4, Claude, Gemini) in transforming user requirements into structured product requirements and corresponding test cases within the context of railway signaling. A custom dataset of client requirements, inspired by realistic signaling scenarios, was developed to enable consistent evaluation across models. Each model's outputs were assessed using defined metrics, including completeness, correctness, consistency, and traceability. The comparative results highlight variations in quality and structure of the generated artifacts, with specific strengths observed for different tasks. While all three models demonstrate promise, their reliability and consistency vary, and human oversight remains essential. This study provides practical insights into the applicability of current LLMs for augmenting early-stage requirements and verification workflows in critical systems engineering.

Author: Mr OȚELEA, Ionuț-Gabriel (National University of Science and Technology POLITEHNICA Bucharest)

Co-authors: Mr PINTEA, Bogdan (Hitachi Romania); Prof. RUGHINIȘ, Răzvan Victor (National University of Science and Technology POLITEHNICA Bucharest)

Presenter: Mr OȚELEA, Ionuț-Gabriel (National University of Science and Technology POLITEHNICA Bucharest)

Session Classification: Pervasive Systems and Computing

Track Classification: Doctoral Symposium