

Formal Verification of a Curated Dataset through Frama-C tools: An Experience Report⁴

Arshad Beg

Principal Investigator(s): Rosemary Monahan and Diarmuid O'Donoghue

1 Introduction

Briefly, the aim of VERIFAI¹ project is to leverage Large Language Models (LLMs) to generate formal specifications, bridging the gap between informal requirements and verifiable code. As part of VERIFAI project, we utilised the power of Frama-C ecosystem to perform the formal verification of an open source recently published dataset².

2 Initial Experiments

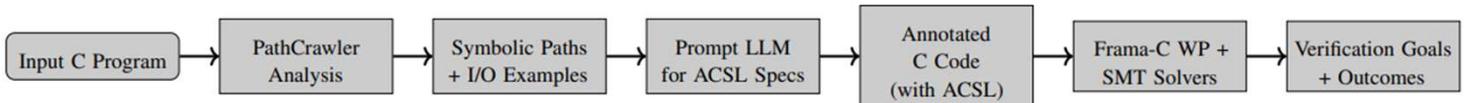


Figure 1: Methodology of the initial experiments following the approach of 2, combining LLM with symbolic analysis tools in the Frama-C ecosystem. The workflow integrates path-based I/O examples and verification outputs to guide the generation of context-aware ACSL specifications.

As a first step, we picked up 36 C programs from iFM2024 tutorial on Frama-C ecosystem and applied the above work pipeline to C programs.

Conclusions drawn from Verifiers Output: (1) Across the evaluated 36 C programs, 28 were evaluated successfully, in which, approximately 94% of all verification goals were successfully proven, demonstrating high reliability of the verification toolchain. (2) Failures were primarily due to timeout or partial prover coverage on complex arithmetic and array manipulation examples. Overall, the combined prover strategy offers robust verification performance with only marginal cases requiring further proof engineering or model refinement.

Research Paper³: The outcome of this stage is reported in 3. It reported: (1) A lightweight literature survey and analysis of forty (40) research articles appeared in 2021 – 2025. (2) Outcome of the experiments mentioned above (3) Challenges and prospects of the field.

3 Tools

- Dataset Description:** these programs were grouped into three predefined categories: (i) **correct** implementations, (ii) implementations exhibiting **obvious** differences between function variants, and (iii) implementations with **subtle** differences in their function bodies. Each of the fifty base programs was analysed across these three categories, yielding a total of 180 evaluated instances. This multi-category organization enables the systematic study of tool behaviour across a controlled spectrum of semantic and syntactic variations.
- Frama-C Tools:** (i) **EVA** for static analysis (ii) **RTE** for ACSL annotation generation (iii) **Pathcrawler** for automated test case generation (iv) **WP** verifiers for formal proofs

4 Empirical Evaluation

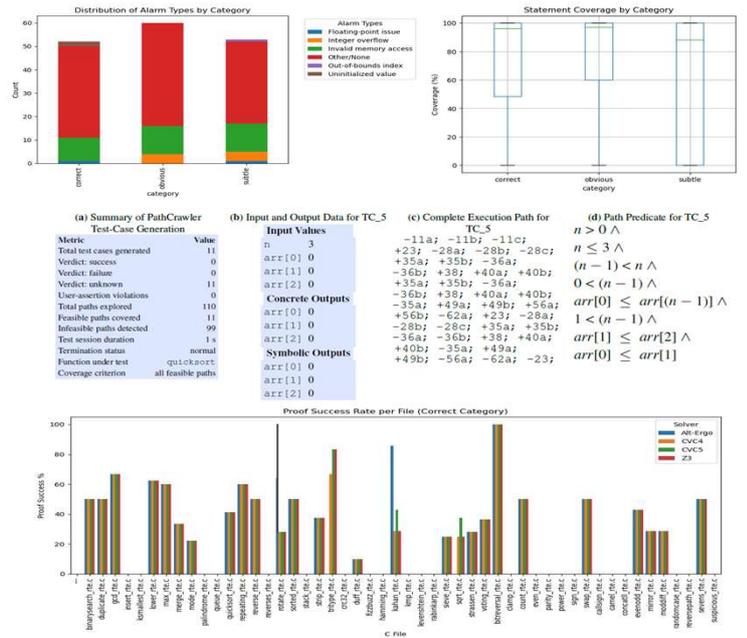


Figure 2: (a) EVA analysis results across the three program categories: (i) distribution of alarm types, (ii) statement coverage of analysed code (b) One of test case generated by PathCrawler for QuickSort Program (c) WP verifiers performance

5 Conclusions

The complete details and results of the work done in this research can be found in 4. We were able to exploit the power of Frama-C tools for formal verification of two different C program sets.

References

- A. Beg, D. O'Donoghue, and R. Monahan, "Formalising Software Requirements with Large Language Models," ADAPT Annual Conference 2025, Dublin City University, May 15, 2025. Available: <https://arxiv.org/abs/2506.10704>.
- G. Granberry, W. Ahrendt, and M. Johansson, "Seeking Specifications: The Case for Neuro-Symbolic Specification Synthesis," *arXiv preprint arXiv:2504.21061*, 2025. Available: <https://arxiv.org/abs/2504.21061> with dataset URL: https://github.com/ggranberry/intent_dataset
- A. Beg, D. O'Donoghue, and R. Monahan, "Leveraging LLMs for Formal Software Requirements: Challenges and Prospects," in *Proc. OVERLAY 2025: Artificial Intelligence and Formal Verification, Logic, Automata, and Synthesis*, CEUR Workshop Proc., vol. 4142, pp. 95–105, 2025. Available: <https://ceur-ws.org/Vol-4142/paper11.pdf>
- A. Beg, D. O'Donoghue, and R. Monahan, "Traceable and verifiable software requirements: A synthesis of AI-enabled formal methods," *submitted* to Journal of Software Testing, Verification and Reliability, Nov. 2025, under review, ver. v1. Available: <https://doi.org/10.5281/zenodo.17772835>

This research was conducted with the financial support of Science Foundation Ireland at ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology at Maynooth University [13/RC/2106_P2].