

## **ABSTRACT**

*The use of AI has become widespread, and one area that has been significantly impacted is application development. Github Copilot is one of the AI-based tools that can simplify application development and improve efficiency in the development process. However, the efficiency of Github Copilot also depends on the AI model it uses. This study aims to quantitatively compare the efficiency of the GPT 4.1 and Claude 3.7 Sonnet AI models using Github Copilot in Agent Mode, with a case study focused on developing a backend REST API application. The study employs a comparative approach by conducting experiments to develop several features in parallel across each model. The measurement used in this study is the Goal Question Metric (GQM) method, with the metrics measured being the number of prompts and the duration required by each prompt to complete each task. The conclusion of this study shows a correlation between speed and the model used for code generation. GPT 4.1 demonstrates the advantages of this model, as seen in the shorter duration required by the model and the fewer prompts needed. However, Claude 3.7 is more effective in handling larger code generation requests in a single prompt.*

**Keywords:** Generative AI, LLM4SE, Github Copilot, Rest API, Backend