CAPSTONE
BYU

Evaluating Open-Weight LLMs for Code Vulnerability

Background

Large legacy codebases often contain vulnerabilities that are difficult and time-consuming to find. Over the past year, we explored how artificial intelligence can improve the efficiency and scalability of vulnerability detection while aiding human analysts.

AI BUGBUSTERS

Objective

To evaluate open-weight LLMs for detecting security vulnerabilities in production code, focusing on accuracy, speed, and privacy, while delivering a report of findings and a benchmark suite for future models.

A model is chosen and vLLM parameters are set

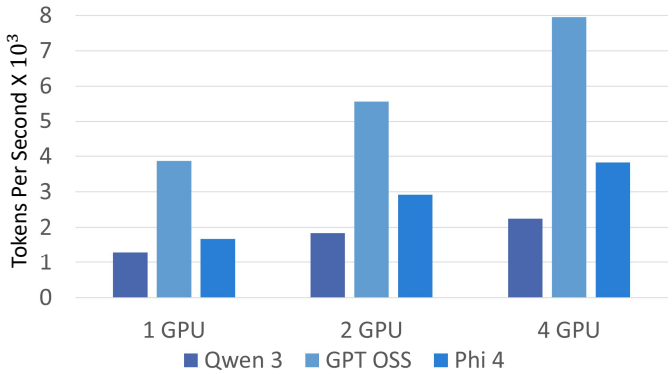
Hardware and job parameters are set, and Slurm job is scheduled

The vLLM server is started and the AI model is loaded

The codebase is scanned for vulnerabilities by the AI model based on the given prompt

The vulnerability and performance data is saved as JSON for further analysis

AI Model Performance



Key Success Measure

Large language models process text as **tokens**, or small pieces of input and output. Our main metric was **tokens per second throughput**, which measures how quickly a model runs. This was a key success measure because it reflects model speed and scalability in real-world vulnerability detection.

Impact

Our project provides a strong starting point for broader implementation across Church Education System (CES). Our findings will guide future AI deployment in improving security across CES services.

BYU Office of Information Technology
2025-2026

AI Bug Busters
Christopher Bingham, Will Hartman, Jinook Kim, Surendra Puppala, Stuart Spear, Ethan Wilhelmsen
Coach: Isaac Teuscher