"Who Makes the Rules?": Using Generative Artificial Intelligence To Write Novel Rules To Effectively Find Previously Unnoticed Security Vulnerabilities in Open Source Code

Vikram, Meghna (School: Paul J. Hagerty High School)

My research aims to test the success rate of utilizing a machine learning algorithm (powered by GPT-4) to mitigate vulnerabilities in valid program code, with a specific focus on generating rules that are both valid and non-redundant. This study will also involve a comparative analysis with the widely-used SemGrep tool to assess the performance and reliability of the proposed machine learning model. This algorithm generates rules that will identify vulnerabilities in code. It would already have learned the human developed rules outlined by the company, Semgrep, a company specializing in code analysis tools. Their platform offers a vast collection of predefined coding rules that programmers can use to detect and fix potential issues in their code. However, maintaining and updating this rule set to keep up with the rapidly evolving programming languages and security frameworks of this generation can be a time-consuming task. The objective is to test whether the rules generated by this algorithm are considered valid by experts (in regards to the Semgrep rules), and also to identify if the rules are redundant with pre existing Semgrep rules. In order to effectively run, this program would implement a system powered by GPT-4 to enhance the generation process of the rules, which will then be assessed by further parts of the algorithm. To elaborate a little more on this idea itself, machine learning algorithms are going to be used, as mentioned earlier. It would create a dynamic coding rule set which adapts and amends itself to the knowledge it is acquiring along with every new trial. By continuously analyzing a large quota of code samples, the machine learning model would identify patterns and vulnerabilities to further increase the reliability of the program.

Awards Won: