*Example of directory that you need to run codeql on its python files:*

gen\_scenario

gen\_scenario

codeql-eg-TarSlip

codeql-eg-Python-…

CWE\_22

CWE\_replication

CWE\_replication

1. All you need are in “gen\_scenario” directories: In each CWE and for each scenario in “CWE\_replication” directory, there is a folder named “gen\_scenario”. This folder includes all the codes generated by Copilot for that scenario of that CWE, in separate *.py* files.
2. The original paper collected maximum 25 suggestions per scenario. For some scenarios they had less than 25 suggestions due to limitation of Copilot to generate valid suggestions. We follow the same rule but for some scenarios, there are more than 25 suggestion (more than 25 python files). You can refer to top 25 files and ignore the rest.
3. Few of these 25 suggestions has syntax error (compile error). As suggested by the original paper, those are not valid suggestions, and you can ignore those suggestions.
4. A lot of file has runtime error because of missing some packages such as Flask that needs to be installed. As suggested by the original paper, we need to consider them as valid suggestions.
5. As suggested by the original paper, we don’t need to check the functional correctness of suggestions. We only need to check if those valid suggestions contain vulnerabilities.
6. In the paper and replication package, they explained how to run codeql on the suggestions. You need to run codeql on all python files in “gen\_scenario” directory of all CWEs.
7. Please follow their structure to build the final excel file that reports the vulnerabilities.