Privacy Concerns: UnitPylot

1. Overview

UnitPylot integrates with GitHub Copilot and other AI services to assist developers with writing and testing Python code. As part of its functionality, it processes user code, test cases, and related context within the development environment.

This document outlines the key privacy concerns identified during the development of UnitPylot, along with the measures taken to mitigate them.

2. Types of Data Processed

UnitPylot may process the following types of user data:

- **Source code and test files**: The extension reads and analyses Python code written by the user to generate and evaluate unit tests.
- **Contextual metadata**: Includes file names, function names, and surrounding code used to provide intelligent suggestions.
- **User interactions**: Such as commands executed within the extension, selected prompts, or toggle settings for features like snapshot history or AI suggestions.

3. Privacy Concerns

a. **Exposure of Sensitive Code**

Users may be working on private, proprietary, or confidential codebases. Sending this code to external AI APIs (e.g., for prompt completion) introduces a risk of unintentional data exposure.

b. **Inadvertent Data Logging**

If logs include file contents, identifiers, or metadata, it could unintentionally capture sensitive data, especially when debugging or monitoring errors.

c. **Third-Party Service Use**

If AI services are used through APIs (e.g., OpenAI), data may be transmitted to third-party servers, where it is subject to those services' privacy policies and data retention practices.

4. Mitigation Strategies

To address the above concerns, we implemented or propose the following safeguards:

- **Local Code Processing with Pytest:** Core functionality such as test execution and coverage analysis is handled locally through the `pytest` framework. This ensures that source code is not transmitted externally during the testing process.
- **Configurable AI Model Sources:** Users can configure the extension to use their own third-party API keys (e.g., OpenAI, Hugging Face) or local LLMs for AI-powered suggestions. This flexibility allows

users to choose services they trust and maintain better control over code privacy.
- **Data Minimisation:** Only minimal, relevant snippets of code (e.g., a specific function or test block) are included in AI prompts to reduce the potential exposure of unrelated or sensitive information.

5. Licensing

This software is released under the **MIT License**. By using UnitPylot, you agree to the terms outlined in this license, including limitations of liability and warranty disclaimers.

6. Contact Information

For questions or concerns regarding privacy or data usage in UnitPylot, please contact the UnitPylot team:

- **Aaditya Kumar**: aaditya.kumar.23@ucl.ac.uk, aadityastyles@gmail.com
- **Asmita Anand**: asmita.anand.23@ucl.ac.uk, asmitaanand04@gmail.com
- **Gughan Ramakrishnan**: gughan.sowndravalli.23@ucl.ac.uk, rs.gughan@gmail.com
- **Swasti Jain**: swasti.jain.23@ucl.ac.uk, swasjn4@gmail.com