Project Number: 2

Project Title: Automated Program Repair System Based on Large Language Models/Artificial Intelligence

Project Clients: Jiaojiao Jiang

Project specializations: Computer Science and Algorithms;Artificial Intelligence (Machine/Deep Learning, NLP);Large language model;

Number of groups: 3 groups

Main contact: Jiaojiao Jiang

Background:

As the scale and complexity of software development continue to increase, fixing code defects and vulnerabilities has become an essential challenge for development teams. Traditional manual repair methods are not only time-consuming and labour-intensive but also prone to omissions, making it challenging to meet the needs of modern software development. In recent years, artificial intelligence technologies, huge language models (LLMs) and deep learning models have made code understanding and generation breakthroughs, providing new possibilities for automated program repair. Currently, the mainstream code repair methods in the industry are mainly categorized as follows: Pattern-matching-based methods: repair using predefined error patterns and repair templates, but the scope of the application is narrow. Test case-based approaches: program validation and repair generation using test cases, but relies on high-quality test suites. Machine learning-based approaches use deep learning models to learn code patterns and repair strategies, but early models have limited repair effectiveness and generalization capabilities. Large language model-based approaches utilize the robust semantic understanding and code generation capabilities of pre-trained language models, which show better potential for repair. In such a context, developing an automated project repair system based on advanced AI models has essential research value and practical significance.

Goal:

Project goals may include but are not limited to（not include all）:

Establishing a complete project remediation process

Integrate multiple AI/large language models/deep learning models for code understanding and analyze program dependencies.

Designing model fusion strategies

Developing model integration framework

Implement collaborative decision-making mechanisms between different models.

Designing dynamic model selection strategies

Optimize the comprehensive evaluation of multi-model outputs

Establish a multi-dimensional restoration quality assessment system

Realize automated testing and verification mechanism

Provide manual audit interface

Requirements and Scope:

This project aims to develop an automated project repair system based on various AI/big language/ models/deep learning/machine learning models to enable the automatic detection and repair of code defects. The automated program repair system must include the modules needed to accomplish the task, and each module must be able to fulfil its responsibilities correctly.

Required Knowledge and skills:

The system developed in this project needs to be able to perform automated program repair tasks. The system must input a buggy program and output code that fixes the bugs and passes validation. The system may need to include tasks such as defect detection of the code, dependency analysis of the code, and generation of program patches. Moreover, the system requires a reasonable test module to test whether the fixed code is fixed correctly or not.

Expected outcomes/deliverables:

Expected deliverables include:

An automated program repair system that is trained and capable of functioning and completing its tasks.

Detailed developer documentation/technical documentation/user's manuals and other documentation

Standards-compliant code comments/system demos/usage resources, etc.