

31 Jan 2026

Holistic Multimodal Intelligent Agent Framework for Software Engineering

Research Statement

Background

I am Nhat-Minh Nguyen, a graduate of the Computer Science and Engineering department at Ho Chi Minh City University of Technology, Vietnam. I have been working as a Research Engineer at the SMU School of Computing and Information Systems under the supervision of [Professor Lingxiao JIANG](#). During this time, I have developed a strong foundation in software engineering (SE) and Machine Learning/Deep Learning (ML/DL), specifically applying Graph Learning and Large Language Models (LLMs) to programming languages. Our team has published four papers on Vulnerability Detection for smart contracts using Graph Neural Networks (GNNs) and LLMs. Based on these research achievements, we are processing a patent filing and have launched a startup, MANDO (mandoscan.com). Following my positive experience in SMU's research environment, I am eager to pursue a PhD to further advance the field of AI4SE.

Research Statement

1. Motivation: From Secure Smart Contracts to Holistic Software Intelligence

To follow up on our achievements at SMU, I would like to go further in intelligent SE fields. Beyond security, I aim to build AI agents that function across the full software lifecycle by processing multiple modalities (multimodal AI agents) and indexing the code context under graph formats (GraphRAG). I believe the approach can bridge the gap between complex SE challenges and current generative AI capabilities that treat software development as a text-processing task.

2. Methodology

a. Multimodal Intelligent Agents

Real-world software engineering is inherently multimodal; developers work with code (text); logic flow and project structure (graphs); system designs and UI/UX (diagrams/visuals). I aim to build agents capable of processing these inputs simultaneously.

Visual to code: I plan to research agents that can ingest visual inputs, such as UML diagrams, whiteboard sketches, or UI screenshots, and generate the corresponding backend logic or frontend components. This requires training models to align visual embeddings with the structural representations of code.

Multimodal Bug Fixing: In GUI apps, debugging often involves correlating a runtime error (log text) and a visual glitch (screenshot). A multimodal agent can utilize **Cross-Modal Attention** to map the visual symptom to the specific node in the underlying Code Property Graph (CPG) that caused the error, offering significantly higher accuracy than text-based models alone.

b. Graph-based Retrieval-Augmented Generation (GraphRAG)

Standard Retrieval-Augmented Generation (RAG) fetches textual code snippets, which often fails to capture the complex logic and file structure of the software. Moreover, unlike natural language, code exhibits **high syntactic variance**—functionally identical code can look completely different due to variable naming and algorithmic choices. The technique is solving the Context Challenge in SE.

Structural Retrieval: Instead of retrieving similar text, the agent will retrieve relevant *sub-graphs* (e.g., Control Flow Graphs, Program Dependence Graphs, or Call Graphs). This allows the agent to access verified "logic patterns" rather than just surface-level syntax for code assistant tasks.

Application: In general SE, it could involve retrieving the correct API usage graphs for complex internal/external libraries, ensuring the agent generates code that is not only syntactically correct but also logically sound. By following the retrieved standard project structure, the application can avoid security issues such as reentrancy patterns.

3. Expected Contributions

My research aims to deliver a unified AI framework adaptable to multiple SE tasks, including **bug detection** to **repair**, and **code generation**. By combining the structural rigor of GraphRAG with the intuitive capabilities of Multimodal Agents, I aim to create AI systems that are not only powerful but also safe and reliable for industrial applications. This

work will directly support the technological growth of the MANDO ecosystem and contribute robust tools to the wider software engineering community.

4. Future Directions and Applications

In the future, I want to develop a holistic AI Agents framework that can cover the full software lifecycle. My ultimate goal is to build a holistic system that automates not just **coding**, but also **CI/CD pipelines** and **operational monitoring**. By embedding industry standards into the agent's graph-retrieval process, we can ensure that AI-generated software is secure by design and ready for the real world.

School of Computing and Information Systems
80 Stamford Road, Singapore, 178902