**Hong Kong Institute of Vocational Education**

**Discipline of Information Technology**

**IT114116 – HD in Data Science and Analytics**

**Final Year Project – Big Data Analytics (ITP4870M)**

**Project Proposal (AY2025/26)**

**Fraud Detection Using Large Language Models and Retrieval-Augmented Generation**

**Group Members:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Name (Student ID)** | **E-Mail** | **Phone Number** |
| 1. | TAN James Anthroi (240350922) | 240350922@stu.vtc.edu.hk | 5911-5923 |
| 2. | AQSA-AHMED (240044574) | 240044574@stu.vtc.edu.hk | 6433 4130 |
| 3. | LIN Yueying (240444846) | 240444846@stu.vtc.edu.hk | 5958-4735 |
| 4. | TAN Xiuhao (240253372) | 240253372@stu.vtc.edu.hk | 6768-6457 |

# **Introduction to the proposed project (Statement of problem to be solved)**

The prevalence of online fraud and scams has grown at an unprecedented rate, posing significant risks to individuals, organizations, and the broader digital economy. Fraudulent activities not only result in financial losses but also undermine trust in digital platforms and communication systems. A key challenge lies in the dynamic nature of fraudulent behavior: as detection systems evolve, so too do the strategies employed by fraudsters. Many users lack the expertise to identify subtle indicators of deception, leaving them vulnerable to exploitation. This project seeks to address this critical issue by investigating the potential of Large Language Models (LLMs), combined with Retrieval-Augmented Generation (RAG), to detect fraud patterns and assist users in identifying scams with greater accuracy and timeliness.

# **Background (Background of the problem)**

The rapid expansion of digital communication and e-commerce has created fertile ground for fraudulent practices. Traditional fraud detection systems, which are often rule-based and reliant on static keyword matching, struggle to adapt to the evolving sophistication of scams. For instance, phishing attempts increasingly mimic legitimate corporate communication, making them difficult to distinguish from authentic messages.

Recent advances in natural language processing (NLP) have demonstrated the capacity of LLMs to capture semantic nuance, contextual meaning, and linguistic subtleties. These capabilities position LLMs as promising tools for fraud detection, particularly in identifying deceptive intent embedded within text. By training models on authentic fraud cases, it becomes possible to enhance their ability to recognize suspicious patterns, thereby improving adaptability to emerging scam techniques. This project builds upon these insights to explore the integration of LLMs and RAG into a fraud detection framework.

# **Outline of proposed solution**

The proposed solution is a web-based fraud detection platform that integrates LLMs with RAG to classify and analyze suspicious content. Fraud-related cases will be collected through web scraping from diverse sources such as Reddit, HK01, Yahoo, SCMP, and the Hong Kong Monetary Authority (HKMA). These cases will be vectorized and stored in ChromaDB, forming a knowledge base for retrieval.

When a user submits potentially fraudulent text (e.g., an email, message, or transaction description), the system will employ RAG to retrieve the most relevant fraud cases from the database. These retrieved examples will provide contextual grounding for the LLM, enabling it to make more accurate and explainable judgments. LangChain agents will orchestrate the workflow, managing prompt structuring, retrieval calls, and function chaining. The platform will provide users with a classification (fraudulent vs. legitimate) and an explanation highlighting the suspicious patterns detected.

**Core Features**

* Integration of LLMs as the central fraud detection engine.
* RAG pipeline for grounding LLM outputs in real-world fraud cases.
* Web scraping from Reddit, HK01, Yahoo, SCMP, and HKMA to build the fraud dataset.
* ChromaDB for vector storage and semantic retrieval.
* LangChain orchestration for prompt management and structured logic.
* User interface for submitting suspicious text and receiving detection results.

**Algorithms**

The system will adopt a layered approach to fraud detection:

* **NER + RAG + LLM (Primary)**: Named Entity Recognition will identify fraud-related entities (e.g., impersonation cues, financial terms). RAG will ground the LLM’s analysis in verified fraud case data.
* **Rule-Based Scoring (Secondary)**: Detected entities will be cross-referenced with a structured fraud case database to assign risk levels.
* **LangChain Agent (Orchestration Layer)**: Manages prompt flow, retrieval, and contextual memory to ensure structured and transparent reasoning.
* **Benchmarking Framework**: Performance will be evaluated using accuracy, precision, recall, and F1-score to ensure reliability.

This hybrid algorithmic structure ensures both feasibility and extensibility, combining the strengths of symbolic and statistical approaches. The integration of RAG with LLMs allows the system to ground its reasoning in real-world fraud cases, reducing hallucinations and improving factual accuracy. The rule-based scoring layer provides additional interpretability, offering structured criteria that complement the LLM’s contextual analysis. LangChain agents serve as the orchestration layer, ensuring that the workflow remains modular, traceable, and adaptable to future enhancements. This design not only addresses the limitations of traditional rule-based systems, which often fail to adapt to evolving scam techniques, but also leverages the semantic and contextual capabilities of LLMs to detect subtle indicators of deception. By training and grounding the model on authentic fraud cases, the system becomes more resilient to emerging scam strategies, while maintaining transparency and user relevance.

**Technologies**

The proposed fraud detection platform will be developed using a modern, modular technology stack that supports scalability, interoperability, and real-time performance. Each component has been carefully selected to align with the project’s objectives of accuracy, adaptability, and ethical responsibility.

**Frontend:**

The user interface will be implemented using **React**, which provides a responsive and interactive web-based environment. This ensures accessibility across devices and allows users to easily input suspicious text and view detection results.

**Backend:**

The backend will be developed with **Node.js**, chosen for its efficiency in handling asynchronous requests and real-time interactions. **MongoDB** will be used for storing user metadata and system logs, while fraud case embeddings will be managed separately in a vector database.

**AI Pipeline:**

At the core of the system is the **RAG + LLM pipeline**, orchestrated through **LangChain**. The LLM provides contextual reasoning, while RAG ensures that outputs are grounded in real-world fraud cases. **Embedding models** such as Sentence-BERT or OpenAI embeddings will be used to vectorize fraud cases, enabling semantic search and retrieval. LangChain agents will manage prompt structuring, retrieval calls, and function chaining, ensuring modularity and traceability.

**Vector Database:**

**ChromaDB** will serve as the vector database, storing embeddings of fraud cases scraped from sources such as Reddit, HK01, Yahoo, SCMP, and HKMA. Its efficient similarity search capabilities make it well-suited for retrieval tasks in the RAG pipeline.

**Speech and Text Processing (Optional):**

If speech input is required, **Whisper** will be used for multilingual speech-to-text transcription, ensuring inclusivity for users across different languages.

**Deployment and CI/CD:**

The system will be containerized using **Docker**, ensuring portability and consistent deployment across environments. **GitHub Actions** will be employed for automated testing and continuous integration, maintaining system reliability throughout development.

**Web Scraping Tools:**

Custom scraping pipelines will be developed to extract fraud cases from online forums, news outlets, and regulatory advisories. These tools ensure that the knowledge base remains dynamic and reflective of evolving fraud tactics.

This technology stack was selected to balance **scalability, performance, and adaptability**. React and Node.js provide a robust foundation for a responsive web platform, while MongoDB and ChromaDB ensure efficient data management. The integration of RAG with LLMs, orchestrated by LangChain, directly addresses the limitations of traditional rule-based fraud detection by grounding outputs in authentic fraud cases. Whisper and multilingual support expand accessibility, while Docker and GitHub Actions streamline deployment and testing. Together, these technologies form a cohesive ecosystem that supports the project’s goal of delivering a reliable, ethical, and user-friendly fraud detection platform.

# **Explanation of why proposed solution is appropriate**

This solution addresses the limitations of traditional fraud detection systems by combining the contextual reasoning capabilities of LLMs with the factual grounding of RAG. By retrieving real-world fraud cases during analysis, the system ensures that outputs are based on verified examples rather than generic or fabricated content. The use of LangChain provides modularity and scalability, while web scraping ensures that the knowledge base remains up-to-date with evolving fraud tactics. The platform’s web-based deployment and multilingual support make it accessible to a wide range of users. Importantly, the system is designed with strict ethical safeguards: it will focus exclusively on fraud detection and will not generate or simulate fraudulent content.

# **Main development phases and Main Deliverables (Main Stages)**

|  |  |  |  |
| --- | --- | --- | --- |
| Phase | Timeline | Task | Assessment Element |
| 1 | Sep 2025 | Define scope, assign roles, finalize RAG + LLM + LangChain stack. *Deliverable: Proposal.* | Proposal (Due: 19 Sep 2025) |
| 2 | Oct–Nov 2025 | Conduct literature review, collect fraud datasets, design user interface. *Deliverable: Initial Report.* | Initial Report (Due: 28 Nov 2025) |
| 3 | Dec 2025 – Jan 2026 | Build RAG pipeline, integrate LLM and LangChain agents. *Deliverable: Interim Report.* | Interim Report (Due: 9 Feb 2026) |
| 4 | Jan–Feb 2026 | Deliver working prototype of fraud detection platform. *Deliverable: Interim Prototype & Demo.* | Interim Prototype & Demo (Due: 9–13 Feb 2026) |
| 5 | Mar–Apr 2026 | Conduct user testing and refine system. *Deliverable: Final Report.* | Final Report (Due: 18 May 2026) |
| 6 | May 2026 | Present fully functional platform with benchmark results. *Deliverable: Final Prototype & Demo.* | Final Prototype & Demo (Due: 18–22 May 2026) |
| Ongoing Performance & Teamwork | Throughout | Continuous team collaboration and performance tracking | Student Performance 10% |

# **Main deliverables**

* Initial, Interim, and Final Reports
* Executable Web-Based Platform
* Source Code with Documentation
* REST API and Backend Database
* Demo Presentation and User Testing Summary
* Fraud Data Analysis with Source References

# **The responsibilities of each member**

* **TAN James Anthroi A. (Team Leader & Backend Developer):** Oversees project coordination, backend API development, and secure data flow.
* **Mobile Developer:** Builds the React Native app, implements the chat interface, OCR integration, and ensures smooth user experience across Android/iOS.
* **Machine Learning Engineer:** Develops the RAG+LLM pipeline for allergen detection, trains CNN models for local cuisine recognition, and integrates AI into the app.
* **TAN Xiuhao (Backend Developer):** **(Backend Developer):** Designs server-side architecture, manages vector database, and supports real-time interaction.
* **LIN Yueying (QA & Testing):** Creates test plans, conducts user trials, documents bugs, and ensures system reliability.
* **Documentation & Reporting Specialist:** Prepares all written deliverables, maintains records, and supports presentations.

# **Executive Summary**

This project proposes a **fraud detection platform** that leverages **Large Language Models (LLMs)**, **Retrieval-Augmented Generation (RAG)**, and **LangChain agents** to identify fraudulent communication with greater accuracy. Fraud cases will be collected through web scraping from diverse sources, including Reddit, HK01, Yahoo, SCMP, and HKMA, and stored in ChromaDB for semantic retrieval. When users submit suspicious text, the system will retrieve relevant fraud cases to ground the LLM’s analysis, producing reliable classifications and explanations. The platform emphasizes scalability, multilingual support, and ethical safeguards, making it a practical and impactful tool for fraud prevention.