**HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**SCHOOL OF INFORMATION & COMMUNICATION TECHNOLOGY**

----- 🙡 🕮 🙣 -----

**A red and yellow logo

Description automatically generated**

**GROUP 9 REPORT**

***Subject:* Chatbot product consulting for e-commerce sites**

|  |  |
| --- | --- |
| Lecturer: | **Ph.D. Trinh Tuan Dat** |
| Class: | 156825 |

|  |  |  |
| --- | --- | --- |
| Student Name | Student ID | Email |
| Vu Minh Dung | 20205179 | [dung.vm205179@sis.hust.edu.vn](mailto:dung.vm205179@sis.hust.edu.vn) |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Ha Noi, 06/2024**

Contents

[**I.** **INTRODUCTION** 4](#_Toc179881830)

[**1.** **Problem description** 4](#_Toc179881831)

[**1.1. Sorting Algorithms** 4](#_Toc179881832)

[**1.1.1. Quick Sort** 4](#_Toc179881833)

[**1.1.2. Selection Sort** 5](#_Toc179881834)

[**1.1.3. Insertion Sort** 6](#_Toc179881835)

[**1.1.4. Merge Sort** 7](#_Toc179881836)

[**1.1.5. Shell Sort** 8](#_Toc179881837)

[**II.** **Diagram** 9](#_Toc179881838)

[**1.** **Use case diagram** 9](#_Toc179881839)

[**2.** **General class diagram** 10](#_Toc179881840)

[**2.1.** **Model** 11](#_Toc179881841)

[**2.2.** **View** 12](#_Toc179881842)

[**2.3.** **Controller** 12](#_Toc179881843)

[**2.4.** **Utility** 13](#_Toc179881844)

[**III.** **Object-Oriented Programming Concepts** 14](#_Toc179881845)

[**1.** **Encapsulation** 14](#_Toc179881846)

[**2.** **Inheritance** 14](#_Toc179881847)

[**3.** **Polymorphism** 15](#_Toc179881848)

[**4.** **Abstraction** 15](#_Toc179881849)

[**5.** **Evaluation** 16](#_Toc179881850)

[**IV.** **Demo & Result** 17](#_Toc179881851)

**ACKNOWLEDGMENTS**

We would like to express our sincere gratitude to our dedicated group members for their unwavering support and collaborative efforts throughout this project. Their commitment and teamwork have been invaluable. We extend our heartfelt appreciation to **Ph.D. Trinh Tuan Dat** for providing guidance, expertise, and encouragement at every stage of our journey. Their insights have significantly enriched our learning experience and contributed to the success of this endeavor. Thank you to both our group members and **Ph.D. Trinh Tuan Dat** for their outstanding contributions to realization of this project.

**ABSTRACT**

In recent years, the increasing volume and complexity of product information on e-commerce platforms have made it difficult for users to efficiently find suitable products based on their specific needs. This project proposes a conversational AI system that leverages Retrieval-Augmented Generation (RAG) combined with Large Language Models (LLMs) to act as a product consultant for smartphones on e-commerce websites. The system can understand natural language queries, retrieving relevant product information from a structured or unstructured database, and generating informative, contextually accurate responses.

The solution is built on a modular architecture, including a data preprocessing pipeline, an embedding-based retrieval mechanism, and an LLM-based response generation module. The system is evaluated based on its relevance, factual accuracy, and fluency in responding to a variety of user queries such as product comparisons, feature explanations, and personalized recommendations. This work demonstrates the potential of RAG-based systems in enhancing user experience on digital commerce platforms and provides a scalable foundation for intelligent product advisors in vertical domains.

# **CHAPTER 1. INTRODUCTION**

* 1. **Motivation**

Trong những năm gần đây, sự phát triển bùng nổ của các nền tảng thương mại điện tử đã dẫn đến sự xuất hiện tràn lan của các sản phẩm điện thoại thông minh, mỗi sản phẩm đều có nhiều tùy chọn về thông số kỹ thuật, tính năng và mức giá khác nhau. Tuy nhiên, phần lớn người dùng cuối thường thiếu kiến thức kỹ thuật hoặc không có thời gian để so sánh và đánh giá các sản phẩm này một cách hiệu quả. Các hệ thống tìm kiếm và lọc dựa trên từ khóa truyền thống không thể nắm bắt được ý định của người dùng khi truy vấn trở nên phức tạp hoặc mang tính hội thoại.

Ví dụ, một người dùng có thể hỏi: “Điện thoại nào dưới 10 triệu có pin và camera tốt?”—một câu hỏi đòi hỏi phải hiểu được ý định, lọc dữ liệu dựa trên nhiều tiêu chí, và đưa ra một khuyến nghị rõ ràng. Các công cụ tìm kiếm hiện tại hoặc các chatbot dựa trên luật không thể xử lý một cách đáng tin cậy những tương tác bằng ngôn ngữ tự nhiên như vậy. Do đó, cần thiết phải phát triển một trợ lý thông minh có khả năng hiểu các truy vấn phức tạp, truy xuất thông tin sản phẩm liên quan, và tạo ra phản hồi chính xác, dễ hiểu cho người dùng. Một hệ thống như vậy sẽ nâng cao trải nghiệm mua sắm, giảm rào cản trong quá trình ra quyết định, và mang lại lợi thế cạnh tranh cho các nền tảng thương mại điện tử..

* 1. **Objectives and scope of the project**

Mục tiêu chính của dự án này là phát triển một hệ thống AI hội thoại đóng vai trò như trợ lý mua sắm điện thoại thông minh cho các trang thương mại điện tử. Trợ lý này được thiết kế để xử lý các truy vấn bằng ngôn ngữ tự nhiên từ người dùng và tạo ra các phản hồi có ngữ cảnh, chính xác về mặt dữ kiện bằng cách kết hợp giữa các Mô hình Ngôn ngữ Lớn (LLM) và kỹ thuật Tạo Sinh Tăng Cường Truy Xuất (Retrieval-Augmented Generation – RAG).

Trong khi một số nền tảng hiện nay đã tích hợp AI vào hệ thống gợi ý sản phẩm, các hệ thống này thường dựa vào lọc cộng tác, các quy tắc mã hóa cứng, hoặc các mẫu hỏi đáp tĩnh. Những cách tiếp cận như vậy có phạm vi hạn chế, không thể khái quát hóa với các truy vấn mới lạ và không đáp ứng được yêu cầu phản hồi chính xác theo thời gian thực dựa trên dữ liệu sản phẩm cập nhật. Ngoài ra, các hệ thống chỉ dựa vào LLM thường gặp vấn đề "ảo giác", tạo ra thông tin nghe có vẻ hợp lý nhưng sai lệch.

Để giải quyết những thách thức này, dự án đề xuất một giải pháp lai: kết hợp khả năng suy luận của LLM với khả năng truy xuất tài liệu thời gian thực từ một cơ sở tri thức sản phẩm có thể là có cấu trúc hoặc không có cấu trúc. Hệ thống nhắm đến ba năng lực cốt lõi: (1) hiểu ý định của người dùng trong ngôn ngữ tự nhiên, (2) truy xuất ngữ nghĩa các thông tin sản phẩm liên quan, và (3) tạo ra phản hồi mạch lạc, chính xác và hữu ích dựa trên dữ liệu đã truy xuất.

* 1. **Tentative solution**

Giải pháp đề xuất áp dụng kiến trúc Tạo sinh tăng cường truy xuất (Retrieval-Augmented Generation – RAG) được tích hợp với Mô hình Ngôn ngữ Lớn (LLM). Hệ thống bao gồm ba thành phần chính:

* **Mô-đun tiền xử lý dữ liệu**: chuyển đổi các thông số kỹ thuật, mô tả sản phẩm và các câu hỏi thường gặp (FAQs) thành các vector đặc trưng dạng đậm (dense vector embeddings) bằng cách sử dụng các bộ mã hóa ở cấp độ câu, sau đó lưu trữ vào một cơ sở dữ liệu vector.
* **Bộ truy xuất ngữ nghĩa**: sử dụng độ tương đồng cosine hoặc tìm kiếm lân cận gần nhất (nearest-neighbor search) để trích xuất nội dung sản phẩm liên quan nhất dựa trên truy vấn của người dùng.
* **Bộ tạo phản hồi**: được vận hành bởi một mô hình LLM đã được huấn luyện trước, sử dụng prompt có cấu trúc kết hợp giữa truy vấn gốc và ngữ cảnh đã truy xuất để tạo ra các phản hồi trôi chảy, chính xác và có căn cứ thực tế.

Dự án này đóng góp một pipeline triển khai hoàn chỉnh cho hệ thống tư vấn sản phẩm dựa trên LLM, bao gồm các mô-đun xử lý dữ liệu, truy xuất và tạo sinh. Hệ thống được thiết kế theo hướng mô-đun và có khả năng mở rộng, dễ dàng thích ứng với các danh mục sản phẩm khác ngoài điện thoại thông minh. Các kết quả thực nghiệm sẽ đánh giá độ chính xác, tính tự nhiên trong ngôn ngữ, và mức độ hài lòng của người dùng trong nhiều kịch bản truy vấn liên quan đến sản phẩm.

# **Chapter 2.**