**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**

In recent years, the exponential growth of e-commerce platforms has resulted in an overwhelming number of smartphone products, each offering a variety of specifications, features, and pricing options. However, most end-users often lack technical knowledge or time to manually compare and evaluate these products effectively. Traditional keyword-based search and filter systems fail to capture user intent when queries become nuanced or conversational in nature.

For example, a user might ask, “Which phone under 10 million VND has good battery life and camera?”—a question that requires understanding intent, filtering based on multiple criteria, and generating a clear recommendation. Current search engines or rule-based bots are not capable of handling such natural language interactions reliably. Therefore, there is an urgent need for an intelligent assistant that can understand complex queries, retrieve relevant product information, and generate accurate, user-friendly responses. Such a system would enhance the shopping experience, reduce friction in decision-making, and bring competitive advantages to e-commerce platforms.

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

The primary objective of this project is to develop a conversational AI system that serves as a smartphone shopping assistant for e-commerce websites. The assistant is designed to process natural language queries from users and generate context-aware, factually grounded responses using a combination of Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) techniques.

While some existing platforms have integrated AI into their recommendation engines, these systems often rely on collaborative filtering, hardcoded rules, or static Q&A templates. Such approaches are limited in scope, fail to generalize to unseen queries, and cannot provide real-time, accurate responses based on updated product data. Additionally, pure LLM-based systems suffer from hallucination issues, producing plausible-sounding but incorrect information.

To address these challenges, this project proposes a hybrid solution: combining the reasoning power of LLMs with real-time document retrieval from a structured or unstructured product knowledge base. The system targets three core capabilities: (1) understanding user intent in natural language, (2) semantically retrieving relevant product information, and (3) generating coherent, accurate, and informative responses grounded in retrieved facts.

* 1. **Tentative solution**

The proposed solution adopts a Retrieval-Augmented Generation (RAG) architecture integrated with a Large Language Model (LLM). The system consists of three main components:

* A data preprocessing module that converts product specifications, descriptions, and FAQs into dense vector embeddings using sentence-level encoders and stores them in a vector database
* A semantic retriever that uses cosine similarity or nearest-neighbor search to extract the most relevant product content based on the user’s query;
* A response generator powered by a pre-trained LLM, which uses a structured prompt combining the original query and retrieved context to generate fluent and factually grounded answers.

This project contributes a complete implementation pipeline for an LLM-based product advisor system, including data ingestion, retrieval, and generation modules. The system is built to be scalable and modular, enabling easy adaptation to other product categories beyond smartphones. Experimental results will evaluate the system’s accuracy, fluency, and user satisfaction in a variety of product-related query scenarios.