

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT



ĐỒ ÁN CHUYÊN NGÀNH
KHOA HỌC MÁY TÍNH

TÊN ĐỀ TÀI:

**XÂY DỰNG NỀN TẢNG OMNIRAG: GIẢI PHÁP ĐA NỀN
TẢNG HỖ TRỢ TỰ KHỞI TẠO TRỢ LÝ ẢO THÔNG MINH
ĐA KÊNH TỪ DỮ LIỆU TÙY CHỈNH CHO CÁ NHÂN VÀ TỔ
CHỨC**

Hà Nội - Năm 2026

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT

ĐỒ ÁN TỐT NGHIỆP

CHUYÊN NGÀNH: KHOA HỌC MÁY TÍNH

GIÁO VIÊN HƯỚNG DẪN:

NGUYỄN THÙY DƯƠNG

BỘ MÔN KHOA HỌC MÁY TÍNH

SINH VIÊN THỰC HIỆN:

NGUYỄN QUỐC ANH

LỚP KHMT ỨNG DỤNG 07B

Hà Nội – Năm 2026

MỤC LỤC

LỜI CẢM ƠN
DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT
DANH MỤC BẢNG BIỂU
DANH MỤC HÌNH ẢNH
MỞ ĐẦU
1. Lý do chọn đề tài.....
2. Mục tiêu của đề tài.....
3. Đối tượng và phạm vi nghiên cứu.....
4. Phương pháp nghiên cứu.....
5. Bố cục đề tài.....
CHƯƠNG 1. TỔNG QUAN VÀ KHẢO SÁT HỆ THỐNG
1.1. Giới thiệu bài toán trợ lý ảo thông minh cho cá nhân và tổ chức.....
1.2. Nhu cầu giao tiếp đa kênh (Omnichannel) trong doanh nghiệp hiện đại.....
1.3. Thách thức trong triển khai trợ lý ảo đa nền tảng và đa nguồn dữ liệu.....
1.4. Khảo sát các giải pháp hiện nay.....
1.4.1. Các nền tảng chatbot đóng gói sẵn (No-code/Low-code Platforms).....
1.4.2. Các hạn chế về tính riêng tư và tùy biến dữ liệu.....
1.4.3. Đánh giá khả năng tích hợp đa kênh của các giải pháp hiện tại.....
1.5. Phân tích yêu cầu hệ thống.....
1.5.1. Yêu cầu chức năng: Quản lý Tenant, Bot Builder, Multi-channel Chat....
1.5.2. Yêu cầu phi chức năng: Bảo mật, Hiệu năng, Khả năng mở rộng.....
1.5.3. Yêu cầu về khả năng tích hợp bên thứ ba (Third-party Integration).....

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT VỀ AI VÀ RAG NÂNG CAO.....

- 2.1. Tổng quan về Large Language Models (LLM) và Generative AI.....
- 2.2. Kiến trúc Retrieval-Augmented Generation (RAG).....
 - 2.2.1. Quy trình RAG cơ bản và các điểm nghẽn (Bottlenecks).....
 - 2.2.2. Phân loại RAG: Naive RAG vs Advanced RAG vs Agentic RAG.....
- 2.3. Các kỹ thuật tối ưu hóa truy xuất nâng cao.....
 - 2.3.1. Kỹ thuật Hybrid Search (Dense & Sparse Retrieval).....
 - 2.3.2. Query Transformation: HyDE và Multi-Query Generation.....
 - 2.3.3. Document Re-ranking và áp dụng Cross-Encoders.....
 - 2.3.4. Các chiến lược Chunking tối ưu: Recursive và Semantic Splitting.....
- 2.4. Khái niệm AI Agents và Agentic Workflow.....
 - 2.4.1. Cơ chế Reasoning và Acting (ReAct Pattern).....
 - 2.4.2. Quản lý trạng thái và luồng hội thoại phức tạp.....
 - 2.4.3. Multi-Agent Collaboration và Task Delegation.....

CHƯƠNG 3. CÁC CÔNG NGHỆ VÀ FRAMEWORK TRỌNG TÂM.....

- 3.1. Ecosystem LangChain cho phát triển AI Applications.....
 - 3.1.1. LangChain: Xây dựng chuỗi (Chains) và tích hợp Tools.....
 - 3.1.2. LangGraph: Điều phối Agents phức tạp với Cyclic Graph.....
 - 3.1.3. LangFuse: Tracing, Observability và Evaluation chất lượng RAG.....
- 3.2. Cơ sở dữ liệu và hạ tầng lưu trữ.....
 - 3.2.1. Vector Database Qdrant và HNSW Indexing.....
 - 3.2.2. SQL (PostgreSQL) và NoSQL (MongoDB) trong hệ thống Hybrid.....
 - 3.2.3. Redis: Caching và Session Management.....

3.2.4. MinIO: Object Storage tương thích S3.....
3.3. Phát triển đa nền tảng và quản trị hệ thống.....
3.3.1. Backend FastAPI: Bất đồng bộ và hiệu năng cao.....
3.3.2. Hạ tầng Docker Microservices và Celery Task Queue.....
3.3.3. Thiết kế Responsive Web với React và Tailwind CSS.....
3.4. Công nghệ tích hợp đa kênh (Omnichannel Integration).....
3.4.1. func.vn Gateway: Adapter cho Zalo và Facebook Messenger.....
3.4.2. Telegram Bot API: Webhook và Long Polling.....
3.4.3. Phát triển Web Widget nhúng thông minh (Embeddable Chat Widget). ...
3.4.4. API Gateway Pattern cho đa nền tảng.....

CHƯƠNG 4. PHÂN TÍCH VÀ THIẾT KẾ NỀN TẢNG OMNIRAG.....

4.1. Phân tích chức năng và luồng nghiệp vụ.....
4.1.1. Quy trình Self-build bot từ PDF, Website và API.....
4.1.2. Cơ chế phân tách dữ liệu đa khách hàng (Isolated Multi-tenancy).....
4.1.3. Luồng xử lý tin nhắn đa kênh (Omnichannel Message Flow).....
4.2. Thiết kế các biểu đồ hệ thống.....
4.2.1. Biểu đồ Use Case và Sequence Diagrams cho RAG Agent.....
4.2.2. Biểu đồ hoạt động (Activity Diagram) của pipeline xử lý dữ liệu.....
4.2.3. Sơ đồ luồng tích hợp đa kênh qua func.vn.....
4.3. Thiết kế kiến trúc phần mềm và cơ sở dữ liệu.....
4.3.1. Thiết kế kiến trúc Cross-platform với API-first approach.....
4.3.2. Thiết kế ERD cho PostgreSQL (Users, Tenants, Bots, Documents).....
4.3.3. Thiết kế cấu trúc Vector Collection trong Qdrant.....

Đồ án chuyên ngành Khoa học máy tính

4.3.4. Thiết kế lớp Connector giao tiếp đa kênh (Omnichannel Connector).....
4.4. Thiết kế giao diện (UI/UX) Dashboard và Chat Playground.....
4.4.1. Dashboard quản trị Tenant và Bot Builder.....
4.4.2. Giao diện cấu hình kênh tích hợp (Channel Configuration).....
CHƯƠNG 5. CÀI ĐẶT, THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ.....
5.1. Triển khai và cài đặt hệ thống.....
5.1.1. Triển khai hạ tầng Microservices trên Docker.....
5.1.2. Cấu hình LangGraph Agent workflow xử lý hội thoại.....
5.1.3. Tích hợp func.vn cho Zalo OA và Facebook Messenger.....
5.1.4. Cài đặt Telegram Bot và Web Widget.....
5.2. Kết quả thực nghiệm và Demo.....
5.2.1. Demo giao diện Web Dashboard và Bot Chat.....
5.2.2. Thủ nghiệm trợ lý ảo trên Zalo OA với dữ liệu tùy chỉnh.....
5.2.3. Thủ nghiệm Web Widget nhúng trên đa nền tảng Website.....
5.2.4. Thủ nghiệm khả năng tự học từ đa định dạng dữ liệu.....
5.3. Đánh giá và kiểm thử chất lượng.....
5.3.1. Đánh giá tốc độ phản hồi (Latency) với Redis Cache.....
5.3.2. Đánh giá độ chính xác (RAG Metrics) thông qua LangFuse.....
5.3.3. Đánh giá tính nhất quán đa kênh (Cross-channel Consistency).....
5.4. Thảo luận và phân tích kết quả.....
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....
DANH MỤC TÀI LIỆU THAM KHẢO.....
PHỤ LỤC.....

Đồ án chuyên ngành Khoa học máy tính

- Phụ lục A: Danh mục API hỗ trợ đa nền tảng và Third-party integration.....
- Phụ lục B: Quy trình thiết lập hệ thống Monitoring với LangFuse.....
- Phụ lục C: Các kịch bản kiểm thử (Test Cases) và kết quả chi tiết.....
- Phụ lục D: Hướng dẫn cấu hình func.vn cho Zalo OA và Facebook.....
- Phụ lục E: Mã nguồn mẫu tích hợp Web Widget và Telegram Bot.....