A picture containing rectangle, screenshot, text, frame

Description automatically generated

**Môn học: Thực tập doanh nghiệp – IS502.P21**

BÁO CÁO THỰC TẬP

ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

KHOA HỆ THỐNG THÔNG TIN

**Giảng viên hướng dẫn**: ThS. Hà Lê Hoài Trung

**Công ty thực tập**: Hitachi Digital Services

**Họ và tên sinh viên:** Nguyễn Hoàng Đăng Khoa

**MSSV:** 21520999

TP. HỒ CHÍ MINH, NĂM 2025

**LỜI CẢM ƠN**

Danh mục

[Chương 1: Giới thiệu 1](#_Toc194306451)

[I. Giới thiệu công ty 1](#_Toc194306452)

[II. Dự án, trách nhiệm công việc 1](#_Toc194306454)

[a. Vị trí thực tập 1](#_Toc194306455)

[b. Lịch làm việc 1](#_Toc194306456)

[Chương 2: Quá trình thực tập 2](#_Toc194306457)

[I. Mốc thời gian giao việc và kết quả 2](#_Toc194306458)

[Chương 3: Cơ sở lý thuyết 3](#_Toc194306459)

[I. Công nghệ sử dụng 3](#_Toc194306460)

[Chương 4: Thông tin chi tiết công việc 4](#_Toc194306461)

[I. Thông tin dự án 4](#_Toc194306462)

[a. Khái quát dự án 4](#_Toc194306463)

[b. Vị trí đảm nhiệm 4](#_Toc194306464)

[II. Công việc thực hiện 4](#_Toc194306465)

[Chương 5: Tổng kết 5](#_Toc194306466)

[I. Tổng kết quá trình thực tập: 5](#_Toc194306467)

[Tài liệu tham khảo 6](#_Toc194306468)

# Chương 1: Giới thiệu

## Giới thiệu công ty

Hitachi Digital Services Vietnam (HitachiDSVN) là một trong những trung tâm giao hàng toàn cầu hàng đầu của Hitachi Digital Services, tập trung vào việc cung cấp các giải pháp chuyển đổi số toàn diện cho doanh nghiệp trên toàn thế giới. Với hơn 1.000 chuyên gia làm việc tại ba địa điểm chính ở Việt Nam, công ty đóng vai trò quan trọng trong việc hỗ trợ các dự án công nghệ quy mô lớn và phức tạp.

HitachiDSVN cung cấp đa dạng dịch vụ và giải pháp công nghệ, bao gồm:

* **Dịch vụ đám mây và dữ liệu**: Hỗ trợ doanh nghiệp trong việc di chuyển, quản lý và tối ưu hóa hạ tầng đám mây và dữ liệu.
* **Dịch vụ IoT**: Triển khai các giải pháp Internet vạn vật để nâng cao hiệu quả hoạt động và kết nối thiết bị.
* **Ứng dụng doanh nghiệp số**: Phát triển và tích hợp các ứng dụng nhằm cải thiện quy trình kinh doanh.
* **Dịch vụ quản lý**: Cung cấp dịch vụ quản lý hệ thống CNTT toàn diện, giúp doanh nghiệp tập trung vào hoạt động cốt lõi.
* **Dịch vụ nhúng**: Phát triển phần mềm nhúng cho các thiết bị và hệ thống chuyên dụng.
* **Trung tâm độ tin cậy ứng dụng (HARC)**: Đảm bảo độ tin cậy và hiệu suất của các ứng dụng doanh nghiệp.



## Dự án, trách nhiệm công việc

### Vị trí thực tập

Vị trí thực tập: Data Intern.

### Lịch làm việc

Lịch làm việc: Thứ 2, thứ 4, thứ 5, thứ 6.

# Chương 2: Quá trình thực tập

## Mốc thời gian giao việc và kết quả

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Timeline | Assigned task(s) | Achieved result(s) | Experience | Further research |
| 15/04/2025 | **Quá trình Onboarding** | | | |
|  | Set up máy tính |  |  |  |
| 16/04/2025 | **Generative Ai tự học** | | | |
|  | * Cài flowise. * Tạo flowchat trong flowise. * Viết test case CV generation cho HR (input và expected output) | * Hoàn thành viết test case. * Flowise đã cài đặt thành công globally nhưng khi bấm icon chat để thử thì bị lỗi UI. |  |  |
| 17/04/2025 |  |  |  |  |
|  | * Sửa lại flowise. * Tạo flowchat trong flowise. * Cài đặt Ollama và test thử API. | * Sử dụng được Flowise local host, không còn bị lỗi UI. * Olama cài thành công và API test thành công. | * Phải downgrade node js từ v22 xuống v20 để install flowise thành công. * LLM OpenAI không sử dụng được trong flowchat do policy network công ty, Ollama thì dùng dược. * Parameter Temperature từ 0 đến 1 dùng để biểu thị cho tính creative của model (Ví dụ 0 thì sẽ very factual còn 1 thì sẽ creative) | * Tìm cách sử dụn OPEN AI LLM Api key |
| 18/04/2025 |  |  |  |  |
|  | * Summarize các loại Prompt. * Tìm cách sử dụng OPEN AI LLM nếu được. | * Hoàn thành summarize các loại prompt. * Không sử dụng được OPEN AI LLM, phải sử dụng Ollama. |  |  |
| 21/04/2025 |  |  |  |  |
|  | * Tìm cách làm 1 chat bot summarize document. | * Thành công tạo được chat bot summarize document. * Tuy nhiên summarize document size nhỏ (vd: CV, PDF) hay các file chứa nhiều hình ảnh trả về kết quả không chính xác bị hallucination cao, còn document size lớn hơn (vs: report 50 trang) thì summarize được. | * Chat flow cần vector store để chưa chứa document được embedding thành vector. | * Tìm cách sử dụng các vector store. * Tìm cách làm chat bot có long term history. |
| 22/04/2025 | **Cài đặt, kiểm thử và báo cáo AI Coder Hỗ trợ TabbyML trên môi trường Vscode.** | | | |
|  | * Thử install tabby trên VS Code để test các feature như chat vs nó, code completion. * Tham gia orientation training của công ty P1. * Cải tiến chat bot có thể summarize được file pdf lớn hơn và implement thêm chat history cho chat bot. | * Các Feature của Tabby hoạt động thành công trên VSCode nhưng quá trình chạy lâu. * Nắm được cấp bậc trong cty, các CoE, các bộ phận trong cty, các quy trình workflow. * Chat Bot đã có thể summarize file report lớn. Chat history đã hoạt động, tuy nhiên chỉ lưu trong memory nên khi khởi động lại flowise thì mất hết. | * Sử dụng Pinecone để làm để làm vector database cho chat flow. Pinecone hỗ trợ tải lên file trong cửa sổ chat. * Buffer Memory có thể giúp lưu lại chat history trong 1 thơi gian trước khi flowise được tắt đi và khởi động lại. | * Có gắng tải lên được file lên vector database. * Thử chat memory khác cho chat flow để có thể lưu được lâu dài. |
| 23/04/2025 |  |  |  |  |
|  | * Tham gia orientation trainging của công ty P2. * Cải tiến chat bot có thể upload file lên vector database và cải thiện chat thông tin trí nhớ. | * Nắm được các chính sách của công ty về bảo hiểm, quyền lợi như gửi xe, ngày nghỉ, etc... * Đã có thể upload lên cloud của pinecone tuy nhiên chưa kéo được dữ liệu về. Container Redis database chạy trên local có thể lưu được lịch sử chat tưu nhiên nếu như đẩy file pdf lớn thì Redis sẽ lưu tất cả vào 1 index gây ra vấn đề overflow, dẫn đến vấn dề Hallucination hay model không nhớ. |  | * Cố gắng cải thiện chat flow để có rag được dữ liệu. * Thử sử dụng cách summarize file trước khi bỏ vào redis chat memory. |
| 24/04/2025 | **Cài đặt và viết hướng dẫn cài đặt TabbyML trên môi trường IntelliJ** | | | |
|  | * Thử install tabby trên Intellij, chạy tabby trên local, và viết báo cáo hướng dẫn install tabby trên Intellij. * Cải tiến chat bot để thể rag dữ liệu từ vector database. | * Sử dụng Tabby thành công trên Intellij và tốc độ respond vẫn chậm như sd trên VS Code. Báo cáo hương dẫn và Demo của Tabby được chấp nhận bởi thành viên IT hỗ trợ. * Thành công cải tiến chat bot mang dữ liệu từ redis stack database chạy trên local. |  | * Cố gắng nâng cấp dữ bộ nhớ để có thể sumamrize file trước khi bỏ vào redis chat memory. |
| 28/04/2025 |  |  |  |  |
|  | * Đưa ra 1 proposal về 1 vấn đề của internal project đề ra: khi có lượng lớn training video, có thể sử dụng GenAI để có thể tìm kiếm video và timestamp. | * Đưa ra được proposal về việc transcribe được phụ đề có timestamps kèm theo thông tin qua sử dụng mô hình dữ liệu lớn. * Đưa ra được kiến trúc hệ thống để giải quết vấn đến. |  |  |
| 29/04/2025 |  |  |  |  |
|  | * Dùng bất cứ tool hay công nghệ nào có thể extract được content từ 1 audio | * Sử dụng ffmpeg để có thể triết xuất âm thanh từ video. * Sử dụng và so sánh giữa hai mô hình PhoWhisper và OpenWhisper để để có thể triết xuất file âm thanh thành dạng chữ bằng ngôn ngữ Python. | * Cần phải có video rõ ràng thì mới extract được transcript (không thể triết xuất được transcript từ video có âm thanh không rõ ràng). | * Lựa chọn vector database để có thể embedding dữ liệu từ hai model |
| 30/4 – 1/5 | **Nghỉ lễ 30/4 – 1/5** | | | |
| 5/5/2025 | **Tham gia project FlexBA để hỗ trợ quản lý TabbyML** | | | |
|  | * Tham gia project FlexBA để hỗ trợ quản lý TabbyML server. |  | * Sử dụng SSH để liên kết tới server host TabbyML. | * TabbyML có bug dẫn đến database bị khóa. Cần phải tìm lý do dẫn đến bug (root cause) |
| 6/5/2025 | **Tham gia project Revoria Cloud Production GenAI** | | | |
|  | * Sử dụng 3 cách tiếp cận Similarity Search: Fuzzywuzzy + Levenshtein, Nano-vectorDB, PandasAI với ngôn ngữ Python và đánh giá kết quả đầu ra. | * PandasAI không sử dụng để similar search được. * Chưa embed dữ liệu dc data cho Nano-vectorDB. | * Khó khăn trong việc embedding khi dữ liệu đầu vào là tiếng Nhật (input\_search\_DB.csv) | * Tìm kiếm phương án embed dữ liệu, đặc biệt khi dữ liệu đầu vào là chữ Nhật. |
| 7/5/2025 |  |  |  |  |
|  | * Tìm phương án embed data và đẩy dữ liệu lên Nano-vectorDB. | * Embed dữ liệu thành công bằng cách sử dụng model ollama nomic-embed-text và multilingual-e4-small. * Upsert lên Nano-vector và query thành công vs thời gian lần lượt là 0.0015s và 0.0004s cho in 100 dòng. |  |  |
| 8/5/2025 |  |  |  |  |
|  | * Đánh giá thời gian đẩy và truy vấn data trên Nano-vectorDB với dữ liệu ~1 triệu dòng. | * Do là sử dụng model nomic-embed-text nên ra đầu ra chiều của vector quá lớn (768 dims) => gắn dữ liệu vào list tốn nhiều thời gian (2h45m), còn model multilingual-e4-small (384 dims) nhanh hơn do có chiều vector nhỏ hơn 15s | * Nếu mỗi bộ dữ liệu nhỏ (Dạng câu, không phải dạng file lớn) nhưng lượng dữ liệu lớn thì có thể giảm chiều dữ liệu vector để có thể giảm thời gian xử lý. |  |
| 12/5/2025 |  |  |  |  |
|  | * Tìm hiểu cấu trúc, thiết kế dataflow, thiết kế database, danh sách RestAPI. * Tìm hiểu về cách dịch vụ AWS sẽ được sử dụng trong project (AWS API Gateway, AWS S3, AWS Aurora, AWS Lambda, AWS Bedrock) * Tìm hiểu về thư viện python RAGAS, công cụ dùng để đánh giá mô hình LLM. |  |  | * Được thêm vào GitLab của dự án. Đọc và tìm hiểu soure code python của dự án. |
| 15/5/2025 |  |  |  |  |
|  | * Thuyết trình về các dịch vụ AWS sẽ được sử dụng trong đồ án cho người hướng dẫn. * Đọc và tìm hiểu source code python của đồ án. | * Hiểu được cách hoạt động, cách tính tiền của các phương án của các dịch vụ. |  | * Tìm hiểu cách các dịch vụ AWS được tích hợp vào dự án. |

# Chương 3: Cơ sở lý thuyết – Công nghệ sử dụng

## Python

## Amazon Web Service

### AWS API Gateway

### AWS Step Function

### AWS Aurora

### AWS S3 Bucket

### AWS Lambda Function

# Chương 4: Thông tin chi tiết công việc

## Thông tin dự án

### Khái quát dự án: **Revoria Cloud Production GenAI**

### Vị trí đảm nhiệm

## Công việc thực hiện

# Chương 5: Tổng kết

## Tổng kết quá trình thực tập:

# Tài liệu tham khảo