

PHIẾU ĐĂNG KÝ

PHƯƠNG THỨC ĐÀO TẠO VÀ ĐỀ TÀI LUẬN VĂN THẠC SĨ

1. Tên đề tài (ghi IN HOA):

- Tên tiếng Việt:
XÂY DỰNG HỆ THỐNG ĐỌC HIỂU VÀ GIẢI THÍCH CÂU TRẢ LỜI MÔN
HÓA HỌC TRUNG HỌC PHỔ THÔNG SỬ DỤNG OPENAI.
- Tên tiếng Anh:
AN OPENAI-BASED EXPLAINABLE QUESTION ANSWERING SYSTEM
FOR HIGH SCHOOL CHEMISTRY.
- Hướng đề tài luận văn: Hướng ứng dụng (12 TC)

2. Ngành học và Mã ngành: 8480101 (Khoa học máy tính)

3. Cán bộ hướng dẫn:

- Họ tên: TS. Dương Việt Hằng
- Email: hangdv@gm.uit.edu.vn
- Điện thoại: (+84) 919 196 708
- Đơn vị công tác: Trường Đại học Công Nghệ Thông Tin – ĐHQG TP.HCM

4. Thời gian thực hiện: 6 tháng. Từ tháng 11/2025

5. Học viên thực hiện:

- Họ tên: Trần Văn Tịnh
- Mã số: 220101039
- Email: tinh.tv.17@grad.uit.edu.vn
- Khóa: 17 Đợt: 2
- Điện thoại: (+84) 358 444 463

TP. HCM, ngày 31 tháng 10 năm 2025

Xác nhận của CBHD

(Ký tên và ghi rõ họ tên)

Học viên

(Ký tên và ghi rõ họ tên)

Dương Việt Hằng

Trần Văn Tịnh

ĐỀ CƯƠNG ĐỀ TÀI LUẬN VĂN THẠC SĨ

1. Giới thiệu

Đề tài "XÂY DỰNG HỆ THỐNG ĐỌC HIỂU VÀ GIẢI THÍCH CÂU TRẢ LỜI MÔN HÓA HỌC TRUNG HỌC PHỔ THÔNG SỬ DỤNG OPENAI" nhằm ứng dụng các công nghệ của OpenAI để xây dựng một hệ thống trí tuệ nhân tạo có khả năng đọc hiểu, trả lời câu hỏi các môn HÓA HỌC bậc trung học phổ thông để hỗ trợ học sinh nhanh chóng, kịp thời. Hệ thống không chỉ cho biết được đáp án mà còn giải thích chi tiết bản chất kiến thức và cách vận dụng từng bước giải quyết bài toán một cách chính xác.

Theo đó, input và output của hệ thống được mô tả như sau:

- Input: gồm 2 thành phần:
 - Module ID: tương ứng là mã môn học (Hóa học – Giáo dục phổ thông 2018). Việc sử dụng ID này cho phép dễ dàng mở rộng sang các môn học khác trong tương lai.
 - Một câu hỏi trong phạm vi môn học đó.
- Output: gồm 2 thành phần:
 - Câu trả lời ngắn gọn cho câu hỏi đã cung cấp.
 - Lời giải thích chi tiết từng bước để suy ra câu trả lời cho câu hỏi, tham chiếu đến các kiến thức cần thiết.

2. Bối cảnh, lý do chọn đề tài và khả năng ứng dụng.

Các ứng dụng công nghệ AI vào hệ thống giáo dục tiên tiến ở nước ta còn nhiều hạn chế. Phần lớn các hệ thống vẫn dựa nhiều vào nguồn lực và sự hỗ trợ bởi con người gây bất cập trong việc học các môn tự nhiên của học sinh. Học sinh phải chờ đợi lâu để được hỗ trợ, giải đáp thắc mắc bởi giáo viên và những người có chuyên môn.

Sự phát triển mạnh mẽ của trí tuệ nhân tạo và các mô hình ngôn ngữ lớn có khả năng lý luận (reasoning) nổi bật là các mô hình của OpenAI (O-series, gpt-5) thúc đẩy việc

ứng dụng trí tuệ nhân tạo trong việc dạy và học phù hợp với xu hướng chuyển đổi số giáo dục Việt Nam trong thời kì công nghệ 4.0.

Hệ thống phát triển theo phương thức mô-đun hóa (theo từng môn học) cho phép tích hợp vào các nền tảng học trực tuyến giúp hỗ trợ quá trình học của học sinh nhanh chóng, hiệu quả. Ngoài ra, hệ thống cũng có thể tích hợp vào các ứng dụng phụ trợ giảng dạy cho giáo viên và nhà trường trong việc tạo câu hỏi, đáp án và giải thích tự động giúp tiết kiệm chi phí và nhân lực.

3. Mục tiêu của đề tài.

- Tìm hiểu về các công nghệ của OpenAI và khả năng ứng dụng của chúng.
- Ứng dụng OpenAI để phát triển mô-đun trả lời câu hỏi kèm giải thích cho môn Hóa Học trung học phổ thông.
- Phát triển một hệ thống nền tảng Web trả lời câu hỏi các môn học bằng cách tích hợp liên tục các mô-đun trả lời cho từng môn học.

4. Nội dung nghiên cứu của đề tài.

- OpenAI platform, các models hỗ trợ và khả năng ứng dụng của chúng.
- Các cơ sở dữ liệu Hóa học (như Pubchem, ChemSpider) và các công cụ hỗ trợ khác (như phương trình phản ứng, cân bằng phương trình). Phương pháp tích hợp chúng vào các OpenAI models.
- Các phương pháp, kỹ thuật và chiến lược Prompting nâng cao như CoT, RAG giúp tối ưu hóa đầu ra của mô hình ngôn ngữ lớn.

5. Phương pháp thực hiện.

- **Xây dựng dataset câu hỏi Hóa Học:**
 - *Thu thập dữ liệu:* Thu thập danh sách câu hỏi từ các nguồn sách bài tập hóa học 10, 11, 12 gồm 3 bộ sách (Kết Nối Tri Thức Với Cuộc Sống, Chân Trời Sáng Tạo, và Cánh Diều). Sau đó sử dụng công cụ Mathpix để trích xuất câu hỏi, công thức hóa học từ PDF để tạo dữ liệu thô.
 - *Xử lý dữ liệu:* Tạo danh sách câu hỏi từ dữ liệu thô và gán nhãn phù hợp. Mỗi câu hỏi gồm 2 nhãn:
 - Theo loại câu hỏi: bao gồm 2 loại trắc nghiệm và tự luận.
 - Theo chủ đề câu hỏi: được phân theo 3 nhóm câu hỏi:

- + *Lý thuyết*: các liên quan đến các khái niệm hóa học như chất, cấu tạo nguyên tử, và các khái niệm cơ bản khác.
- + *Chất – Hợp chất*: các câu hỏi liên quan đến các đặc điểm và ứng dụng của một chất. Gồm các chủ đề nhỏ: phân tích, nhận biết, vận dụng.
- + *Phản ứng – chuỗi phản ứng hóa học*: các câu hỏi liên quan đến phản ứng hóa học. Gồm các chủ đề nhỏ hơn: phân tích, vận dụng, tính toán.

Bộ dataset sẽ được dùng để thử nghiệm và đánh giá ở các bước sau đó.

- **Xây dựng mô-dun (OpenAI agent) trả lời câu hỏi Hóa Học:**

- *Xây dựng OpenAI workflow cơ bản 3 giai đoạn*: Classification Agent → Reasoning and Solver Agent → Synthesizes and Generation Agent.

Trong đó:

- **Classification Agent**: chịu trách nhiệm xác định câu hỏi. Bao gồm xác định loại câu hỏi và chủ đề câu hỏi như trong bộ dữ liệu.
- **Reasoning and Solver Agent**: chịu trách nhiệm phân tích câu hỏi, thực hiện chi tiết suy luận và giải quyết câu hỏi.
- **Synthesizes and Generation Agent**: thực hiện tổng hợp và tạo câu trả lời thân thiện cho người dùng.
- *Lựa chọn mô hình*: thực hiện đánh giá hiệu suất của tập dữ liệu đã thu thập trên toàn bộ workflow và từng thành phần với các mô hình OpenAI khác nhau (gpt-5, gpt-4, O-series) và chọn mô hình phù hợp nhất.
- *Cải thiện với các công cụ hỗ trợ hóa học*: tích hợp các công cụ hỗ trợ hóa học (cơ sở dữ liệu Hóa học, phương trình phản ứng, cân bằng phương trình) . Đánh giá và so sánh kết quả.

- **Xây dựng Web cơ sở tích hợp mô-dun trả lời câu hỏi Hóa Học**: sử dụng các framework hiện đại Flask cho server API và React cho client UI.

6. Kết quả, sản phẩm dự kiến

- Bộ dataset trả lời câu hỏi môn Hóa học ngôn ngữ tiếng Việt.
- Module trả lời câu hỏi kèm giải thích môn Hóa học ngôn ngữ tiếng Việt.

- Ứng dụng Web cơ sở để cho phép tích hợp các mô đun trả lời câu hỏi theo một cách thống nhất.

7. Tài liệu tham khảo.

[1] OpenAI, “OpenAI Platform,” *Openai.com*, 2025.

<https://platform.openai.com/docs/overview>

[2] “OpenAI Cookbook,” *Openai.com*, 2025. <https://cookbook.openai.com/>.

[3] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. v, & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837.

[4] “Introduction - PubChemPy 1.0.5 documentation,” *Pubchempy.org*, 2025. <https://docs.pubchempy.org/en/latest/guide/introduction.html> (accessed Oct. 30, 2025).

[5] “ChemSpiPy — ChemSpiPy 2.0.0 documentation,” *Readthedocs.io*, 2018. <https://chemspipy.readthedocs.io/en/latest/index.html> (accessed Oct. 30, 2025).

8. Kế hoạch

Công việc	Thời gian Dự kiến (Ngày)	Ghi Chú
1. Thu thập danh sách câu hỏi Hóa học.	3	- Câu hỏi từ sách bài tập Hóa 10-12 3 bộ sách theo chương trình giáo dục mới. (2 ngày) - Sử dụng Mathpix snip để tạo dữ liệu thô. (1 ngày)
2. Xử lý dữ liệu	24	- Xử lý phân tách dữ liệu thô để phân chia thành danh sách các câu hỏi. (15 ngày, mỗi bộ sách 5 ngày) - Gán nhãn loại câu hỏi và chủ đề câu hỏi. (9 ngày, mỗi bộ sách 3 ngày)
3. Xây dựng OpenAI workflow	10	- Xây dựng OpenAI workflow cơ bản 3 giai đoạn. (7 ngày)

cơ bản cho mô-đun trả lời câu hỏi Hóa Học		- Thực hiện các tối ưu lời nhắc và hướng dẫn (3 ngày).
4. Đánh giá hiệu suất với các mô hình OpenAI khác nhau.	13	- Đánh giá Gpt-5 (3 ngày) - Đánh giá Gpt-4 (3 ngày) - Đánh giá O-series (7 ngày)
5. Nghiên cứu cải thiện với các công cụ hỗ trợ Hóa Học.	23	- Tìm hiểu các cơ sở dữ liệu Hóa Học PubChem, ChemSpider (3 ngày) - Tích hợp cơ sở dữ liệu Hóa Học và đánh giá hiệu quả. (15 ngày) - Nghiên cứu các công cụ về phương trình phản ứng, cân bằng phương trình (5 ngày - nếu cần thiết).
6. Xây dựng ứng dụng web cơ sở.	14	- Xây dựng Server API (7 ngày). - Xây dựng Client UI app (7 ngày)
7. Phân tích kết quả và viết luận văn.	27	- Thu thập tài liệu cần thiết cho việc viết luận văn (5 ngày). - Viết bản thảo luận văn (15 ngày) - Hoàn thiện và sửa đổi luận văn dựa trên phản hồi của giảng viên hướng dẫn (5 ngày) - Hoàn thiện và nộp bản thảo luận văn (2 ngày)
8. Làm Slides báo cáo	9	- Làm slides báo cáo và nhận phản hồi từ giảng viên hướng dẫn. (5 ngày) - Hoàn thiện slides báo cáo dựa trên phản hồi từ giảng viên hướng dẫn (4 ngày)
Tổng cộng	123	

Biểu đồ Grant

[illegible]

GIẢI TRÌNH CHỈNH SỬA

Các HV ghi ý kiến của ĐVCM trong thông báo kết quả xét duyệt và điền giải trình chỉnh sửa vào đây.