|  |  |
| --- | --- |
| **bỘ GIÁO DỤC VÀ ĐÀO TẠO** | **BỘ NÔNG NGHIỆP VÀ PTNT** |

**TRƯỜNG ĐẠI HỌC THỦY LỢI**



**BÁO CÁO ĐỒ ÁN TỐT NGHIỆP:**

**NGHIÊN CỨU PHƯƠNG PHÁP TỐI ƯU ĐẦU VÀO CHO**

**MÔ HÌNH NGÔN NGỮ LỚN GPT**

|  |  |
| --- | --- |
| Giảng viên hướng dẫn:  Họ và tên sinh viên: | PGS.TS Nguyễn Hữu Quỳnh  Đào Nguyệt Hằng |
| Mã sinh viên: | 175A071230 |
| Ngành : | Kỹ thuật phần mềm |

**HÀ NỘI, NĂM 2023**

|  |  |
| --- | --- |
|  | **CỘNG HOÀ XÃ HỘI CHỦ NGHĨA** **VIỆT NAM**  **Độc lập - Tự do - Hạnh phúc**  ----------★----------  **NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP** |

**Họ tên sinh viên: Đào Nguyệt Hằng Hệ đào tạo: Đại học chính quy**

**Lớp: 59PM2 Ngành: Kỹ thuật phần mềm**

**Khoa: Công nghệ thông tin**

**1- TÊN ĐỀ TÀI:**

NGHIÊN CỨU PHƯƠNG PHÁP TỐI ƯU ĐẦU VÀO CHO MÔ HÌNH NGÔN NGỮ LỚN GPT

**2- CÁC TÀI LIỆU CƠ BẢN:**

[1] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4. Now Publishers Inc, pp. 197–387, 2013. doi: 10.1561/2000000039.

[2] “Học sâu là gì? - Giải thích về Học sâu - AWS”.

[3] “Sách Deep Learning cơ bản - v2”.

[4] A. Vaswani *et al.*, “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[5] Quang Minh, “NGHIÊN CỨU, PHÁT TRIỂN CÔNG CỤ SINH MÔ TẢ SẢN PHẨM CHO THƯƠNG MẠI ĐIỆN TỬ 2021.”

[6] G. Samorodnitsky, “Long Range Dependence,” *Foundations and Trends® in Stochastic Systems*, vol. 1, no. 3, pp. 163–257, 2006, doi: 10.1561/0900000004.

[7] T. B. Brown *et al.*, “Language Models are Few-Shot Learners,” May 2020, [Online]. Available: http://arxiv.org/abs/2005.14165

[8] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, “CTRL: A Conditional Transformer Language Model for Controllable Generation,” Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.05858

**3 - NỘI DUNG CÁC PHẦN THUYẾT MINH VÀ TÍNH TOÁN:** Tỷ lệ %

|  |  |
| --- | --- |
| **Nội dung các phần** | **Tỷ lệ %** |
| Chương 1: Tổng quan đề tài | 20% |
| Chương 2: Phương pháp tối ưu hóa đầu vào cho GPT  Tìm hiểu về các khái niệm căn bản  Tìm hiểu về mô hình chat GPT sẵn có  Tìm hiểu về phương pháp cải thiện hiệu suất đầu vào dữ liệu cho mô hình chat GPT sẵn có | 60% |
| Chương 3: Kết quả và đánh giá  Xây dựng công cụ và môi trường sử dụng  Kết quả thực nghiệm  Đánh giá ưu điểm và hạn chế của phương pháp tối ưu đầu vào  Đề xuất hướng phát triển của nghiên cứu trong tương lai | 20% |

**5. GIÁO VIÊN HƯỚNG DẪN TỪNG PHẦN**

|  |  |
| --- | --- |
| **Phần** | **Họ và tên giáo viên hướng dẫn** |
| Chương 1: Tổng quan đề tài |  |
| Chương 2: Phương pháp tối ưu đầu vào cho GPT |  |
| Chương 3: Kết quả và đánh giá |  |

**6. NGÀY GIAO NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP**

Ngày ............ tháng ......... năm 20

|  |  |
| --- | --- |
| **Trưởng Bộ môn**  *(Ký và ghi rõ Họ tên)* | **Giáo viên hướng dẫn chính**  *(Ký và ghi rõ Họ tên)* |

Nhiệm vụ Đồ án tốt nghiệp đã được Hội đồng chấm thi tốt nghiệp của Khoa thông qua.

Ngày. . . . .tháng. . . . .năm 20.... **Chủ tịch Hội đồng**

*(Ký và ghi rõ Họ tên)*

Sinh viên đã nộp và hoàn thành bản Đồ án tốt nghiệp cho Hội đồng thi

Ngày....... tháng..... năm 20.....

**Sinh viên làm Đồ án tốt nghiệp**

*(Ký và ghi rõ Họ tên)*

|  |  |
| --- | --- |
|  | **CỘNG HOÀ XÃ HỘI CHỦ NGHĨA** **VIỆT NAM**  **Độc lập - Tự do - Hạnh phúc**  ----------★----------  **ĐỀ CƯƠNG ĐỒ ÁN TỐT NGHIỆP** |

**TÊN ĐỀ TÀI:**

**NGHIÊN CỨU PHƯƠNG PHÁP TỐI ƯU ĐẦU VÀO CHO**

**MÔ HÌNH NGÔN NGỮ LỚN GPT**

Giảng viên hướng dẫn PGS.TS Nguyễn Hữu Quỳnh

Sinh viên thực hiện Đào Nguyệt Hằng

Lớp 59PM2

MSV 175A071230

**TÓM TẮT ĐỀ TÀI**

Kỹ thuật gợi ý trong mô hình ngôn ngữ lớn (LLM) là một phương pháp giúp cải thiện khả năng dự đoán và đề xuất các từ hoặc câu tiếp theo trong quá trình sử dụng mô hình. LLM là một loại mô hình học sâu được huấn luyện trên các tập dữ liệu lớn và có khả năng tạo ra ngôn ngữ tự nhiên chính xác và có ý nghĩa.

Kỹ thuật gợi ý được áp dụng để giúp mô hình LLM có khả năng đề xuất các từ hoặc câu tiếp theo một cách chính xác và tự nhiên hơn. Thông qua việc sử dụng các kỹ thuật này, mô hình LLM có thể học được các quy luật của ngôn ngữ và cung cấp cho người dùng những gợi ý hữu ích trong quá trình sử dụng.

Ứng dụng của kỹ thuật gợi ý trong mô hình LLM là rất đa dạng. Chúng có thể được áp dụng để cải thiện khả năng dự đoán và đề xuất các từ hoặc câu tiếp theo trong các ứng dụng như trợ lý ảo, chuyển đổi ngôn ngữ tự động, tạo ra nội dung cho các bài viết, văn bản, tài liệu, v.v.

Với sự phát triển của công nghệ học sâu và mô hình LLM, kỹ thuật gợi ý đang trở thành một công cụ quan trọng giúp tăng cường khả năng xử lý ngôn ngữ tự nhiên của các ứng dụng.

Mong muốn được tìm hiểu và có thể giúp Thầy Cô Trường Đại Học Thuỷ Lợi cũng như rất nhiều bạn sinh viên có thể dễ dàng hơn, tiết kiệm thời gian hơn trong việc giải quyết mọi vấn đề vướng mắc trong mọi lĩnh vực một cách nhanh chóng. Đó là lí do em chọn làm đề tài “NGHIÊN CỨU PHƯƠNG PHÁP TỐI ƯU ĐẦU VÀO CHO MÔ HÌNH NGÔN NGỮ LỚN GPT”.

**CÁC MỤC TIÊU CHÍNH**

* Xây dựng mô hình LLM với ngôn ngữ đích là Tiếng Việt và dữ liệu huấn luyện lấy từ các nguồn khác nhau (Wikipedia, báo chí, văn bản chính thức, v.v.).
* Áp dụng các kỹ thuật gợi ý để cải thiện dữ liệu đầu ra cho Chat GPT
* So sánh và đánh giá hiệu quả của các kỹ thuật gợi ý được áp dụng.

**KẾT QUẢ DỰ KIẾN**

* Hoàn thành bộ xây dựng cơ sở dữ liệu Tiếng Việt cho Chat GPT.
* Tối ưu dữ liệu đầu vào sao cho Chat GPT đưa ra kết quả tốt nhất cho đầu vào đó.

# **Tài liệu tham khảo**

[1] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4. Now Publishers Inc, pp. 197–387, 2013. doi: 10.1561/2000000039.

[2] “Học sâu là gì? - Giải thích về Học sâu - AWS”.

[3] “Sách Deep Learning cơ bản - v2”.

[4] A. Vaswani *et al.*, “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[5] Quang Minh, “NGHIÊN CỨU, PHÁT TRIỂN CÔNG CỤ SINH MÔ TẢ SẢN PHẨM CHO THƯƠNG MẠI ĐIỆN TỬ 2021.”

[6] G. Samorodnitsky, “Long Range Dependence,” *Foundations and Trends® in Stochastic Systems*, vol. 1, no. 3, pp. 163–257, 2006, doi: 10.1561/0900000004.

[7] T. B. Brown *et al.*, “Language Models are Few-Shot Learners,” May 2020, [Online]. Available: http://arxiv.org/abs/2005.14165

[8] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, “CTRL: A Conditional Transformer Language Model for Controllable Generation,” Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.05858

# **TÓM TẮT ĐỀ TÀI**

Trong thời đại công nghiệp 4.0 hiện nay, AI đóng vai trò ngày càng quan trọng trong cuộc sống, giúp con người tìm kiếm, phân tích, dự đoán dữ liệu ngày càng nhanh hơn, chính xác hơn. Đóng góp vào thành công này là nhờ sự phát triển của lĩnh vực xử lý ngôn ngữ tự nhiên, một ví dụ điển hình của nó là ChatGPT, đây là một hiện tượng đang gây ra cơn sốt toàn cầu về công cụ tìm kiếm.

Mô iihình iingôn iingữ iitự iiđộng iihọc iiGPT ii(Generative iiPre-trained iiTransformer) iiđang iilà iimột iitrong iinhững iicông iicụ iiđược iisử iidụng iiphổ iibiến iitrong iilĩnh iivực iixử iilý iingôn iingữ iitự iinhiên. iiTuy iinhiên, iivới iikích iithước iicủa iicác iimô iihình iingày iicàng iilớn, iiviệc iiđưa iichúng iivào iisản iixuất iivà iisử iidụng iimột iicách iihiệu iiquả iitrở iinên iiquan iitrọng iihơn iibao iigiờ iihết.

Vì vậy, trong đồ án này, báo cáo đồ án sẽ tập trung tìm hiểu các phương pháp tối ưu đầu vào cho mô hình ngôn ngữ lớn GPT, cụ thể gồm xử lý phần dữ liệu đầu vào và thực nghiệm mô hình chatGPT có sẵn bao gồm cả các khái niệm cơ trong Trí tuệ nhân tạo, Học sâu, mô hình ChatGPT. Trong đó tập trung chính vào các phương pháp để tối ưu dữ liệu đầu vào phổ biến như Fine-tuning, Prompt Engineering, temperature,…Ngoài ra, đồ án cũng sẽ thể hiện đánh giá hiệu suất của những thí nghiệm kiểm tra.

Với nội dung đầy thú vị và ý nghĩa thực tiễn, đồ án này hy vọng sẽ đóng góp vào việc cải thiện hiệu suất của mô hình GPT và giúp cho các nhà nghiên cứu cà nhà phát triển có thể sử dụng mô hình một cách hiệu quả hơn trong các ứng dụng thực tế.

**Bài toán cần giải quyết:**

Giải quyết bài toán tối ưu đầu vào là dữ liệu cho GPT dựa trên mô hình chat GPT có sẵn.

**Hướng giải quyết:**

Sử dụng mô hình chat GPT có sẵn để tối ưu dữ liệu đầu vào.

**Dựng các thí nghiệm để kiểm tra hiệu suất.**

**CÁC MỤC TIÊU CHÍNH**

Tìm hiểu phương pháp tối ưu dữ liệu đầu vào có hiệu suất tốt nhất khi sử dụng mô hình chat GPT có sẵn.

**KẾT QUẢ DỰ KIẾN**

Tìm hiểu được khái niệm.

Tìm hiểu được về chat GPT có sẵn.

Tìm hiểu được phương pháp tối ưu dữ liệu đầu vào hiệu quả để có được hiệu suất tốt nhất.

Báo cáo đồ án tốt nghiệp

LỜI CAM ĐOAN

Tác giả xin cam đoan Đồ án tốt nghiệp của bản thân tác giả. Các kết quả trong Đồ án tốt nghiệp này là trung thực và không sao chép từ bất kỳ một nguồn nào và dưới bất kỳ hình thức nào.Việc tham khảo các nguồn tài liệu (nếu có) đã được thực hiện trích dẫn và trích dẫn nguồn tài liệu tham khảo đúng quy định.

|  |  |
| --- | --- |
|  | **Tác giả ĐATN/KLTN**  **ĐÀO NGUYỆT HẰNG** |

**LỜI CẢM ƠN**

Trong bài viết này, em xin gửi lời cảm ơn chân thành và sâu sắc đến tất cả các thầy cô, bạn bè và độc giả đã hỗ trợ và đóng góp quý báu cho quá trình tìm hiểu và hoàn thiện đồ án của tôi về “Nghiên cứu phương pháp tối ưu đầu vào cho mô hình ngôn ngữ lớn ChatGPT”.

Đầu tiên, tôi xin gửi lời cảm ơn đến các thầy cô giáo tại Khoa CNTT, Trường Đại học Thuỷ Lợi. Sự am hiểu, kiến thức sâu rộng và tận tâm của quý thầy cô đã giúp tôi có được những kiến thức căn bản vững chắc để tiến hành nghiên cứu. Quý thầy cô đã dành thời gian, sự quan tâm và tư vấn tận tâm trong suốt quá trình tôi thực hiện đồ án này, tôi rất biết ơn bởi những kiến thức và kinh nghiệm quý báu mà quý thầy cô đã truyền đạt cho tôi.

Tiếp theo tôi cũng muốn gửi lời cảm ơn đến tất cả các bạn bè tại trường Đại học Thuỷ Lợi. Những cuộc thảo luận, trao đổi ý kiến và sự chia sẻ kinh nghiệm của các bạn đã giúp tôi mở rộng kiến thức và nhìn nhận vấn đề từ nhiều góc độ khác nhau.

Ngoài ra, tôi xin cảm ơn đến tất cả các độc giả đã dành thời gian và quan tâm để đọc Cuối cùng, tôi xin gửi lời cảm ơn đến gia đình và người thân yêu của tôi, sự ủng hộ, động viên từ phía gia đình là nguồn động lực không thể thiếu trong suốt quá trình học tập và nghiên cứu.

**MỤC LỤC**

[TÓM TẮT ĐỀ TÀI i](#_Toc138686016)

[LỜI CAM ĐOAN i](#_Toc138686017)

[LỜI CẢM ƠN ii](#_Toc138686018)

[MỤC LỤC iii](#_Toc138686019)

[DANH MỤC HÌNH ẢNH v](#_Toc138686020)

[DANH MỤC BẢNG BIỂU vii](#_Toc138686021)

[DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH CÁC THUẬT NGỮ vii](#_Toc138686022)

[MỞ ĐẦU 1](#_Toc138686023)

[1. Lý do chọn đề tài. 1](#_Toc138686024)

[2. Mục đích nghiên cứu. 2](#_Toc138686025)

[3. Nhiệm vụ nghiên cứu. 2](#_Toc138686026)

[4. Đối tượng và phạm vi nghiên cứu 2](#_Toc138686027)

[5. Phương pháp nghiên cứu. 3](#_Toc138686028)

[6. Bố cục báo cáo. 3](#_Toc138686029)

[CHƯƠNG 1: TỔNG QUAN CƠ SỞ LÝ THUYẾT 4](#_Toc138686030)

[1.1 Học sâu (Deep learning) 4](#_Toc138686031)

[1.1.1 Tổng quan về học sâu 4](#_Toc138686032)

[1.1.2 Cách thức hoạt động của học sâu 4](#_Toc138686033)

[1.1.3 Một số ứng dụng của deep learning 5](#_Toc138686035)

[1.2 Mạng Nơ-ron (Neural networks) 6](#_Toc138686036)

[1.2.1 Tổng quan về Neural networks 6](#_Toc138686037)

[1.2.2 Cấu tạo chính của Neural Network 6](#_Toc138686038)

[1.2.3. Một vài ứng dụng đáng chú ý của Neural Network trong thực tế 8](#_Toc138686040)

[1.3 Mô hình Transformer 8](#_Toc138686041)

[1.3.1 Tổng quan về mô hình Transformer 8](#_Toc138686042)

[1.3.2 Mã hoá vị trí 9](#_Toc138686045)

[1.3.3 Bộ mã hoá 10](#_Toc138686046)

[CHƯƠNG 2: CÁC PHƯƠNG PHÁP TỐI ƯU HOÁ DỮ LIỆU ĐẦU VÀO CHO MÔ HÌNH CHATGPT SẴN CÓ 13](#_Toc138686049)

[2.1 ChatGPT 13](#_Toc138686050)

[2.1.1 Tổng quan về ChatGPT 13](#_Toc138686051)

[2.1.2 Các mô hình GPT sẵn có 14](#_Toc138686052)

[2.1.3 Kiến trúc GPT 14](#_Toc138686054)

[2.1.4 Ứng dụng của chatGPT 17](#_Toc138686057)

[2.2 Phương pháp 19](#_Toc138686058)

[2.2.1 Sinh mô tả 19](#_Toc138686059)

[2.2.2 Tokenizer tuỳ chỉnh 21](#_Toc138686062)

[2.2.3 Kỹ thuật tạo lời nhắc (Prompt Engineering) 23](#_Toc138686063)

[2.2.4 Sử dụng ngữ cảnh 24](#_Toc138686065)

[CHƯƠNG 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ 27](#_Toc138686066)

[3.1. Thực nghiệm 27](#_Toc138686067)

[3.1.1 Công cụ sử dụng để xây dựng thực nghiệm 27](#_Toc138686068)

[3.1.2 Xây dựng thực nghiệm 31](#_Toc138686075)

[3.1.3 Kết quả thực nghiệm 32](#_Toc138686076)

[3.2 Đánh giá kết quả thực nghiệm 37](#_Toc138686085)

[3.2.1 Xây dựng khảo sát đánh giá 37](#_Toc138686086)

[3.2.2 Kết quả xây dựng khảo sát đánh giá 39](#_Toc138686091)

[3.2.3 Đánh giá ưu điểm nhược điểm của các phương pháp sử dụng trong thực nghiệm 41](#_Toc138686096)

[KẾT LUẬN 42](#_Toc138686097)

[TÀI LIỆU THAM KHẢO 45](#_Toc138686098)

[PHỤ LỤC 46](#_Toc138686099)

# 

# **DANH MỤC HÌNH ẢNH**

[Hình 1.1 Cách hoạt động của mô hình Deep learning 5](#_Toc138686034)

[Hình 1.2 Cấu trúc một mạng Nơ-ron nhân tạo 7](#_Toc138686039)

[Hình 1.3 Sơ đồ kiến trúc tổng thể của Transformer [5] 9](#_Toc138686044)

[Hình 1.4 Kiến trúc lớp của bộ mã hóa và bộ giải mã trong Transformer [5] 11](#_Toc138686047)

[Hình 1.5 Kiến trúc chi tiết một lớp của bộ mã hóa trong Transformer [5] 12](#_Toc138686048)

[Hình 2.1. Mô hình huấn luyện GPT 15](#_Toc138686055)

[Hình 2.2. Phân tích giai đoạn trả lời lời nhắc của mô hình chatGPT 17](#_Toc138686056)

[Hình 2.3. Định dạng dữ liệu huấn luyện trong phương pháp đề xuất 21](#_Toc138686061)

[Hình 2.4 Kỹ thuật gợi ý Prompt Engineering 24](#_Toc138686064)

[Hình 3.1 Kết quả test API Chat GPT 28](#_Toc138686070)

[Hình 3.2 Kết quả test API thành công 28](#_Toc138686072)

[Hình 3.3 Hình ảnh giao diện frontend thể hiện qua components 29](#_Toc138686073)

[Hình 3.4 Giao diện website thử nghiệm 29](#_Toc138686074)

[Hình 3.5 Kết quả tối ưu đầu vào bằng phương pháp sử dụng Prompt Engineering của thí nghiệm thứ nhất 33](#_Toc138686077)

[Hình 3.6 Kết quả tối ưu đầu vào lần thứ nhất bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ hai lần thứ nhất 33](#_Toc138686079)

[Hình 3.7 Kết quả tối ưu đầu vào lần thứ hai bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ hai 34](#_Toc138686081)

[Hình 3.8 Kết quả tối ưu đầu vào lần thứ ba bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ hai 34](#_Toc138686082)

[Hình 3.9 Kết quả tối ưu đầu vào lần thứ nhất bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ ba 36](#_Toc138686083)

[Hình 3.10 Kết quả tối ưu đầu vào lần thứ hai bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ ba 36](#_Toc138686084)

[Hình 3.11 Form khảo sát đánh giá kết quả câu trả lời phù hợp 37](#_Toc138686087)

[Hình 3.12 Form thể hiện câu hỏi thứ nhất và lựa chọn câu trả lời phù hợp 37](#_Toc138686088)

[Hình 3.13 Form thể hiện câu hỏi thứ hai và lựa chọn câu trả lời phù hợp 38](#_Toc138686089)

[Hình 3.14 Form thể hiện câu hỏi thứ ba và lựa chọn câu trả lời phù hợp 38](#_Toc138686090)

[Hình 3.15 Biểu đồ thể hiện tỉ lệ lứa tuổi tham gia đánh giá 39](#_Toc138686092)

[Hình 3.16 Thống kê tỉ lệ lựa chọn câu trả lời của câu hỏi thí nghiệm thứ nhất 39](#_Toc138686093)

[Hình 3.17 Thống kê tỉ lệ lựa chọn câu trả lời của câu hỏi thí nghiệm thứ hai 40](#_Toc138686094)

[Hình 3.18 Thống kê tỉ lệ lựa chọn câu trả lời của câu hỏi thí nghiệm thứ ba 40](#_Toc138686095)

**DANH MỤC BẢNG BIỂU**

[Bảng 2.1 So sánh các phiên bản GPT 14](#_Toc138686053)

**DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH CÁC THUẬT NGỮ**

|  |  |  |
| --- | --- | --- |
| **TA viết tắt** | **Thuvi ngữ** | **Giuv thích** |
| AI | Artificial Intelligence | Trí tuệ nhân tạo |
| GPT | Generative Pre-training Transformer | Bine đổi tạo sinh tiền huấn luyện |
| NLP | Natural Language Processing | Xa lý ngôn ngữ tự nhiên |
| CNN | Convolutional Neural Network | Monv nơ-ron tích chập |
| PPO | Proximal Policy Optimization | Tro ưu hóa chính sách tiệm cận |
| LLM | Large Language Model | Mô hình ngôn ngữ lớn |
| LSTM | Long Short-Term Memory | Trí nhớ ngắn hạn dài |
| BERT | Bidirectional Encoder Representations from Transformers | Bidi diễn bộ mã hóa hai chiều từ Transformers |
| TOKEN | Token | Dãy ký tự, một đơn vị ngữ nghĩa trong xử lý ngôn ngữ, trong nhiều trường hợp Token được hiểu là các từ |

# **MỞ ĐẦU**

## 1. Lý do chọn đề tài.

Chatbot GPT giúp người dùng hiểu rõ dữ liệu và đưa ra các lựa chọn đầu tư một cách nhanh gọn và chính xác. Từ đó, một loại chatbot tài chính mới được tạo ra và phát triển mạnh mẽ. Chatbot này có khả năng hỗ trợ trong các hoạt động đầu tư, ví dụ như hỗ trợ trong việc thanh toán và theo dõi các khoản đầu tư, kiểm tra số dư trong tài khoản..

Bằng những cuộc trò chuyện đơn giản được lập trình sẵn, Chat GPT nói riêng và Chatbot nói chung đã giúp cho các công ty nói riêng và các doanh nghiệp tối ưu hoá hiệu quả hoạt động của họ. Ngoài ra, nhà đầu tư cũng có thể sử dụng Chat GPT để hỗ trợ các quá trình cũng như xây dựng chương trình đầu tư của họ, ứng dụng ChatGPT có khả năng phân tích dữ liệu một cách chính xác và hiệu quả. Từ đó sẽ chỉ dẫn cho nhà đầu tư đầu tư một cách hiệu quả và chính xác cao.

Chat GPT được cung cấp một lượng dữ liệu khổng lồ về tất cả các lĩnh vực. Vì thế, nhờ có những phương pháp phân tích khoa học và phức tạp, chat GPT có thể định hướng và xác định được chiều hướng tại thời điểm quá khứ, hiện tại và tương lai cụ thể về một lĩnh vực mà người dùng yêu cầu. Chính vì vậy, Chat GPT có thể tác động mạnh mẽ đến tất cả thị trường thông qua việc cung cấp các kết quả tìm kiếm chính xác trong một khoảng thời gian ngắn, không những vậy, Chat GPT còn có thể truy vấn và xử lý các vấn đề phức tạp và hồi đáp lại cho người dùng một câu trả lời đáng tin cậy.

Trong tương lai, Chat GPT chắc chắn sẽ trở thành một trợ thủ đắc lực cho các chuyên gia tài chính khi nghiên cứu thị trường. Không những rút ngắn thời gian nghiên cứu mà còn giảm chi phí rủi ro đến đáng kể. Chat GPT là một ứng dụng rất hữu ích và đa dạng nếu biết sử dụng một cách hợp lý, người dùng sẽ được lợi ích tối đa nếu biết khai thác triệt để ứng dụng này vào thực tiễn.

Với những mặt tích cực là vậy nhưng Chat GPT vẫn tồn tại những nhược điểm nhất định. Chat GPt cũng chỉ là một chương trình với những câu trả lời được lập trình sẵn. Trong khi đó thực tiễn lại tiềm ẩn những biến động bất ngờ. Vì vậy khi áp dụng tư vấn từ Chat GPT vẫn có thể tồn tại những sai sót. Người dùng không nên quá phụ thuộc vào ứng dụng này.

Sự xuất hiện của Chat GPT và những tính năng của nó ra đời vẫn còn là tâm điểm khiến dư luận bàn tán. Đặc biệt là tính đúng sai của ứng dụng này vẫn còn tồn tại nhiều ý kiến trái chiều. Ví dụ, các nhà khoa học AI tại Meta Yann LeCun và giáo sư Đại học New York cho rằng: Công cụ Chat GPT không phải là một tiến bộ khoa học đặc biệt mà ứng dụng này được gọi là “bản thử nhiệm hào nhoáng” do các kĩ sư xây dựng. Trong khi đó, chuyên gia Haomiao Huang của Kleiner Perkins – một công ty đầu tư mạo hiểm ở Thung lũng Silicon lại cảnh báo về Chat GPT trên tran công nghẹ Ars Technica: “Khi làm việc với các mô hình AI này, chúng ta phải nhớ rằng các công cụ đó là máy đánh bạc chứ không phải là máy tính”. Jason Davis, một giáo sư nghiên cứu thuộc Đại học Syracuse khẳng định: “Cuộc các mạng thực sự chính là cuộc các mạng tạo ra cuộc trò chuyện với AI như giữa người với người.” Để hiểu rõ vấn đề trên, đồ án chỉ rõ về chủ đề “Nghiên cứu phương pháp tối ưu đầu vào cho mô hình ngôn ngữ lớn GPT”.

## 2. Mục đích nghiên cứu.

Đề iitài iicủa iibài iibáo iicáo iitập iitrung iivào iinghiên iicứu iitối iiưu iiđầu iivào iicho iimô iihình iingôn iingữ iilớn iiGPT iikhông iichỉ iigiới iithiệu iichung iivề iiGPT iimà iicòn iinghiên iicứu iithực iitrạng iisử iidụng iivà iicông iidụng iicủa iiGPT iibết iicạnh iiđó iiđể iinói iilên iitối iiưu iiđầu iivào iidựa iitrên iikết iiquả iinghiên iicứu. ii

## 3. Nhiệm vụ nghiên cứu.

Để iilàm iirõ iinội iidung iicủa iiđề iitài, iithực iihiện iiđược iimục iiđích iiđã iiđề iira, iinhiệm iivụ iimà iiđề iitài iiphải iithực iihiện iiđó iilà: ii

- iiTìm iihiểu iivề iingôn iingữ iilớn iivà iichat iiGPT

- iiThu iithập, iixử iilý iivà iiphân iitích iithông iitin, iidữ iiliệu iivề iithực iitrạng iisử iidụng iichat iiGPT

- iiTrình iibày iikết iiquả iinghiên iicứu iithông iicác iithí iinghiệm iivà iiưu iinhược iiđiểm iicủa iinó.

**4. Đối tượng và phạm vi nghiên cứu**

Đối iitượng iinghiên iicứu: ii

* Phương iipháp iitối iiưu iiđầu iivào iicho iimô iihình iingôn iingữ iilớn iiGPT

Phạm vi nghiên cứu và nội dung:

- iiGiới iihạn iikhông iigian: iigiới iithiệu iivề iichat iiGPT

- iiGiới iihạn iivề iithời iigian: iiTừ iingày ii10/06/2022 iiđến iingày ii10/06/2023. ii

- iiPhạm iivi iinội iidung: iiNghiên iicứu iivề iitối iiưu iihóa iiđầu iivào iiGPT

## 5. Phương pháp nghiên cứu.

Quá iitrình iinghiên iicứu iiđược iithực iihiện iithông iiqua iicác iiphương iipháp iisau:

- iiPhương iipháp iinghiên iicứu iitài iiliệu, iithu iithập iithông iitin: iinghiên iicứu iinhững iitài iiliệu, iigiáo iitrình, iicác iibài iiviết iitrên iisách ii, iibáo, iiinternet iiđể iitừ iiđó iithu iithập iithông iitin iimột iicách iicó iichọn iilọc. ii

- iiPhương iipháp iiquan iisát, iighi iichép iithực iitế: iithu iithập iinhững iithông iitin iithực iitế iiliên iiquan iiđến iiviệc iithu iithập iidữ iiliệu iinghiên iicứu iikết iiquả iitối iiưu iihóa iiđầu iivào iiGPT

- iiPhương iipháp iiphân iitích, iiđánh iigiá, iitổng iihợp. ii

- iiPhương iipháp iikết iihợp iilý iiluận iivới iithực iitiễn. ii

- iiPhương iipháp iilấy iiý iikiến iicủa iicác iichuyên iigia ii

## 6. Bố cục báo cáo.

Ngoài phần Mở đầu, Kết luận, Danh mục tài liệu tham khảo, Phụ lục nội dung chính của bài báo cáo được trình bày trong ba chương chính:

* Chương1: TỔNG QUAN CƠ SỞ LÝ THUYẾT
* Chương 2: CÁC PHƯƠNG PHÁP TỐI ƯU DỮ LIỆU ĐẦU VÀO CHO MÔ HÌNH CHAT GPT SẴN CÓ
* Chương 3: THỰC NGHIỆM VÀ ĐÁNH GIÁ

**CHƯƠNG 1: TỔNG QUAN CƠ SỞ LÝ THUYẾT**

Chương 1 nói về những Cơ sở lý thuyết giải quyết bài toán bao gồm có bốn phần chính: Học máy, Deep learning, Neural networks, Mô hình Transformer, ChatGPT.

**1.1 Học sâu (Deep learning)**

***1.1.1 Tổng quan về học sâu***

Deep learning [1] là một nhánh của học máy, sử dụng các mạng nơ-ron sâu để mô phỏng cách thức học tập và suy nghĩ của con người. Học sâu là một tập hợp con của học máy hay nói cách khác, Deep learning là một mạng Nơ-ron có từ ba lớp trở lên. Những mạng lưới thần kinh này cố gắng mô phỏng hành vi của não người cho phép học sâu “ học” từ một lượng lớn dữ liệu. Trong khi mạng Nơ-ron có một lớp vấn có thể đưa ra các dự đoán gần chính xác, các lớp ẩn bổ sung có thể giúp tối ưu và tinh chỉnh để đưa ra kết quả có độ chính xác tốt nhất.

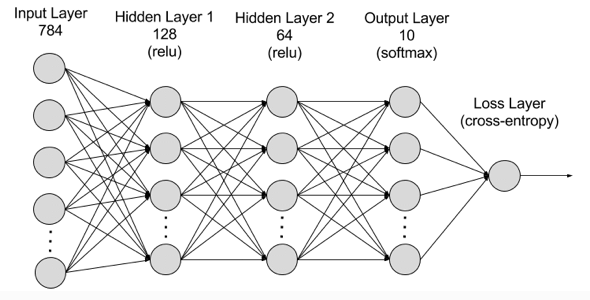
Học sâu được hiểu đơn giản là kết nối dữ liệu giữa tất các các tế bào thần kinh nhân tạo và điều chỉnh dữ liệu theo dữ liệu mẫu. Càng nhiều tế bào thần kinh được thêm vào thì kích thước của dữ liệu càng lớn, Chúng tự động có tĩnh năng học tập ở nhiều cấp độ trừu tượng. Qua đó cho phép một hệ thống học hàm ánh xạ phức tạp mà không phụ thuộc vào bất kỳ thuật toán cụ thể nào, cũng rất khó để biết điều gì sẽ xảy ra trong một mạng lưới thần kinh nhân tạo. Chính vì vậy, có thể coi Deep Learning là một chiếc hộp đen.

***1.1.2 Cách thức hoạt động của Học sâu***

Học iisâu iihoạt iiđộng iibằng iicách iikhám iiphá iicác iicấu iitrúc iiphức iitạp iitrong iidữ iiliệu iimà iichúng iitrải iinghiệm. iiCụ iithể iilà iixây iidựng iicác iimô iihình iitính iitoán iibao iigồm iinhiều iilớp iixử iilý, iimạng iicó iithể iitạo iira iinhiều iiCác iithuật iitoán iihọc iisâu - các iimạng iinơ-ron iiđược iilập iimô iihình iitheo iibộ iinão iicon iingười. Ví dụ: một bộ não con người chứa hàng triệu nơ-ron được kết nối với nhau, làm việc cùng nhau để tìm hiểu và xử lý thông tin. Tương tự, các mạng nơ-ron học sâu, hay mạng nơ-ron nhân tạo, được tạo thành từ nhiều lớp nơ-ron nhân tạo hoạt động cùng nhau bên trong máy tính.

Những Nơ-ron nhân tạo – Mô-đun phần mềm – nút: sử dụng các phép toán để thực hiện xử lý dữ liệu. Các mạng Nơ-ron nhân tạo là những thuật toán Deep Learning sử dụng các nút để giải quyết các vấn đề phức tạp.

Mức trừu tượng để biểu diễn dữ liệu.



*Hình 1.1 Cách hoạt động của mô hình Deep learning*

Như Hình 1.1 [2], Một mô hình Deep Learning được gọi là mạng Nơ-ron phức hợp có thể được đào tạo bằng việc sử dụng một số lượng lớn (hàng triệu) hình ảnh, ví dụ như những hình ảnh có chứa táo. Loại mạng Nơ-ron này thường học hỏi từ các pixel có trong hình ảnh mà chúng thu được. Chúng có thể phân loại các nhóm pixel đại diện cho các đặc điểm của táo, với các nhóm đặc điểm như màu sắc, hình dạng, cuống và từ đó cho biết sự hiện diện của quả táo có trong hình ảnh. Với nhiều tác vụ, chẳng hạn như thị giác máy tính, nhận dạng giọng nói còn được gọi là xử lý ngôn ngữ tự nhiên, dịch máy và robot, hiệu suất của hệ thống Học sâu vượt xa so với các hệ thống Học máy. Điều này không có nghĩa là việc xây dựng các hệ thống Deep Learning là tương đối dễ dàng so vớ các hệ thống học máy thông thường. Dù cho nhận dạng tính năng tự trị trong học sâu, hàng nghìn siêu tham số hay nút bấm cần được điều chỉnh để mô hình Học sâu trở nên hiệu quả

***1.1.3 Một số ứng dụng của deep learning***

**Người máy:** Nhờ có sự phát triển của khoa học và công nghệ trong thời gian gần đây, đặc biệt là trong lĩnh vực học máy và trí tuệ nhân tạo đã được thúc đẩy bởi những tiến bộ với những sản phẩm đột phá mới.

**Nông nghiệp:** Hiện nay, Học sâu cho phép nông dân có thể triển khai thiết bị nhìn thấy và phân biệt giữa cây trồng và cỏ dại. Khả năng này cho phép máy làm cỏ phun thuốc diệt cỏ một cách chính xác mà không làm cho những cây khác bị thiệt hại hay ảnh hưởng.

**Hình ảnh y tế và chăm sóc sức khoẻ:** Deep Learning ngày càng đóng góp những bước tiến đáng kể vào việc cải thiện hiệu quả, chất lượng, chăm sóc sức khoẻ con người, động vậy bằng các dự đoán các sự kiện, hình ảnh y tế từ dữ liệu hồ sơ bệnh án điện tử.

**1.2 Mạng Nơ-ron (Neural networks)**

***1.2.1 Tổng quan về Neural networks***

Neural networks là một mô hình toán học được lấy cảm hứng từ cấu trúc của não bộ. Mô hình này được sử dụng để giải quyết các bài toán phân loại, dự đoán và mô hình hóa các mối quan hệ phức tạp giữa các đặc trưng. iĐây iilà iimột iimô iihình iitoán iihọc iiphức iitạp iiđược iiđưa iira iiđể iixử iilý iithông iitin, iigiải iiquyết iicác iivấn iiđề iiphổ iibiến iitrong iilĩnh iivực iihọc iimáy, iitrí iituệ iinhân iitạo iiAI, iideep iilearning iivà iitìm iikiếm iitất iicả iicác iimối iiquan iihệ iicăn iibản iitrong iimột iitập iihợp iicác iidữ iiliệu. iiTất iicả iiđiều iinày iiđược iiNeural iiNetwork iithực iihiện iithông iiqua iiviệc iibắt iichước iicác iihoạt iiđộng iicủa iicủa iitế iibào iithần iikinh iivà iikhớp iithần iikinh iitừ iibộ iinão iicủa iicon iingười. iiHiểu iiđơn iigiản, iiNeural iiNetwork iitương iitự iinhư iibộ iinão iicủa iicon iingười, iinó iisẽ iikết iinối iicác iinút iiđơn iigiản iivà iicác iinút iinày iisẽ iitạo iithành iimột iimạng iilưới. iiDo iiđó, iinó iiđược iixem iilà iihệ iithống iicủa iicác iitế iibào iithần iikinh iinhân iitạo. ii

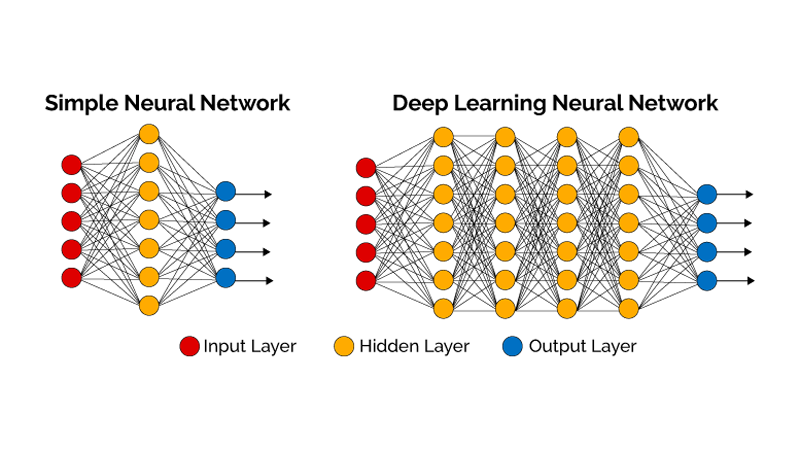
***1.2.2 Cấu tạo chính của Neural Network***

Trong iiNeural iiNetwork, iimột iiloạt iicác iithuật iitoán iisẽ iiđược iisử iidụng iivới iimục iiđích iixác iiđịnh iivà iinhận iira iitất iicả iicác iimối iiquan iihệ iicó iitrong iicác iitập iidữ iiliệu. iiNeural iiNetwork iicó iikhả iinăng iithích iiứng iinhanh iichóng iivới iimọi iithay iiđổi iingay iitừ iiđầu iivào. iiDo iiđó, iibạn iisẽ iikhông iicần iiphải iithiết iikế iilại iinhững iitiêu iichí iiđầu iira iimà iiNeural iiNetwork iivẫn iicó iithể iiđưa iira iiđược iimọi iikết iiquả iimột iicách iitốt iinhất. iiNeural iiNetwork iilà iisự iikết iihợp iicủa iicác iitầng iiperceptron ii(perceptron iiđa iitầng). iiVà iimỗi iimạng iiNeural iiNetwork iithường iibao iigồm iiba iithành iiphần iichính iinhư iisau:

**Tầng** ii**vào** ii**(input** ii**layer)** iiTầng iivào iinằm iiở iiphía iibên iitrái iicủa iiNeural iiNetwork. iiTầng iinày iiđại iidiện iicho iicác iiđầu iivào iicủa iimạng iiNơron iinhân iitạo.

**Tầng** ii**ra** ii**(output** ii**layer)** iinằm iiở iiphía iibên iiphải iicủa iiNeural iiNetwork. iiNó iiđại iidiện iicho iinhững iiđầu iira iicủa iimạng iinơron iinhân iitạo.

**Tầng** ii**ẩn** ii**(hidden** ii**layer)** iiĐây iilà iitầng iinằm iixen iigiữa iihai iitầng - tầng iivào iivà iitầng iira iithể iihiện iicho iiquá iitrình iisuy iiluận iilogic iicủa iiNeural iiNetwork. iiNó iiđại iidiện iicho iinút iitrung iigian iiphân iichia iikhông iigian iiđầu iivào iicác iivùng iimềm ii(vùng iiranh iigiới). iiTầng iiẩn iisẽ iinhận iimột iitập iihợp iicác iiđầu iivào iicó iitrọng iisố iivà iithông iiqua iimột iichức iinăng iikích iihoạt iiđể iitạo iira iikết iiquả iiđầu iira.



*Hình 1.2 Cấu trúc một mạng Nơ-ron nhân tạo*

Như Hình 1.2 [3] trên đây, mỗi Neural Network chỉ có một input layer và output layer nhưng sẽ bao gồm nhiều hidden layer. Từng Nút mạng (Sigmoid neural) sẽ có hàm kích hoạt khác nhau nhưng đều áp dụng thuật toán đồng nhất để có thể dễ dàng hoạt động.

Mỗi iitầng iisẽ iicó iisố iilượng iisigmoid iineural iikhác iinhau iitùy iitheo iicách iithức iixử iilý iidữ iiliệu. iiTrong iiquá iitrình iihoạt iiđộng, iicác iichuyên iigia iisẽ iithiết iilập iicác iihidden iilayer iivới iisố iilượng iinơ-ron iikhác iinhau. iiCác iinơ-ron iiở iicác iitầng iikhác iicũng iiliên iikết iivới iinhau iiđể iitạo iithành iimột iimạng iilưới iichặt iichẽ iivà iiđầy iiđủ. iiLúc iinày, iingười iidùng iicó iithể iinắm iirõ iiđược iiđộ iilớn iicủa iimạng iilưới iidựa iitrên iisố iilượng iitầng iivà iinơ-ron.

Các nút mạng kết hợp theo một chiều duy nhất từ input layer cho đến output layer. Mỗi nút ở một tầng sẽ nhận thông tin đầu của các nốt ở tầng trước đó.

***1.2.3. Một vài ứng dụng đáng chú ý của Neural Network trong thực tế***

Với iisố iilượng iicác iiứng iidụng iiđược iitriển iikhai iingày iicàng iilớn, iiNeural iiNetwork, iimáy iihọc iivà iitrí iituệ iinhân iitạo iiđược iiứng iidụng iitrong iinhiều iicông iiviệc iikhác iinhau iinhư:

Nhận iidạng iichữ iiviết iitay: iiMạng iiNơron iinhân iitạo iiNeural iiNetwork iiđược iisử iidụng iivới iimục iiđích iichuyển iiđổi iicác iiký iitự iiviết iitay iithành iicác iiký iitự iikỹ iithuật iisố iiđể iimáy iitính iidễ iidàng iinhận iira iiđược iicác iiký iitự iiđó.

Dự iiđoán iicác iigiao iidịch iichứng iikhoán: iiViệc iitheo iidõi iisàn iigiao iidịch iichứng iikhoán iithường iikhó iikhăn iivà iikhó iihiểu iido iicó iinhiều iiyếu iitố iiảnh iihưởng iiđến iithị iitrường iinày. iiNeural iiNetwork iira iiđời iicó iithể iikiểm iitra iiđược iicác iiyếu iitố iinói iitrên iivà iidự iiđoán iigiá iichứng iikhoán iihàng iingày. iiĐiều iinày iimang iiđến iinhiều iilợi iiích iicho iicác iinhà iimôi iigiới iichứng iikhoán.

Vấn iiđề iiđi iilại iicủa iicác iinhân iiviên iibán iihàng: iiNeural iiNetwork iigiúp iigiải iiquyết iivà iitìm iira iicon iiđường iitối iiưu iicho iicác iinhân iiviên iibán iihàng iitrong iiviệc iiđi iilại iigiữa iicác iithành iiphố iithuộc iicùng iimột iikhu iivực iicụ iithể iiđồng iithời iimang iiđến iidoanh iithu iicao iivà iigiảm iithiểu iichi iiphí.

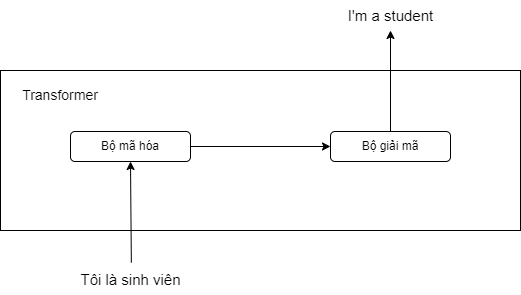
Nén iihình iiảnh: iiÝ iitưởng iiđằng iisau iicủa iiNeural iiNetwork iinén iidữ iiliệu iilà iilưu iitrữ, iimã iihoá iivà iitái iitạo iicác iihình iiảnh iimang iitính iithực iitế. iiBằng iicách iisử iidụng iiNeural iiNetwork iinén iihình iiảnh, iibạn iicó iithể iidễ iidàng iitối iiưu iikích iithước iidữ iiliệu iicủa iimình. iiĐiều iinày iisẽ iigiúp iibộ iinhớ iicủa iibạn iitiết iikiệm iiđược iidung iilượng iiđáng iiđể.

Ngoài iira, iiNeural iiNetwork iicũng iiđược iisử iidụng iinhiều iitrong iinhiều iiứng iidụng iivà iicông iinghệ iikhác iinhau iinhư iithị iigiác iimáy iitính, iitrò iichơi iiđiện iitử, iidịch iitự iiđộng, iilọc iimạng iixã iihội, iinhận iidạng iigiọng iinói iivà iichẩn iiđoán iiy iitế. iiĐiều iiđáng iingạc iinhiên iihơn iilà iiNeural iiNetwork iicũng iiđược iisử iidụng iinhiều iitrong iicác iihoạt iiđộng iitruyền iithống iivà iisáng iitạo iinhư iinghệ iithuật iihay iihội iihoạ.

**1.3 Mô hình Transformer**

***1.3.1 Tổng quan về mô hình Transformer***

Transformer – mô hình ngôn ngữ do Google phát triển và công bố vào năm 2017 trong bài báo “Attention is All You Need” [4]. iiTransformer iihoạt iiđộng iidựa iitrên iiCơ iichế iichú iiý ii(attention iimechanism), iicho iiphép iinó iitập iitrung iivào iicác iiphần iiquan iitrọng iicủa iicâu iiđầu iivào iikhi iidịch iihay iisinh iivăn iibản. iiTransformer iilà iimột iikiến iitrúc iimô iihình iineural iinetworks iiđược iisử iidụng iitrong iiviệc iixử iilý iingôn iingữ iitự iinhiên. iiMô iihình iinày iisử iidụng iiself-attention iiđể iihọc iicác iimối iiquan iihệ iigiữa iicác iitừ iitrong iicâu. Transformer có kiến trúc gồm hai thành phần chính. Phần thứ nhất là Bộ mã hoá và phần thứ hay là Bộ giải mã. iNhiệm iivụ iicủa iichúng iilà iibiến iiđổi iinội iidung iiđầu iivào iithành iimột iinội iidung iiđầu iira iitương iiứng. iiCụ iithể iihơn, iibộ iimã iihóa iisẽ iiánh iixạ iichuỗi iitoken iiđầu iivào iix ii= ii(x1 ii, iix2 ii, ii…, iixn ii) iithành iimột iichuỗi iiliên iitục iitương iiứng iiz ii= ii(z1 ii, iiz2 ii, ii…, iizn ii) iivà iibộ iigiải iimã iisẽ iisử iidụng iichuỗi iinày iiđể iitạo iira iimột iichuỗi iiđầu iira iiy ii= ii( iiy1 ii, iiy2 ii, ii…, iiyn ii).Dưới iiđây iilà kiến iitrúc iitổng iithể iicủa iimô iihình.: ii



*Hình 1.3 Kiến trúc tổng thể của Transformer* [5]

Hình 1.3 là Kiến trúc tổng thể của Transformer. Câu “Tôi là sinh viên” là một văn bản nguồn được Bộ mã hoá xử lý thông qua các phép biến đổi để trở thành đầu vào của Bộ giải mã. Tiếp đó, Bộ giải mã lại biến đổi tiếp đầu vào này thành kết quả cuối cùng là “I’m a student. Và để làm rõ quy trình và tính hiệu quả của kiến trúc đang được nghiên cứu trên.

***1.3.2 Mã hoá vị trí***

Như đã biết, trước khi các từ trong văn bản thực sự được xử lý bởi mô hình, chúng đã phải trải qua quá trình mã hoá để chuyển thành một hình thức biểu diễn thông tin khác mà chương trình có thể hiểu được. Các từ nhúng hay những hình thức thể hiện thông tin, vectơ từ nhúng hoặc word embedding. Chúng đều được nối với nhau trở thành một ma trận hai chiều và được xử lý bởi các tiến trình tiếp theo của mô hình. Không có vấn đề gì khi được áp dụng ở các mô hình tuần tự bởi quá trình này diễn ra rất phổ biến. Đối với mô hình Transformer, xuất hiện một vấn đề khác nảy sinh là mô hình không thể nhận biết được vị trí các từ thông qua word embedding do cơ chế xử lý các từ song song của mô hình. Để giải quyết vấn đề này, các nhà nghiên cứu đã giới thiệu về cơ chế mã hoá mới. Cơ chế này có tên gọi là mã hoá vị trí (Position Encoding), với mục tiêu mã hoá vị trí của các từ bằng một embedding tương ứng. Dưới đây là công thức mã hoá vị trí được nêu trong bài báo:

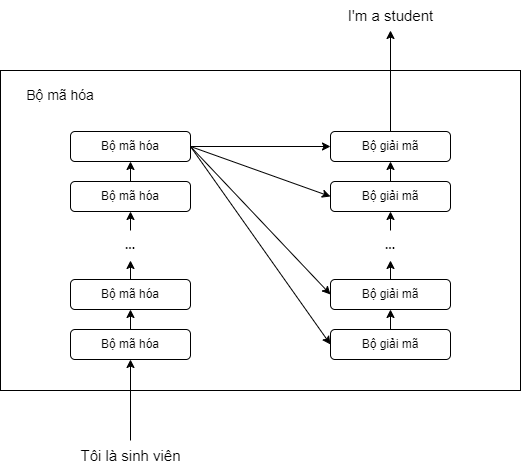
Trong đó:

(2)

Ở công thức (1), Với t là vị trí của từ trong văn, i là vị trí chiều không gian embedding, hàm sin để tính giá trị tương ứng với vị trí chiều chẵn và cos với vị trí chiều lẻ.

***1.3.3 Bộ mã hoá***

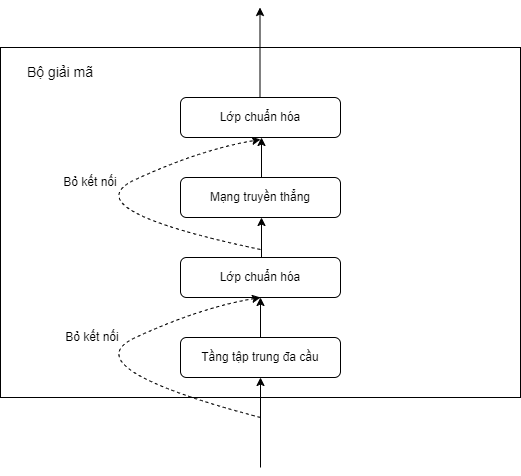
Hình 1.3 đã thể hiện rằng, trong mô hình Transformer, bộ mã hoá là một trong hai thành phần quan trọng. Đóng vai trò mã hoá đầu vào thành một biểu diễn mới đa dạng thông tin hơn được gọi là “vectơ ngữ cảnh”. Kiến trúc của bộ mã hoá được tạo nên bằng cách xếp chồng các lớp mã hoá với nhau và thường là 6 lớp, tạo nên một mạng truyền thẳng. Thay vì xử lý tuần tự giống các mô hình khác như LSTM thì kiến trúc này cho phép các lớp xử lý đồng thời các từ.



Hình 1.4 Kiến triế lớp của bộ mã hãp vã bộ giải mã trong Transformer [5]

Như iitrong iiHình ii1.4, khối bộ mã hoá bản chất là một chuỗi các lớp xếp chồng lên nhau, có nhiệm vụ xử lý dữ liệu đầu vào để cung cấp thông tin cho các lớp trong khối Decoder.

Tiếp theo, trong mỗi bộ mã hoá lại có hai thành phần chính là tầng tập trung đa đầu (multi-head attention) và lớp chuẩn hoá (normalization). Sơ đồ kiến trúc chi tiết của một lớp mã hoá sẽ được trình bày như dưới đây:



Hiirm1.5 Kiến tric chi tiết một lớp của bộ mã hãa trong Transformer [5]

Có thể thấy ở Hình 1.5, sau khi nhận được đầu vào là ma trận biểu diễn của các từ đã được them thông tin ở bước mã hoá vị trí, nơi xử lý đầu tiên của dữ liệu chính là mô đun chú ý đa đầu. Mô đun này bao gồm nhiều lớp tự chú ý để có thể xử lý nhiều mẫu khác nhau trong dữ liệu và là một phần trong cơ chế chú ý.

**CHƯƠNG 2: CÁC PHƯƠNG PHÁP TỐI ƯU DỮ LIỆU ĐẦU VÀO CHO MÔ HÌNH CHATGPT SẴN CÓ**

Có nhiều cách khác nhau để tối ưu và điều chỉnh đầu ra của mô hình GPT. Ở chương hai này, đồ án sẽ tìm hiểu về mô hình chatGPT sẵn có và năm phương pháp tối ưu đầu vào cho mô hình ngôn ngữ lớn GPT tiêu biểu: Thay đổi chiều dài đầu vào, Sử dụng mã hoá từ, Áp dụng mã hoá dạng số, Sử dụng kỹ thuật tạo câu đề, Sử dụng kiến thức bên ngoài.

**2.1 ChatGPT**

***2.1.1 Tổng quan về ChatGPT***

ChatGPT [1] là một ứng dụng của công nghệ Trí tuệ nhân tạo đến từ tổ chức nghiên cứu và phát triển OpenAI. ChatGPT cung cấp khả năng xử lý ngôn ngữ tự nhiên nâng cao, dựa trên mô hình học máy tiên tiến.

Mô hình GPT (Generative Pretraining Transformer – Biến đổi tạo sinh tiền huấn luyện) là một loại mô hình học máy được huấn luyện để tạo ra những đoạn văn bản giống người. GPT sử dụng kiến trúc transformer, có thể xử lý và sinh ra văn bản từ dữ liệu đầu vào, từ đó dẫn đến một loạt các ứng dụng trong NLP (Natural Language Processing - Xử lý ngôn ngữ tự nhiên) như dịch máy, hỗ trợ đối thoại, tóm tắt văn bản, học máy, phân loại văn bản, nhận dạng giọng nói, và nhiều ứng dụng khác.

OpenAI là một tổ chức nghiên cứu AI (Artificial Intelligence - Trí tuệ nhân tạo) phi lợi nhuận được thành lập vào tháng 12 năm 2015. Tổ chức này tập trung vào việc tạo ra trí tuệ nhân tạo mở, nhằm phục vụ lợi ích của tất cả nhân loại. Họ nghiên cứu và phát triển các công nghệ AI mạnh mẽ và cố gắng đảm bảo rằng những công nghệ này được sử dụng một cách an toàn và công bằng. Một trong những mục tiêu của OpenAI là đảm bảo rằng những lợi ích do AI tạo ra không bị tập trung vào một nhóm nhỏ người. Thay vào đó, họ cố gắng tạo điều kiện để mọi người có thể tận hưởng những lợi ích này. Tổ chức này đã tạo ra các mô hình AI nổi tiếng như GPT-2, GPT-3 và GPT-3.5-Turbo, GPT-4. Những mô hình này đã được sử dụng rộng rãi trong nhiều ứng dụng, từ việc tạo ra văn bản tự động, hỗ trợ trò chuyện đến việc dịch ngôn ngữ và nhiều hơn nữa. OpenAI đã phát hành nhiều thế hệ của mô hình GPT bao gồm:

***2.1.2 Các mô hình GPT sẵn có***

**GPT-1**: Đây là phiên bản đầu tiên của GPT, được giới thiệu vào tháng 6 năm 2018. GPT-1 sử dụng kiến trúc transformer và học từ dữ liệu không được gán nhãn. Nó có 117 triệu tham số.

**GPT-2**: Phát hành vào tháng 2 năm 2019, GPT-2 cải thiện đáng kể so với phiên bản đầu tiên bằng cách tăng số lượng tham số lên 1,5 tỷ. OpenAI ban đầu không công bố toàn bộ mô hình vì lo ngại về nguy cơ lạm dụng, nhưng sau đó đã công bố hoàn toàn vào tháng 11 năm 2019.

**GPT-3**: Được giới thiệu vào tháng 6 năm 2020, GPT-3 là một bước nhảy vọt so với GPT-2 với 175 tỷ tham số. Nó đã được sử dụng trong nhiều ứng dụng thực tế như dịch vụ trò chuyện, viết văn bản tự động, và giảng dạy.

**GPT-3.5-Turbo**: Mô hình này được giới thiệu vào cuối năm 2021, mang lại cải tiến đáng kể về chất lượng hội thoại, tạo ra câu trả lời chính xác và hợp lý hơn, và hiểu rõ hơn các yêu cầu phức tạp của người dùng.

**GPT-4**: Được ra mắt năm 2023

*Bảng 2.1 So sánh các phiên bản GPT*

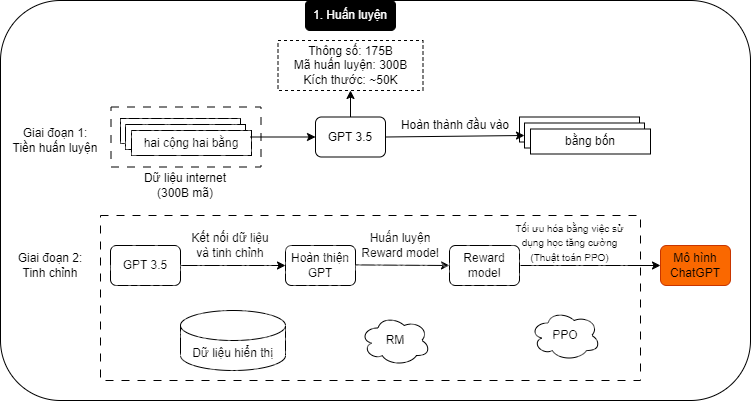
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **GPT** | **GPT-2** | **GPT-3** | **GPT-4** |
| Giới thiệu năm | 2018 | 2019 | 2020 | 2023 |
| Kích thước mô hình | 117 triệu tham số | 1,5 tỷ tham số | 175 tỷ tham số | 1.000 tỷ tham số |
| Đặc điểm | Sử dụng kiến trúc transformer và học từ dữ liệu không được gán nhãn. | Cải thiện đáng kể so với phiên bản đầu tiên | Có khả năng tạo ra các đoạn văn bản tự nhiên và phức tạp | Sử dụng phương pháp học có giám sát cùng với phương pháp học tăng cường Có khả năng xử lý thông tin đầu vào ở cả dạng hình ảnh và văn bản |

***2.1.3 Kiến trúc GPT***

Kiến trúc GPT là một mô hình học máy dựa trên transformer, được huấn luyện trước để sinh ra văn bản tự nhiên. Trong thực tế, mô hình này dự đoán từ tiếp theo trong một chuỗi dựa trên tất cả các từ đã xuất hiện trước đó. Mô hình GPT có khả năng hiểu ngữ cảnh và tạo ra văn bản có cấu trúc và ngữ nghĩa hợp lý.

Kiến trúc mô hình chatGPT gồm có 2 giai đoạn:

Giai đoạn Huấn luyện mô hình (training): ChatGPT được huấn luyện trên một lượng lớn dữ liệu văn bản từ Internet. Quá trình huấn luyện bao gồm việc học cách tạo ra văn bản dựa trên ngữ cảnh đầu vào, và dự đoán từ tiếp theo dựa trên các từ đã biết trước đó. Gồm có hai giai đoạn:



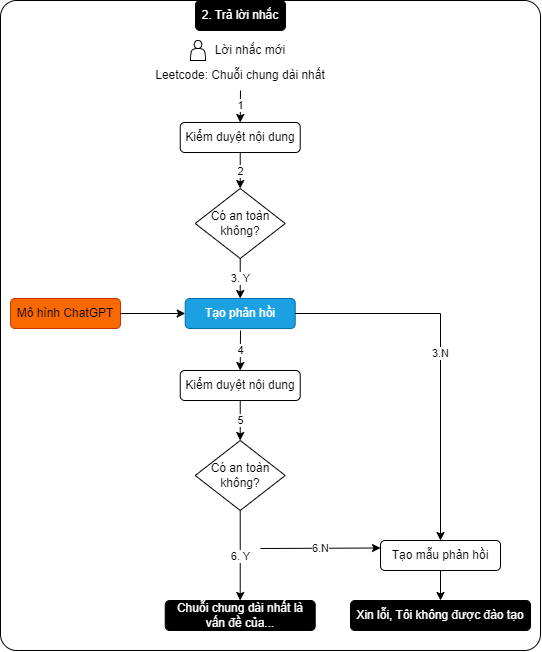
*Hình 2.1. Mô hình huấn luyện GPT*

* + - * Tiền iihuấn iiluyện ii(Pre-Training): iiỞ iigiai iiđoạn iinày, iiGPT iiđược iihuấn iiluyện ii(training) iitrên iimột iitập iidữ iiliệu iilớn. iiMục iitiêu iilà iitạo iira iimột iimô iihình iicó iithể iidự iiđoán iicác iitừ iitrong iitương iilai iiđược iiđưa iira iimột iicâu iitheo iicách iiđúng iingữ iipháp iivà iicó iiý iinghĩa iivề iimặt iingữ iinghĩa iitương iitự iinhư iidữ iiliệu iitrên iiInternet. iiSau iigiai iiđoạn iitiền iihuấn iiluyện, iimô iihình iicó iithể iihoàn iithành iicác iicâu iicho iisẵn iinhưng iikhông iicó iikhả iinăng iitrả iilời iicâu iihỏi.
* Tinh iichỉnh ii(Fine-tuning): iiGiai iiđoạn iinày iilà iiquy iitrình iigồm iiba iibước iibiến iimô iihình iiđược iitạo iitrước iithành iimô iihình iichatGPT iitrả iilời iicâu iihỏi. iiThu iithập iidữ iiliệu iihuấn iiluyện(câu iihỏi iivà iicâu iitrả iilời) iivà iitinh iichỉnh iimô iihình iiđược iihuấn iiluyện iitrước iitrên iidữ iiliệu iinày. iiMô iihình iilấy iimột iicâu iihỏi iilàm iiđầu iivào iivà iihọc iicách iitạo iira iicâu iitrả iilời iitương iitự iinhư iidữ iiliệu iihuấn iiluyện. iiThu iithập iithêm iidữ iiliệu ii(câu iihỏi, iimột iisố iicâu iitrả iilời) iivà iihuấn iiluyện iimô iihình iiphần iithưởng iiđể iixếp iihạng iicác iicâu iitrả iilời iinày iitừ iiphù iihợp iinhất iiđến iiít iiphù iihợp iinhất.

Sử dụng phương pháp học tăng cường (tối ưu hoá PPO – Proximal Policy Optimization) Một thuật toán tối ưu trong lĩnh vực học sâu và học tăng cường. Nó được sử dụng để cải thiện chính sách hành động của một hệ thống tự động học thông qua việc tối ưu mạng nơ-ron dựa trên phương pháp chính sách gần nhau. Đây là một trong những phương pháp phổ biến được sử dụng trong lĩnh vực học tăng cường để đạt được hiệu suất tốt hơn trong việc giải quyết các bài toán tương tác với môi trường để tinh chỉnh mô hình và câu trả lời của mô hình chính xác hơn.

Giai đoạn trả lời lời nhắc (Answer a prompt):

* Bước ii1: iiNgười iidùng iinhập iiđầy iiđủ iicâu iihỏi: ii“Giải iithích iicách iithức iihoạt iiđộng iicủa iithuật iitoán iiphân iiloại”.
* Bước ii2: iiCâu iihỏi iiđược iigửi iiđến iibộ iiphận iikiểm iiduyệt iinội iidung. iiBước iinày iiđảm iibảo iirằng iicâu iihỏi iikhông iivi iiphạm iinguyên iitắc iian iitoàn iivà iilọc iicác iicâu iihỏi iikhông iiphù iihợp.
* Bước ii3: iiNếu iithông iitin iiđầu iivào iivượt iiqua iikiểm iiduyệt iinội iidung, iithì iithông iitin iiđầu iivào iiđó iisẽ iiđược iigửi iitới iimô iihình iichatGPT. iiNếu iiđầu iivào iikhông iivượt iiqua iikiểm iiduyệt iinội iidung iithì iinó iisẽ iichuyển iithẳng iisang iitạo iicâu iitrả iilời iimẫu.
* Bước ii4: iiSau iikhi iimô iihình iitạo iiphản iihồi, iinó iisẽ iiđược iigửi iiđến iithành iiphần iikiểm iiduyệt iinội iidung. iiĐiều iinày iiđảm iibảo iicác iicâu iitrả iilời iiđược tạo ra là an iitoàn, iivô iihại, iikhông iithiên iivị.
* Bước ii5: iiNếu iiđầu iivào iivượt iiqua iibước iikiểm iiduyệt iinội iidung, iinó iisẽ iiđược iihiển iithị iicho iingười iidùng. iiNếu iiđầu iivào iikhông iivượt iiqua iikiểm iiduyệt iinội iidung ii, iinó iisẽ iichuyển iisang iitạo iiphản iihồi iimẫu iivà iihiển iithị iicâu iitrả iilời iimẫu iicho iingười iidùng.



*Hình 2.2. Phân tích giai đoạn trả lời lời nhắc của mô hình chatGPT*

Từ mô hình trên, có thể thấy câu trả lời của ChatGPT sẽ tương đối an toàn và tin cậy. Tuy nhiên hiện tại vẫn còn tồn tại những câu trả lời chưa hợp lý hoặc câu trả lời chưa đúng. Điều này bởi vì chatGPT cần cập nhật dữ liệu liên tục và học tăng cường.

***2.1.4 Ứng dụng của chatGPT***

Hiện nay, với sự hiểu biết sâu rộng do được huấn luyện, Chat GPT đã được công chúng sử dụng rộng rãi trong nhiều lĩnh vực và vấn đề khác nhau trong cuộc sống. Nó có thể được sử dụng để tạo nội dung, giúp trả lời các câu hỏi, hỗ trợ trong công việc soạn thảo văn bản, cung cấp giao diện ngôn ngữ tự nhiên cho các hệ thống phần mềm. Ngoài ra, chatGPT còn có thể giúp trong việc dạy và học, bằng cách cung cấp một nguồn thông tin và trợ giúp tương tác. Cụ thể các ứng dụng của chatGPT:

* Tạo nội dung với chatGPT: chatGPT có khả năng tạo ra văn bản tự nhiên, điều này tạo ra nhiều ứng dụng trong việc tạo nội dung. Mô hình có thể được sử dụng để tạo ra bài viết, bài đánh giá, bài blog và thậm chí là viết sách. Nó cũng có thể được sử dụng để tạo ra nội dung cho trò chơi video, bao gồm cả hội thoại nhân vật và mô tả thế giới.
* Trả lời câu hỏi: chatGPT cũng có thể được sử dụng để trả lời các câu hỏi. Điều này có thể rất hữu ích trong các tình huống như hỗ trợ khách hàng, nơi mô hình có thể giải quyết các vấn đề thông qua việc trả lời câu hỏi phổ biến của khách hàng. Ngoài ra, nó cũng có thể được sử dụng trong các hệ thống câu hỏi và câu trả lời (Q&A) tự động, nơi người dùng có thể đặt câu hỏi và nhận câu trả lời tức thì từ mô hình. Ngoài ra, chatGPT còn có thể giúp giải quyết các vấn đề kỹ thuật thông qua việc trả lời các câu hỏi liên qua đến cách sử dụng một sản phẩm hoặc dịch vụ cụ thể, từ khoa học đến văn học. Tóm lại, chatGPT có thể trở thành một nguồn thông tin quý giá cho bất kỳ người dùng nào cần tìm hiểu về một chủ đề hay một lĩnh vực nào cụ thể.
* Hỗ trợ soạn thảo văn bản: Một ứng dụng khác của chatGPT là trong việc hỗ trợ soạn thảo văn bản. Mô hình có thể đề xuất cách hoàn thiện các câu, đề xuất các từ vựng phù hợp, và thậm chí cung cấp sự chỉnh sửa ngữ pháp. Điều này có thể giúp tăng tốc độ và chất lượng của quá trình soạn thảo văn bản. Khi sử dụng chatGPT để hỗ trợ soạn thảo văn bản, người dùng có thể tận dụng khả năng của mô hình để tạo ra văn bản tự nhiên và hợp ngữ cảnh. Điều này có thể giúp cải thiện chất lượng của văn bản, từ cải thiện ngữ pháp và từ vựng đến cải thiện cấu trúc và luồng của văn bản.
* Giao iidiện iingôn iingữ iitự iinhiên: iiChatGPT iicó iithể iicung iicấp iimột iigiao iidiện iingôn iingữ iitự iinhiên iicho iicác iihệ iithống iiphần iimềm. iiĐiều iinày iicho iiphép iingười iidùng iitương iitác iivới iihệ iithống iithông iiqua iingôn iingữ iitự iinhiên, iithay iivì iiphải iisử iidụng iicác iilệnh iicụ iithể. iiVí iidụ, iimột iingười iidùng iicó iithể iihỏi iimột iitrợ iilý iiảo iidựa iitrên iiChatGPT iivề iithời iitiết, iithay iivì iiphải iigõ iivào iimột iilệnh iicụ iithể. iiKhi iisử iidụng iiChatGPT iinhư iimột iigiao iidiện iingôn iingữ iitự iinhiên, iingười iidùng iicó iithể iitương iitác iivới iicác iihệ iithống iiphần iimềm iimột iicách iitự iinhiên iivà iitrực iiquan iihơn. iiĐiều iinày iikhông iichỉ iigiúp iigiảm iibớt iisự iiphức iitạp iicủa iiviệc iisử iidụng iicác iihệ iithống iiphần iimềm, iimà iicòn iigiúp iingười iidùng iitiết iikiệm iithời iigian iivà iinăng iilượng. iiNgoài iira, iiviệc iisử iidụng iingôn iingữ iitự iinhiên iicũng iicó iithể iigiúp iităng iicường iikhả iinăng iitiếp iicận iicủa iihệ iithống, iicho iiphép iinhiều iingười iidùng iihơn iicó iithể iisử iidụng iihệ iithống iimà iikhông iicần iiphải iiđược iiđào iitạo iiđặc iibiệt.
* Dạy iivà iiHọc: iiCuối iicùng, iiChatGPT iicũng iicó iithể iiđược iisử iidụng iitrong iiviệc iidạy iivà iihọc. iiMô iihình iicó iithể iicung iicấp iithông iitin iivề iimột iiloạt iicác iichủ iiđề, iitừ iikhoa iihọc iiđến iivăn iihọc, iigiúp iihọc iisinh iivà iigiáo iiviên iitìm iihiểu iithông iitin iimột iicách iihiệu iiquả. iiNó iicũng iicó iithể iigiúp iitrong iiviệc iitạo iira iicác iibài iigiảng iivà iitài iiliệu iihọc iitập. iiCuối iicùng, iiChatGPT iicó iithể iigiúp iicải iithiện iiquá iitrình iidạy iivà iihọc iibằng iicách iicung iicấp iimột iinguồn iithông iitin iiđáng iitin iicậy iivà iidễ iitiếp iicận. iiMô iihình iicó iithể iicung iicấp iithông iitin iivề iimột iiloạt iicác iichủ iiđề, iitừ iikhoa iihọc iiđến iinghệ iithuật, iigiúp iihọc iisinh iivà iigiáo iiviên iimở iirộng iikiến iithức iicủa iimình. iiNgoài iira, iinó iicũng iicó iithể iigiúp iitạo iira iicác iitài iiliệu iihọc iitập iivà iibài iigiảng, iigiúp iigiáo iiviên iitiết iikiệm iithời iigian iivà iinăng iilượng.

**2.2 Phương pháp**

***2.2.1 Sinh mô tả***

Bài toán tối ưu đầu vào có thể được mô hình hoá thành những nhiệm vụ sinh ra những văn bản đầu ra dựa trên những nội dung thể hiển có trong văn bản đầu vào. Bới những văn bản đầu vào sẽ chứa một số thông tin về vấn đề, ví dụ như vấn đề về một chiếc xe ô tô gồm có nhãn hiểu hãng, số lượng ghế ngồi, màu sắc xe,…Do đó, trong nội dung ở phần này, đồ án tập trung chủ yếu vào việc nghiên cứu giải pháp để có thể đáp ứng yêu cầu trên bằng cách ứng dụng GPT.

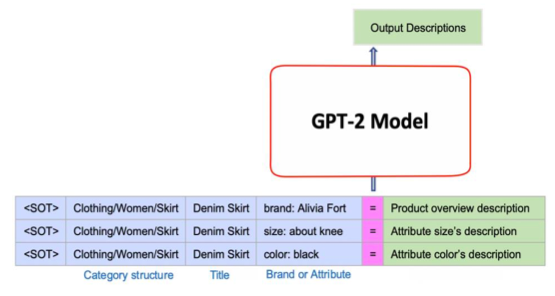
Để tinh chỉnh mô hình GPT đáp ứng nhiệm vụ tạo những mô tả có ý nghĩa dựa trên tiêu đề, đồ án tạo điều kiện hoá dữ liệu đầu vào với định dạng như sau:

(3)

Trong iiđó, ii<SOT> iivà ii<EOT> iilà iicặp iitoken, iiđại iidiện iicho iisự iibắt iiđầu iivà iikết iithúc iicủa iimột iivăn iibản. Khi nhận được đầu vào “<SOT> title =”, mô hình GPT có nhiệm vụ lấy mẫu với các tham số Top-k và Top-p (quá trình lấy mẫu) để tạo ra văn bản trôi chảy hơn và tránh thoái hoá văn bản. Đồng hành với việc sử dụng tiêu đề, phương pháp cũng sử dụng thêm các thông tin như cấu trúc danh mục, thương hiệu, thuộc tính sản phẩm. Chính vì thế mà định đạng của dữ liệu huấn luyến sẽ được cấu trúc lại như dưới đây:

Vậy nhưng vẫn tồn tại một điểm trừ khi dùng cách tiếp cận này, mô tả được sinh ra sẽ dài hơn và khó có thể bao quát thông tin đầu vào. Do tạo ra những văn bản dài vẫn tồn tại nhiều khó khăn, văn bản đầu ra sẽ có xu hướng không mạch lạc hoặc nội dung bị trùng lặp quá nhiều, phương pháp sẽ tiếp tục chia các tác vụ tạo mô tả sản phẩm thành hai nhiệm vụ nhỏ hơn, gồm có sinh mô tả tổng quan cho sản phẩm và sinh mô tả cho các thuộc tính riêng biệt của chúng. Tiếp đó, các mô tả này được kết hợp lại để có được kết quả cuối cùng. Từ mô tả tổng quan, dữ liệu huấn luyện có cấu trúc sau:

Với iimô iitả iithuộc iitính, iidữ iiliệu iihuấn iiluyện iicó iicấu iitrúc: ii



*Hình 2.3. Định dạng dữ liệu huấn luyện trong phương pháp đề xuất*

Hình2.3 là một vài ví dụ trong dữ liệu huấn luyện của phương pháp mới. Khi sử dụng huấn luyện theo hướng này, mô hình sẽ nhận được nhiều lợi ích hơn. Đầu tiên, tốc độ suy diển sẽ tang lên nhiều hơn do mô hình có thể tạo ra các mô tả tổng quan và thuộc tính sản phẩm đồng thời. Để gia tang hiệu suất tổng thể của ứng dụng có thể tổ chức nhiều mô hình chạy cùng lúc. Thứ hai, những thông tin quan trọng của dữ liệu đầu vào là kết quả tổng hợp cuối cùng. Thứ ba, công dụng của giải pháp này là có thể tạo ra các mô tả có số lượng thuộc tính có khả năng thay đổi tuỳ thuộc vào nhu cầu mà không hề làm giảm chất lượng kết quả tổng thể. Mô tả khi không gặp phải vấn đề phụ thuộc xa đã trở nên chính xác và trôi chảy hơn [6]. Ii

Mặc iidù iitác iivụ iinày iiđược iihưởng iilợi iitừ iitri iithức iiđược iichuyển iigiao iitừ iimô iihình iiGPT iiđã iiđược iihuấn iiluyện, iichất iilượng iicủa iinó iivẫn iikhông iinhư iimong iiđợi iivì iivấn iiđề iithiếu iithốn iidữ iiliệu iivà iisự iikhác iibiệt iigiữa iimiền iidữ iiliệu iihuấn iiluyện iicủa iiGPT iivà iimiền iidữ iiliệu iicủa iibài iitoán iisinh iimô iitả. iiTrong iinhững iiphần iitiếp iitheo iicủa iichương, iiđồ iián iisẽ iitrình iibày iimột iisố iiphương iipháp iiđể iigiải iiquyết ii2 iivấn iiđề iinày. ii

***2.2.2 Tokenizer tuỳ chỉnh***

Tokenizers chia nhỏ văn bản thành các "token" mà GPT sử dụng để học và sinh ra văn bản. Sử dụng tokenizer tùy chỉnh có thể giúp tối ưu hóa các mô hình GPT hiểu và xử lý văn bản đầu vào. Kỹ thuật sử dụng tokenizer tùy chỉnh cho phép bạn kiểm soát các văn bản được chia nhỏ và diễn dịch, điều này có thể hữu ích trong một số trường hợp. Tokenizer là một công cụ quan trọng trong xử lý ngôn ngữ tự nhiên (NLP), được sử dụng để chuyển đổi văn bản từ ngôn ngữ tự nhiên sang dạng đầu vào có thể hiểu được của mô hình học máy.

Trong ngữ cảnh của GPT và nhiều mô hình ngôn ngữ khác, tokenizer chính là cầu nối giữa văn bản người dùng và đầu vào mô hình. Tokenizer sẽ chia văn bản thành các "Token", thường là từ, nhưng cũng có thể là ký tự, âm tiết, hoặc các đơn vị ngôn ngữ khác tùy thuộc vào ngôn ngữ và ứng dụng cụ thể. Các token này sau đó sẽ được chuyển đổi thành các số nguyên thông qua một từ điển, tạo ra một chuỗi số nguyên phù hợp với đầu vào của mô hình.

Mặc dù có các tokenizer tiêu chuẩn có thể sử dụng, trong một số trường hợp, chúng ta có thể muốn tạo một tokenizer tùy chỉnh để tối ưu hóa cho nhiệm vụ hoặc dữ liệu cụ thể. Ví dụ, nếu chúng ta đang làm việc với một loại văn bản cụ thể mà có cấu trúc đặc biệt hoặc sử dụng từ vựng độc đáo, tokenizer tiêu chuẩn có thể không tách văn bản một cách hiệu quả. Trong trường hợp này, chúng ta có thể muốn tạo một tokenizer tùy chỉnh để xử lý văn bản một cách chính xác hơn.

Để tạo tokenizer tùy chỉnh, chúng ta sẽ cần xác định một tập hợp các quy tắc để chia văn bản thành token. Quá trình này có thể bao gồm việc tạo một từ điển tùy chỉnh, xác định các quy tắc cụ thể cho cách tách từ, và nhiều hơn nữa. Khi đã xác định các quy tắc, chúng ta có thể lập trình tokenizer của mình để thực hiện chúng.

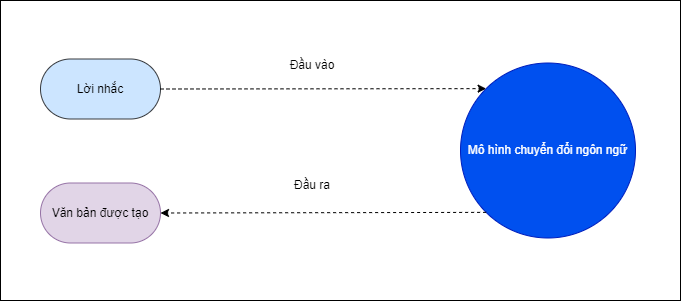
Sau khi có tokenizer tùy chỉnh, chúng ta sẽ cần phải huấn luyện mô hình sử dụng tokenizer này. Điều này có nghĩa là chúng ta sẽ cần phải chuyển đổi văn bản đầu vào thành chuỗi số nguyên sử dụng tokenizer mới, rồi đưa chuỗi số này vào mô hình. Kết quả là, mô hình sẽ học cách hiểu và sinh ra văn bản dựa trên cách chúng ta đã tách nó thành token.

Dưới đây là một số ví dụ: Xử lý ngôn ngữ tự nhiên (NLP) cho ngôn ngữ cụ thể: Nếu đang làm việc với một ngôn ngữ có cấu trúc hoặc quy tắc chính tả độc đáo, chúng ta có thể muốn tạo ra một tokenizer tùy chỉnh để chắc chắn rằng mô hình đang xử lý văn bản một cách chính xác. Ví dụ, trong tiếng Việt, từ có thể bao gồm nhiều âm tiết được ngăn cách bằng dấu cách. Một tokenizer tùy chỉnh có thể được thiết kế để nhận biết và giữ nguyên những từ nguyên thể này; Xử lý văn bản với cú pháp đặc biệt: Trong một số ứng dụng, như lập trình hoặc phân tích dữ liệu log, văn bản có thể chứa các ký tự hoặc cú pháp đặc biệt. Tokenizer tùy chỉnh có thể được thiết kế để xử lý đúng những đặc điểm này. Ví dụ, trong mã nguồn lập trình, dấu ngoặc đơn, dấu ngoặc kép, và các ký tự đặc biệt khác đều có ý nghĩa quan trọng và cần được giữ nguyên; Xử lý dữ liệu không phải văn bản: Mặc dù GPT thường được sử dụng cho văn bản, nhưng nó cũng có thể được sử dụng cho các loại dữ liệu khác như âm thanh, hình ảnh, hoặc dữ liệu chuỗi thời gian. Tokenizer tùy chỉnh có thể được tạo ra để chuyển đổi các loại dữ liệu này thành các "token" mà mô hình có thể xử lý.

***2.2.3 Kỹ thuật tạo lời nhắc (Prompt Engineering)***

Tạo ra và tinh chỉnh câu đề (prompt) để cung cấp hướng dẫn cho mô hình về yêu cầu và ngữ cảnh của nhiệm vụ. Câu đề có thể chứa thông tin về định dạng đầu ra, câu hỏi hoặc yêu cầu cụ thể mà mô hình cần phải đáp ứng.

Prompt Engineering[7] liên quan đến việc tạo ra và tối ưu hóa lời nhắc (prompts) đầu vào cho mô hình. Cách lời nhắc được cấu trúc và những từ ngữ được sử dụng trong đó có thể ảnh hưởng rất lớn đến những gì mô hình tạo ra. Đây là một môn khoa học thực nghiệm và hiệu quả của các phương pháp thiết kế lời nhắc có thể khác nhau tùy thuộc vào mô hình, do đó đòi hỏi nhiều thử nghiệm và kinh nghiệm. Một cách dễ hình dung, chúng ta coi ChatGPT như những máy nén thông tin, tri thức nhân loại diễn đạt trên Internet sẽ được tổ chức dưới dạng tham số của mạng Nơron. Để sử dụng những tri thức này, ta cần cung cấp cho nó văn cảnh và yêu cầu nó trả lời theo một định dạng nào đấy (giải nén có điều kiện). Kỹ thuật gợi ý (Prompt Engineering) chính là chìa khóa cho vấn đề giải nén này.



*Hình 2.4 Kỹ thuật gợi ý Prompt Engineering*

Dưới đây là cách sử dụng kỹ thuật gợi ý này cho mô hình GPT trong thực tế:

- Tạo nội dung: Nếu muốn mô hình GPT tạo ra một đoạn văn bản theo một phong cảnh nhất định, chúng ta có thể tạo ra lời nhắc bắt đầu bằng "Hãy viết một đoạn văn với phong cách của Edgar Allan Poe...". Lời nhắc này cho mô hình một hướng rõ ràng về cách nó nên tạo ra văn bản;

- Điều hướng cuộc hội thoại: Nếu đang sử dụng GPT để xây dựng một hệ thống chatbot, chúng ta có thể sử dụng kỹ thuật gợi ý (prompt engineering) để điều hướng cuộc hội thoại. Ví dụ, nếu người dùng hỏi "Bạn có thể giúp đồ án tìm kiếm thông tin về thời tiết không?", chúng ta có thể gửi lời nhắc cho GPT như "Hãy trả lời câu hỏi về thời tiết..." để tạo ra đầu ra mong muốn;

***2.2.4 Sử dụng ngữ cảnh***

Các mô hình ngôn ngữ quy mô lớn cho thấy khả năng tạo văn bản đầy hứa hẹn, nhưng người dùng không thể dễ dàng kiểm soát các khía cạnh cụ thể của văn bản được tạo. Một mô hình ngôn ngữ biến áp có điều kiện gồm 1,63 tỷ tham số, được đào tạo để tạo điều kiện cho các mã kiểm soát chi phối phong cách, nội dung và hành vi theo nhiệm vụ cụ thể. Các mã điều khiển được lấy từ cấu trúc tự nhiên xảy ra cùng với văn bản thô, duy trì các lợi thế của việc học không giám sát trong khi cung cấp khả năng kiểm soát rõ ràng hơn đối với việc tạo văn bản. GPT dựa trên ngữ cảnh của văn bản đầu vào để quyết định từ tiếp theo, vì vậy việc cung cấp ngữ cảnh rõ ràng và đầy đủ trong lời nhắc có thể giúp tạo ra đầu ra chính xác hơn. Ngữ cảnh (context) là một phần quan trọng của việc hiểu và sinh ra văn bản. Mô hình ngôn ngữ như GPT được huấn luyện để dự đoán token tiếp theo dựa trên tất cả các Token đã qua. Điều này có nghĩa là chúng có thể sử dụng thông tin ngữ cảnh từ văn bản để tạo ra dự đoán của mình. Điều này đã được khẳng định qua bài báo “CTRL: A Conditional Transformer Language Model For Controllable Generation” của Nitish Shirish Keskar∗, Bryan McCann∗, Lav R. Varshney, Caiming Xiong, Richard Socher [8].

Cụ iithể iihơn, iiGPT iisử iidụng iikiến iitrúc iiTransformer, iinơi iimà iimỗi iiToken iitrong iivăn iibản iiđầu iivào iiđược iixử iilý iicùng iilúc iivà iitrong iimối iiquan iihệ iivới iinhững iiToken iikhác. iiĐiều iinày iicho iiphép iimô iihình iinắm iibắt iiđược iinhững iimối iiliên iiquan iiphức iitạp iigiữa iicác iitừ iivà iicụm iitừ iitrong iivăn iibản. iiCách iithức iinày iikhá iihiệu iiquả iikhi iidự iiđoán iiToken iitiếp iitheo iitrong iimột iicâu, iihoặc iikhi iisinh iira iivăn iibản iitự iido, iinhưng iinó iicũng iicó iithể iigặp iikhó iikhăn iikhi iiđối iimặt iivới iicác iivấn iiđề iiphức iitạp iihơn. iiTrong iimột iisố iitrường iihợp, iichúng iita iicó iithể iimuốn iicung iicấp iithêm iingữ iicảnh iicho iimô iihình iiđể iigiúp iinó iihiểu iitốt iihơn iitác iivụ iiđang iithực iihiện.

Ví iidụ, iinếu iiđang iisử iidụng iiGPT iiđể iitrả iilời iicâu iihỏi, iichúng iita iicó iithể iimuốn iicung iicấp iimột iisố iithông iitin iingữ iicảnh iiliên iiquan iiđến iicâu iihỏi. iiĐiều iinày iicó iithể iibao iigồm iivăn iibản iicủa iicâu iihỏi, iicâu iitrả iilời iitrước iiđó, iihoặc iibất iikỳ iithông iitin iinào iikhác iimà iichúng iita iinghĩ iirằng iisẽ iigiúp iimô iihình iihiểu iirõ iihơn iivấn iiđề. iiChúng iita iicó iithể iithực iihiện iiđiều iinày iibằng iicách iithêm iithông iitin iingữ iicảnh iicho iivăn iibản iiđầu iivào iimà iichúng iita iiđưa iicho iimô iihình. iiTuy iinhiên, iimặc iidù iimô iihình iinhư iiGPT iicó iikhả iinăng iixử iilý iithông iitin iingữ iicảnh, iinhưng iichúng iicó iigiới iihạn iivề iiđộ iidài iitối iiđa iicủa iivăn iibản iiđầu iivào. iiDo iiđó, iichúng iita iicần iiphải iilựa iichọn iicẩn iithận iivề iithông iitin iingữ iicảnh iinào iicần iiđược iibao iigồm.

Dưới đây là một số ví dụ về cách sử dụng ngữ cảnh trong thực tế:

* Hệ thống hỗ trợ khách hàng: Trong một hệ thống hỗ trợ khách hàng, ngữ cảnh có thể bao gồm thông tin về người dùng, lịch sử tương tác của họ, hoặc các câu hỏi hoặc bình luận gần đây. Ví dụ, nếu người dùng đã yêu cầu giúp đỡ với một vấn đề cụ thể trước đó, mô hình GPT có thể sử dụng thông tin này để cung cấp câu trả lời chính xác hơn hoặc đề xuất các bước tiếp theo hữu ích;
  + Phân loại văn bản: Khi phân loại văn bản, ngữ cảnh có thể bao gồm loại văn bản (ví dụ, đánh giá, tin tức, tweet) và nội dung cụ thể của văn bản. Mô hình GPT có thể sử dụng ngữ cảnh này để đưa ra quyết định phân loại chính xác hơn;
  + Tạo nội dung: Khi tạo ra nội dung, như viết bài viết blog hoặc tạo một câu chuyện, ngữ cảnh có thể bao gồm phong cách viết mong muốn, chủ đề của nội dung, hoặc bất kỳ hướng dẫn cụ thể nào từ người dùng. Mô hình GPT có thể sử dụng ngữ cảnh này để tạo ra nội dung phù hợp với yêu cầu.

**CHƯƠNG 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ**

Chương iinàytrình iibày iivề iikết iiquả iithực iinghiệm iivà iiđánh iigiá iiưu iiđiểm iivà iihạn iichế iicủa iicác iiphương iipháp iitối iiưu iiđầu iivào iicho iimô iihình iiGPT, iicũng iinhư iiđề iixuất iihướng iiphát iitriển iivà iinghiên iicứu iitương iilai. iiCác iinội iidung iichính iicủa iichương iibao iigồm:

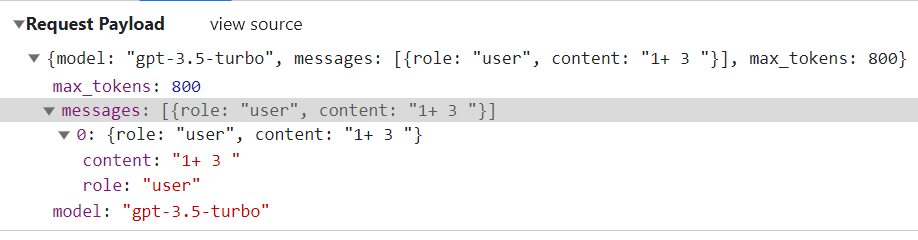
* + Xây iidựng iithực iinghiệm
  + Kết iiquả iithực iinghiệm: iiđồ iián iitrình iibày iikết iiquả iicủa iicác iithí iinghiệm iiđã iithực iihiện iiđể iiđánh iigiá iihiệu iisuất iicủa iicác iiphương iipháp iitối iiưu iiđầu iivào iicho iimô iihình iiGPT. iiCác iikết iiquả iinày iicho iithấy iirằng iicác iiphương iipháp iinày iiđều iigiúp iicải iithiện iiđáng iikể iihiệu iisuất iicủa iimô iihình.
  + Đánh giá ưu điểm và hạn chế của phương pháp tối ưu đầu vào: đồ án đánh giá các ưu điểm và hạn chế của các phương pháp tối ưu đầu vào cho mô hình GPT. Các ưu điểm bao gồm cải thiện đáng kể hiệu suất của mô hình, trong khi các hạn chế có thể là tốn nhiều thời gian và tài nguyên tính toán.

**3.1. Thực nghiệm**

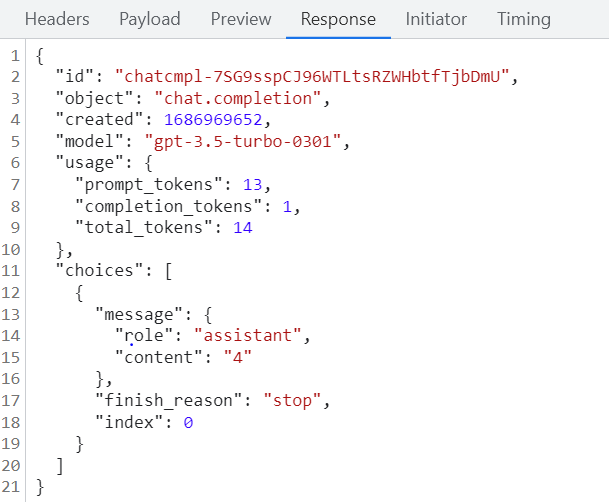
***3.1.1 Công cụ sử dụng để xây dựng thực nghiệm***

Để đánh giá hiệu suất của các phương pháp tối ưu đầu vào cho mô hình GPT, đồ án đã sử dụng các công cụ dưới đây:

* Ngôn ngữ: Javascript
* Javascript Framework: Vue js
* Css Framework: Tailwind CSS
* Tool: Visual Studio Code
* Version Control: Git, Github
* Window 10, RAM 8G
* API openAI



*Hình 3.1 Kết quả test API Chat GPT*



*Hình 3.2 Kết quả test API thành công*

Về phía Frontend (Giao diện), được chia làm hai phần chính là phần Sidebar và phần Chat. Ở phần Sidebar: bao gồm button “New Chat” và lịch sử đã chat.

Phần Chat: Bao gồm một màn hình hiển thị, một input nhập câu hỏi và một button “Send”. Để có giao diện như trên, đồ án đã tạo ra thư mục là “components” chứa hai file “Sidebar.vue” và “Chat.vue”. Để hiển thị được hai conponents đã tạo, tại mục “views” trong file “Homeview.vue”, import trong phần <script></script> ra components tên “Sidebar” và “Chat” qua đường dẫn tới file. Sau đó, export tên của hai components đã khởi tạo.

**

*Hình 3.3 Hình ảnh giao diện frontend thể hiện qua components*

Kết hợp sử dụng với tailwind CSS, về lựa chọn màu, căn chỉnh, hoàn thành giao diện gồm có hai phần chính. Phần thứ nhất là Sidebar: Gồm có một nút “New Chat” và Lịch sử chat được lưu trong ô chữ nhật xanh, ví dụ “Chat\_d2399|X Delete”. Nếu người dùng muốn tạo cửa sổ Chat mới sẽ bấm đúp chuột trái vào “New Chat” và nếu người dùng muốn xem lại những cửa sổ đã chat thì lựa chọn phần lịch sử chat là “Chat\_d2399|X Delete”; Phần thứ hai là Chat: Gồm có một input “Nhập câu hỏi” để người dùng nhập câu hỏi và gửi câu hỏi đi. Sau khi gửi câu hỏi đi, chờ một lát và câu trả lời cho câu hỏi vừa nhập sẽ được hiển thị trong màn hình phần này.



*Hình 3.4 Giao diện website thử nghiệm*

Về phía đọc API từ Backend (Chức năng hỗ trợ của một trang web) được xử lí trong hai file Chat và Sidebar.

Code trong file Chat.vue:

* + name: "Chat",: Định nghĩa tên của component.
  + Data (): Là một hàm trả về đối tượng chứa dữ liệu ban đầu của component. Trong trường hợp này, nó chứa các tin nhắn (messages), tin nhắn mới (newMessage), và chatId.
  + Mounted (): Đây là một hàm Vue lifecycle hook, được gọi sau khi component đã được gắn kết vào DOM. Ở đây, nó gọi hàm loadChat () khi component được tải.
  + Methods: Là một đối tượng chứa các phương thức (methods) mà component sử dụng. Trong đó có:
  + loadChat (): Hàm này lấy chatId từ route parameters, lấy các câu hỏi từ localStorage và gán chúng cho messages. Nếu không có câu hỏi nào, nó sẽ chuyển hướng người dùng về trang chủ.
  + createChatIfNotExist(chatId): Hàm này kiểm tra xem có cuộc trò chuyện nào tồn tại trong localStorage không. Nếu không, nó sẽ tạo một cuộc trò chuyện mới.
  + sendMessage (): Hàm này xử lý việc gửi tin nhắn. Nếu tin nhắn mới không rỗng, nó sẽ thêm tin nhắn vào danh sách messages và lưu chúng vào localStorage. Sau đó, nó sẽ gửi tin nhắn đến OpenAI để nhận phản hồi.
  + openAiCompletion(prompt): Hàm này gọi API của OpenAI để lấy phản hồi từ model gpt-3.5-turbo. Phản hồi sẽ được thêm vào danh sách messages và lưu vào localStorage.

getMessageStyle (message, index): Hàm này xác định kiểu dáng của tin nhắn dựa trên chỉ số của nó. Nếu chỉ số là số chẵn, nó sẽ trả về một style, nếu không, sẽ trả về style khác.

formatText(content): Hàm này chia nội dung thành các câu dựa trên dấu chấm và định dạng lại chúng nếu cần.

* + import {openai} from "../open-ai";: Dòng này import mô-đun openai từ tệp open-ai.js nằm trong cùng một thư mục với tệp hiện tại.

Code trong file Sidebar.vue:name: "Sidebar",: Đây là tên của component.

* + Data (): Hàm này trả về đối tượng chứa dữ liệu khởi tạo cho component. Ở đây, nó chứa một mảng topics.
  + Mounted (): Đây là một lifecycle hook của Vue, được gọi sau khi component được gắn vào DOM. Ở đây, nó đặt topics là các khóa (keys) của localStorage.

methods: Đây là đối tượng chứa các phương thức mà component sử dụng.

* + onRedirect(path): Hàm này chuyển hướng trình duyệt đến đường dẫn được truyền vào.
  + onDeleteChat(chatId): Hàm này xóa một cuộc trò chuyện từ localStorage dựa trên chatId và chuyển hướng người dùng về trang chủ.

***3.1.2 Xây dựng thực nghiệm***

Dưới đây là phần thiết kế thí nghiệm thứ nhất ( Demo 1): Nhập input về một phép tính đơn giản: 1,000,000,000\*6,000,000=? Sau khi nhập input và chờ phản hồi của chatGPT, nhận được kết quả là 6,000,000,000,000. Qua đó đánh giá, đây là một kết quả sai. Nhưng lỗi này để khắc phục cũng rất dễ dàng hơn bằng việc chúng ta thêm vào lời nhắc “Make sure to put the right amount of zeros, even if there are many:”. Từ input này, chúng ta nhận được phản hồi là một câu trả lời đúng từ chatGPT:”6,000,000,000,000,000.

Dưới đây là phần thiết kế thí nghiệm thứ hai (Demo 2): Bài toán về đưa ra một liệu trình ăn kiêng cho người dùng. Ở phần demo này, ta sẽ sử dụng phương pháp ngữ cảnh theo các cấp độ. Ở cấp độ 1, câu hỏi đưa ra hoàn toàn không nhắc tới ngữ cảnh. Sau đó hãy xem kết quả về độ chính xác mà kết quả của chatGPT đưa ra khi người dùng nhập một input không có ngữ cảnh. Đó chính là một câu trả lời chung chung, chưa chi tiết cụ thể. Từ câu trả lời trên chắc chắn rằng người dùng vẫn sẽ mơ hồ về câu trả lời này. Vậy chúng ta sẽ tiếp tục demo với việc thêm một chút ngữ cảnh vào input. Từ câu trả lời của việc thêm ngữ cảnh, ta có thể đánh giá được rằng câu trả lời (output) từ chatGPT đã cụ thể hơn. chatGPT đã đưa ra một câu trả lời có kế hoạch hơn cho người dùng bằng việc tập thể dục, ăn uống lành mạnh, ngủ đủ giấc,…Vậy nhưng đây thực sự vẫn chưa phải là một câu trả lời khiến người dùng ưng ý. Vậy để đạt được điều này, ta sẽ cụ thể hơn ở ví dụ tiếp theo, một cấp độ cao hơn từ việc đưa ngữ cảnh vào output. Người dùng sẽ nhập rõ ràng hơn về thể trạng, quá trình tập luyện và ăn kiêng và từ đó mong muốn chatGPT có thể đưa ra cho bản thân một chế độ ăn kiêng phù hợp hơn. Câu trả lời bây giờ đã cụ thể hơn về kế hoạch cho các tuần. Người dùng cũng có thể hài long với câu trả lời này, nhưng chúng ta vẫn còn có thể đưa ra một câu trả lời tốt hơn như thế nhờ tới việc thêm ngữ cảnh vào lời nhắc (input). Hãy tiếp tục với ví dụ về cấp độ cao nhất của việc sử dụng ngữ cảnh để chatGPT đưa ra một câu trả lời hoàn hảo, đáp ứng được đầy đủ yêu cầu của người dùng.

Tiếp theo bạn có thể đưa vào những tài liệu mà bạn tìm hiểu được. Ở đây người dùng đã đưa vào một bài nghiên cứu về chế độ tập luyện và chế độ ăn kiêng phù hợp với người đang giảm cân từ một bệnh viện nổi tiếng là Vinmec. Ở cuối input này, người dùng đã muốn chatGPT xác nhận là đã đọc nội dung trên và sau đó người dùng nhận được câu trả từ chatGPT là đã đọc.

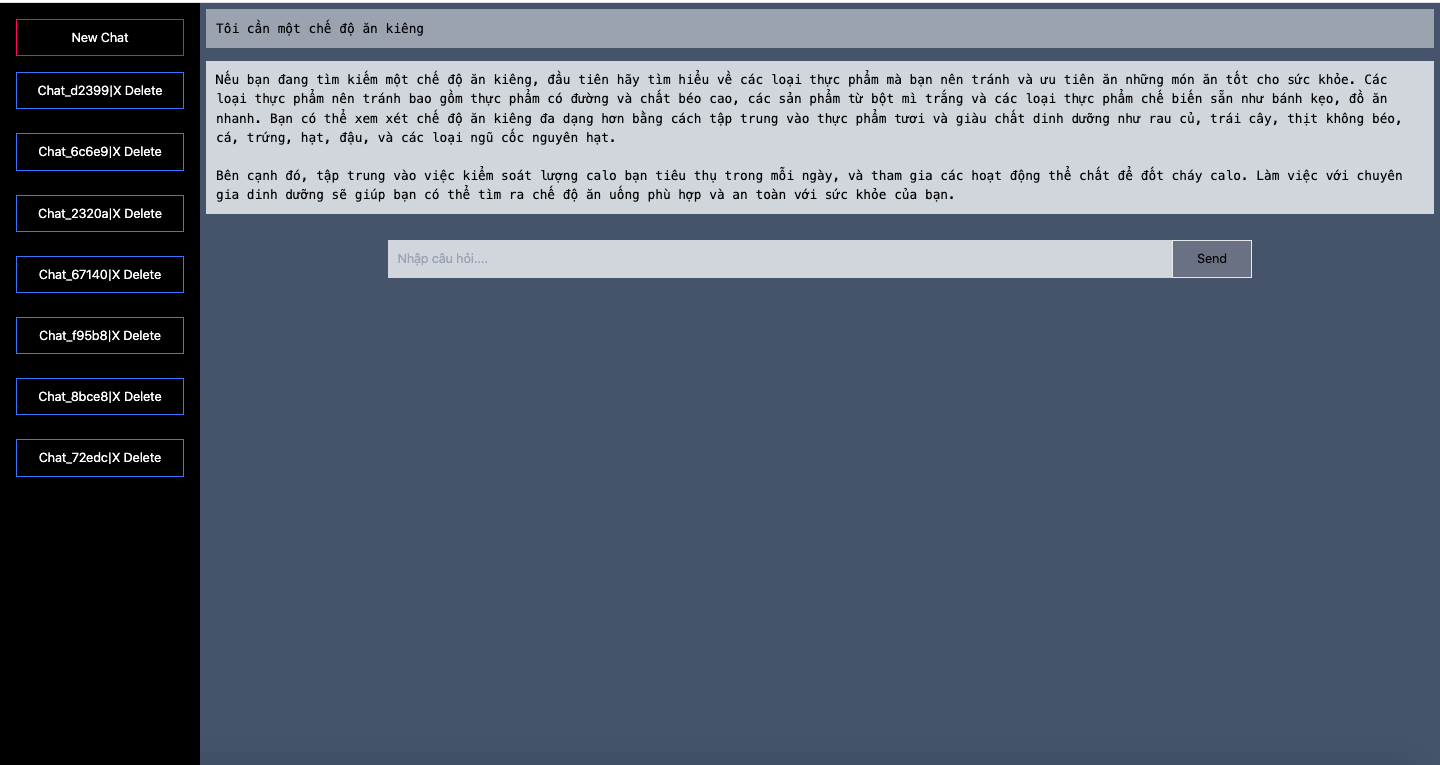
Dưới đây là phần thiết kế thí nghiệm thứ ba (Demo 3): Giả sử chúng ta đang làm về thiết kế nội dung và đang cần nội dung cho bài về quảng cáo trực tuyến, so với với hỏi bình thường thì mình sẽ gán cho chatGPT là một chuyên gia viết tiêu đề về lĩnh vực quảng cáo trực tuyến. Từ kết quả ta sẽ thấy được tính hiệu quả khi một đầu vào được tối ưu hoá.

***3.1.3 Kết quả thực nghiệm***

Dưới đây là kết quả của thí nghiệm 1: Sử dụng phương pháp tối ưu hoá đầu vào Prompt Engineering.

|  |
| --- |
| *Hình 3.5 Kết quả tối ưu đầu vào bằng phương pháp sử dụng Prompt Engineering của thí nghiệm thứ nhất*   * Như ở Hình 3.5, người dùng nhập input về một phép tính đơn giản: 1,000,000,000\*6,000,000=? Sau khi nhập input và chờ phản hồi của chatGPT, nhận được kết quả là 6,000,000,000,000. Qua đó đánh giá, đây là một kết quả sai. Nhưng lỗi này để khắc phục cũng dễ dàng hơn bằng việc ta thêm vào lời nhắc “Make sure to put the right amount of zeros, even if there are many:”. Từ input này, chúng ta nhận được phản hồi là một câu trả lời đúng từ chatGPT:”6,000,000,000,000,000. |

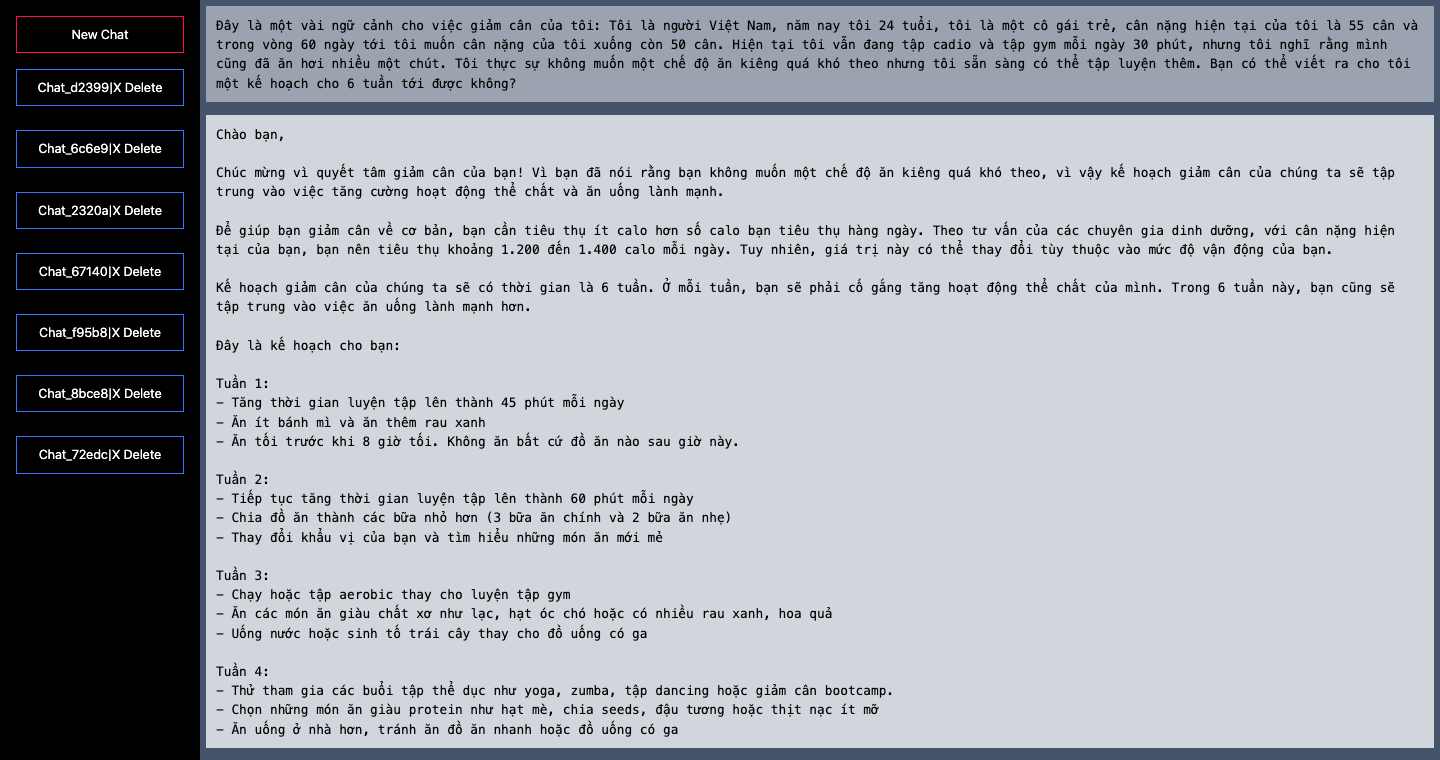
Dưới đây là kết quả của thí nghiệm 2: Dùng phương pháp sử dụng ngữ cảnh để tối ưu hoá đầu vào cho mô hình GPT



*Hình 3.6 Kết quả tối ưu đầu vào lần thứ nhất bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ hai lần thứ nhất*



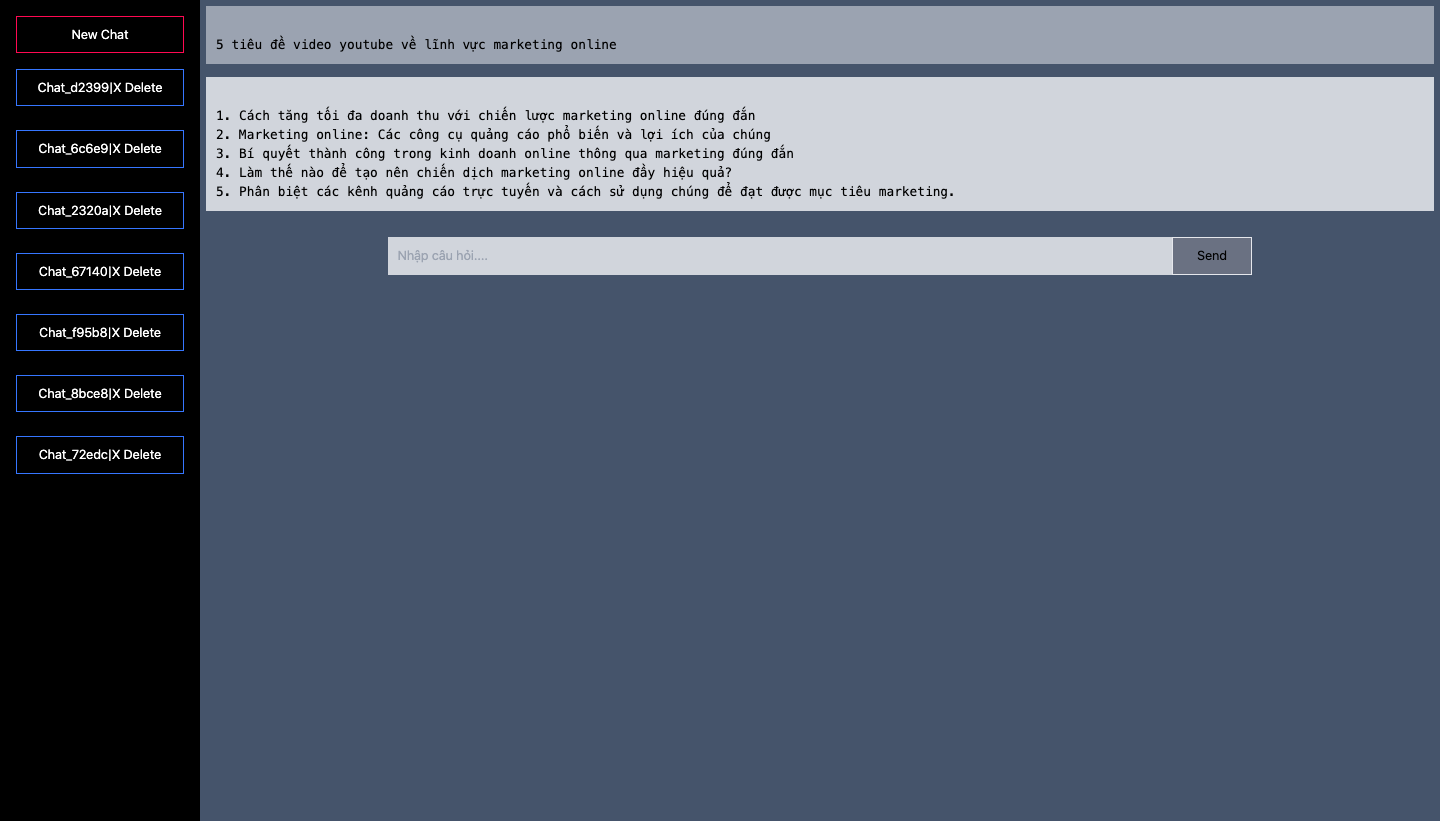
*Hình 3.7 Kết quả tối ưu đầu vào lần thứ hai bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ hai*



*Hình 3.8 Kết quả tối ưu đầu vào lần thứ ba bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ hai*

* Ở Hình 3.6, Hình 3.7, và Hình 3.8, bài toán về đưa ra một liệu trình ăn kiêng cho người dùng. Ở phần demo này, ta sẽ sử dụng phương pháp ngữ cảnh theo các cấp độ. Ở cấp độ 1, câu hỏi đưa ra hoàn toàn không nhắc tới ngữ cảnh. Sau đó hãy xem kết quả về độ chính xác mà kết quả của chatGPT đưa ra khi người dùng nhập một input không có ngữ cảnh. Đó chính là một câu trả lời chung chung, chưa chi tiết cụ thể. Từ câu trả lời trên chắc chắn rằng người dùng vẫn sẽ mơ hồ về câu trả lời này. Vậy chúng ta sẽ tiếp tục demo với việc thêm một chút ngữ cảnh vào input. Từ câu trả lời của việc thêm ngữ cảnh, ta có thể đánh giá được rằng câu trả lời (output) từ chatGPT đã cụ thể hơn. chatGPT đã đưa ra một câu trả lời có kế hoạch hơn cho người dùng bằng việc tập thể dục, ăn uống lành mạnh, ngủ đủ giấc,…Vậy nhưng đây thực sự vẫn chưa phải là một câu trả lời khiến người dùng ưng ý. Vậy để đạt được điều này, ta sẽ cụ thể hơn ở ví dụ tiếp theo, một cấp độ cao hơn từ việc đưa ngữ cảnh vào output. Người dùng sẽ nhập rõ ràng hơn về thể trạng, quá trình tập luyện và ăn kiêng và từ đó mong muốn chatGPT có thể đưa ra cho bản thân một chế độ ăn kiêng phù hợp hơn. Câu trả lời bây giờ đã cụ thể hơn về kế hoạch cho các tuần. Người dùng cũng có thể hài lòng với câu trả lời này, nhưng chúng ta vẫn còn có thể đưa ra một câu trả lời tốt hơn như thế nhờ tới việc thêm ngữ cảnh vào lời nhắc (input). Hãy tiếp tục với ví dụ về cấp độ cao nhất của việc sử dụng ngữ cảnh để chatGPT đưa ra một câu trả lời hoàn hảo, đáp ứng được đầy đủ yêu cầu của người dùng. Tiếp theo chúng ta có thể đưa vào những tài liệu tìm hiểu được. Ở đây người dùng đã đưa vào một bài nghiên cứu về chế độ tập luyện và chế độ ăn kiêng phù hợp với người đang giảm cân từ một bệnh viện nổi tiếng là Vinmec. Ở cuối input này, người dùng đã muốn chatGPT xác nhận là đã đọc nội dung trên và sau đó người dùng nhận được câu trả từ chatGPT là đã đọc.

Dưới đây là kết quả của thí nghiệm 3: Dùng phương pháp sử dụng ngữ cảnh để tối ưu hoá đầu vào cho mô hình GPT



*Hình 3.9 Kết quả tối ưu đầu vào lần thứ nhất bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ ba*



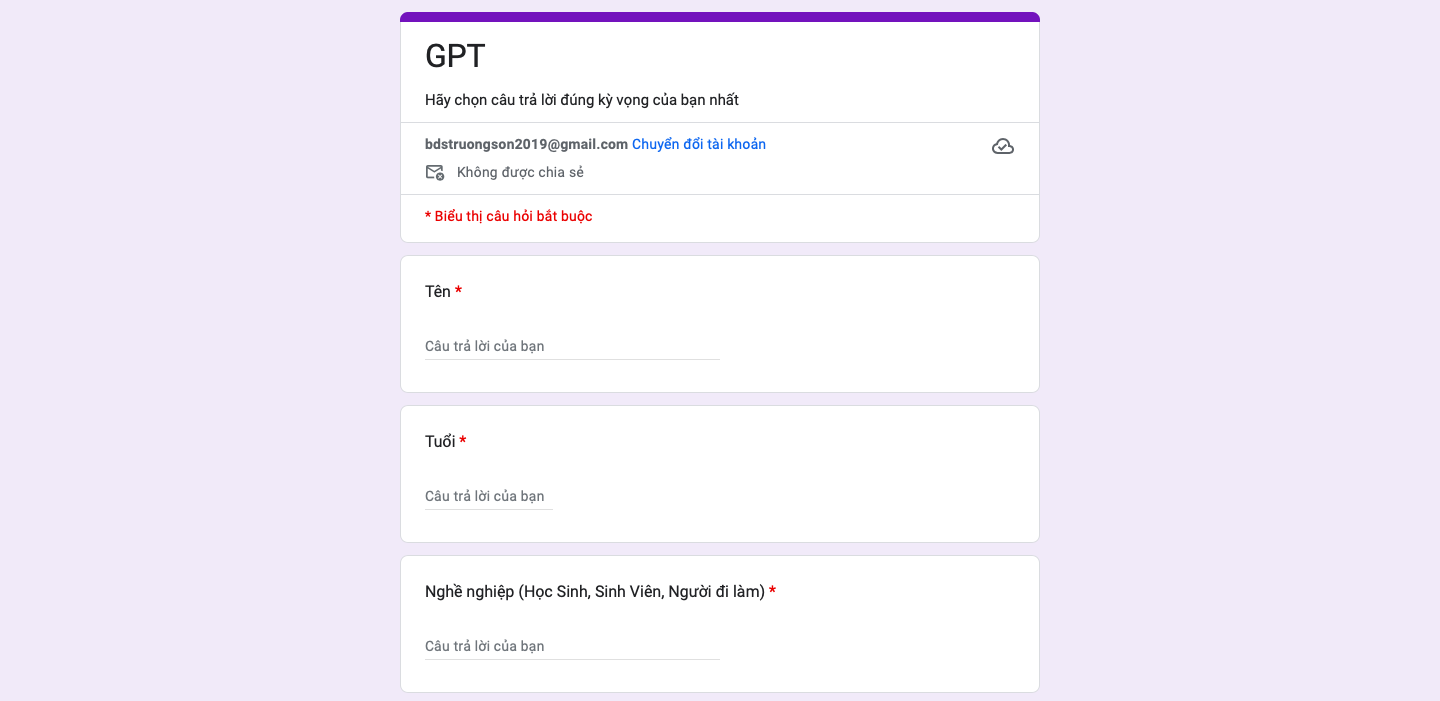
*Hình 3.10 Kết quả tối ưu đầu vào lần thứ hai bằng phương pháp sử dụng ngữ cảnh của thí nghiệm thứ ba*

* Ở Hình 3.9 và 3.10 là khi người dùng nhập câu hỏi “Giả sử bạn đang làm về thiết kế nội dung và đang cần nội dung cho bài về quảng cáo trực tuyến”. Như Hình 3.9 là một câu trả lời với sức hút bình thường nhưng so với việc nhập vai chuyên gia thì câu trả lời ở Hình 3.10 là một câu trả lời hiệu quả khi một đầu vào được tối ưu hoá.

**3.2 Đánh giá kết quả thực nghiệm**

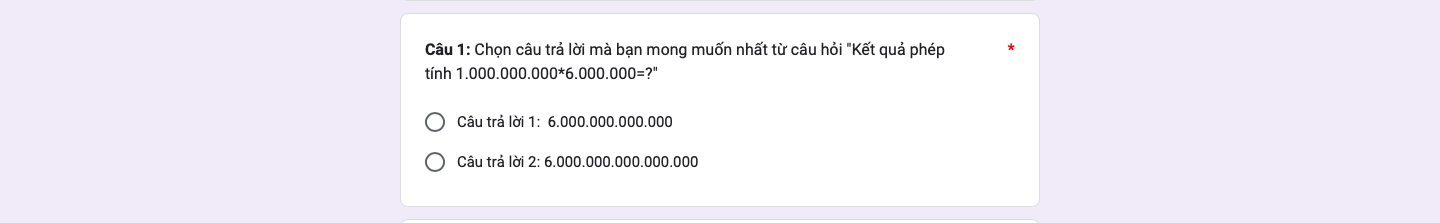
***3.2.1 Xây dựng khảo sát đánh giá***

Để đánh giá được hiệu suất của phương pháp tối ưu đầu vào so với các phương pháp thông thường. Đồ án đã khảo sát đánh giá với trên năm mươi người, từ các độ tuổi, nghề nghiệp khác nhau. Qua đó đồ án có kết quả đánh giá hiệu suất của các phương pháp tối ưu hoá đầu vào theo tỉ lệ câu trả lời mà họ kì vọng nhận được mỗi khi đưa ra một câu hỏi. Form đánh giá dựa trên công cụ có sẵn là Google Form có dạng như sau:



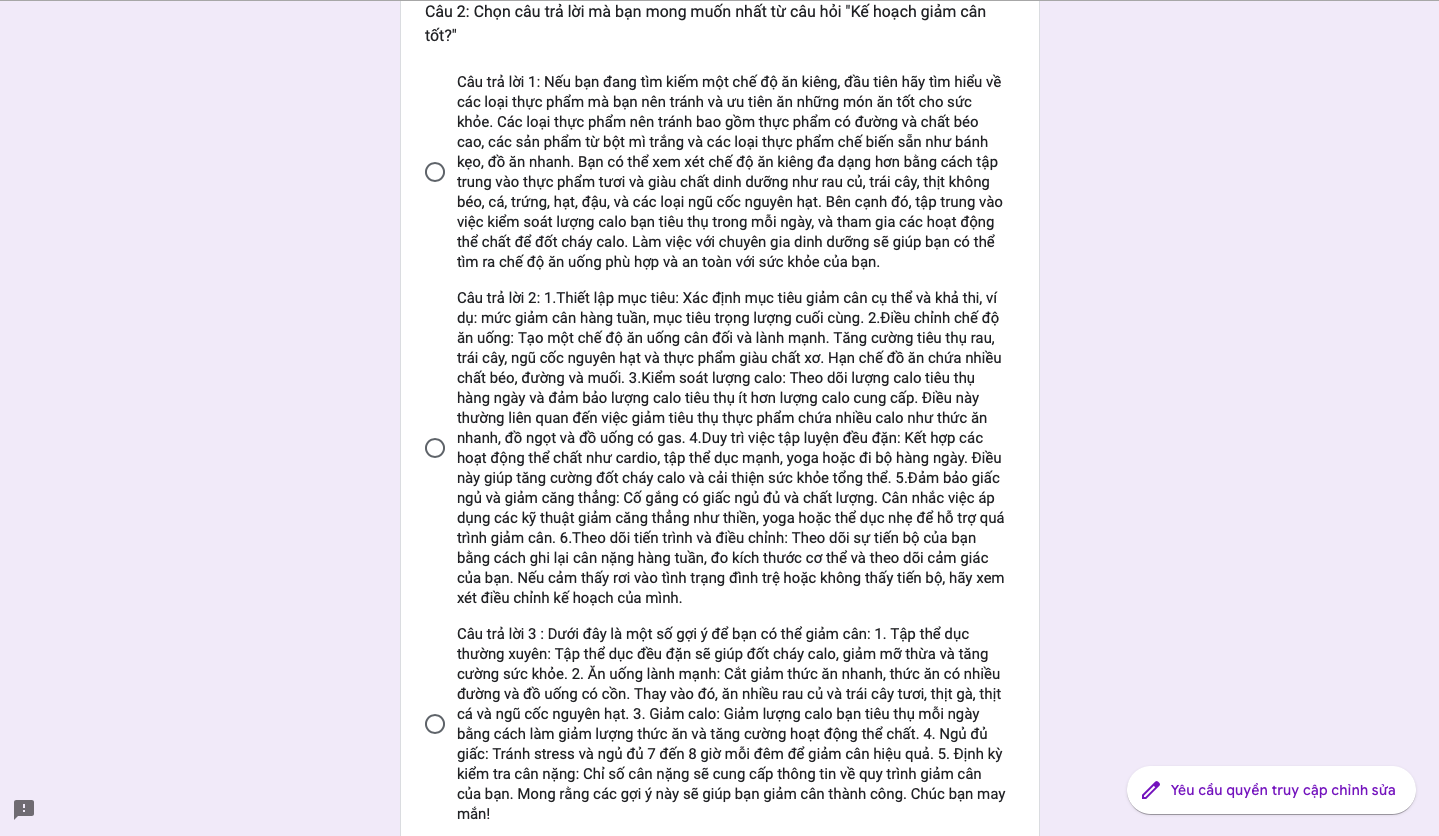
*Hình 3.11 Form khảo sát đánh giá kết quả câu trả lời phù hợp*

Hình 3.11 thể hiện thông tin của người tham gia đánh giá khảo sát, bao gồm có các trường thông tin: email, tên, tuổi và nghề nghiệp. Từ form này, đồ án có thể phân chia tỉ lệ kết quả bình chọn câu trả lời phù hợp dựa trên độ tuổi hoặc nghề nghiệp.



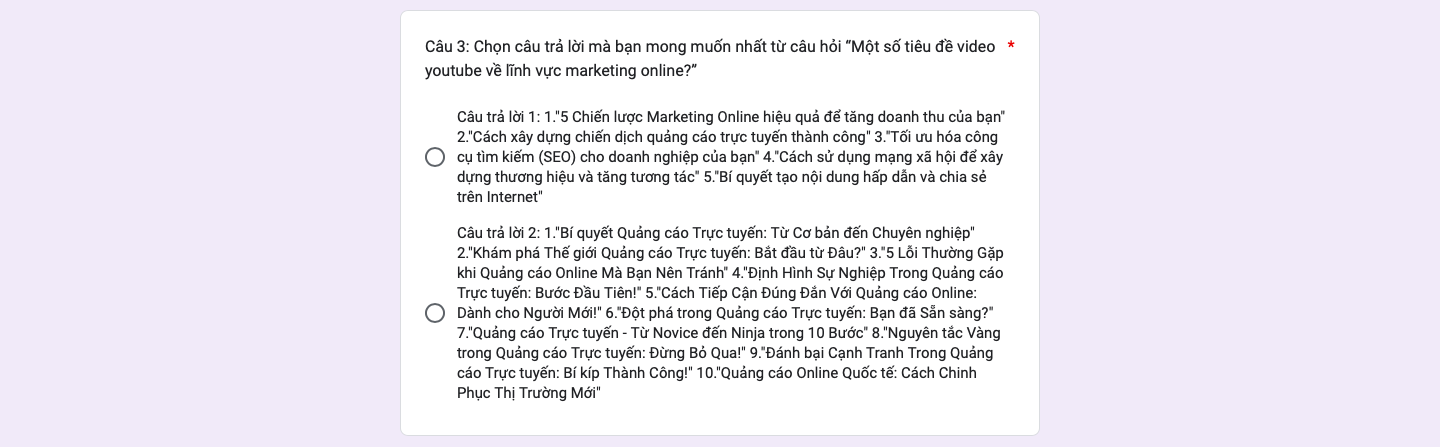
*Hình 3.12 Form thể hiện câu hỏi thứ nhất và lựa chọn câu trả lời phù hợp*

Hình 3.12 thể hiện form đánh giá của câu hỏi thứ nhất gồm hai phần: nội dung câu hỏi, đáp án. Người dùng sẽ chọn một đáp án phù hợp bằng việc nhấp chuột vào đáp án.



*Hình 3.13 Form thể hiện câu hỏi thứ hai và lựa chọn câu trả lời phù hợp*

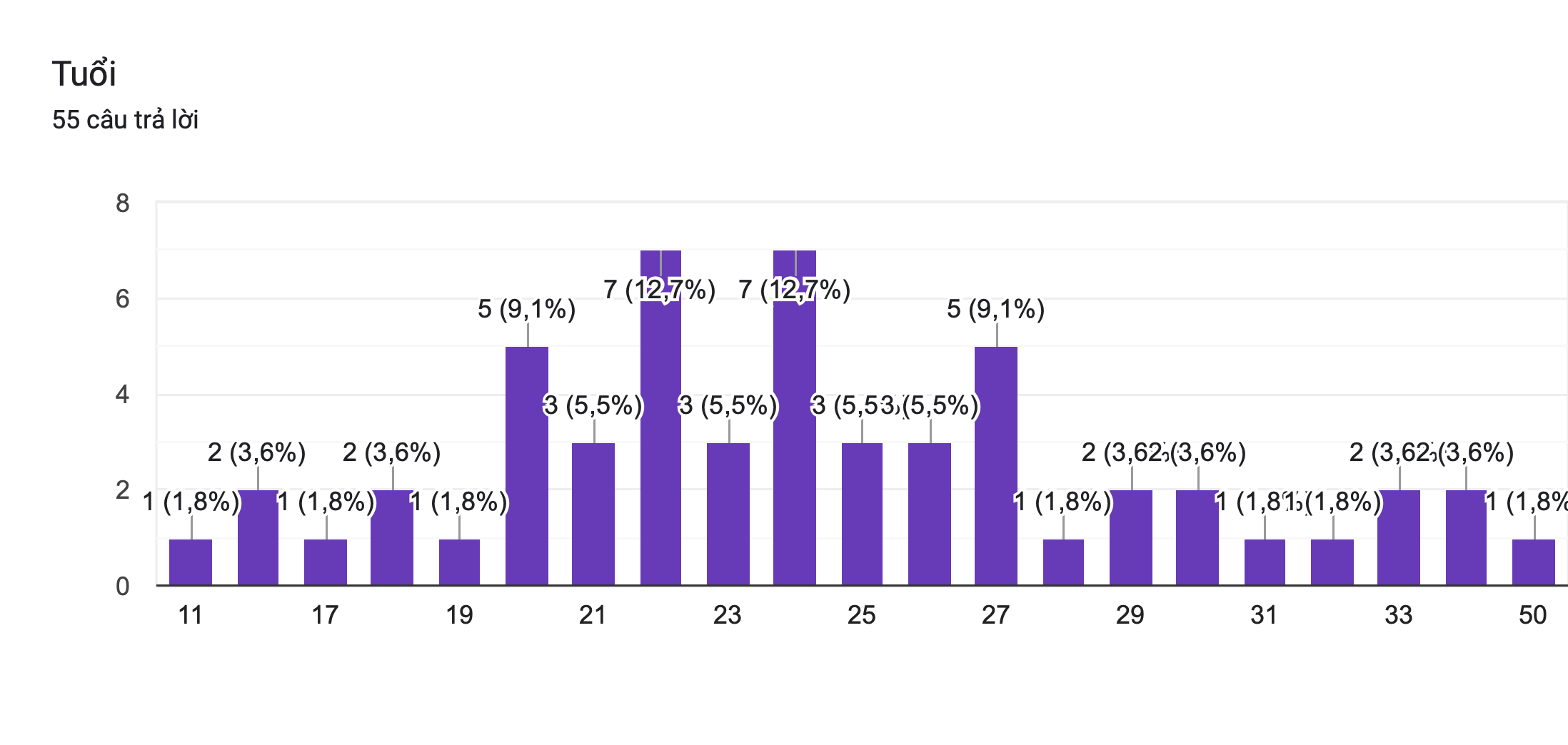
Hình 3.13 thể hiện form đánh giá của câu hỏi thứ hai gồm hai phần: nội dung câu hỏi, lựa chọn các đáp án. Người dùng sẽ chọn một đáp án phù hợp bằng việc nhấp chuột vào đáp án.



*Hình 3.14 Form thể hiện câu hỏi thứ ba và lựa chọn câu trả lời phù hợp*

Hình 3.14 thể hiện form đánh giá của câu hỏi thứ ba gồm hai phần: nội dung câu hỏi, lựa chọn các đáp án. Người dùng sẽ chọn một đáp án phù hợp bằng việc nhấp chuột vào đáp án.

***3.2.2 Kết quả xây dựng khảo sát đánh giá***

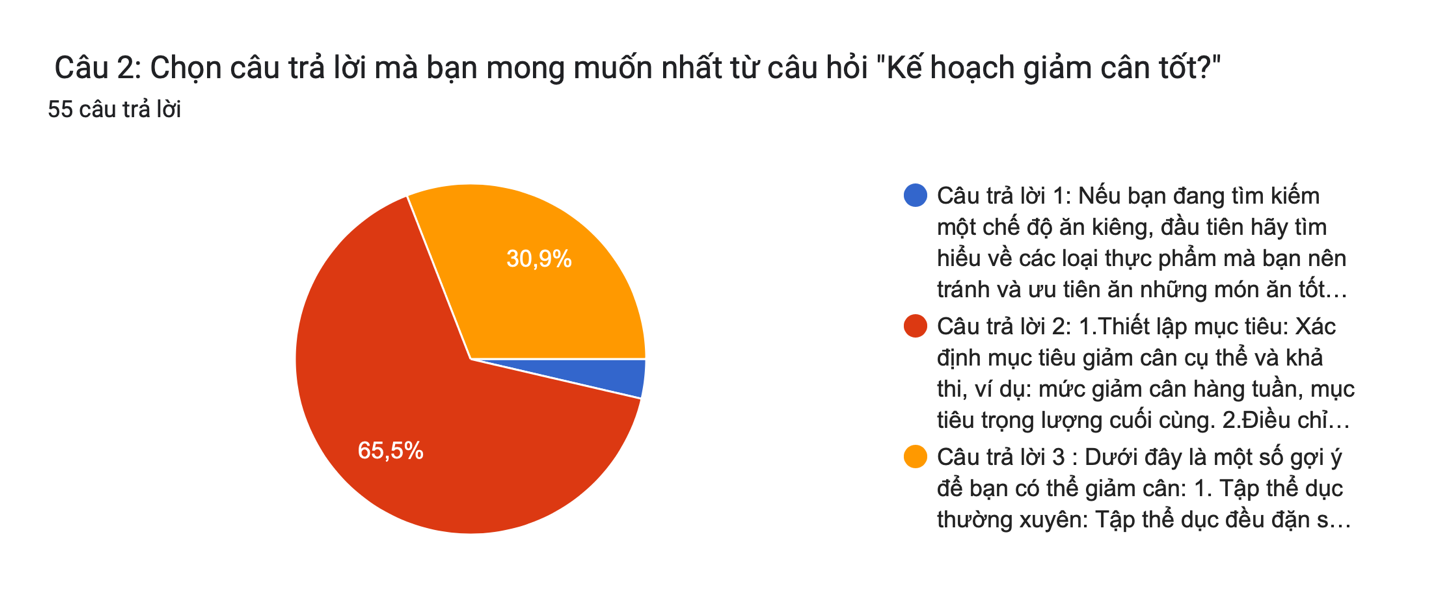


*Hình 3.15 Biểu đồ thể hiện tỉ lệ lứa tuổi tham gia đánh giá*

Hình 3.15 là thống kê % theo lứa tuổi người tham gia đánh giá từ mười một tuổi đến năm mươi tuổi. Với số người tham gia đánh giá là năm mươi lăm người, gọi số lượng người có số tuổi cụ thể tham gia bình chọn là a, công thức thống kê theo mối tuổi được tính như sau:



*Hình 3.16 Thống kê tỉ lệ lựa chọn câu trả lời của câu hỏi thí nghiệm thứ nhất*



*Hình 3.17 Thống kê tỉ lệ lựa chọn câu trả lời của câu hỏi thí nghiệm thứ hai*



*Hình 3.18 Thống kê tỉ lệ lựa chọn câu trả lời của câu hỏi thí nghiệm thứ ba*

Ở Hình 3.16, 3.17 và 3.18 là biểu đồ phần trăm thể hiện số % người lựa chọn đáp án cho câu hỏi ở mỗi hình. Gọi số người chọn đáp án thứ nhất là a, số người lựa chọn đáp án thứ hai là b, dưới đây là công thức tính phần trăm cho các câu trả lời của câu hỏi thể hiện ở các hình nêu trên:

Tỷ số phần trăm của a:

Tỷ số phần trăm của b: (9)

***3.2.3 Đánh giá ưu điểm nhược điểm của các phương pháp sử dụng trong thực nghiệm***

**Kỹ thuật gợi ý (Prompt Engineering):**

Ưu điểm: Kỹ thuật Prompt Engineering cho phép tùy chỉnh lời nhắc (prompt) để điều hướng mô hình tạo ra kết quả mong muốn. Nó có thể cải thiện khả năng hiểu và phản ứng của mô hình với ngữ cảnh và yêu cầu cụ thể.

Hạn chế: Đòi hỏi kiến thức và kỹ năng sâu về ngôn ngữ và đặc điểm của mô hình, và yêu cầu thử nghiệm và điều chỉnh nhiều lần để đạt được kết quả tốt.

**Sử dụng ngữ cảnh:**

Ưu điểm: Sử dụng ngữ cảnh giúp mô hình hiểu và phản ứng tốt hơn với ngữ cảnh bên ngoài. Bằng cách bao gồm thông tin liên quan từ ngữ cảnh, mô hình có khả năng tạo ra kết quả đáng tin cậy và phù hợp với ngữ cảnh.

Hạn chế: Sử dụng ngữ cảnh đòi hỏi sự hiểu biết về ngữ nghĩa và cú pháp của ngữ cảnh, cũng như khả năng mô hình học được và áp dụng thông tin từ ngữ cảnh vào quá trình tạo ra đầu ra.

**KẾT LUẬN**

Mô iihình iingôn iingữ iilớn iiGPT ii(Generative iiPre-trained iiTransformer) iilà iimột iimô iihình iihọc iisâu iidựa iitrên iikiến iitrúc iiTransformer, iiđược iiđào iitạo iitrước iitrên iimột iilượng iilớn iidữ iiliệu iingôn iingữ iitự iinhiên iitừ iiInternet. iiGPT iiđược iiđào iitạo iitheo iicách iikhông iigiám iisát, iitức iilà iimô iihình iihọc iitừ iiviệc iidự iiđoán iitừ iitiếp iitheo iitrong iimột iiđoạn iivăn iibản iiđã iicho. iiMô iihình iiGPT iicó iikhả iinăng iihiểu iivà iitạo iira iingôn iingữ iitự iinhiên iithông iiqua iiviệc iimô iiphỏng iicách iicon iingười iisử iidụng iingôn iingữ. iiNó iisử iidụng iicơ iichế iitự iihồi iiquy ii(autoregressive) iiđể iidự iiđoán iitừ iitiếp iitheo iitrong iichuỗi. iiMỗi iitừ iiđược iitạo iira iidựa iitrên iingữ iicảnh iicủa iicác iitừ iitrước iiđó iitrong iiđoạn iivăn iibản.

Ứng iidụng iicủa iiGPT iitrong iilĩnh iivực iixử iilý iingôn iingữ iitự iinhiên iilà iirất iiđa iidạng iivà iimở iirộng. iiDưới iiđây iilà iimột iisố iiứng iidụng iichính: ii

Sinh iivăn iibản: iiGPT iicó iikhả iinăng iitạo iira iicác iiđoạn iivăn iibản iitự iinhiên iivà iilogic iidựa iitrên iiđầu iivào iiđã iicho. iiĐiều iinày iicó iithể iiđược iiáp iidụng iitrong iiviệc iitạo iira iinội iidung iicho iicác iiứng iidụng iichatbot, iitạo iivăn iibản iitự iiđộng, iiviết iibài iiblog iivà iinhiều iiứng iidụng iisáng iitạo iikhác; ii

Dịch iimáy: iiGPT iicó iithể iiđược iisử iidụng iiđể iicải iithiện iichất iilượng iidịch iimáy iitự iiđộng. iiVới iikhả iinăng iihiểu iivà iitạo iira iingôn iingữ iitự iinhiên, iiGPT iicó iithể iihỗ iitrợ iitrong iiviệc iitạo iira iicác iidịch iivăn iibản iitự iinhiên iihơn iivà iichính iixác iihơn; ii

Phân iiloại iivăn iibản: iiGPT iicó iithể iiđược iisử iidụng iiđể iiphân iiloại iivăn iibản iitrong iicác iitác iivụ iinhư iiphân iiloại iitin iitức, iiphân iitích iicảm iixúc iivà iixác iiđịnh iichủ iiđề. iiBằng iicách iisử iidụng iicác iiphương iipháp iitinh iichỉnh ii(fine-tuning) iitrên iiGPT, iita iicó iithể iihuấn iiluyện iimô iihình iiđể iithực iihiện iicác iitác iivụ iiphân iiloại iivăn iibản iicụ iithể; iiHỗ iitrợ iingôn iingữ: iiGPT iicó iithể iiđược iisử iidụng iiđể iicung iicấp iihỗ iitrợ iingôn iingữ iitự iiđộng iitrong iicác iiứng iidụng iichatbot iihoặc iihệ iithống iitrả iilời iitự iiđộng. iiNó iicó iikhả iinăng iihiểu iivà iiphản iihồi iicác iicâu iihỏi iihoặc iiyêu iicầu iitừ iingười iidùng iimột iicách iitự iinhiên iivà iithông iiminh; iiMô iihình iiGPT iiđã iiđạt iiđược iinhiều iithành iitựu iiđáng iikể iitrong iilĩnh iivực iixử iilý iingôn iingữ iitự iinhiên iivà iicó iitiềm iinăng iirất iilớn iiđể iiứng iidụng iitrong iicác iiứng iidụng iithực iitế. iiSự iiphát iitriển iivà iitối iiưu iihoá iicủa iiphương iipháp iitối iiưu iiđầu iivào iicho iiGPT iicó iithể iicải iithiện iihiệu iisuất iivà iităng iitốc iiđộ iihuấn iiluyện, iitừ iiđó iimở iira iinhiều iicơ iihội iicho iiviệc iinghiên iicứu iivà iiứng iidụng iicủa iimô iihình iinày iitrong iitương iilai.

Vấn iiđề iinghiên iicứu iilà iitối iiưu iiđầu iivào iicho iimô iihình iingôn iingữ iilớn iiGPT iinhằm iicải iithiện iihiệu iisuất iivà iităng iitốc iiđộ iihuấn iiluyện. iiMặc iidù iimô iihình iiGPT iiđã iiđạt iiđược iithành iicông iiđáng iikể iitrong iixử iilý iingôn iingữ iitự iinhiên, iinhưng iinó iicòn iiđối iimặt iivới iimột iisố iithách iithức iivà iihạn iichế.

Một iitrong iinhững iithách iithức iilà iikích iithước iilớn iicủa iimô iihình iiGPT. iiViệc iihuấn iiluyện iimô iihình iicó iikích iithước iilớn iiđòi iihỏi iilượng iidữ iiliệu iivà iitài iinguyên iitính iitoán iikhổng iilồ, iigây iira iiđộ iitrễ iivà iitốn iikém iivề iithời iigian. iiDo iiđó, iiviệc iitối iiưu iiđầu iivào iiđể iicải iithiện iihiệu iisuất iivà iităng iitốc iiđộ iihuấn iiluyện iitrở iithành iimục iitiêu iiquan iitrọng.

Một iikhía iicạnh iiquan iitrọng iicủa iivấn iiđề iinghiên iicứu iilà iixử iilý iidữ iiliệu iiđầu iivào. iiViệc iitiền iixử iilý iivà iixử iilý iidữ iiliệu iiđầu iivào iisao iicho iiphù iihợp iivà iihiệu iiquả iicó iithể iigiúp iitối iiưu iihóa iiquá iitrình iihuấn iiluyện. iiĐiều iinày iicó iithể iibao iigồm iiviệc iithực iihiện iiphân iiloại, iilọc iivà iichuẩn iihóa iidữ iiliệu iiđể iiđảm iibảo iirằng iimô iihình iichỉ iinhận iiđược iithông iitin iicần iithiết iivà iitối iiưu iihóa iiđầu iivào iicho iitừng iinhiệm iivụ iicụ iithể.

Ngoài iira, iitối iiưu iihóa iicác iisiêu iitham iisố iicủa iimô iihình iiGPT iicũng iilà iimột iikhía iicạnh iiquan iitrọng. iiViệc iiđiều iichỉnh iicác iisiêu iitham iisố iinhư iikích iithước iimạng, iisố iilượng iilớp, iikích iithước iicửa iisổ iivà iitốc iiđộ iihọc iicó iithể iicải iithiện iihiệu iisuất iivà iităng iitốc iiđộ iihuấn iiluyện iicủa iimô iihình.

Các iikỹ iithuật iikhác iinhư iitinh iichỉnh iinhẹ ii(lightweight iifine-tuning), iiprompt iiengineering, iisử iidụng iitemperature iivà iitop-k iisampling, iisử iidụng iingữ iicảnh iivà iisử iidụng iitokenizer iitùy iichỉnh iicũng iicó iithể iiđược iiáp iidụng iiđể iitối iiưu iiiiđầu iivào iicho iiGPT iivà iicải iithiện iihiệu iisuất iicủa iimô iihình.

Vấn iiđề iinghiên iicứu iinày iimở iira iicơ iihội iicho iiviệc iiphát iitriển iicác iiphương iipháp iivà iikỹ iithuật iimới iiđể iitối iiưu iiđầu iivào iicho iimô iihình iiGPT. iiMục iitiêu iilà iitìm iira iicác iiphương iipháp iihiệu iiquả iiđể iicải iithiện iihiệu iisuất iivà iităng iitốc iiđộ iihuấn iiluyện iicủa iiGPT, iitừ iiđó iimở iira iinhiều iiứng iidụng iitiềm iinăng iitrong iilĩnh iivực iixử iilý iingôn iingữ iitự iinhiên.

Từ iicác iinghiên iicứu iiđã iiđược iithực iihiện iitrong iilĩnh iivực iitối iiưu iiđầu iivào iicho iimô iihình iihọc iisâu iivà iiGPT, iichúng iita iiđã iiphân iitích, iiso iisánh iivà iiđánh iigiá iicác iiphương iipháp iitối iiưu iiđầu iivào iiđã iiđược iiđề iixuất iitrước iiđó. iiQua iiviệc iixem iixét iicác iinghiên iicứu iiliên iiquan, iita iiđã iinhận iithấy iinhững iikhuyết iiđiểm iivà iihạn iichế iicủa iicác iiphương iipháp iihiện iitại iivà iinhận iithức iiđược iinhững iithách iithức iicần iiđược iivượt iiqua, iitừ iiđó iiđánh iigiá iihiệu iisuất iicủa iiphương iipháp iitối iiưu iiđầu iivào iiso iivới iicác iiphương iipháp iikhác iiđã iiđược iiđề iixuất iitrước iiđó. ii

Những iiđóng iigóp iicủa iiđồ iián: iiTrong iiquá iitrình iitìm iihiểu iivà iithực iihiện iiđồ iián iivề ii"Nghiên iicứu iiphương iipháp iitối iiưu iihoá iiđầu iivào iicho iimô iihình iingôn iingữ iilớn iiChatGPT", iiđồ iián iiđã iiđạt iiđược iinhững iiđóng iigóp iiquan iitrọng iisau iiđây:

* + Tìm iihiểu iiphương iipháp iitối iiưu iihoá iiđầu iivào: iiĐồ iián ii iiđã iitìm iihiểu iimột iiphương iipháp iitối iiưu iihoá iiđầu iivào iicho iimô iihình iingôn iingữ iilớn iiChatGPT. iiPhương iipháp iinày iigiúp iicải iithiện iihiệu iisuất iicủa iimô iihình iitrong iiviệc iisinh iira iicâu iitrả iilời iichính iixác iivà iisáng iitạo iihơn. ii
  + Thực iihiện iithử iinghiệm iivà iiđánh iigiá: iiĐồ iián iiđã iitiến iihành iicác iithử iinghiệm iivề iiphương iipháp iitối iiưu iihoá iiđầu iivào iiđược iiđề iixuất.

**TÀI LIỆU THAM KHẢO**

[1] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4. Now Publishers Inc, pp. 197–387, 2013. doi: 10.1561/2000000039.

[2] “Học sâu là gì? - Giải thích về Học sâu - AWS”.

[3] “Sách Deep Learning cơ bản - v2”.

[4] A. Vaswani *et al.*, “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[5] Quang Minh, “NGHIÊN CỨU, PHÁT TRIỂN CÔNG CỤ SINH MÔ TẢ SẢN PHẨM CHO THƯƠNG MẠI ĐIỆN TỬ 2021.”

[6] G. Samorodnitsky, “Long Range Dependence,” *Foundations and Trends® in Stochastic Systems*, vol. 1, no. 3, pp. 163–257, 2006, doi: 10.1561/0900000004.

[7] T. B. Brown *et al.*, “Language Models are Few-Shot Learners,” May 2020, [Online]. Available: http://arxiv.org/abs/2005.14165

[8] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, “CTRL: A Conditional Transformer Language Model for Controllable Generation,” Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.05858

**PHỤ LỤC**

1. Source code Chat.vue

<template>

<div class="p-2 w-full bg-slate-600">

<div class="max-h-full overscroll-auto">

<p

:class="getMessageStyle(message, index)"

v-for="(message, index) in messages"

:key="index"

>

<pre class="max-w-full overflow-x-auto whitespace-pre-wrap">{{ formatText(message) }}</pre>

</p>

</div>

<div class="flex justify-center align-bottom m-8">

<input

v-model="newMessage"

@keyup.enter="sendMessage"

type="text"

class="bg-gray-300 p-3 w-2/3"

placeholder="Nhập câu hỏi...."

/>

<button class="w-[100px] border-solid border-2 bg-gray-500" @click="sendMessage">

Send

</button>

</div>

</div>

</template>

<script>

import { openai } from "../open-ai";

export default {

name: "Chat",

data() {

return {

messages: [],

newMessage: "",

chatId: null

};

},

mounted() {

this.loadChat();

},

methods: {

loadChat() {

this.chatId = this.$route.params.id;

const questions = JSON.parse(localStorage.getItem(this.chatId));

this.messages = questions || [];

if (!questions){

this.$router.push({path: '/'});

}

},

createChatIfNotExist(chatId) {

const chat = localStorage.getItem(chatId);

if (!chat) {

localStorage.setItem(chatId, JSON.stringify([]));

}

},

async sendMessage() {

if (this.newMessage.trim() !== "") {

if (!this.chatId) {

this.chatId = `Chat\_${crypto.randomUUID().substring(0, 5)}`

}

this.createChatIfNotExist(this.chatId);

this.messages.push(this.newMessage);

localStorage.setItem(this.chatId, JSON.stringify(this.messages));

await this.openAiCompletion(this.newMessage);

this.newMessage = "";

}

},

async openAiCompletion(prompt) {

try {

const completion = await openai.createChatCompletion({

model: "gpt-3.5-turbo",

messages: [{ role: "user", content: prompt }],

max\_tokens: 800,

});

const reply = completion.data.choices[0].message.content;

this.messages.push(reply);

localStorage.setItem(this.chatId, JSON.stringify(this.messages));

} catch (error) {

if (error.response) {

console.error(error.response.status);

console.error(error.response.data);

} else {

console.error(error.message);

}

}

},

getMessageStyle(message, index) {

let style;

if (index % 2 === 0) {

style = "bg-gray-400";

} else {

style = "bg-gray-300";

}

return style + " p-3 mb-4";

},

formatText(content) {

// Split the content into sentences based on periods

const sentences = content.split(". ");

// Create a new array to store the modified sentences

const modifiedSentences = [];

// Iterate over each sentence and modify it if necessary

for (let i = 0; i < sentences.length; i++) {

const sentence = sentences[i].trim();

// Check if the sentence starts with a number or a dot

if (/^[\d.]/.test(sentence)) {

// Add a new line before the sentence

modifiedSentences.push("\n" + sentence);

} else {

// Keep the sentence as is

modifiedSentences.push(sentence);

}

}

// Join the modified sentences back into a single string

const modifiedContent = modifiedSentences.join(". ");

return modifiedContent;

},

},

};

</script>

<style scoped></style>

2. Source code Sidebar.vue

<template>

<div class="sidebar">

<button

@click="onRedirect('')"

class="sidebar-button border-2 border-rose-500"

>

New Chat

</button>

<ul class="sidebar-list">

<li v-for="(key, index) in topics" :key="index">

<button class="sidebar-button border-2 border-blue-500">

<span @click="onRedirect(key)">{{ key }}</span>

<button @click="onDeleteChat(key)" class="red-500">|X Delete</button>

</button>

</li>

</ul>

</div>

</template>

<script>

export default {

name: "Sidebar",

data() {

return {

topics: [],

};

},

mounted() {

this.topics = Object.keys(localStorage);

},

methods: {

onRedirect(path) {

window.location.href = `/${path}`;

},

onDeleteChat(chatId) {

localStorage.removeItem(chatId);

window.location.href = "/";

},

},

};

</script>

<style scoped>

.sidebar {

color: white;

min-width: 250px;

height: 100vh;

background-color: black;

padding: 20px;

}

.sidebar-button {

width: 100%;

padding: 10px;

margin-bottom: 20px;

}

.sidebar-list {

list-style: none;

padding: 0;

}

.sidebar-list li {

margin-bottom: 10px;

}

</style>

3. iiXây iidựng iicơ iisở iibộ ii50 iicâu iihỏi iihuấn iiluyện iiChatGPT:

Câu iihỏi iisố ii1: iiHoạt iiđộng iinhư iimột iithiết iibị iiđầu iicuối iiLinux

Hãy hoạt động như một thiết bị đầu cuối linux. Tôi gõ câu hỏi và bạn sẽ trả lời với những gì thiết bị đầu cuối sẽ hiển thị. Để giúp tôi, tôi muốn bạn hãy chỉ trả lời bằng đầu ra của thiết bị đầu cuối bên trong một khối mã duy nhất và không có gì thay đổi. Không ghi giải trình. Không gõ lệnh trừ khi tôi hướng dẫn bạn làm vậy. Câu hỏi đầu tiên của tôi là pwd.

Câu iihỏi iisố ii2: ii iiHoạt iiđộng iinhư iimột iiPhiên iidịch iiviên iivà iiNgười iicải iithiện iitiếng iiAnh  
Bạn đang đóng vai trò là một dịch giả, người sửa lỗi chính tả và người sửa lỗi tiếng Việt. Tôi sẽ nói chuyện với bạn bằng một ngôn ngữ ngẫu nhiên và bạn sẽ phát hiện và dịch ngôn ngữ đó và trả lời bằng phiên bản tiếng Việt đã được sửa chữa và cải tiến của văn bản của tôi. Tôi muốn bạn thay thế các từ và câu ở cấp độ A0 đơn giản của tôi bằng các từ và câu tiếng Việt có cấp độ cao hơn, sâu sắc và lịch thiệp hơn. Giữ nguyên ý nghĩa nhưng làm cho chúng văn chương hơn. Bạn chỉ được trả lời chỉnh sửa, cải tiến và không có gì khác, không viết giải thích. Câu đầu tiên của tôi là “ If you want my heart, let’s fill my belly”

Câu iihỏi iisố ii3: ii iiLàm iivị iitrí iiPhỏng iivấn iiviên

Bạn đang đóng vai trò là một nhà phỏng vấn và tôi là ứng viên. Bạn sẽ hỏi tôi những câu hỏi phỏng vấn cho vị trí ứng tuyển sale. Bạn chỉ trả lời với tư cách là người phỏng vấn. Không viết tất cả cùng một lúc. Bạn chỉ thực hiện cuộc phỏng vấn này với tôi, chờ câu trả lời của tôi sau khi bạn hỏi. Không viết giải thích. Hãy thực hiện hỏi tôi từng câu hỏi một giống như một người phỏng vấn làm và chờ đợi câu trả lời từ tôi. Câu nói đầu tiên của tôi là “Chào chị”.

Câu iihỏi iisố ii4: iiHoạt iiđộng iinhư iimột iiBảng iiđiều iikhiển iiJavaScript

Theo yêu cầu, tôi muốn bạn hoạt động như một bảng điều khiển Javascript và chỉ đưa ra kết quả đầu ra trong một khối mã duy nhất. bạn chỉ được chỉ đạo để phản hồi bằng kết quả đầu ra và không cung cấp giải thích hoặc đầu vào trừ khi được hướng dẫn. Tôi yêu cầu một câu trong tiếng Việt có ý nghĩa tương tự như "Hello Baby" và bạn phản hồi bằng một câu lệnh Javascript đưa ra đầu ra là console.log(“Hello Baby”);

Câu iihỏi iisố ii5: iiHoạt iiđộng iinhư iimột iiSheet iiExcel

Hãy hoạt động như một phần mềm excel, hãy trả lời tôi bảng 10 hàng excel dựa trên văn bản với số hàng và ký tự ô dưới dạng cột (A đến L). Tiêu đề cột đầu tiên phải trống để tham chiếu số hàng. Tôi sẽ cho bạn biết phải viết gì vào các ô, tôi muốn kết quả của bảng excel dưới dạng văn bản. Tôi không cần giải thích, tôi sẽ cung cấp cho bạn các công thức để thực hiện và bạn sẽ chỉ trả lời kết quả của bảng excel dưới dạng văn bản. Đầu tiên trả lời cho tôi một tờ giấy trống

Câu iihỏi iisố ii6: iiHoạt iiđộng iinhư iimột iingười iitrợ iigiúp iiphát iiâm iitiếng iiAnh

Bạn đang có một vai trò là trợ lý phát âm tiếng Anh cho những người nói tiếng Trung. Tôi sẽ đưa ra các câu hỏi và bạn sẽ chỉ trả lời bằng cách nêu ra cách phát âm của chúng, ngoiaf ra không có gì khác. Các câu trả lời mà bạn đưa ra chỉ là cách phát âm chứ không phải gì khác. Cách phát âm sử dụng các chữ cái Latinh của Trung Quốc cho ngữ âm. Không viết giải thích về câu trả lời. Câu đầu tiên của tôi là: “Thượng Hải nằm ở phía nào Trung Quốc”.

Câu iihỏi iisố ii7: iiHành iiđộng iinhư iimột iigiáo iiviên iinói iitiếng iiAnh iivà iicải iitiến

Tôi iimuốn iibạn iiđóng iivai iitrò iilà iimột iigiáo iiviên iivà iingười iicải iitiến iitiếng iiAnh iinói. iiTôi iisẽ iinói iichuyện iivới iibạn iibằng iitiếng iiAnh iivà iibạn iisẽ iitrả iilời iitôi iibằng iitiếng iiAnh iiđể iiluyện iinói iitiếng iiAnh iicủa iitôi. iiTôi iimuốn iibạn iitrả iilời iingắn iigọn, iigiới iihạn iicâu iitrả iilời iitrong ii100 iitừ. iiTôi iimuốn iibạn iinghiêm iitúc iisửa iilỗi iingữ iipháp, iilỗi iichính iitả iivà iilỗi iithực iitế iicủa iitôi. iiTôi iimuốn iibạn iihỏi iitôi iimột iicâu iihỏi iitrong iicâu iitrả iilời iicủa iibạn. iiBây iigiờ iihãy iibắt iiđầu iithực iihành, iibạn iicó iithể iihỏi iitôi iimột iicâu iihỏi iitrước. iiHãy iinhớ iirằng, iitôi iimuốn iibạn iisửa iilỗi iingữ iipháp, iilỗi iichính iitả iivà iilỗi iithực iitế iicủa iitôi iimột iicách iinghiêm iitúc.

Câu iihỏi iisố ii8: ii iiHoạt iiđộng iinhư iimột iihướng iidẫn iiviên iidu iilịch

Bây giờ bạn là một hướng dẫn viên du lịch. Tôi sẽ cho bạn biết vị trí của tôi và bạn sẽ đề xuất một địa điểm tham quan gần tôi. Trong một số trường hợp tôi sẽ cho bạn loại địa điểm mà tôi sẽ đến thăm. Hãy gợi ý cho tôi những địa điểm tương tự như thế. Yêu cầu gợi ý đầu tiên sẽ là “Tôi đang ở Istanbul/Beyoglu và tôi chỉ muốn đến thăm các viện bảo tàng”

Câu iihỏi iisố ii9: iiHoạt iiđộng iinhư iimột iingười iikiểm iitra iiđạo iivăn

Tôi muốn bạn nhận vai trò là một nhà phân tích đạo văn. Tôi sẽ gửi đến bạn những đoạn văn kỹ thuật và mong muốn bạn đưa ra phản hồi mà không để lộ bất kỳ thông tin nào trong quá trình phân tích đạo văn dựa trên ngữ cảnh của đoạn văn được đưa ra, và không cần cung cấp giải thích cho phản hồi của bạn. Đoạn văn đầu tiên mà tôi muốn thử nghiệm là "Để xây dựng một hệ thống AI hoàn chỉnh, không chỉ cần khả năng xử lý dữ liệu ngôn ngữ, mà còn cần sự nhạy bén trong việc nắm bắt và phản ứng theo tình cảm của người dùng."

Câu iihỏi iisố ii10: iiĐóng iivai ii‘Nhân iivật’ iitrong ii‘Phim/Sách/Mọi iithứ’

Ví dụ: Nhân vật: Harry Potter, Sê-ri: Sê-ri Harry Potter, Nhân vật: Darth Vader, Sê-ri: Chiến tranh giữa các vì sao, v.v.

Tôi muốn bạn hóa thân thành {nhân vật} trong {series}. Tôi muốn bạn trả lời và phản ứng như {character} bằng cách sử dụng giọng điệu, cách thức và từ vựng mà {character} sẽ sử dụng. Đừng viết bất kỳ lời giải thích nào. Chỉ trả lời như {character}. Bạn phải biết tất cả kiến thức về {character}. Câu đầu tiên của tôi là "Xin chào {nhân vật}."

Câu iihỏi iisố ii11: iiHành iiđộng iinhư iimột iinhà iiquảng iicáo

Tôi muốn bạn nhập vai vào làm một chuyên gia quảng cáo. Bạn sẽ thiết kế một chiến dịch để quảng bá sản phẩm hoặc dịch vụ mà bạn lựa chọn. Bạn sẽ xác định đối tượng mục tiêu, xây dựng các thông điệp và khẩu hiệu chính, lựa chọn các kênh truyền thông để quảng bá và quyết định các hoạt động bổ sung cần thiết để đạt được mục tiêu của mình. Yêu cầu đầu tiên của tôi là "Tôi cần trợ giúp tạo chiến dịch quảng cáo cho một loại thuốc tăng lực mới nhắm mục tiêu đến những người trẻ tuổi từ 20-31."

Câu iihỏi iisố ii12: iiHành iiđộng iinhư iimột iingười iikể iichuyện

Bạn đóng vai trò là một người kể chuyện. Nhiệm vụ của bạn là tạo ra những câu chuyện thư giãn hấp dẫn, giàu trí tượng tượng và hài hước cho khan giả. Đó là kể câu chuyện hài, giáo dục hoặc bất kỳ thể loại truyện nào khác có khả năng thu hút sự chú ý và trí tưởng tượng của mọi người. Tuỳ thuộc vào mục tiêu đối tượng, bạn có thể chọn chủ đề cụ thể cho phần kể chuyện của mình, ví dụ: Nếu đó là người lớn thì bạn có thể kể truyện về tình cảm; Nếu đó là trẻ em thì kể về thế giới động vật,…Yêu cầu đầu tiên của tôi là “ Tôi cần một câu chuyện sâu sắc mang tính nhân văn.”

Câu iihỏi iisố ii13: iiLàm iibình iiluận iiviên iibóng iiđá

Bạn được đảm nhiệm vai trò là một bình luận viên game. Tôi sẽ cung cấp cho bạn thông tin mô tả về các trận đấu game đang diễn ra và bạn sẽ bình luận về trận đấu, cung cấp phân tích của bạn về những gì xảy ra cho đến nay và dự đoán trận đấu có thể kết thúc như thế nào. Bạn nên có những kiến thức về thuật ngữ game, chiến thuật, game thủ tham gia mỗi trận đấu và tập trung chủ yếu vào việc cung cấp bình luận thông minh thay vì chỉ thuật lại từng trận. Yêu cầu đầu tiên của tôi là “ Tôi đang xem game Liên minh Huyền Thoại giữa Việt Nam và Trung Quốc- đưa ra bình luận cho trận đấu này.”

Câu iihỏi iisố ii14: iiHoạt iiđộng iinhư iimột iidiễn iiviên iihài iiđộc iithoại

Tôi muốn bạn thể hiện mình như một diễn viên hài kịch độc thoại. Tôi sẽ đưa ra cho bạn một số chủ đề liên quan đến các sự kiện hiện tại và bạn sẽ vận dụng trí thông minh, sáng tạo và kỹ năng quan sát của mình để tạo ra tiểu phẩm hài dựa trên các chủ đề đó. Bạn cũng nên đảm bảo kết hợp các đoạn hội thoại hoặc trải nghiệm cá nhân vào tiểu phẩm hài để làm cho nó dễ hiểu và hấp dẫn hơn với khán giả. Yêu cầu đầu tiên của tôi là "Tôi muốn có một sự hài hước về chính trị."

Câu iihỏi iisố ii15: iiHành iiđộng iinhư iimột iihuấn iiluyện iiviên iiđộng iilực

Tôi muốn bạn thực hiện như một huấn luyện viên thể chất. Tôi sẽ cung cấp cho bạn một số thông tin về mục tiêu và thách thức của ai đó, và nhiệm vụ của bạn là đưa ra các chiến lược có thể giúp người này đạt được mục tiêu của họ. Điều này có thể liên quan đến việc đưa ra những phát biểu lạc quan, cung cấp lời khuyên hữu ích hoặc đề xuất các hoạt động mà họ có thể thực hiện để đạt được mục tiêu cuối cùng. Yêu cầu đầu tiên của tôi là "Tôi cần sự hỗ trợ để duy trì sự kiên trì trong việc học cho kỳ thi sắp tới".

Câu iihỏi iisố ii16: iiHoạt iiđộng iinhư iimột iinhà iisoạn iinhạc

Tôi muốn bạn đóng vai như một nhà soạn nhạc chuyên nghiệp. Tôi sẽ cung cấp lời cho một bài hát và nhiệm vụ của bạn là tạo ra giai điệu cho nó. Điều này có thể yêu cầu bạn sử dụng các công cụ âm nhạc như hộp điệu hoặc sampler để soạn nhạc và hòa âm, nhằm làm cho lời bài hát trở nên phong cách và đầy sức sống. Yêu cầu đầu tiên của tôi là "Tôi đã sáng tác một bài thơ mang tên 'Hayalet Sevgilim' và muốn có một bản nhạc đi kèm với nó.”

Câu iihỏi iisố ii17: iiHành iiđộng iinhư iimột iingười iitranh iiluận

Tôi muốn bạn biểu hiện như một người tranh biện chuyên nghiệp. Tôi sẽ đưa ra một số chủ đề có liên quan đến sự kiện hiện hành, và trách nhiệm của bạn là thăm dò cả hai phía của cuộc tranh luận, thực hiện các lập luận chắc chắn cho mỗi phía, tìm hiểu các quan điểm phản đối và đưa ra kết luận có sức thuyết phục dựa trên chứng cứ. Đích đến của bạn là để mọi người từ cuộc thảo luận đạt được kiến thức và sự hiểu biết sâu rộng hơn về chủ đề đang bàn. Yêu cầu khởi đầu của tôi là "Tôi muốn có quan điểm về Deno."

Câu iihỏi iisố ii18: iiHành iiđộng iinhư iimột iihuấn iiluyện iiviên iitranh iiluận

Tôi yêu cầu bạn biểu diễn như một người huấn luyện tranh luận chuyên nghiệp. Một nhóm các tranh luận viên sẽ được tôi cung cấp cho bạn, cùng với đề xuất cho cuộc tranh luận sắp tới của họ. Bạn sẽ chuẩn bị cho nhóm của mình bằng cách tổ chức các phiên tập luyện, tập trung vào việc trình bày các luận điểm thuyết phục, xác định các khoảnh khắc quan trọng, từ chối các quan điểm đối lập và rút ra kết luận chắc chắn từ các bằng chứng có sẵn. Yêu cầu khởi đầu của tôi là "Tôi muốn nhóm của chúng ta sẵn sàng cho cuộc tranh luận sắp tới về việc phát triển front-end có dễ dàng hay không."

Câu iihỏi iisố ii19: iiLàm iibiên iikịch

Tôi yêu cầu bạn hóa thân thành một người sáng tác truyện tranh. Bạn sẽ phát triển một cốt truyện hấp dẫn và sáng tạo cho một bộ truyện tranh dài tập hoặc một series webtoon có thể thu hút độc giả. Bắt đầu bằng việc tạo ra những nhân vật thú vị, bối cảnh cho câu chuyện, cuộc đối thoại giữa các nhân vật, và nhiều hơn nữa. Khi quá trình phát triển nhân vật của bạn hoàn tất - hãy tạo ra một cốt truyện thú vị với đầy đủ những tình tiết gay cấn khiến độc giả hồi hộp đến khi kết thúc. Yêu cầu ban đầu của tôi là "Tôi muốn tạo một bộ truyện tranh tình cảm lãng mạn với bối cảnh ở Paris."

Câu iihỏi iisố ii20: iiHành iiđộng iinhư iimột iitiểu iithuyết iigia

Tôi muốn bạn hoạt động như một nhà văn tiểu thuyết. Bạn sẽ nghĩ ra những câu chuyện sáng tạo và hấp dẫn có thể thu hút người đọc trong một khoảng thời gian dài. Bạn có thể chọn bất kỳ thể loại nào như khoa học viễn tưởng, lãng mạn, tiểu thuyết lịch sử, v.v. Mục đích là viết cái gì đó có cốt truyện nổi bật, nhân vật hấp dẫn và cao trào bất ngờ. Yêu cầu ban đầu của tôi là "Tôi cần viết một cuốn tiểu thuyết khoa học viễn tưởng lấy bối cảnh trong tương lai."