ỨNG DỤNG WINDOWS CHỈNH SỬA VÀ TRÍCH XUẤT THÔNG TIN TỪ ẢNH

LUẬN VĂN CỬ NHÂN

Cao Thành Danh – 1912836

Giảng viên hướng dẫn

ThS. Nguyễn Khánh Lợi

|  |  |
| --- | --- |
|  | ĐẠI HỌC QUỐC GIA THÀNH PHỐ HÒ CHÍ MINH  TRƯỜNG ĐẠI HỌC BÁCH KHOA  KHOA ĐIỆN – ĐIỆN TỬ, BỘ MÔN VIỄN THÔNG |

11 – 2023

ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC BÁCH KHOA

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh Phúc

Số: \_\_\_\_\_\_ /BKĐT

Khoa: **Điện – Điện tử**

Bộ Môn: **Viễn Thông**

**NHIỆM VỤ LUẬN VĂN TỐT NGHIỆP**

1. Họ và tên: Cao Thành Danh MSSV: 1912836
2. Ngành: Điện – Điện tử Chuyên ngành: Kỹ thuật Điện tử - Truyền thông
3. Đề tài: Ứng dụng windows chỉnh sửa và trích xuất thông tin từ ảnh
4. Nhiệm vụ:

* Thiết kế mạch … thiết kế hệ thống …
* Đo đạc thực nghiệm … Mô phỏng hệ thống …
* Phân tích và so sánh kết quả …

1. Ngày giao nhiệm vụ luận văn: 02/10/2023
2. Ngày hoàn thành nhiệm vụ: 04/12/2023
3. Họ và tên người hướng dẫn: Phần hướng dẫn

ThS. Nguyễn Khánh Lợi,

BM Viễn Thông, Khoa Điện – Điện Tử 100%

Nội dung và yêu cầu LVTN đã được thông qua Bộ Môn.

*TP.HCM, ngày 04 tháng 12 năm 2023*

|  |  |
| --- | --- |
| **CHỦ NHIỆM BỘ MÔN** | **NGƯỜI HƯỚNG DẪN CHÍNH** |
|  |  |
| PGS. TS. Hà Hoàng Kha | ThS. Nguyễn Khánh Lợi |

**PHẦN DÀNH CHO KHOA, BỘ MÔN:**

Người duyệt (chấm sơ bộ):

Đơn vị:

Ngày bảo vệ :

Điểm tổng kết:

Nơi lưu trữ luận văn:

# 

# LỜI CÁM ƠN

Lời đầu tiên em xin được cảm ơn tập thể các thầy cô của trường Đại học Bách Khoa, các thầy cô của khoa Điện – Điện tử và các thầy cô ở bộ môn Viễn thông vì đã tận tình chỉ dạy và trang bị cho em những kiến thức về các môn đại cương cũng như các môn chuyên ngành đề làm nền tảng cho việc thực hiện Đồ án cũng như những kinh nghiệm quý báu để nhóm em vững tin hơn trong môi trường làm việc sau này.

Đặc biệt, em xin chân thành cảm ơn thầy Nguyễn Khánh Lợi, người đã tận tình hướng dẫn, giúp đỡ, định hướng, góp ý và cung cấp ý tưởng cũng như tài liệu tham khảo trong thời gian làm Đồ án.

Vì kiến thức bản thân còn nhiều hạn chế, trong quá trình học tập, hoàn thiện báo cáo này em không tránh khỏi những sai lầm. Em rất mong nhận được sự chỉ bảo, đóng góp ý kiến của các thầy cô để em có điều kiện bổ sung, nâng cao ý thức của mình, phục vụ tốt hơn công tác thực tế sau này.

Em xin chân thành cảm ơn .

|  |
| --- |
| TP. HCM, ngày 04, tháng 12 năm 2023 |
|  |
| Cao Thành Danh |

# LỜI CAM ĐOAN

Tôi tên: Cao Thành Danh là sinh viên chuyên ngành Kỹ thuật Điện tử - Truyền thông, khóa 2019, tại Đại học Quốc gia thành phố Hồ Chí Minh – Trường Đại học Bách Khoa. Tôi xin cam đoan những nội dung sau đều là sự thật: (i) Công trình nghiên cứu này hoàn toàn do chính tôi thực hiện; (ii) Các tài liệu và trích dẫn trong luận văn này được tham khảo từ các nguồn thực tế, có uy tín và độ chính xác cao; (iii) Các số liệu và kết quả của công trình này được tôi tự thực hiện một cách độc lập và trung thực.

|  |
| --- |
| TP. HCM, ngày 04, tháng 12 năm 2023 |
|  |
| Cao Thành Danh |

# TÓM TẮT LUẬN VĂN

Tóm tắt luận văn bằng tiếng Việt. Giới hạn trong 1 trang. Nội dung tóm tắt bao gồm: bài toán nghiên cứu (mục tiêu). Phương pháp nghiên cứu. Kết quả nghiên cứu.

# ABSTRACT

Tóm tắt luận văn bằng tiếng Anh. Giới hạn trong 1 trang. Nội dung tóm tắt bao gồm: bài toán nghiên cứu (mục tiêu). Phương pháp nghiên cứu. Kết quả nghiên cứu.

**MỤC LỤC**

[LỜI CÁM ƠN i](#_Toc143557022)

[LỜI CAM ĐOAN ii](#_Toc143557023)

[TÓM TẮT LUẬN VĂN iii](#_Toc143557024)

[ABSTRACT iv](#_Toc143557025)

[DANH SÁCH BẢNG vii](#_Toc143557026)

[DANH SÁCH HÌNH VẼ viii](#_Toc143557027)

[DANH SÁCH TỪ VIẾT TẮT ix](#_Toc143557028)

[CHƯƠNG 1. GIỚI THIỆU 1](#_Toc143557029)

[1.1 Đặt vấn đề 1](#_Toc143557030)

[1.2 Phạm vi và phương pháp nghiên cứu 2](#_Toc143557031)

[1.2.1 Phạm vi nghiên cứu 2](#_Toc143557032)

[1.2.2 Phương pháp nghiên cứu 3](#_Toc143557033)

[1.3 Các đóng góp của luận văn 3](#_Toc143557034)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 5](#_Toc143557035)

[2.1 Chatbot và ứng dụng 5](#_Toc143557036)

[2.1.1 Khái niệm về Chatbot 5](#_Toc143557037)

[2.1.2 Một số ứng dụng của Chatbot 6](#_Toc143557038)

[2.2 Cơ sở lý thuyết về GPT 8](#_Toc143557039)

[2.2.1 GPT - Generative Pre-trained Transformer 8](#_Toc143557040)

[2.2.2 GPT và Mô Hình Transformer 9](#_Toc143557041)

[2.3 Một số kỹ thuật embedding word 11](#_Toc143557042)

[2.3.1 Sentence-BERT 13](#_Toc143557043)

[2.3.2 OpenAI 14](#_Toc143557044)

[2.4 Một số thư viện sử dụng 14](#_Toc143557045)

[2.4.1 Streamlit – build API 14](#_Toc143557046)

[2.4.2 Fitz – tương tác với PDF 14](#_Toc143557047)

[CHƯƠNG 3. KẾT QUẢ VÀ PHÂN TÍCH 14](#_Toc143557048)

[3.1 Phương pháp tiếp cận 14](#_Toc143557049)

[3.1.1 Pipeline data 14](#_Toc143557050)

[3.2 Kết quả và phân tích 15](#_Toc143557051)

[3.2.1 Khảo sát thông số A (ghi chú: các mục này nên ghi theo câu hỏi nghiên cứu) 15](#_Toc143557052)

[3.2.2 Kết quả mô phỏng thông số B 15](#_Toc143557053)

[3.3 Kết luận chương 15](#_Toc143557054)

[CHƯƠNG 4. KẾT LUẬN 16](#_Toc143557055)

[4.1 Tóm tắt và kết luận chung 16](#_Toc143557056)

[4.2 Hướng phát triển 16](#_Toc143557057)

[PHỤ LỤC A 17](#_Toc143557058)

[A.1 Code chương trình giao tiếp Arduino 17](#_Toc143557059)

[A.2 Code chương trình xử lý dữ liệu dùng Matlab 17](#_Toc143557060)

[TÀI LIỆU THAM KHẢO 18](#_Toc143557061)

# DANH SÁCH BẢNG

**No table of figures entries found.**

# DANH SÁCH HÌNH VẼ

[Hình 2.3‑1 Giới thiệu về Chroma 13](#_Toc143557014)

# DANH SÁCH TỪ VIẾT TẮT

|  |  |
| --- | --- |
| SNR | Signal to noise ratio |
| Ghi chú | Ghi các từ viết tắt sử dụng trong luận văn vào bảng này theo thứ tự alphabet |

# GIỚI THIỆU

## Đặt vấn đề

Trong thời đại số hóa ngày nay, việc chỉnh sửa ảnh và trích xuất thông tin từ hình ảnh đã trở nên vô cùng phổ biến và quan trọng. Cùng với sự phát triển của công nghệ, các công cụ chỉnh sửa ảnh đã trở nên dễ dàng sử dụng hơn bao giờ hết, cho phép người dùng tạo ra những tác phẩm sáng tạo, tinh chỉnh hình ảnh theo ý thích của mình. Đồng thời, việc trích xuất thông tin từ hình ảnh cũng vô cùng quan trọng, đặc biệt trong các lĩnh vực như trí tuệ nhân tạo, khoa học dữ liệu, y học, an ninh mạng và nhiều lĩnh vực khác. Tuy nhiên việc trích xuất chỉnh sửa ảnh qua các phần mềm phức tạp và trích xuất thông tin thủ công lại thường gây nhiều khó khăn.

Trong vai trò là một người dùng cần sự đơn giản và hiệu quả trong công việc, cần có một công cụ hỗ trợ các thao tác xử lý liên quan đến hình ảnh nhanh chóng. Cần khắc phục được các nhược điểm thường gặp sau:

* Quá trình xử lý quá phức tạp: Việc chỉnh sửa một bức ảnh như xóa một vật thể, chỉnh màu, cũng như là thay đổi kích thước ảnh cần cài một phần mềm chuyên biệt, quà trình cài đặt lâu và cần phải nghiên cứu thao tác sử dụng nhiều. Việc này gây trở ngại với nhu cầu xử lý nhanh gọn
* Xử lý thủ công mất thời gian: Việc trích xuất thông tin từ ảnh bằng mắt thường đã là một điều tốn rất nhiều thơi gian và công sức, bên cạnh đó còn kết hợp với việc nhập liệu và lưu trữ bằng tay khiến các công việc này lặp đi lặp lại quà nhiều lần. Gây lãng phí thời gian và nhân lực cho việc này.
* Không hiệu quả: Việc thực hiện các công việc trên vừa mất nhiều thời gian nhưng đôi lúc lại không thật sự chính xác và kịp thời cho các nhu cầu tức thì.

Trước những thách thức nêu trên, sự kết hợp giữa xử lý ảnh và thị giác máy tính đã mở ra một hướng giải quyết tiềm năng. Cụ thể, việc sử dụng kĩ thuật Inpaint thông qua OpenCV và OCR để phát triển một ứng dụng trên Windows dùng WinUI giúp chỉnh sửa và trích xuất thông tin từ ảnh giúp giải quyết các vấn đề trên một cách hiệu quả.

OpenCV là một thư viện mã nguồn mở rộng lớn và mạnh mẽ dành cho xử lý ảnh và thị giác máy tính. Viết tắt của "Open Source Computer Vision Library," OpenCV cung cấp một loạt các công cụ và chức năng để xử lý ảnh số, nhận diện vật thể, phát hiện đối tượng, xử lý video và nhiều ứng dụng khác trong lĩnh vực thị giác máy tính. OpenCV giúp người dùng dễ dàng thực hiện các tác vụ xử lý ảnh và thị giác máy tính một cách hiệu quả, từ các thao tác cơ bản đến các ứng dụng phức tạp.

OCR là viết tắt của "Optical Character Recognition" (Nhận diện ký tự quang học). Đây là một công nghệ dùng để chuyển đổi các hình ảnh hoặc các tài liệu chứa văn bản, ký tự thành dạng văn bản có thể chỉnh sửa được trên máy tính. Công nghệ OCR có nhiều ứng dụng rộng rãi, từ việc chuyển đổi tài liệu giấy sang dạng điện tử, tạo điều kiện cho việc tìm kiếm và chỉnh sửa văn bản, đến việc xử lý tự động thông tin từ các tài liệu, hóa đơn, bằng lái xe, thẻ ID, và nhiều ứng dụng khác trong các lĩnh vực như y tế, ngân hàng, công nghiệp, văn phòng và quản lý tài liệu.

Mục tiêu của đề tài là phát triển một ứng dụng chỉnh sửa ảnh hiệu quả và trích xuất thông tin nhanh chóng từ ảnh trong các trường hợp sử dụng phổ biến như: Xóa bỏ vật thể, làm mờ vùng ảnh, thực hiện thay đổi màu ảnh qua các filter, thay đổi kích thước ảnh, trích xuất thông tin từ ảnh có chứa văn bản nhanh chóng v.v… Ứng dụng được kỳ vọng sẽ giải quyết các vấn đề trên tiết kiệm thời gian và nâng cao trải nghiệm trong công việc.

Để đạt được mục tiêu của đề tài, chúng ta sẽ tiến hành các bước sau:

1. Nghiên cứu OpenCV và nguyên lý hoạt động của nó.

2. Tìm hiểu các kĩ thuật xử lý ảnh như: Inpaint (xóa vật thể khỏi ngữ cảnh), xử lý màu.

3. Nghiên cứu OCR và cách vận hành trong C Sharp

4. Xây dựng ứng dụng dùng .Net 6.0 sử dụng Framwork WinUI 3 để xây dựng ưng dụng trên Windows

5. Thực hiện kiểm tra, đánh giá và cải thiện hiệu quả thực hiện của ứng dụng.

## Phạm vi và phương pháp nghiên cứu

### Phạm vi nghiên cứu

Đề tài tập trung vào việc phát triển một ứng dụng hỗ trợ người dùng có nhu cầu chỉnh sửa ảnh và trích xuất thông tin nhanh chóng. Phạm vi của nghiên cứu bao gồm:

Xây dựng giao diện và tương tác người dùng: Thiết kế một giao diện trực quan và thân thiện giúp người dùng tương tác với ứng dụng trên Windows quen thuộc.

Xây dựng kiến trúc ứng dụng: Phát triển kiến trúc phần mềm phù hợp cho việc quản lý hệ thống và xử lý dữ liệu ảnh và văn bản.

Xây dựng các đầu ra phù hợp: Lực chọn các dạng đầu ra của phần mềm từ ảnh, văn bản dưới nhiều hình thức phù hợp với nhu cầu người dùng.

### Phương pháp nghiên cứu

Trong quá trình thực hiện đề tài, chúng ta sẽ áp dụng một loạt các phương pháp để đạt được mục tiêu đã đề ra:

Nghiên cứu và tìm hiểu OpenCV: Đầu tiên, tiến hành nghiên cứu chi tiết về OpenCV, hiểu cách thư viện này hoạt động và khả năng của nó trong việc xử lý ảnh.

Thiết kế giao diện dùng WinUI 3 Framwork: Dựa trên giao diện Windows phổ biến tiến hành nghiên cứu sử dụng các thành phần giao diện kết hợp với nhau để ra được giao diện phù hợp. Bên cạnh đó cũng cần chú trọng việc tổ chức thuận tiện cho việc phát triển và bảo trì.

Nghiên cứu thành phần chỉnh sửa ảnh: Sử dụng các các thành phần hỗ trợ của OpenCv và pilgram, chúng ta sẽ xây dựng một số tính năng xử lý ảnh có khả năng xử lý yêu cầu xóa vật thể và thay đổi màu ảnh.

Nghiên cứu trích xuất thông tin thông qua OCR: Tìm hiểu về lý thuyết, nguyên lý và cách thức triển khai của OCR. Đưa kiến thức này vào ứng dụng để trích xuất thông từ đơn giản đến phức tạp. Giúp quá trình trích xuất thông tin đỡ tốn quá nhiều chi phí.

Thử nghiệm và đánh giá: Cuối cùng, chúng ta sẽ tiến hành các thử nghiệm với trường hợp xử lý ảnh và trích xuất thông tin thực tế để đánh giá hiệu suất của ứng dụng Windows. Mọi vấn đề gặp phải sẽ được ghi nhận và cải thiện ứng dụng.

## Các đóng góp của luận văn

Các đóng góp chính của đề tài này bao gồm:

* Xứ lý ảnh đơn giản, nhanh gọn và chính xác

Ứng dụng Windows trong đề tài sẽ giúp cải thiện trải nghiệm xử lý ảnh đơn. Thay vì phải tốn thời gian và công sức để sử dụng một ứng dụng phức, người dùng có thể tương tác với ứng dụng đơn giản và thuận tiện để xóa vật thể và chỉnh sửa ảnh.

* Sử dụng OCR trích xuất tự động:

Sử dụng kĩ thuật OCR, đề tài này giúp kết hợp xử lý ảnh và thị giác máy để xây dựng một ứng dụng thông minh có khả năng trích xuất thông tin tự động theo nhu cầu của người dùng. Điều này giúp tiết kiệm thời gian và công sức trong việc trích xuất thông tin bằng thủ công.

* Môi trường xây dựng quen thuộc và dễ sử dụng:

Việc lựa chọn xây dựng ứng dụng trên Windows vì đây là một trong những hệ điều hành được sử dụng phố biến hiện nay. Qua đó người dùng không cần mất quá nhiều thời gian tìm hiểu và vận dụng công việc để mang lại hiệu quả kịp thời.

# CƠ SỞ LÝ THUYẾT

## Chatbot và ứng dụng

### Khái niệm về Chatbot

Trong bài báo [[1](#_ENREF_1)], tác giả có đề cập “Chatbot là một phần mềm hỗ trợ giao tiếp giữa người và máy. Các chatbot được cài đặt một lượng thông tin và có khả năng đưa ra câu trả lời hoặc các quyết định khi người dùng truy vấn.” Với chatbot, người dùng có thể giao tiếp với các hệ thống bằng ngôn ngữ tự nhiên (natural language) mà không cần phải sử dụng các ngôn ngữ lập trình phức tạp. Chatbot đóng vai trò như một trợ lý ảo, có thể được sử dụng trong nhiều lĩnh vực khác nhau như kinh doanh, y tế và cả trong giáo dục.

* Phân loại chatbot
* Theo mục đích sử dụng:
  + Chatbot hỗ trợ khách hàng (Customer Support Chatbot): Thiết kế để giải quyết các vấn đề, trả lời câu hỏi và cung cấp hỗ trợ cho khách hàng trong thời gian thực.
* Chatbot giáo dục (Educational Chatbot): Cung cấp thông tin và kiến thức cho người dùng, thường được sử dụng trong môi trường học tập.
* Chatbot giải trí (Entertainment Chatbot): Tạo nội dung giải trí như đùa cợt, kể chuyện, hoặc thực hiện trò chơi cùng người dùng.
* Chatbot thương mại (Commerce Chatbot): Hỗ trợ quá trình mua sắm trực tuyến, tư vấn sản phẩm, và giúp khách hàng tìm kiếm thông tin về sản phẩm, dịch vụ.
* Theo khả năng:
* Chatbot dựa trên nguyên tắc (Rule-Based Chatbot): Hoạt động dựa trên các quy tắc và kịch bản được lập trình trước. Các phản hồi của chatbot bị giới hạn trong phạm vi những gì đã được xác định trước.
* Chatbot dựa trên trí tuệ nhân tạo (Artificial Intellinge Chatbot): Sử dụng trí tuệ nhân tạo tự động học tập từ dữ liệu người dùng và cải thiện khả năng hoạt động.
* Theo cách thức hoạt động:
* Chatbot dựa trên văn bản (Text-Based Chatbot): Giao tiếp với người dùng thông qua văn bản, thường là qua các ứng dụng nhắn tin, trang web, hoặc ứng dụng di động.
* Chatbot dựa trên giọng nói (Voice-Based Chatbot): Sử dụng nhận dạng giọng nói để tương tác với người dùng thông qua cuộc trò chuyện giọng nói.
* Theo loại công nghệ:
* Chatbot truyền thống: Dựa trên ngôn ngữ và các thuật toán xử lý ngôn ngữ tự nhiên truyền thống.
* Chatbot học sâu (Deep Learning Chatbot): Sử dụng các kiến thức từ học sâu và mạng nơ-ron để hiểu và tạo ra câu trả lời tự nhiên hơn.
* Theo ngữ cảnh ứng dụng:
* Chatbot dành cho doanh nghiệp (Enterprise Chatbot): Được triển khai trong môi trường doanh nghiệp để hỗ trợ nhiều tác vụ như quản lý dự án, tư vấn nhân sự, quản lý hỗ trợ kỹ thuật.
* Chatbot xã hội: Được sử dụng trong các nền tảng mạng xã hội để tương tác với người dùng trong môi trường xã hội.

### Một số ứng dụng của Chatbot

* Dịch vụ khách hàng: Chatbot có thể giúp cung cấp hỗ trợ tức thì cho các câu hỏi thường gặp và vấn đề kỹ thuật. Chúng có khả năng cung cấp thông tin, hướng dẫn sửa lỗi và giải quyết vấn đề của khách hàng.

Ví dụ: [[2](#_ENREF_2)] Answer Bot là một chatbot hỗ trợ khách hàng được tích hợp vào hệ thống Zendesk. Khi khách hàng gửi một câu hỏi, Answer Bot sẽ tự động tìm kiếm trong cơ sở dữ liệu hỗ trợ để cung cấp câu trả lời thích hợp

* Bán hàng và Thương mại điện tử: Chatbot có thể giúp tư vấn sản phẩm, giải đáp câu hỏi về thông tin sản phẩm, quy trình đặt hàng và thanh toán, tạo ra trải nghiệm mua sắm trực tuyến thuận tiện.

Ví dụ: [[3](#_ENREF_3)] Sephora Virtual Artist là một chatbot cho phép người dùng thử nghiệm mỹ phẩm ảo trên khuôn mặt của họ thông qua camera điện thoại, giúp họ chọn sản phẩm mỹ phẩm phù hợp.

* Lĩnh vực Sức khỏe: Chatbot có thể cung cấp thông tin về triệu chứng bệnh, chế độ dinh dưỡng, lời khuyên về sức khỏe và thậm chí lên lịch hẹn với bác sĩ.

Ví dụ: [[4](#_ENREF_4)] Babylon Health là một ứng dụng hỗ trợ sức khỏe cá nhân có chatbot tích hợp, giúp người dùng tư vấn về triệu chứng bệnh, đặt lịch hẹn với bác sĩ và cung cấp lời khuyên về sức khỏe

* Ngân hàng và Tài chính: Chatbot có thể cung cấp thông tin về tài khoản, giao dịch, lãi suất, và thậm chí thực hiện một số giao dịch cơ bản như chuyển khoản.

Ví dụ: [[5](#_ENREF_5)] Erica là chatbot của Bank of America, giúp người dùng kiểm tra số dư tài khoản, thực hiện chuyển khoản, đặt lịch hẹn với chuyên viên tài chính và cung cấp thông tin về tài chính cá nhân.

* Giáo dục: Chatbot có thể giúp đáp án các câu hỏi học tập, cung cấp tài liệu tham khảo, hướng dẫn về chương trình học, và cung cấp hỗ trợ cho việc học từ xa.

Ví dụ: Duolingo sử dụng chatbot trong quá trình học ngôn ngữ. Người dùng có thể tương tác với các chatbot để thực hành giao tiếp và rèn luyện khả năng ngôn ngữ

* Lĩnh vực Du lịch: Chatbot có thể tư vấn về điểm đến, đặt vé máy bay, đặt khách sạn, và cung cấp thông tin về lịch trình du lịch.

Ví dụ: [[6](#_ENREF_6)] Expedia sử dụng ChatGPT trong ứng dụng di động của họ để hỗ trợ người dùng tìm kiếm thông tin về chỗ ở, vé máy bay, và lựa chọn điểm đến du lịch.

* Lĩnh vực Nhân sự: Chatbot có thể thực hiện vòng phỏng vấn sơ bộ, giúp tìm kiếm và tuyển dụng ứng viên, cung cấp thông tin về chính sách nhân sự và quy trình làm việc trong công ty.

Ví dụ: [[7](#_ENREF_7)] Mya là một chatbot dành cho quá trình tuyển dụng và nhân sự. Nó có khả năng tương tác với ứng viên, thu thập thông tin, và tự động lên lịch phỏng vấn.

* Lĩnh vực Nghiên cứu thị trường: Chatbot có thể thực hiện khảo sát, thu thập dữ liệu từ người dùng về thị trường, xu hướng, và ý kiến về sản phẩm/dịch vụ.

Ví dụ: [[8](#_ENREF_8)] SurveySparrow cho phép người dùng tạo các khảo sát bằng chatbot, giúp tăng khả năng tham gia và cung cấp thông tin chính xác hơn về thị trường.

* Học tập trực tuyến: Chatbot có thể cung cấp bài giảng, lời giải cho bài tập, và hỗ trợ học tập cho sinh viên trong khóa học trực tuyến.

Ví dụ: Coursera sử dụng chatbot AI để cung cấp thông tin về khóa học, hỗ trợ học tập, và giải đáp câu hỏi của sinh viên trong thời gian thực.

* Giải trí và Thư giãn: Chatbot có thể tương tác với người dùng qua việc chơi trò chơi, kể truyện, cung cấp gợi ý về phim và âm nhạc, giúp người dùng giải trí và thư giãn.

Ví dụ: [[9](#_ENREF_9)] Woebot là một chatbot dành cho hỗ trợ tâm lý và thư giãn. Nó sử dụng kỹ thuật dựa trên dữ liệu và tâm lý học để tương tác với người dùng và cung cấp lời khuyên về tâm lý

## Cơ sở lý thuyết về GPT

### GPT - Generative Pre-trained Transformer

#### Định nghĩa GPT và bối cảnh ra đời của nó trong lĩnh vực xử lý ngôn ngữ tự nhiên.

GPT là viết tắt của Generative Pre-trained Transformer, là một mô hình ngôn ngữ tự hồi quy (autoregressive) được phát triển bởi OpenAI – nơi tạo ra DALLE-2, trình chuyển văn bản thành hình ảnh (cũng là một mô hình GPT). ChatGPT và DALLE-2 được đào tạo trên cơ sở của GPT-3, phiên bản thứ ba của mô hình Generative AI được đào tạo dựa trên dữ liệu văn bản từ các nguồn trực tuyến như Wiki, các trang web, bài báo, sách…

[[10](#_ENREF_10)] Theo TS. Đặng Minh Tuấn, người đứng đầu Lab Blockchain tại PTIT và là Viện trưởng của Viện nghiên cứu Ứng dụng công nghệ CMC, việc ra đời của ChatGPT thể hiện sự cần thiết không thể tránh trong bối cảnh công nghệ thông tin đang phát triển mạnh mẽ hiện nay.

GPT-1 (2018): GPT-1, phiên bản đầu tiên của GPT, đã ra mắt vào năm 2018. Mặc dù có khả năng tạo ra văn bản tự nhiên và trả lời các câu hỏi một cách logic, nhưng nó vẫn có những hạn chế về độ chính xác và hiểu biết ngôn ngữ. Tuy vậy, GPT-1 đã tạo ra sự chú ý đáng kể và đánh dấu sự bắt đầu của cuộc hành trình phát triển của Chat GPT.

GPT-2 (2019): Phiên bản GPT-2 tiếp theo được công bố vào năm 2019. Với hơn 1,5 tỷ tham số, GPT-2 đã nâng cao đáng kể khả năng tạo ra văn bản tự nhiên và giao tiếp với con người. GPT-2 đạt được độ chính xác cao hơn so với phiên bản trước đó, thu hút sự quan tâm của cộng đồng trí tuệ nhân tạo.

GPT-3 (2020): GPT-3 đã gây ra một cú sốc lớn trong lĩnh vực trí tuệ nhân tạo khi ra mắt vào năm 2020. Với hơn 175 tỷ tham số, GPT-3 trở thành một trong những mô hình mạng nơ-ron lớn nhất và mạnh mẽ nhất từ trước đến nay. Khả năng linh hoạt của GPT-3 cho phép nó thực hiện nhiều tác vụ khác nhau, từ tạo văn bản tự nhiên đến giải quyết các vấn đề phức tạp.

Chat GPT-3.5: Chat GPT-3.5 là phiên bản mới nhất và mạnh mẽ của GPT, phát triển bởi OpenAI. Mô hình này có khả năng xử lý và hiểu ngôn ngữ tự nhiên với mức độ chân thực cao. Với khả năng hiểu ngữ cảnh, Chat GPT-3.5 có thể tạo ra văn bản tự nhiên và chính xác, đáp ứng các câu hỏi, cung cấp thông tin, thậm chí tạo nội dung mới.

GPT đặc biệt ấn tượng với khả năng học từ dữ liệu không gian mẫu ngôn ngữ lớn và chuyển giao kiến thức này cho các tác vụ cụ thể. Mô hình này thường được "pre-trained" trước đó trên một lượng lớn văn bản để học cấu trúc ngôn ngữ và kiến thức tổng quát. Sau đó, nó có thể được "fine-tuned" trên các nhiệm vụ cụ thể như phân loại văn bản, tạo nội dung, hay trả lời câu hỏi.

### GPT và Mô Hình Transformer

Transformer là một thuật toán mạnh mẽ cho xử lý bảng dữ liệu liên tục (sequence data) trong Deep Learning. Nó được giới thiệu bởi Vaswani et al. trong bài báo [[11](#_ENREF_11)] năm 2017.

Trong Transformer, các từ trong một câu hoặc văn bản được biểu diễn dưới dạng một ma trận vector và các quan hệ giữa các từ được xử lý bằng cách sử dụng một mechanism attention. Điều này cho phép mô hình hoạt động độc lập với độ dài của đầu vào và có thể hiểu được mối quan hệ giữa các từ trong câu một cách linh hoạt hơn so với các mô hình truyền thống như RNN hoặc LSTM.

Đối với chatbot GPT, Transformer được sử dụng để xử lý dữ liệu văn bản và tìm ra đáp án phù hợp nhất cho câu hỏi của người dùng. Mô hình Transformer của GPT được huấn luyện trên lượng dữ liệu lớn để có thể hiểu được ngữ cảnh và tìm ra câu trả lời phù hợp nhất cho câu hỏi của người dùng.

* Khác biệt giữa GPT và Mô Hình Transformer:

Mục tiêu ứng dụng: Mô hình Transformer được thiết kế để xử lý ngôn ngữ tự nhiên và dịch máy. GPT tập trung vào việc tạo văn bản tự nhiên và tạo ra các dãy từ logic và hợp lý.

Bộ giải mã: Mô hình Transformer thông thường bao gồm cả bộ mã hóa và bộ giải mã. Trong khi đó, GPT chỉ sử dụng bộ mã hóa để tạo dữ liệu đầu vào cho việc dự đoán từ tiếp theo.

Tính sáng tạo: GPT có khả năng tạo ra văn bản mới và sáng tạo dựa trên thông tin đầu vào, trong khi mô hình Transformer thường được sử dụng để thực hiện nhiệm vụ dịch máy hoặc phân loại.

Công dụng ứng dụng: GPT thường được sử dụng trong các ứng dụng tạo nội dung, chatbot, viết tự động và tương tác với người dùng, trong khi mô hình Transformer có thể được áp dụng rộng rãi trong nhiều tác vụ xử lý ngôn ngữ tự nhiên.

Tóm lại, GPT và mô hình Transformer có quan hệ chặt chẽ với nhau, trong đó GPT là một biến thể của kiến trúc mạng Transformer, tập trung vào việc tạo ra văn bản tự nhiên và tương tác ngôn ngữ tự nhiên với người dùng.

## Một số kỹ thuật embedding word

[[12](#_ENREF_12)] Embedding là một kỹ thuật đưa một vector có số chiều lớn, thường ở dạng thưa, về một vector có số chiều nhỏ, thường ở dạng dày đặc. Phương pháp này đặc biệt hữu ích với những đặc trưng hạng mục có số phần tử lớn ở đó phương pháp chủ yếu để biểu diễn mỗi giá trị thường là một vector dạng one-hot.

Embeddings là cách để biểu diễn dữ liệu dưới góc nhìn đặc biệt của trí tuệ nhân tạo, giúp chúng trở thành lựa chọn tuyệt vời cho việc làm việc với các công cụ và thuật toán dựa trên trí tuệ nhân tạo. Chúng có thể ánh xạ dữ liệu từ văn bản, hình ảnh và trong tương lai có thể áp dụng cho âm thanh và video. Việc tạo embeddings có nhiều phương pháp khả thi, có thể thực hiện cục bộ bằng cách sử dụng thư viện đã cài đặt hoặc thông qua việc gọi các API.

Các mô hình ngôn ngữ lớn có tính sinh như GPT, PaLM và các mô hình tương tự được đào tạo trên khối lượng lớn dữ liệu. Nhưng chúng không sử dụng văn bản từ tập dữ liệu một cách trực tiếp, bởi vì máy tính không thể hiểu văn bản như con người, chúng chỉ hiểu số. Đây là lý do tại sao chúng ta cần đến khái niệm "embeddings" hoặc "vector embeddings". Embeddings chính là cách biểu diễn của văn bản, nhưng trong định dạng số hóa.

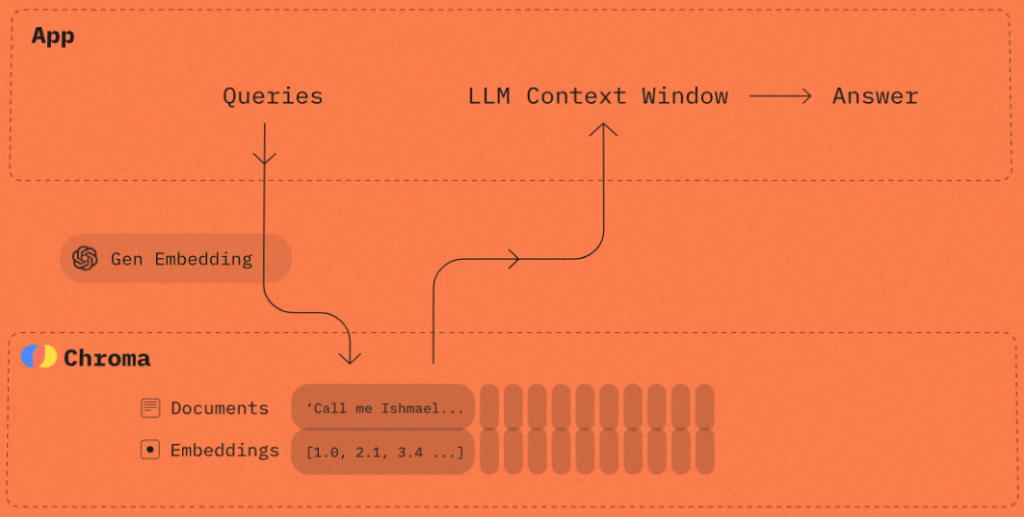
[[13](#_ENREF_13)] Tất cả thông tin gửi và nhận từ các mô hình ngôn ngữ lớn đều thông qua những embeddings này. Tuy nhiên, việc truy cập trực tiếp vào các embeddings này yêu cầu thời gian và tốn kém. Đó là lý do tại sao chúng ta cần đến khái niệm "Cơ sở dữ liệu Vector" hoặc "Vector Databases", những hệ thống lưu trữ được thiết kế đặc biệt để hiệu quả lưu trữ và truy xuất các vector embeddings.

Hệ thống quản lý cơ sở dữ liệu liên hệ truyền thống (RDMS - Relational Database Management Systems) không thể dùng để lưu trữ các vector embeddings này. Đây là lý do tại sao Vector Store / Vector Databases được phát triển để đảm nhận vai trò này. Vector Databases được thiết kế để hiệu quả lưu trữ và truy xuất các vector embeddings. [[13](#_ENREF_13)] Hiện nay, có nhiều Vector Stores khác nhau, chúng có sự khác biệt về các mô hình embedding mà chúng hỗ trợ và loại thuật toán tìm kiếm chúng sử dụng để tìm các vector tương tự.

Hãy xem xét một ví dụ về Chatbot dựa trên một tập tin PDF. Khi người dùng nhập một truy vấn, bước quan trọng đầu tiên là trích xuất nội dung liên quan từ tập tin PDF cho truy vấn đó và cung cấp thông tin này cho Chatbot. Điều này cho phép Chatbot sử dụng thông tin liên quan đến truy vấn để cung cấp câu trả lời thích hợp cho người dùng. Vậy làm thế nào chúng ta có thể tìm nội dung liên quan từ tập tin PDF dựa trên truy vấn của người dùng? Câu trả lời là phương pháp tìm kiếm sự tương đồng đơn giản.

Khi dữ liệu được biểu diễn dưới dạng các vector embeddings, chúng ta có thể tìm sự tương đồng giữa các phần khác nhau của dữ liệu và trích xuất dữ liệu tương tự dựa trên một vector embedding cụ thể. Trước hết, truy vấn được chuyển đổi thành embeddings bởi một mô hình embedding, sau đó Vector Store sử dụng vector embedding này và thực hiện tìm kiếm sự tương đồng (qua các thuật toán tìm kiếm) với các vector embeddings khác mà nó đã lưu trữ trong cơ sở dữ liệu, sau đó truy xuất tất cả dữ liệu liên quan. Các vector embeddings liên quan này sau đó được truyền đến Mô hình Ngôn ngữ Lớn, tức là chatbot, mà sử dụng thông tin này để tạo ra một câu trả lời cuối cùng cho người dùng.

Chroma DB là một cơ sở dữ liệu phổ biến và open source, được sử dụng rộng rãi để giải quyết vấn đề lưu trữ và truy xuất các vector embeddings. [[14](#_ENREF_14)] Chroma đơn giản hóa việc sử dụng các nhà cung cấp embedding phổ biến thông qua việc cung cấp các bao bọc nhẹ. Điều này giúp ta dễ dàng tích hợp chúng vào ứng dụng của mình. Có thể thiết lập một hàm embedding khi tạo một Chroma collection. Hàm này sẽ được tự động áp dụng khi cần, hoặc ta có thể tự mình gọi chúng theo yêu cầu.



Hình ‑ Giới thiệu về Chroma

### Sentence-BERT

Mặc định, Chroma sử dụng mô hình all-MiniLM-L6-v2 dựa theo [*Sentence Transformers*](https://arxiv.org/pdf/1908.10084.pdf) để tạo ra các vector embeddings. Mô hình embedding này có thể tạo ra các vector embedding cho câu và văn bản, có thể được sử dụng cho nhiều nhiệm vụ khác nhau. Chức năng embedding này chạy cục bộ trên máy tính của bạn và có thể yêu cầu bạn tải xuống các tệp mô hình (điều này sẽ xảy ra tự động).

Bài báo [[15](#_ENREF_15)] giới thiệu một phương pháp tạo ra các biểu diễn (embeddings) cho các câu sử dụng mạng BERT (Bidirectional Encoder Representations from Transformers) và kiến trúc mạng Siamese.

Phương pháp được đề xuất tập trung vào việc tạo ra các biểu diễn cho các câu sao cho rằng các biểu diễn này cần phải chứa thông tin về sự tương đồng giữa các câu. Để làm được điều này, các mạng BERT được sử dụng để biểu diễn các từ trong các câu. Sau đó, mạng Siamese được áp dụng để tạo ra biểu diễn tổng hợp cho toàn bộ câu dựa trên biểu diễn từ của chúng.

Kiến trúc mạng Siamese là một mạng nơ-ron gồm hai nhánh (branches), mỗi nhánh biểu diễn một câu riêng. Hai nhánh này chia sẻ các trọng số, điều này giúp tăng cường khả năng học thông tin về sự tương đồng giữa các cặp câu.

Bài báo [[15](#_ENREF_15)] cũng đề cập đến việc huấn luyện mô hình dựa trên một bộ dữ liệu chứa các cặp câu tương tự và không tương tự để mô hình có thể học cách tạo ra các biểu diễn thích hợp cho cả hai trường hợp.

Phương pháp "Sentence-BERT" chứng minh khả năng tạo ra các biểu diễn câu mạnh mẽ, có khả năng tương đồng và phản ánh độ tương tự giữa các câu một cách hiệu quả. Điều này có thể ứng dụng rộng rãi trong các nhiệm vụ liên quan đến xử lý ngôn ngữ tự nhiên, như phân loại văn bản, tạo nội dung và phân tích tương đồng giữa các văn bản.

### OpenAI

## Một số thư viện sử dụng

### Streamlit – build API

Streamlit là một framework mạnh mẽ mã nguồn mở để xây dựng các ứng dụng web tương tác, và nó cũng có thể được sử dụng để xây dựng các API đơn giản. Nó được thiết kế để dễ sử dụng và trực quan, giúp các nhà khoa học dữ liệu và kỹ sư nhanh chóng tạo ra các ứng dụng web có thể chia sẻ ý tưởng của họ với người khác. Khi sử dụng Streamlit để xây dựng một ứng dụng web, thường ta có thể tạo ra các endpoint để giao tiếp với ứng dụng của mình thông qua API.

Streamlit sử dụng Python làm ngôn ngữ lập trình chính, vì vậy các nhà khoa học dữ liệu và kỹ sư có thể tận dụng các kỹ năng và kiến thức hiện có của họ. Streamlit cũng cung cấp một bộ công cụ tích hợp sẵn để tạo các thành phần ứng dụng phổ biến, chẳng hạn như biểu đồ, bảng và thanh trượt.

#### Ưu điểm

**Dễ sử dụng**: Streamlit được thiết kế để dễ sử dụng, ngay cả đối với những người mới bắt đầu. Các nhà khoa học dữ liệu và kỹ sư có thể nhanh chóng tạo ra các ứng dụng web có thể chia sẻ ý tưởng của họ với người khác.

**Trực quan**: Streamlit sử dụng giao diện người dùng trực quan, giúp người dùng dễ dàng tương tác với các ứng dụng web.

**Tích hợp sẵn**: Streamlit cung cấp một bộ công cụ tích hợp sẵn để tạo các thành phần ứng dụng phổ biến, giúp tiết kiệm thời gian và công sức.

**Mở rộng**: Streamlit có thể được mở rộng bằng cách sử dụng Python, cho phép các nhà khoa học dữ liệu và kỹ sư tạo ra các ứng dụng web có tính năng cao.

#### Ứng dụng chính

Các ứng dụng chính của Streamlit:

* Ứng dụng web hiển thị kết quả của mô hình học máy
* Ứng dụng web cho phép người dùng tương tác với dữ liệu
* Ứng dụng web để tạo đồ thị và biểu đồ
* Ứng dụng web để chia sẻ ý tưởng với người khác

### Fitz – tương tác với PDF

Fitz là một thư viện Python được sử dụng để làm việc với tài liệu PDF. Thư viện này cho phép bạn đọc, viết và chỉnh sửa các tập tin PDF bằng cách sử dụng mã nguồn Python. Fitz dựa trên thư viện Poppler, một thư viện mã nguồn mở cho việc xử lý tài liệu PDF.

Với Fitz, bạn có thể thực hiện các tác vụ như trích xuất văn bản từ PDF, trích xuất hình ảnh, chèn hình ảnh và văn bản vào tập tin PDF, tạo ra các trang PDF mới và thậm chí làm việc với các tệp tin PDF mã hóa bảo mật.

#### Ứng dụng chính

**Trích xuất văn bản từ PDF**: Fitz cho phép bạn trích xuất văn bản từ các trang trong tài liệu PDF. Bạn có thể sử dụng chức năng **get\_text()** để lấy nội dung văn bản và xử lý nó theo cách bạn muốn.

**Trích xuất hình ảnh từ PDF**: Fitz cung cấp khả năng trích xuất hình ảnh từ các trang PDF. Bạn có thể sử dụng các phương thức như **get\_pixmap()** để lấy hình ảnh từ trang và lưu chúng dưới dạng các đối tượng hình ảnh.

**Chèn hình ảnh và văn bản vào PDF**: Bạn có thể sử dụng Fitz để thêm hình ảnh và văn bản vào tài liệu PDF. Việc này rất hữu ích khi bạn muốn tạo các trang bìa, chú thích hoặc bất kỳ phần tử nào khác vào tài liệu.

**Chỉnh sửa tài liệu PDF**: Fitz cho phép bạn thực hiện các thay đổi trong tài liệu PDF, chẳng hạn như xoay trang, thay đổi kích thước trang, thay đổi font chữ và màu sắc.

**Tạo trang PDF mới**: Bạn có thể sử dụng Fitz để tạo các trang PDF mới và thêm nội dung vào chúng. Điều này rất hữu ích khi bạn muốn tạo tài liệu PDF động từ dữ liệu tự động sinh ra.

**Xử lý tệp PDF mã hóa bảo mật**: Fitz cũng hỗ trợ làm việc với các tài liệu PDF được mã hóa bảo mật. Bạn có thể thực hiện các thao tác như trích xuất nội dung từ tài liệu mã hóa, cung cấp mật khẩu để mở tài liệu, và thậm chí là gỡ bỏ mật khẩu khỏi tài liệu.

**Trích xuất metadata**: Fitz cho phép bạn truy cập và trích xuất thông tin metadata từ tài liệu PDF, chẳng hạn như tiêu đề, tác giả, ngày tạo và nhiều thông tin khác.

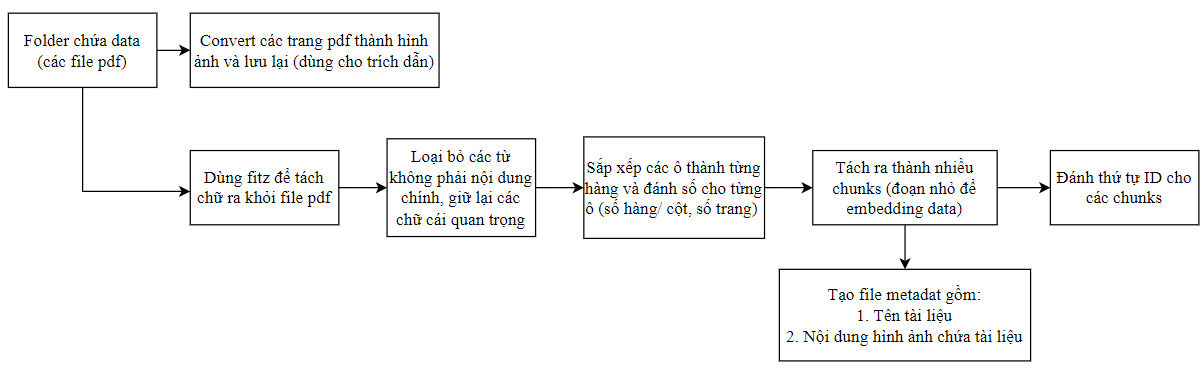
**Chuyển đổi PDF sang hình ảnh**: Fitz cung cấp các phương thức để chuyển đổi các trang PDF thành các hình ảnh, giúp bạn tạo ra các tệp hình ảnh PNG, JPEG hoặc TIFF từ tài liệu PDF.

# KẾT QUẢ VÀ PHÂN TÍCH

## Phương pháp tiếp cận

### Pipeline data

Thu thập data từ các file pdf (quy chế, quy định được tải trên web của trường)



## Kết quả và phân tích

### Khảo sát thông số A (ghi chú: các mục này nên ghi theo câu hỏi nghiên cứu)

Đưa các hình vẽ, bảng biểu kết quả tương tự như đã hướng dẫn ở trên. Lưu ý hình vẽ và bảng biểu bắt buộc phải đi kèm phân tích. Các yếu tố có thể phân tích là ý nghĩa các cực trị, nguyên nhân tăng giảm, nguyên nhân khác biệt giữa các kết quả.

### Kết quả mô phỏng thông số B

Tương tự như trên cho các vấn đề nghiên cứu đã đặt ra.

## Kết luận chương

Kết luận ngắn chương 3. Tóm tắt lại các ý phân tích và trả lời các câu hỏi nghiên cứu.

# KẾT LUẬN

## Tóm tắt và kết luận chung

Trong phần kết luận, trả lời các câu hỏi nghiên cứu đã đặt ra ở chương 1 dựa vào kết quả thực hiện va phân tích trong chương 3. Nêu các điểm đạt, chưa đạt của luận văn. Nguyên nhân đạt, nguyên nhân chưa đạt và cách khắc phục nếu có.

## Hướng phát triển

Phần hướng phát triển nêu ngắn gọn, không tràn lan.

# PHỤ LỤC A

(Ghi chú: Code chương trình có thể để vào phụ lục A và các chứng minh toán học hỗ trợ có thể để vào phụ lục B)

## A.1 Code chương trình giao tiếp Arduino

(Copy code vào đây)

## A.2 Code chương trình xử lý dữ liệu dùng Matlab

(Copy code vào đây)

# TÀI LIỆU THAM KHẢO

[1] S. S. S. R. Rautela, "Chatbot: A Bridge Between Technology and Learn," *2022 International Conference on Cyber Resilience (ICCR),* 2022.

[2] R. Mursalzade. "Introducing Answer Bot." <https://www.zendesk.com/blog/introducing-answer-bot/> (accessed.

[3] S. Singapore. "Sephora Virtual Artist | Sephora Singapore." <https://www.sephora.sg/pages/virtual-artist> (accessed.

[4] "Babylon." <https://www.babylonhealth.com/> (accessed.

[5] "Erica - Bank of America Mobile Banking app." <https://promotions.bankofamerica.com/digitalbanking/mobilebanking/erica> (accessed.

[6] "Expedia launches conversational trip planning powered by ChatGPT to inspire members to dream about travel in new ways." <https://www.expediagroup.com/investors/news-and-events/financial-releases/news/news-details/2023/Chatgpt-Wrote-This-Press-Release--No-It-Didnt-But-It-Can-Now-Assist-With-Travel-Planning-In-The-Expedia-App/default.aspx> (accessed.

[7] K. Johnson. "Recruitment chatbot Mya automates 75% of hiring process." <https://venturebeat.com/business/recruitment-chatbot-mya-automates-75-of-hiring-process/> (accessed.

[8] "Survey Sparrow Home page." <https://surveysparrow.com/> (accessed.

[9] K. Brown. "Something Bothering You? Tell It to Woebot." <https://www.nytimes.com/2021/06/01/health/artificial-intelligence-therapy-woebot.html> (accessed.

[10] Đ. Hưng. "ChatGPT: Những quan điểm "sáng" và "tối" trong quá trình “tự động hoá” tri thức." <https://ictvietnam.vn/chatgpt-nhung-quan-diem-sang-va-toi-trong-qua-trinh-tu-dong-hoa-tri-thuc-56494.html> (accessed.

[11] N. S. Ashish Vaswani, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, "Attention Is All You Need," 2017.

[12] T. Vu. "Embedding." <https://machinelearningcoban.com/tabml_book/ch_embedding/embedding.html#:~:text=Embedding%20l%C3%A0%20m%E1%BB%99t%20k%E1%BB%B9%20thu%E1%BA%ADt,m%E1%BB%99t%20vector%20d%E1%BA%A1ng%20one%2Dhot>. (accessed.

[13] A. K. Reddy. "Guide to Chroma DB | A Vector Store for Your Generative AI LLMs." <https://www.analyticsvidhya.com/blog/2023/07/guide-to-chroma-db-a-vector-store-for-your-generative-ai-llms/> (accessed.

[14] "Embeddings." <https://docs.trychroma.com/embeddings> (accessed.

[15] I. G. Nils Reimers, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *EMNLP 2019,* 2019.