TRƯỜNG ĐẠI HỌC KIẾN TRÚC HÀ NỘI

**KHOA CÔNG NGHỆ THÔNG TIN**

**----- o0o -----**

A blue and white logo

Description automatically generated with low confidence

**ĐỒ ÁN TỐT NGHIỆP**

**Tên đề tài:**

**NGHIÊN CỨU XỬ LÝ NGÔN NGỮ TỰ NHIÊN VÀ XÂY DỰNG ỨNG DỤNG TRẢ LỜI VĂN BẢN HÀNH CHÍNH BẰNG PYTHON**

|  |  |
| --- | --- |
| **GIÁO VIÊN HƯỚNG DẪN:** | **TH.SĨ NGUYỄN THỊ HUỆ** |
| **SINH VIÊN THỰC HIỆN:** | **HOÀNG TUẤN VŨ** |
| **MÃ SINH VIÊN :**  **LỚP:** | **2055010293**  **20CN5** |

**Hà Nội, tháng 10 năm 2024**

**Lời nói đầu**

Ngày nay với sự phát triển của khoa học cộng nghệ và đặc biệt là sự bùng nổ trong lĩnh vực Trí tuệ nhân tạo, đã mang lại cho con người nhiều tiện tích trong cuộc sống. Dựa trên tiêu chí đó để có thể giúp người dùng trong lĩnh vực văn bản hành chính trong công việc và đời sống, đồ án này nhằm nghiên cứu về lĩnh vực Xử lý Ngôn ngữ Tự nhiên với mục tiêu xây dựng một ứng dụng hỗ trợ trả lời các văn bản hành chính.

Để có thể xây dựng ứng dụng

Thông qua việc nghiên cứu chi tiết về đề tài ứng dụng xử lý ngôn ngữ tự nhiên về văn bản hành chính, với mục đích để có thể hiểu hơn về kiến thức trí tuệ nhân tạo nói chung cũng như mong muốn có thể giúp đỡ được người dùng.

Em muốn bày tỏ lòng biết ơn chân thành đến cô Nguyễn Thị Huệ

đã hỗ trợ và hướng dẫn nhóm em suốt quá trình thực hiện đề tài này. Sự

đóng góp của cô đã giúp nhóm em hiểu rõ hơn về vấn đề này và hoàn

thiện đề tài của mình.

# Chương 1: Cơ sở lý thuyết

## 1.1. Tổng quan về trí tuệ nhân tạo

### 1.1.1. Khái niệm

Trí tuệ nhân tạo (AI) là một lĩnh vực trong khoa học máy tính và công nghệ thông tin tập trung vào việc phát triển máy tính và hệ thống có khả năng thực hiện các nhiệm vụ thông minh mà trước đây chỉ có con người có thể thực hiện. AI có mục tiêu tạo ra các chương trình máy tính hoặc máy tính thông minh có khả năng học, tự điều chỉnh và thực hiện các nhiệm vụ mà yêu cầu sựhiểu biết, lý thuyết và khả năng giải quyết vấn đề.

### 1.1.2. Các hướng nghiên cứu

Có nhiều hướng nghiên cứu quan trọng về trí tuệ nhân tạo (AI) mà các nhà nghiên cứu trên khắp thế giới đang tiến hành. Dưới đây là một số hướng nghiên cứu quan trọng trong lĩnh vực này:

- Học máy và Học sâu (Machine Learning and Deep Learning): Nghiêncứu về các thuật toán học máy và học sâu để cải thiện khả năng học và tự điều chỉnh của các hệ thống AI. Điều này bao gồm việc tối ưu hóa các mô hình học máy hiện có và phát triển các mô hình mới.

- Thị giác máy tính (Computer Vision): Nghiên cứu về khả năng của máy tính để nhận diện, hiểu và xử lý hình ảnh và video. Các ứng dụng bao gồm nhận diện khuôn mặt, phân tích hình ảnh y tế và xe tự hành.

- Xử lý ngôn ngữ tự nhiên (NLP): Nghiên cứu về khả năng của máy tính để hiểu và tạo ra ngôn ngữ tự nhiên. Điều này liên quan đến chatbots, dịch máy, phân tích cảm xúc và nhiều ứng dụng khác.

- Robotics và Trí tuệ nhân tạo trong Robot (Robotics and AI in Robotics): Nghiên cứu về cách tạo ra và cải thiện robot có khả năng học và tự điều chỉnh, để chúng có thể thực hiện nhiều nhiệm vụ khác nhau, từ công nghiệp đến y tế và dịch vụ khách hàng.

### 1.1.3. Phân loại trí tuệ nhân tạo

Bốn loại của trí tuê nhân tạo đó là: Máy phản ứng, Bô nhớ hạn chế, Lý thuyết tâm trí, Tự nhận thức.

- Máy phản ứng (Reactive Machines):

Máy phản ứng là cấp độ đơn giản nhất của AI. AI sẽ có khả năng phân tích những động thái khả nghi nhất của mình và đối thủ. Sau đó, sẽ đưa ra giải pháp tốt nhất.

Deep Blue của IBM, một cỗ máy được thiết kế để chơi cờ vua với con người. Deep Blue đánh giá các quân cờ trên bàn cờ và phản ứng với chúng, dựa trên các chiến lược cờ vua được mã hóa trước. Deep Blue không học hỏi hoặc cải thiện khi chơi – nó chỉ đơn giản là ‘phản ứng’. Và nó đánh bại kiện tướng cờ vua Garry Kasparov vào năm 1997.

- Bộ nhớ hạn chế (Limited Memory):

Máy có bộ nhớ hạn chế, có thể giữ lại một số thông tin học được từ việc quan sát các sự kiện hoặc dữ liệu trước đó. AI có thể xây dựng kiến thức bằng cách sử dụng bộ nhớ đó kết hợp với dữ liệu được lập trình sẵn.

VD: Đối với xe không người lái, nhiều cảm biến được trang bị xung quanh xe và ở đầu xe để tính toán khoảng cách với các xe phía trước, công nghệ AI sẽ dự đoán khả năng xảy ra va chạm, từ đó điều chỉnh tốc độ xe phù hợp để giữ an toàn cho xe.

- Lý thuyết tâm trí (Theory of Mind):

Con người có những suy nghĩ và cảm xúc, ký ức hoặc các mô hình não khác điều khiển và ảnh hưởng đến hành vi của họ.

Dựa trên tâm lý này, các nhà nghiên cứu lý thuyết về tâm trí hy vọng phát triển các máy tính có khả năng bắt chước các mô hình tinh thần của con người. Máy móc có thể hiểu rằng con người và động vật có những suy nghĩ và cảm xúc có thể ảnh hưởng đến hành vi của chính chúng.

Lý thuyết về máy móc tâm trí sẽ được yêu cầu sử dụng thông tin thu được từ con người và học hỏi từ nó, sau đó sẽ thông báo bằng cách máy móc giao tiếp hoặc phản ứng với một tình huống khác.

- Tự nhận thức (Self-awareness):

Công nghệ AI này có khả năng tự nhận thức về bản thân, có ý thức và hành xử như con người. Thậm chí, chúng còn có thể bộc lộ cảm xúc cũng như hiểu được những cảm xúc của con người. Đây được xem là bước phát triển cao nhất của công nghệ AI và đến thời điểm hiện tại, công nghệ này vẫn chưa khả thi.

### 1.1.4. Ứng dụng

- Xử lý ngôn ngữ tự nhiên (NLP):

* Chatbots và Virtual Assistants: AI được sử dụng để xây dựng cáctrợ lý ảo và chatbot để tương tác với người dùng và cung cấp thông tin hoặc giải quyết vấn đề.
* Phân tích tình cảm: AI có thể phân tích tình cảm trong văn bản, giúp doanh nghiệp hiểu đánh giá của khách hàng về sản phẩm hoặc dịch vụ.

- Thị giác máy tính (Computer Vision):

* Nhận diện khuôn mặt: AI có thể nhận diện và xác định khuôn mặt trong hình ảnh hoặc video, thường được sử dụng trong hệ thống đảm bảo an ninh và xác thực.
* Xử lý hình ảnh y tế: Giúp bác sĩ chẩn đoán bệnh dựa trên hình ảnh y tế như X-quang, MRI, CT scans.

- Tự động hóa và Robotica:

* Ô tô tự lái: AI được sử dụng trong xe ô tô tự lái để phát hiện và tránh các vật thể, biển báo giao thông, và điều khiển xe.
* Robot công nghiệp: Robot được trang bị trí tuệ nhân tạo để thực hiện các nhiệm vụ trong môi trường sản xuất, từ lắp ráp đến kiểmtra chất lượng.

- Dự đoán Tài chính:

* Dự đoán thị trường tài chính: AI được sử dụng để dự đoán xu hướng thị trường chứng khoán và thực hiện giao dịch tự động.
* Quản lý rủi ro: Các công ty bảo hiểm sử dụng AI để định giá và quản lý rủi ro.

- Y tế và Chăm sóc sức khỏe:

* Chẩn đoán bệnh: AI có thể phân tích dữ liệu y tế và hình ảnh để hỗ trợ bác sĩ trong việc chẩn đoán và điều trị bệnh.
* Quản lý dữ liệu bệnh nhân: Hệ thống AI có thể quản lý và xử lý thông tin liên quan đến bệnh nhân, giúp cải thiện chất lượng dịch vụ y tế.

- Giáo dục:

* Học tập trực tuyến và cá nhân hóa giáo dục: AI có thể tạo ra nội dung giảng dạy cá nhân hóa dựa trên nhu cầu và khả năng của từng học sinh.
* Đánh giá học sinh: Hệ thống AI có thể phân tích kết quả học tập và đề xuất cải thiện cho học sinh.

- Điều khiển và Tự động hóa nhà thông minh:

* Quản lý năng lượng: Hệ thống AI có thể điều khiển thiết bị để tiết kiệm năng lượng trong nhà thông minh.
* Bảo mật: AI có thể giúp giám sát và bảo vệ nhà thông minh khỏi các mối đe dọa an ninh.

- Quảng cáo và Tiếp thị:

* Tối ưu hóa quảng cáo trực tuyến: AI giúp tối ưu hóa chiến dịch quảng cáo trực tuyến dựa trên dữ liệu về tương tác của người dùng.

- Khám phá dữ liệu và Nghiên cứu khoa học:

* Xử lý dữ liệu lớn (Big Data): AI giúp phân tích và rút ra thông tin quan trọng từ các nguồn dữ liệu lớn và phức tạp.

### 1.1.5. Mặt tích cực và hạn chế của trí tuệ nhân tạo

Trí tuệ nhân tạo (AI) mang lại nhiều lợi ích và tiềm năng cách mạng hóa nhiều lĩnh vực khác nhau. Tuy nhiên, cũng có những tích cực và hạn chế cần xem xét:

\*Tích cực:

- Tăng năng suất và hiệu suất công việc: AI có thể thực hiện nhiều nhiệmvụ một cách nhanh chóng và chính xác, giúp tăng năng suất và hiệu suất làm việc trong các ngành công nghiệp.

- Dự đoán và ứng dụng trong quản lý tài chính: AI có khả năng phân tích dữ liệu tài chính phức tạp và dự đoán xu hướng thị trường, giúp đưa ra quyết định tài chính hiệu quả.

- Quản lý y tế và chẩn đoán bệnh: Trong lĩnh vực y tế, AI có thể hỗ trợ các bác sĩ trong việc chẩn đoán bệnh và lập kế hoạch điều trị dựa trên phân tích hình ảnh và dữ liệu bệnh lý.

- Ô tô tự hành: AI có tiềm năng thúc đẩy phát triển xe tự hành, giảm tai nạn giao thông và giúp người già hoặc khuyết tật di chuyển dễ dàng hơn.

- Dự đoán thời tiết và khí hậu: AI có khả năng xử lý dữ liệu khí hậu lớn và dự đoán thời tiết một cách chính xác, giúp cảnh báo thiên tai và khắc phục hậu quả.

- Hỗ trợ trong giáo dục: AI có thể cung cấp các công cụ học tập thông minh, cá nhân hóa giáo dục và tạo điều kiện học tập tốt hơn cho học sinh.

\*Hạn chế:

- Sự lo ngại về đạo đức và quyền riêng tư: Sử dụng AI có thể dẫn đến các vấn đề về quyền riêng tư và đạo đức, bao gồm việc thu thập và sử dụng dữ liệu cá nhân một cách không đúng mục đích.

- Thất nghiệp và thay thế công việc: Các hệ thống tự động hóa dự kiến sẽ thay thế một số công việc, gây ra lo ngại về thất nghiệp và sự bất ổn kinh tế.

- Giới hạn trong việc hiểu và giải quyết bài toán phức tạp: AI hiện tại vẫn gặp khó khăn trong việc hiểu và giải quyết bài toán phức tạp mà con người có thể làm.

- Nguy cơ trục trặc và lỗi hệ thống: AI có thể gặp trục trặc và lỗi hệ thống, đặc biệt khi dựa vào dữ liệu không chính xác hoặc không đủ lớn.

- Khả năng phân biệt đạo đức và quyết định etic: AI hiện tại không có khả năng phân biệt đạo đức và quyết định theo tiêu chuẩn etic, gây ra các vấn đề đạo đức liên quan đến quyết định của chúng.

- Phụ thuộc vào dữ liệu lớn: AI cần dữ liệu lớn để hoạt động hiệu quả, vàđiều này có thể tạo ra vấn đề về riêng tư và bảo mật dữ liệu.

## 1.2. Tổng quan về đề tài

Ứng dụng trả lời văn bản trả lời được những câu hỏi:

- Tên của văn bản hành chính

- Nội dung của văn bản hành chính

- Ngày ban hành, cơ quan ban hành của văn bản hành chính

### 1.3. Công nghệ sử dụng

# Chương 2: Quy trình

## 2.1. Chuẩn bị tập dữ liệu

## 2.2. Tiền xử lý dữ liệu

## 2.3. Lựa chọn mô hình

## 2.4. Huấn luyện mô hình

## 2.5. Đánh giá kết quả

# Chương 3: Triển khai và thực nghiệm

## 3.1 Công cụ sử dụng

## 3.2 Phân tích thiết kế hệ thống

## 3.3 Quy trình thực nghiệm

## 3.4 Đánh giá mô hình

## 3.5 Kết luận

TÀI LIỆU THAM KHẢO

[1] Giáo trình Nhập môn Trí tuệ nhân tạo – Thầy Từ Minh Phương

[2] https://docs.python.org/3

[3] https://www.pygame.org/docs

36

[question\_answering\_bartpho\_phobert/code.ipynb at main · phkhanhtrinh23/question\_answering\_bartpho\_phobert](https://github.com/phkhanhtrinh23/question_answering_bartpho_phobert/blob/main/code.ipynb)

[abaoxomtieu/RAG-API-VietnamesePDF: Build API for Q&A on private PDF.](https://github.com/abaoxomtieu/RAG-API-VietnamesePDF)

[Báo cáo bài tập lớn - TRƯỜNG ĐẠI HỌC KIẾN TRÚC HÀ NỘI KHOA CÔNG NGHỆ THÔNG TIN ĐỒ ÁN MÔN TRÍ TUỆ - Studocu](https://www.studocu.com/vn/document/truong-dai-hoc-kien-truc-ha-noi/tri-tue-nhan-tao/bao-cao-bai-tap-lon/74287674)