**Bài tập đề tài 1**

Đặng Công Tuấn - 21110709

1/

Báo cáo "The National AI Research and Development Strategic Plan" của Mỹ năm 2020 đã đề ra một số nguyên tắc đạo đức quan trọng trong việc phát triển trí tuệ nhân tạo. Các nguyên tắc này bao gồm:

+Trách nhiệm đạo đức: Các tổ chức và nhà nghiên cứu phát triển trí tuệ nhân tạo cần phải chịu trách nhiệm đạo đức và phải đảm bảo rằng trí tuệ nhân tạo được sử dụng một cách có đạo đức.

+Kính trọng quyền riêng tư: Các nhà nghiên cứu phát triển trí tuệ nhân tạo cần phải kính trọng quyền riêng tư và bảo vệ dữ liệu người dùng.

+Kính trọng đa dạng: Các nhà nghiên cứu phát triển trí tuệ nhân tạo cần phải kính trọng đa dạng và đảm bảo rằng trí tuệ nhân tạo không gây ra sự phân biệt đối xử dựa trên giới tính, chủng tộc, tuổi tác, v.v.

+Bảo đảm tính minh bạch: Các nhà nghiên cứu phát triển trí tuệ nhân tạo cần đảm bảo tính minh bạch của quá trình phát triển và sử dụng trí tuệ nhân tạo.

+Bảo đảm an toàn: Các nhà nghiên cứu phát triển trí tuệ nhân tạo cần đảm bảo an toàn trong việc phát triển và sử dụng trí tuệ nhân tạo.

Một trong những nguyên tắc đạo đức quan trọng trong báo cáo này là trách nhiệm đạo đức. Điều này quan trọng đối với việc phát triển trí tuệ nhân tạo vì trí tuệ nhân tạo có thể ảnh hưởng đến nhiều khía cạnh của cuộc sống, bao gồm cả quyền lợi và sự an toàn của con người. Nếu không có trách nhiệm đạo đức, các ứng dụng trí tuệ nhân tạo có thể gây ra những hậu quả không mong muốn, gây ra sự phân biệt đối xử và vi phạm quyền riêng tư.

Do đó, các nhà phát triển trí tuệ nhân tạo cần phải chịu trách nhiệm đạo đức và đảm bảo rằng trí tuệ nhân tạo được sử dụng phụ thuộc vào một nền tảng đạo đức vững chắc. Điều này có nghĩa là họ cần phải xem xét cẩn thận các hậu quả đạo đức của trí tuệ nhân tạo và đảm bảo rằng các quyết định được đưa ra phù hợp với các giá trị đạo đức.

Đối với các công ty và tổ chức phát triển trí tuệ nhân tạo, trách nhiệm đạo đức còn đòi hỏi họ phải thực hiện các phản hồi nhanh chóng và có hiệu quả nếu xảy ra các vấn đề liên quan đến đạo đức. Nếu có bất kỳ hậu quả nào về mặt đạo đức của các ứng dụng trí tuệ nhân tạo, các nhà phát triển cần phải chịu trách nhiệm và đưa ra giải pháp để giải quyết vấn đề.

Để đảm bảo trách nhiệm đạo đức, các nhà phát triển trí tuệ nhân tạo cần phải có một tầm nhìn rõ ràng và sự cam kết với các giá trị đạo đức. Họ cũng cần phải thực hiện các tiêu chuẩn và quy định đạo đức cụ thể trong việc phát triển và triển khai trí tuệ nhân tạo. Bằng cách tuân thủ các nguyên tắc đạo đức này, các nhà phát triển trí tuệ nhân tạo có thể đảm bảo rằng trí tuệ nhân tạo được sử dụng một cách có đạo đức và đóng góp vào sự tiến bộ và phát triển của xã hội.

2/

Một ví dụ về việc sử dụng trí tuệ nhân tạo không đạo đức là việc sử dụng kỹ thuật deepfake để tạo ra những video giả mạo hoặc những hình ảnh nhằm lừa đảo hoặc gây ra những hậu quả đáng tiếc cho những người bị ảnh hưởng.

Deepfake là một kỹ thuật trí tuệ nhân tạo được sử dụng để tạo ra các nội dung giả mạo bằng cách thay đổi nội dung ban đầu và kết hợp các thành phần khác nhau. Mặc dù kỹ thuật này có thể được sử dụng để tạo ra những nội dung giải trí hấp dẫn nhưng nó cũng có thể được sử dụng để gây hại cho những người bị ảnh hưởng.

Ví dụ, deepfake có thể được sử dụng để tạo ra các video giả mạo của các nhân vật nổi tiếng để lừa đảo hoặc gây ảnh hưởng đến quyết định của công chúng. Nó cũng có thể được sử dụng để tạo ra các hình ảnh khiêu dâm giả mạo của những người khác nhằm lạm dụng và bóc lột họ. Những hậu quả của việc sử dụng deepfake có thể là gây tổn thương tâm lý, phá hỏng danh tiếng và tước đoạt quyền riêng tư của những người bị ảnh hưởng.

Đó là lý do tại sao chúng ta cần có các quy định đạo đức trong phát triển trí tuệ nhân tạo. Những quy định này sẽ giúp đảm bảo rằng các ứng dụng trí tuệ nhân tạo sẽ được sử dụng một cách có đạo đức và an toàn, tránh gây hại đến những người bị ảnh hưởng. Các quy định đạo đức này cũng giúp đảm bảo rằng các nhà phát triển trí tuệ nhân tạo đang làm việc với sự cam kết đúng đắn và tôn trọng đạo đức trong việc triển khai và sử dụng trí tuệ nhân tạo.

3/

Trong việc sử dụng trí tuệ nhân tạo, bảo mật thông tin là một trong những rủi ro chính. Những mô hình và hệ thống trí tuệ nhân tạo phải sử dụng rất nhiều dữ liệu nhạy cảm và thông tin cá nhân của người dùng, do đó việc bảo mật thông tin là vô cùng quan trọng để đảm bảo sự tin tưởng và sự an toàn của người dùng.

Tuy nhiên, việc sử dụng trí tuệ nhân tạo cũng có nhiều rủi ro về bảo mật thông tin, bao gồm:

1. Tấn công mạng: Các kẻ tấn công có thể tấn công vào mạng của một hệ thống trí tuệ nhân tạo để truy cập dữ liệu nhạy cảm của người dùng.
2. Lỗ hổng bảo mật: Các lỗ hổng bảo mật có thể được tìm thấy trong mô hình và hệ thống trí tuệ nhân tạo, cho phép những kẻ tấn công tiếp cận dữ liệu.
3. Sử dụng sai mục đích: Trí tuệ nhân tạo có thể được sử dụng để thu thập thông tin cá nhân của người dùng một cách không đúng đắn và không được cho phép.
4. Lợi dụng thông tin: Những người không đúng đắn có thể sử dụng thông tin cá nhân của người dùng để gây hại hoặc lợi dụng.

Để giảm thiểu những rủi ro về bảo mật thông tin, các quy định như CPRA (California Privacy Rights Act) và NAIIA (National Artificial Intelligence Initiative Act) là cần thiết. CPRA là một đạo luật tại tiểu bang California, Mỹ, nơi có nhiều công ty công nghệ lớn và được coi là tiên phong trong lĩnh vực quyền riêng tư. Đạo luật này tập trung vào việc bảo vệ quyền riêng tư của người dùng trực tuyến, bao gồm việc giới hạn việc thu thập, sử dụng và chia sẻ thông tin cá nhân.

NAIIA là một đạo luật mới được thông qua vào năm 2021 bởi Chính phủ Mỹ, đặt mục tiêu đẩy mạnh phát triển trí tuệ nhân tạo và đảm bảo an toàn và đạo đức trong việc sử dụng nó. Đạo luật này yêu cầu các đối tác chính phủ phải đảm bảo rằng các hệ thống trí tuệ nhân tạo được sử dụng một cách đúng đắn và đảm bảo tính riêng tư, an toàn và bảo mật của người dùng.

Cả CPRA và NAIIA đều đưa ra các quy định và nghĩa vụ để đảm bảo bảo mật thông tin trong việc sử dụng trí tuệ nhân tạo. Chúng yêu cầu các tổ chức thu thập và sử dụng dữ liệu cá nhân phải có sự đồng ý rõ ràng của người dùng và giải thích rõ ràng về mục đích sử dụng dữ liệu. Ngoài ra, chúng cũng yêu cầu các tổ chức phải bảo vệ dữ liệu của người dùng và bảo mật các hệ thống trí tuệ nhân tạo của mình.

Tóm lại, các rủi ro về bảo mật thông tin là rất đáng lo ngại trong việc sử dụng trí tuệ nhân tạo. Việc đưa ra các quy định và nghĩa vụ như CPRA và NAIIA là rất cần thiết để đảm bảo an toàn, bảo mật và đạo đức trong việc sử dụng trí tuệ nhân tạo và tăng cường sự tin tưởng của người dùng đối với công nghệ này.

**Bài tập đề tài 2**

Đặng Công Tuấn - 21110709

Câu 1:

Trong trường hợp này, công ti đã vi phạm đạo đức AI bởi vì việc sử dụng hệ thống AI để loại bỏ những ứng viên có khả năng bị trầm cảm hoặc mang thai trong tương lai là một hành động kỳ thị và phân biệt đối xử.

Việc sử dụng các thuật toán AI để loại bỏ ứng viên dựa trên một số thông tin như vậy không chỉ không công bằng mà còn có thể dẫn đến những hậu quả xấu cho những người bị loại bỏ. Ngoài ra, việc sử dụng dữ liệu nhạy cảm như thông tin về tâm lý và sức khỏe của ứng viên để đưa ra quyết định tuyển dụng cũng là việc vi phạm quyền riêng tư của ứng viên.

Vì vậy, việc sử dụng hệ thống AI để loại bỏ những ứng viên dựa trên các tiêu chí không liên quan đến năng lực là một hành động không đạo đức và phải được tránh. Thay vào đó, công ty nên sử dụng các phương pháp tuyển dụng công bằng và đúng đắn để đảm bảo tính đạo đức và công bằng trong quá trình tuyển dụng.

Câu 2:

Trung Quốc đã đặt ra một loạt các mục tiêu chiến lược AI trong nhiều lĩnh vực khác nhau như kinh tế, an ninh quốc gia, y tế và giáo dục. Một số mục tiêu đáng chú ý bao gồm:

1. Trở thành một cường quốc về trí tuệ nhân tạo vào năm 2030: Trung Quốc đang nỗ lực để trở thành một trong những quốc gia hàng đầu thế giới về trí tuệ nhân tạo vào năm 2030. Mục tiêu này được đặt ra nhằm thúc đẩy sự phát triển kinh tế và cải thiện đời sống của người dân.
2. Tăng cường nghiên cứu và phát triển AI: Trung Quốc đã đặt mục tiêu tăng cường nghiên cứu và phát triển trí tuệ nhân tạo để nhanh chóng đạt được các mục tiêu chiến lược. Trong năm 2017, Trung Quốc đã đầu tư 2,12 tỷ USD cho nghiên cứu và phát triển AI, gấp đôi số tiền đầu tư của Mỹ.
3. Sử dụng AI để cải thiện dịch vụ y tế: Trung Quốc đặt mục tiêu sử dụng trí tuệ nhân tạo để cải thiện chất lượng dịch vụ y tế và tiếp cận dịch vụ y tế tốt hơn cho những người dân ở các vùng khó khăn.
4. Sử dụng AI để cải thiện hệ thống an ninh: Trung Quốc đặt mục tiêu sử dụng trí tuệ nhân tạo để cải thiện hệ thống an ninh và giảm tội phạm. Các công nghệ AI được sử dụng để giám sát, nhận diện khuôn mặt và phát hiện hành vi bất thường của con người.

Tuy nhiên, việc Trung Quốc đặt ra các mục tiêu chiến lược AI cũng đặt ra một số lo ngại về các vấn đề như quyền riêng tư, an ninh mạng và an ninh quốc gia. Vì vậy, cần có các quy định đạo đức và các biện pháp an ninh mạng để đảm bảo rằng các ứng dụng AI được phát triển và sử dụng một cách an toàn và bảo mật.

Câu 3;

Trong số các luật về AI được áp dụng ở Trung Quốc từ năm 2020 đến nay, có một bộ luật nổi bật là Luật về Quản lý Dữ liệu, được áp dụng từ tháng 9 năm 2021. Bộ luật này đặt ra các quy định về quản lý và bảo vệ dữ liệu cá nhân, đặc biệt là trong lĩnh vực trí tuệ nhân tạo. Bộ luật này được coi là tiến bộ và đáng chú ý nhất, vì nó bao gồm các điều khoản bảo vệ quyền riêng tư và bảo mật dữ liệu của người dùng, đồng thời cung cấp các yêu cầu cụ thể về phân loại và quản lý dữ liệu.

Nếu bộ luật này được áp dụng ở Việt Nam, cần có một số cải tiến để phù hợp với hoàn cảnh hiện tại. Ví dụ, Việt Nam có thể cần bổ sung thêm quy định về quyền sở hữu trí tuệ, đặc biệt là với các công nghệ AI được phát triển trong nước. Việt Nam cũng có thể cần đưa ra các quy định cụ thể hơn về trách nhiệm của các tổ chức sử dụng trí tuệ nhân tạo đối với người dùng và xã hội nói chung. Đồng thời, cần có các biện pháp thúc đẩy sự phát triển của ngành công nghiệp trí tuệ nhân tạo tại Việt Nam, bao gồm các chính sách hỗ trợ và đào tạo nguồn nhân lực chất lượng cao trong lĩnh vực này.