

CHIẾN LƯỢC XÂY DỰNG ĐẠO ĐỨC TRÍ TUỆ NHÂN TẠO TẠI CHLB ĐỨC VÀ MỘT SỐ GỢI MỞ CHO VIỆT NAM

Hoa Hữu Cường*

Chủ Thị Nhuận**

Tóm tắt: Trí tuệ nhân tạo (AI) đóng vai trò trung tâm trong cuộc cách mạng công nghiệp lần thứ tư. Tuy nhiên, AI cũng làm phát sinh các vấn đề kinh tế xã hội mới như bảo vệ dữ liệu và quyền riêng tư; thành kiến không công bằng..., tất cả những vấn đề đó cần một cơ chế mới để điều chỉnh. Trong bối cảnh đó, một số nước thành viên EU trong đó có Đức đã xây dựng những quy định riêng trong một số lĩnh vực ứng dụng AI. Cụ thể, vào tháng 11 năm 2018, chiến lược quốc gia về AI của CHLB Đức đã được ra đời, đề ra một loạt các biện pháp về đạo đức. Bài viết phân tích và làm rõ các nguyên tắc về đạo đức AI và chiến lược xây dựng cơ chế để vấn đề đạo đức AI từ doanh nghiệp đến cơ quan quản lý nhà nước đi vào cuộc sống, từ đó thấy được những giá trị tham khảo cho Việt Nam.

Từ khóa: trí tuệ nhân tạo, đạo đức trí tuệ nhân tạo, AI, Cộng hòa Liên bang Đức

Abstract: Artificial intelligence (AI) plays a central role in the fourth industrial revolution. However, AI also gives rise to new socioeconomic issues such as data protection and privacy; unfair bias, etc, all of which need a new mechanism to correct. In that context, some EU member states, including Germany, have developed their own regulations in some areas of AI application. Specifically, in November 2018, Germany's national AI strategy was born, setting out a series of ethical measures. The article analyzes and clarifies the principles of AI ethics and the strategy to build a mechanism for AI ethics issues from businesses to state management agencies to come to life, thereby seeing the values of reference for Vietnam.

Keywords: Artificial intelligence, Artificial intelligence ethics, AI, Federal Republic of Germany

* TS., Viện Nghiên cứu Châu Âu

** Ths., Viện Nghiên cứu Châu Âu

Nhận bài ngày: 17/1/2022

Phản biện xong: 24/1/2022

Chấp nhận đăng: 31/1/2022

1. Chiến lược xây dựng đạo đức trí tuệ nhân tạo của CHLB Đức

CHLB Đức là quốc gia đi đầu trong phát triển trí tuệ nhân tạo (AI), đặc biệt Đức rất coi trọng vấn đề đạo đức trí tuệ nhân tạo “lấy con người làm trung tâm” thông qua việc triển khai nhiều biện pháp để thực hiện vấn đề này.

1.1. Thiết lập các tổ chức để nghiên cứu và tư vấn cho chính phủ về đạo đức trí tuệ nhân tạo

**** Thành lập Ủy ban đạo đức về xe tự hành***

“Ủy ban đạo đức về xe tự hành” trực thuộc Bộ Giao thông vận tải và Cơ sở hạ tầng kỹ thuật số Liên bang được thành lập vào tháng 9 năm 2016. Đây là một Ủy ban liên ngành bao gồm các chuyên gia cao cấp đến từ các lĩnh vực như: triết học, luật, khoa học xã hội, công nghệ, bảo vệ người tiêu dùng, ngành công nghiệp ô tô và kinh tế kỹ thuật số. Đây là ủy ban đầu tiên trên thế giới giải quyết các vấn đề liên quan đến khía cạnh đạo đức của xe tự hành.

**** Thành lập Ủy ban Enquete***

Ủy ban này trực thuộc Hạ viện Liên bang Đức, được thành lập vào tháng 6 năm 2018 với nhiệm vụ là điều tra những ảnh hưởng trong tương lai của AI đối với các vấn đề đời sống xã hội, nền kinh tế và việc làm. Ủy ban bao gồm các thành viên của Hạ viện Đức (tính theo tỷ lệ phần trăm đại diện của nhóm nghị sĩ tương ứng trong Quốc hội) và các chuyên gia bên ngoài.

**** Thành lập Ủy ban Đạo đức Dữ liệu***

Chính phủ Liên bang Đức thành lập Ủy ban Đạo đức Dữ liệu vào ngày 18 tháng 7 năm 2018 để phát triển các hướng dẫn và khuyến nghị về mặt đạo đức nhằm bảo vệ “cá nhân, duy trì sự gắn kết xã hội, bảo vệ và thúc đẩy sự thịnh vượng trong thời đại thông tin.” Ủy ban gồm 16 thành viên là các học giả, chuyên gia bảo vệ dữ liệu và một số đại diện của một số ngành. Theo quan điểm của Ủy ban Đạo đức Dữ liệu, AI chỉ đơn thuần là một trong số nhiều biến thể có thể có của hệ thống thuật toán và có nhiều điểm chung về các vấn đề về đạo đức và pháp lý mà nó đặt ra. Với quan điểm như vậy, Ủy ban Đạo đức Dữ liệu đã triển khai các hoạt động theo hai vấn đề là: dữ liệu và hệ thống thuật toán (theo nghĩa rộng hơn). Kể từ khi thành lập cho đến nay, Ủy ban đã đưa ra một loạt các ý kiến về thuật toán và trí tuệ nhân tạo cũng như đưa ra các khuyến nghị và xây dựng một loạt các nguyên tắc chung liên quan đến việc sử dụng dữ liệu và xây dựng các hệ thống thuật toán:

Đối với dữ liệu, ủy ban đề xuất trách nhiệm giải trình, tôn trọng quyền của các bên liên quan, sử dụng dữ liệu và chia sẻ dữ liệu vì lợi ích chung, chất lượng dữ liệu phù hợp với mục đích, mức độ an toàn thông tin và dữ liệu hướng đến lợi ích minh bạch.

Đối với các thuật toán, Ủy ban khuyến nghị các thuật toán trong các phần mềm phải được thiết kế lấy con người làm trung tâm và phải tương thích với các giá trị xã hội, mang tính bền vững, hướng tới chất lượng và hiệu

suất mạnh mẽ, an toàn, giảm thiểu tác hại và phân biệt đối xử khi sử dụng.

1.2. Xây dựng các hướng dẫn và ban hành bộ quy tắc về đạo đức trí tuệ nhân tạo

** Bộ quy tắc về đạo đức cho chương trình xe tự hành*

Đức đã ban hành bộ nguyên tắc đạo đức đầu tiên trên thế giới về chương trình xe tự lái. Các nguyên tắc do Ủy ban Đạo đức về xe tự hành phát triển vào tháng 6/2017. Bộ quy tắc này gồm 19 quy tắc dành cho các nhà thiết kế phần mềm, ưu tiên các vấn đề về “*an toàn, phẩm giá con người, quyền tự do lựa chọn của cá nhân và quyền tự chủ về dữ liệu của cá nhân*” (Ethics Commission, 2027). Cụ thể; 1) Nguyên tắc tự chủ cá nhân, có nghĩa là các cá nhân được hưởng quyền tự do hành động mà bản thân họ phải chịu trách nhiệm; 2) Việc bảo vệ người sử dụng phải là ưu tiên quan trọng nhất trong phát triển xe tự hành; 3) Các cơ quan quản lý nhà nước phải có trách nhiệm đảm bảo sự an toàn của xe tự hành khi được cấp phép lưu thông trên đường; 4) Các quyết định quản lý của chính phủ về xe tự hành phải thúc đẩy quyền tự do và bảo vệ các cá nhân cũng như đảm bảo sự an toàn cho xã hội; 5) Trong khi thiết kế công nghệ xe tự hành phải lường hết các tình huống nguy cấp có thể phát sinh ngay từ đầu; 6) Việc bảo vệ tính mạng con người phải được ưu tiên hàng đầu; 7) Trong quá trình lập trình cho hệ thống xe tự hành, các lập trình viên cần phải tham khảo một cơ quan quản lý nhà nước độc lập; 8) Khi lập trình cho xe tự hành nghiêm cấm sự phân biệt dựa

trên đặc điểm cá nhân như: tuổi tác, giới tính, thể chất hoặc tinh thần; 9) Cần phải điều chỉnh các quy định pháp lý hiện hành cũng như xây dựng và ban hành các quy định pháp lý mới để quản lý hệ thống xe tự hành; 10) Trách nhiệm pháp lý đối với những thiệt hại do hệ thống xe tự hành gây ra phải được xử lý theo các nguyên tắc tương tự như trách nhiệm pháp lý đối với hệ thống lái xe có người lái hiện nay; 11) Người dân có quyền được thông báo về các công nghệ xe tự hành mới và việc triển khai chúng các công nghệ này; 12) Khả năng kết nối và điều khiển cho tất cả các phương tiện tự hành phải được thiết kế để đảm bảo an toàn cho người tham gia giao thông; 13) Hệ thống xe tự hành phải được thiết kế để chống lại những cuộc tấn công mạng cũng như hệ số sai số phần mềm điều khiển phải cực kỳ nhỏ gần bằng không; 14) Các hình thức kinh doanh sử dụng dữ liệu được tạo ra từ hệ thống xe tự hành phải được sự đồng ý của người sử dụng phương tiện cũng như việc cung cấp dữ liệu cá nhân sẽ do cá nhân chủ động thực hiện một cách tự nguyện; 15) Hệ thống xe tự hành phải được thiết kế sao cho có thể phân biệt được tránh nhiệm giải trình giữa người với công nghệ tự hành trong các trường hợp có người lái xe và sử dụng công nghệ tự hành cho xe; 16) Phần mềm và công nghệ trên các phương tiện tự hành phải được thiết kế sao cho trong những trường hợp khẩn cấp, quyền điều khiển xe phải được trao cho người lái xe; 17) Hệ thống tự học trong xe tự hành phải được kết nối với cơ sở dữ liệu ở trung tâm điều khiển và trung tâm điều khiển này

phải do một tổ chức độc lập vận hành dựa trên các tiêu chuẩn về tính an toàn, tính bảo mật...; 18) Trong các tình huống khẩn cấp, phần mềm tự lái trên xe được kích hoạt mà không có sự trợ giúp của con người phải điều khiển xe một cách an toàn; 19) Cần phải dạy việc sử dụng hệ thống tự động trên xe tự hành cho người dân, đặc biệt là trong các trường dạy lái xe.

** Ban hành bộ quy tắc về đạo đức trí tuệ nhân tạo*

Ủy ban Đạo đức Dữ liệu của CHLB Đức đưa ra quan điểm về đạo đức kỹ thuật số nói chung và đạo đức AI nói riêng như sau: “Khi thiết kế các công nghệ kỹ thuật số nói chung và AI nói riêng cần phải dựa trên khuôn khổ đạo đức ở các khía cạnh về các giá trị, quyền và tự do cơ bản của con người được ghi trong Hiến pháp Đức và trong Hiến chương các quyền cơ bản của Liên minh Châu Âu”. Với quan điểm này, Ủy ban Đạo đức Dữ liệu đã xây dựng và công bố bộ tắc về đạo đức và pháp lý cho kỹ thuật số nói chung và AI nói riêng (Ethics Commission, 2017) như sau: 1) “Phẩm giá con người”: nghiêm cấm các hành vi như: giám sát kỹ thuật số hoàn toàn đối với cá nhân hoặc các hoạt động lừa dối, chi phối, sỉ nhục... của công nghệ kỹ thuật số đối với con người; 2) “Quyền tự quyết” là một biểu hiện cơ bản của tự do và bao gồm quyền tự quyết về mặt thông tin; 3) “Sự riêng tư”: các mối đe dọa tiềm ẩn đối với quyền riêng tư bao gồm việc thu thập, đánh giá, sử dụng dữ liệu cá nhân cần bị cấm; 4) “Bảo vệ”:

quy tắc này không chỉ liên quan đến sự an toàn về thể chất và tình cảm của con người mà còn liên quan đến bảo vệ môi trường cũng như việc bảo quản các tài sản cực kỳ quan trọng; 5) “Dân chủ”: các công nghệ kỹ thuật số phải ngăn chặn được sự thao túng và cực đoan hóa làm tổn hại đến nền dân chủ của xã hội; 6) “Công lý và đoàn kết”: việc bảo vệ quyền tiếp cận công bằng về công nghệ và dữ liệu cũng như công bằng trong phân phối các dữ liệu và công nghệ là một nhiệm vụ cấp bách; 7) “Sự bền vững”: công nghệ kỹ thuật số phải góp phần đạt được các mục tiêu bền vững về kinh tế, sinh thái và xã hội.

** Các nguyên tắc đạo đức trí tuệ nhân tạo tại công ty*

Hiện nay, các doanh nghiệp tại CHLB Đức cũng quan tâm đến vấn đề đạo đức trí tuệ nhân tạo, đặc biệt là những doanh nghiệp trong lĩnh vực công nghệ thông tin, minh chứng là các công ty đã xây dựng và triển khai các nguyên tắc đạo đức trí tuệ nhân tạo cho các sản phẩm mà mình cung cấp ra thị trường và nguyên tắc đạo đức trí tuệ nhân tạo đều hướng tới các khía cạnh như: 1) “Công bằng và có thể tiếp cận được với tất cả mọi người”; 2) “Trí tuệ nhân tạo phải phục vụ xã hội”; 3) “Các khía cạnh về mặt đạo đức phải được thiết lập cho AI”; 4) “Việc sử dụng dữ liệu cần phải minh bạch và phải được bảo vệ”; “AI phải đảm bảo an toàn cho con người”; 6) “Con người đóng vai trò quyết định trong phán quyết về mặt đạo đức của AI” (France24, 2018).

1.3. Xây dựng và ban hành Chiến lược quốc gia về AI

Ngày 15/11/2018, CHLB Đức đã thông qua Chiến lược quốc gia về AI do Bộ Giáo dục và Nghiên cứu Liên bang, Bộ Kinh tế và Năng lượng Liên bang, Bộ Lao động và Xã hội Liên bang đồng thực hiện. Chiến lược này hướng tới việc phát triển AI có đạo đức, cụ thể: “Tích hợp AI với các điều kiện đạo đức, luật pháp, văn hóa và thể chế dựa trên đối thoại xã hội rộng rãi và các biện pháp chính trị tích cực”.

Để triển khai mục tiêu này, Chính phủ Liên bang đã giao nhiệm vụ cho Ủy ban Đạo đức Dữ liệu nghiên cứu và ban hành các hướng dẫn cho việc phát triển và sử dụng AI, trong đó đặc biệt chú trọng vào khía cạnh đạo đức trí tuệ nhân tạo. Triển khai nhiệm vụ được giao, Ủy ban Đạo đức dữ liệu đã tiến hành xây dựng bộ hướng dẫn về đạo đức trí tuệ nhân tạo và vào tháng 10 năm 2019. Dựa trên hướng dẫn của Ủy ban Đạo đức Dữ liệu, Chính phủ Liên bang Đức đã công bố báo cáo tạm thời trình bày các biện pháp chính đã được thực hiện của Chiến lược AI của Đức vào tháng 11/2019. Đến tháng 10/2020, Ủy ban Nghiên cứu về Trí tuệ Nhân tạo - Trách nhiệm Xã hội, Tiềm năng Kinh tế, Xã hội và Sinh thái của Hạ viện Đức đã trình bày báo cáo cuối cùng với các khuyến nghị hành động cụ thể. Tháng 12/2020, Chính phủ Liên bang Đức đã áp dụng Chiến lược AI cập nhật. Trong Chiến lược AI cập nhật, Chính phủ Liên bang ủng hộ việc sử dụng phương pháp tiếp

cận “đạo đức theo thiết kế” trong tất cả các giai đoạn phát triển và sử dụng các ứng dụng dựa trên AI cũng như đề cao sự tham gia đối thoại với các nước và khu vực hàng đầu về phát triển AI để đạt được thỏa thuận về các hướng dẫn chung và tiêu chuẩn đạo đức cho AI.

1.4. Hợp tác quốc tế trong thúc đẩy thực hiện đạo đức trí tuệ nhân tạo

Để đẩy mạnh việc thực hiện đạo đức trí tuệ nhân tạo trên phạm vi quốc tế, CHLB Đức đã tạo ra một liên minh với Canada để phát triển một AI có đạo đức bằng cách thành lập Hội nghị Canada và Đức (GCC) vào năm 2012. Nhiệm vụ của liên minh là “xây dựng các hoạt động hợp tác AI xuyên Đại Tây Dương” với mục tiêu tạo ra và duy trì sức mạnh nhằm thúc đẩy AI có đạo đức. Sự hợp tác giữa Đức và Canada khác với Trung Quốc và Mỹ là cả Trung Quốc và Mỹ đều không có ý định tạo ra một đạo đức cho AI. Sự quan tâm của Canada và Đức đối với một tiêu chuẩn đạo đức về AI đã tạo nên một liên minh để tăng tính khả dụng của thông tin và tạo tiền lệ cho sự sẵn sàng cộng tác của quốc tế.

1.5. Hoàn thiện cơ sở pháp lý hiện hành để điều chỉnh các hoạt động đạo đức trí tuệ nhân tạo

Hiện nay, CHLB Đức chưa xây dựng và ban hành một đạo luật cụ thể trực tiếp nào liên quan đến vấn đề đạo đức trí tuệ nhân tạo. Tuy nhiên CHLB Đức đã xây dựng được một số đạo luật có thể điều chỉnh một số khía cạnh của đạo đức trí tuệ nhân tạo, cụ thể:

** Luật Bảo vệ dữ liệu*

Luật Bảo vệ dữ liệu của CHLB Đức (German Bundesdatenschutzgesetz -BDSG) hướng tới việc đảm bảo dữ liệu cá nhân chỉ có thể được sử dụng khi có sự đồng ý của cá nhân. Giá trị đạo đức cơ bản được luật pháp bảo vệ về mặt này là quyền riêng tư và quyền tự chủ cá nhân. Một số khía cạnh trong luật có thể được sử dụng để thực hiện đạo đức trí tuệ nhân tạo như: Cấm việc thu thập, xử lý và sử dụng dữ liệu cá nhân, trừ khi được pháp luật cho phép hoặc người có liên quan đồng ý (khoản 4 điều I, BDSG); Dữ liệu cá nhân phải được thu thập trực tiếp từ người có liên quan (khoản 4 điều III, BDSG); Mọi hệ thống xử lý dữ liệu phải đạt được mục tiêu là không sử dụng (hoặc càng ít càng tốt) dữ liệu nhận dạng cá nhân; Nếu dữ liệu cá nhân được thu thập, đơn vị chịu trách nhiệm phải thông báo cho người bị ảnh hưởng về danh tính của họ và mục đích của việc thu thập, xử lý hoặc sử dụng (khoản 4 điều III, BDSG).

** Luật Viễn thông*

Vào ngày 17 tháng 2 năm 2017, Cơ quan Mạng lưới Liên bang Đức đã ra phán quyết cấm bán một con búp bê có tên là Cayla và ra lệnh tiêu hủy tất cả các thiết bị đã được bán (German Federal Network Agency, 2017). Cơ sở pháp lý của quyết định này là Số 2 Điều 90, Luật Viễn thông Đức. Lý do là vì búp bê có kết nối với nhà sản xuất (bắt buộc vì búp bê đã được kích hoạt AI), búp bê thực sự là một gián điệp theo dõi đứa trẻ, ghi lại tất cả dữ liệu mà đứa trẻ nói

với các thiết bị, bao gồm cả những bí mật của đứa trẻ. Tương tự như vậy, cơ quan này lo ngại rằng các thiết bị này có thể bị hack khiến trẻ em phải đối mặt với các mối đe dọa như ấu dâm hoặc ảnh hưởng tư tưởng (Kay Firth-Butterfield, 2018). Kể từ đó, cơ quan quản lý tại CHLB Đức đã sử dụng Luật Viễn thông để cấm các thiết bị tương tự như đồng hồ thông minh. Phương pháp nghiêm ngặt này được áp dụng để bảo vệ trẻ em, một trong những đối tượng dễ bị tổn thương nhất.

Như vậy có thể thấy, với một loạt các hoạt động nhằm củng cố và hoàn thiện từ cơ sở pháp lý, thể chế, các nguyên tắc đạo đức trong ứng dụng AI đến nay, Đức đã hoàn thiện về cơ bản những điều kiện cần thiết để đẩy mạnh và ứng dụng đạo đức AI vào trong cuộc sống.

2. Một số gợi mở cho Việt Nam trong triển khai ứng dụng trí tuệ nhân tạo trong cuộc sống

Chính phủ Việt Nam đã xác định trí tuệ nhân tạo là một trong những công nghệ đột phá, mũi nhọn của cuộc cách mạng công nghiệp 4.0. Từ năm 2014, AI đã được đưa vào danh sách công nghệ cao được ưu tiên đầu tư phát triển. Nhiều năm qua, Chính phủ đã phê duyệt danh mục sản phẩm công nghệ cao ưu tiên đầu tư và danh mục sản phẩm công nghệ cao khuyến khích phát triển, trong đó AI được đưa vào danh mục công nghệ cao được ưu tiên đầu tư phát triển. Ý chí phát triển AI của Chính phủ còn được thể hiện qua hàng loạt văn bản như: (i) Chỉ thị số 16/CT-TTg ngày 4/5/2017 của Thủ tướng

Chính phủ về việc tăng cường khả năng tiếp cận Công nghiệp 4.0 xác định công nghệ AI là một trong những khâu đột phá và là công nghệ mũi nhọn; (ii) Nghị quyết phiên họp Chính phủ tháng 3/2018 đặt ra nhiệm vụ nghiên cứu, đánh giá tác động và xây dựng Chiến lược quốc gia về công nghiệp 4.0. Trong quá trình xây dựng Chiến lược, công nghệ AI sẽ được đánh giá, nghiên cứu kỹ lưỡng để hình thành chiến lược, lộ trình, giải pháp phát triển và ứng dụng AI cụ thể phù hợp nhất với thực tế và tiềm năng của các doanh nghiệp Việt Nam; và (iii) đặc biệt là Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030, ban hành ngày 26/01/2021 với mục tiêu trở thành trung tâm đổi mới và phát triển, dẫn đầu về các giải pháp và ứng dụng trí tuệ nhân tạo trong Khu vực ASEAN và thế giới (Thủ tướng Chính phủ, 2021).

Thực tế phát triển AI ở Việt Nam trong thời gian qua cho thấy ứng dụng AI đã góp phần tạo ra những sản phẩm công nghệ mới, thúc đẩy sự phát triển của nhiều lĩnh vực như công nghệ thông tin và truyền thông, y tế, du lịch... Nó mang lại rất nhiều lợi ích trong việc chăm sóc sức khỏe, cải thiện hiệu quả của hệ thống sản xuất, tăng cường an ninh và theo nhiều cách khác nữa mà chúng ta chỉ có thể bắt đầu hình dung. Tuy nhiên, trí tuệ nhân tạo cũng tạo ra một số rủi ro tiềm ẩn như ra quyết định không rõ ràng, thiên vị và phân biệt đối xử, ảnh hưởng đến cuộc sống cá nhân của con người hoặc bị sử dụng cho mục đích phạm tội. Do đó, ngoài việc xây

dựng một chiến lược phát triển AI, Việt Nam cần xác định rõ một số yếu tố trọng tâm trong quá trình xây dựng và phát triển AI như:

- *Về xây dựng hệ thống văn bản quy phạm pháp luật và hành lang pháp lý về AI.* Có thể thấy, AI và các vấn đề chính sách liên quan đến AI còn rất mới không chỉ đối với Việt Nam - một quốc gia non trẻ trong lĩnh vực AI này, mà còn đối với thế giới. Việt Nam đã có những động thái ban đầu trong quá trình thiết lập hệ thống pháp lý cho việc phát triển và ứng dụng trí tuệ nhân tạo trong cuộc sống. Tuy nhiên, vẫn còn nhiều vấn đề về trí tuệ nhân tạo mà con người chưa hiểu rõ, chưa thống nhất, khó có thể khẳng định hay lường trước được tác động xã hội, cũng như những tác động đến phạm trù đạo đức trong ứng dụng trí tuệ nhân tạo. Trí tuệ nhân tạo có thể dẫn đến gia tăng bất bình đẳng xã hội, tạo nên một thị trường lao động chia tách, phân mảnh với mức lương quá chênh lệch, làm trầm trọng thêm phân hóa và phân tầng xã hội... Do đó Việt Nam cần phải sớm hoàn thiện khuôn khổ pháp lý, cơ chế vận hành đồng bộ cho việc phát triển và ứng dụng trí tuệ nhân tạo vào cuộc sống.

- *Thiết lập được bộ quy tắc về đạo đức trí tuệ nhân tạo*

Đức là quốc gia đi đầu trong công nghệ AI, việc xuất hiện các thách thức mới từ phát triển AI là không thể tránh khỏi. Những thách thức này làm nảy sinh những vấn đề liên quan đến đạo đức của trí tuệ nhân tạo. Quá trình thiết kế, ứng dụng và triển khai AI đang được cho là có mối nguy

hiểm tiềm tàng làm xói mòn những quyền cơ bản của con người hay AI có thể gây ra những ảnh hưởng đến cách nhận thức của con người về thế giới bên ngoài nếu thiếu đi sự minh bạch, trách nhiệm giải trình, sự công bằng, tính trung thực và đạo đức. Với việc CHLB Đức đưa ra những nguyên tắc đạo đức để nhằm đảm bảo thực hiện được mục tiêu kép là thúc đẩy sự hấp thụ sử dụng AI và giải quyết những rủi ro liên quan đến việc sử dụng một số ứng dụng mới này rất hữu ích cho Việt Nam học hỏi kinh nghiệm. Bởi trong quá trình xử lý dữ liệu tự động về sức khỏe, công việc, phúc lợi của công dân, AI có thể đưa ra những quyết định có kết quả phân biệt đối xử và không công bằng. “Mặt tối” của các thuật toán trong quá trình đưa ra quyết định được giải quyết thông qua các nguyên tắc này sẽ đảm bảo các quy trình ra quyết định tự động tương thích với quyền con người và đảm bảo dân chủ.

Tuy nhiên, khái niệm liên quan đến vấn đề đạo đức có thể được hiểu khác nhau giữa các nền văn hóa hay các cộng đồng khác nhau. Truyền thống của phương Tây coi trọng quyền riêng tư của cá nhân nhưng điều này không được cho là vấn đề quá quan trọng ở phương Đông. Nho giáo nhấn mạnh đến lợi ích tập thể nhiều hơn đến lợi ích cá nhân, do vậy quan niệm về quyền riêng tư cá nhân theo truyền thống ít được chú ý hơn và đôi khi mang tính tiêu cực. Hiện nay, các cuộc tranh luận về vấn đề đạo đức đều mong muốn rằng các hệ thống AI được xây dựng dựa trên giá trị của con

người. Tuy nhiên, ai sẽ là người quyết định giá trị của con người? Trong bối cảnh của mỗi dân tộc, mỗi quốc gia, khu vực lại có những quan niệm khác nhau về đạo đức, chuẩn mực xã hội. Do đó, trong quá trình phát triển và ứng dụng công nghệ AI, Việt Nam cần phải chủ động xây dựng cho mình một bộ quy tắc về đạo đức riêng phù hợp với văn hóa và luật pháp Việt Nam, bên cạnh đó cũng cần thúc đẩy hợp tác với các quốc gia trong khu vực và quốc tế xây dựng một chương trình khung toàn cầu để đảm bảo các công nghệ trí tuệ nhân tạo hoạt động dựa trên tiêu chuẩn đạo đức chung.

Việt Nam cũng cần có những bước chuẩn bị cho việc phát triển các thể chế có liên quan, quy hoạch, đầu tư phát triển công nghệ AI trước khi công nghệ đi vào đời sống và ảnh hưởng đến đời sống của con người. Sự thiếu kiến thức về trí tuệ nhân tạo hay sự chậm trễ trong đầu tư phát triển lĩnh vực công nghệ này có thể làm gia tăng khoảng cách về công nghệ giữa Việt Nam và thế giới, đặc biệt là các nước phát triển. Do đó, Việt Nam cần phải có các giải pháp xây dựng cơ sở hạ tầng công nghệ trí tuệ nhân tạo đáp ứng tốt yêu cầu phát triển khoa học công nghệ. Cần có mục tiêu và quy hoạch cụ thể đối với các vùng để có các hướng xây dựng tập trung chuyển đổi cơ sở hạ tầng và chuyển đổi cơ cấu thị trường lao động có giá trị gia tăng thấp sang thị trường sản xuất có giá trị gia tăng cao với các khoản đầu tư có mục tiêu vào nghiên cứu và phát triển về công nghệ trí tuệ nhân tạo. Có như vậy, Việt

Nam mới có thể bắt kịp công nghệ và ứng dụng AI với các quốc gia trong khu vực và thế giới. Ngoài ra, cần có thể chế đối phó với những vấn đề đạo đức đặt ra trong tương lai do sự khác biệt về ý thức hệ. Mặt khác, cần xác định rõ vai trò của các chủ thể: nhà nước hay phi nhà nước trong quản lý và phát triển công nghệ trí tuệ nhân tạo.

Tài liệu tham khảo

1. “Begriff und Geschichte des Datenschutzes”, 28 May 2014.
2. Data Ethics Commission (2019), “Opinion of the Data Ethics Commission”, https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=1
3. Ethics Commission (2017) “Automated and Connected Driving”, https://www.bmvi.de/SharedDocs/EN/Documents/G/ethic-commission-report.pdf?__blob=publicationFile.
4. Kay Firth-Butterfield (2018), “*Generation AI: What happens when your child's friend is an AI toy that talks back?*”, World Economic Forum, <http://www.weforum.org/agenda/2018/05/generation-ai-whathappens-when-your-childs-invisible-friend-is-an-ai-toy-that-talks-back/>.
5. German Federal Network Agency (2017), Bundesnetzagentur Removes Children’s Doll, “Cayla” From the Market, Bundesnetzagentur [BNetzA].
6. France24 (2018), German firm's 7 commandments for ethical AI, tải ngày: 20/1/2022, <https://www.france24.com/en/20180918-german-firms-7-commandments-ethical-ai>
7. Thủ tướng Chính phủ (2021), “*Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030*”.