AI ĐÁNG TIN CẬY VÀ CÁC NGUYÊN TẮC THỰC THI

ThS. CHU THỊ THẮM, TS. NGUYỄN ĐỨC THỦY
Viện Công nghệ số và Chuyển đổi số quốc gia, Bộ TT&TT

Ngày nay, với những tiềm năng ứng dụng khổng lồ và tác động to lớn của AI, một điều đặc biệt quan trọng là phải đảm bảo rằng tất cả các hệ thống AI phải hoạt động một cách đáng tin cậy "như con người" hoặc

Tóm tắt:
- Hiện nay cần có cơ chế để có thể giám sát các quyết định do AI đưa ra, đảm bảo rằng chúng đáng tin cậy, phù hợp với luật pháp, thông lệ...
- Khái niệm tính tin cậy của AI.
- Bốn nguyên tắc thực thi AI đáng tin cậy:
+ Nguyên tắc tôn trọng quyền tự chủ của con người;
+ Nguyên tắc ngăn ngừa tác hại;
+ Nguyên tắc giải thích;
+ Nguyên tắc công bằng.

Tầm quan trọng và sự cần thiết đảm bảo AI tin cậy

Thị trường AI được dự báo là rất lớn, trị giá khoảng 200 tỷ đô la Mỹ vào năm 2023 và dự kiến sẽ tăng trưởng vượt mức đó lên hơn 1.800 tỷ đô la Mỹ vào năm 2030 [1]. Những con số xét về khía cạnh thị trường này cho thấy AI đang tác động đến xã hội của chúng ta như thế nào trong những năm tới, khi mà ứng dụng của nó lan tỏa trong hầu hết các lĩnh vực kinh tế, xã hội của một quốc gia. Việc ra quyết định bằng thuật toán và trí tuệ nhân tạo (AI) đang đóng vai trò quan trọng trong cuộc sống hàng ngày của chúng ta.

Với lượng dữ liệu khổng lồ, sức mạnh tính toán, hiệu năng thực thi của các thuật toán ngày càng được nâng cao, AI đã và sẽ cung cấp cho chúng ta nhiều giải pháp hữu ích, mang lại lợi ích cho xã hội. Tuy nhiên, đi cùng với những lợi ích đó, sự phát triển, những tiến bộ của AI cũng đang gây ra những mối lo ngại về độ phức tạp, khó hiểu để con người có thể kiểm soát nó.

Stephen Hawking từng nói *"Tác động của AI có thể là thảm họa trừ phi sự phát triển nhanh chóng của nó được kiểm soát"* [2]. Hệ thống AI có thể nguy hiểm và có hại nếu không thực hiện các biện pháp nghiêm ngặt trong quá trình thiết kế, phát triển, triển khai và giám sát chúng.

Các chuyên gia cho rằng, hiện nay những hệ thống này thường được thiết kế quá phức tạp và không rõ ràng, nên rất khó để đánh giá và lý giải liệu quyết định mà chúng đưa ra có công bằng và đáng tin cậy? Một điều nữa là phải dựa vào các tiêu chuẩn hoặc cơ chế nào để quản lý và kiểm tra tính tin cậy của các hệ thống này?

Thực tế cho thấy một số hệ thống AI được triển khai đã phát sinh nhiều vấn đề. Ví dụ, ô tô tự lái đã giết chết một người đi bộ vì hệ thống tự lái của nó quyết định không thực hiện bất kỳ hành động nào sau khi phát hiện người đi bộ trên đường [3]. AI chat-bot có biểu hiện phân biệt chủng tộc trong việc thu thập tri thức để trả lời câu hỏi trên Twitter [4]; hoặc là ở Mỹ, thuật toán đánh giá mức độ tái phạm tội từ các tù nhân được thả, sử dụng ở phạm vi toàn liên bang cũng có biểu hiện thiên vị về chủng tộc [5]. Thuật toán tuyển dụng được một công ty đa quốc gia sử dụng được phát hiện là có thành kiến với phụ nữ [6]. Chương trình xác định các đối tượng khó khăn để hỗ trợ hóa đơn tiền điện (BOSCO) của chính phủ Tây Ban Nha triển khai trên phạm vi toàn quốc đã nhận được rất nhiều những khiếu nại, xuất phát từ sự thiếu minh bạch của hệ thống [7].

Những ví dụ trên cho thấy rằng, AI có thể không đáng tin cậy và nguy hiểm như thế nào nếu sự phát triển của chúng không được kiểm soát.

Do vậy, nhu cầu hiện nay là cần có cơ chế để có thể giám sát các quyết định do AI đưa ra, đảm bảo rằng chúng đáng tin cậy, phù hợp với luật pháp, thông lệ, thậm chí là phù hợp với văn hóa, bản sắc, thể chế chính trị của từng quốc gia, dân tộc.

Khái niệm AI đáng tin cậy

Tính đáng tin cậy (Trustworthy AI) có nghĩa là một chủ thể đáng được tin cậy hoặc có thể tin cậy được. Hiện nay, khái niệm về tính đáng tin cậy trong AI được nhiều tổ chức, cá nhân nêu ra, tùy thuộc vào cách thức họ nhìn nhận AI ở các khía cạnh thực thi hoặc sự thể hiện khác nhau của nó. Tuy nhiên, một định nghĩa mang tính tổng quan và khá toàn diện, thể hiện những thuộc tính chung nhất về tính tin cậy của AI được Hiệp hội Internet công nghiệp (IIC) cũng như Viện Tiêu chuẩn và công nghệ quốc gia của Hoa Kỳ (NIST) đưa ra và được phát biểu như sau [8]:

"Tính tin cậy là mức độ tin tưởng một hệ thống thực hiện như mong đợi. Các đặc điểm của nó bao gồm tính an toàn, tính bảo mật, tính riêng tư, độ tin cậy và khả năng phục hồi khi hệ thống phải đối mặt với những xáo trộn của môi trường, lỗi của con người, lỗi hệ thống và các cuộc tấn công".

Khái niệm này chỉ ra năm đặc điểm thể hiện các thuộc tính của tính đáng tin cậy và bốn loại

4 nguyên tắc thực thi của hệ thống AI được coi là đáng tin cậy.

hình đe dọa đối với tính đáng tin cậy của AI là: xáo trộn môi trường, tấn công, lỗi con người và lỗi hệ thống. Những mối đe dọa này có thể dẫn đến các nguy cơ nảy sinh các hệ quả tiêu cực do sự xuất hiện những sự kiện không mong muốn hoặc không có trong kế hoạch.

Bốn nguyên tắc thực thi AI đáng tin cậy

Trí tuệ nhân tạo được sử dụng để đưa ra quyết định trong các ứng dụng quan trọng như chăm sóc sức khỏe, giao thông vận tải, hệ thống tư pháp và nhiều ứng dụng khác. Với sự gia tăng sử dụng hệ thống AI, đòi hỏi cần phải có các hướng dẫn và chính sách để đảm bảo rằng AI sẽ không gây ra bất kỳ tác hại vô tình hoặc cố ý cho cả xã hội và người dùng sử dụng nó. Vấn đề này ngày càng trở nên quan trọng và cấp thiết, được các tổ chức quốc tế, các cơ quan quản lý của các quốc gia, khu vực trên thế giới nỗ lực quyết tâm, tập trung tiềm lực nghiên cứu, tìm các biện pháp giải quyết.

Trên cơ sở đó, các cơ quan chuyên môn của các quốc gia, tổ chức chịu trách nhiệm đã xúc tiến các hoạt động nghiên cứu, đề xuất các hướng dẫn và chính sách hướng tới mục tiêu đảm bảo rằng các hệ thống, sản phẩm, dịch vụ AI được sử dụng mang tính tin cậy. Đến thời điểm hiện tại, các chuyên gia, nhà nghiên cứu của các tổ chức, doanh nghiệp từ các quốc gia, khu vực đã thống nhất với nhau 4 nguyên tắc chính để AI hoạt động có tính tin cậy [9], bao gồm: (i) Nguyên tắc tôn trọng quyền tự chủ của con người; (ii) Nguyên tắc ngăn ngừa tác hại; (iii) Nguyên tắc giải thích và (iv) Nguyên tắc công bằng (hình trên).

1. Nguyên tắc tôn trọng quyền tự chủ của con người

Đây là nguyên tắc quan trọng nhất trong thiết kế, triển khai, vận hành và sử dụng AI đáng tin cậy. Nó đảm bảo rằng các hệ thống AI phải được thiết kế để bổ sung, tăng cường kỹ năng và nâng cao khả năng nhận thức của con người chứ không phải thay thế cho con người. Kỷ nguyên phát triển AI ngày nay đòi hỏi hoạt động tư duy mang tính hợp tác, con người và máy móc làm việc cùng nhau hướng tới những mục tiêu chung.

Sự kết hợp làm việc giữa con người và máy móc giúp giảm thiểu những kết quả không mong muốn, không chính xác mang tính chủ quan mà con người dễ mắc phải, tránh được những rủi ro có thể xảy ra nhiều khi rất nghiêm trọng. Tôn trọng quyền tự chủ của con người còn có nghĩa là việc thiết kế các hoạt động của máy móc dựa trên AI phải lấy con người làm trung tâm, con người phải tham gia vào các cấp độ, giai đoạn khác nhau trong vòng đời hoạt động của mọi hệ thống, sản phẩm, dịch vụ AI.

Cụ thể, con người tham gia vào giai đoạn lập kế hoạch, thiết kế, phát triển và giám sát hệ thống AI dựa trên các yêu cầu ứng dụng cụ thể của chúng. Con người phải là nhân tố trung tâm đặt ra các phạm vi, giới hạn cảnh báo các lỗi do máy móc gây ra. Trong các tình huống ứng dụng quan

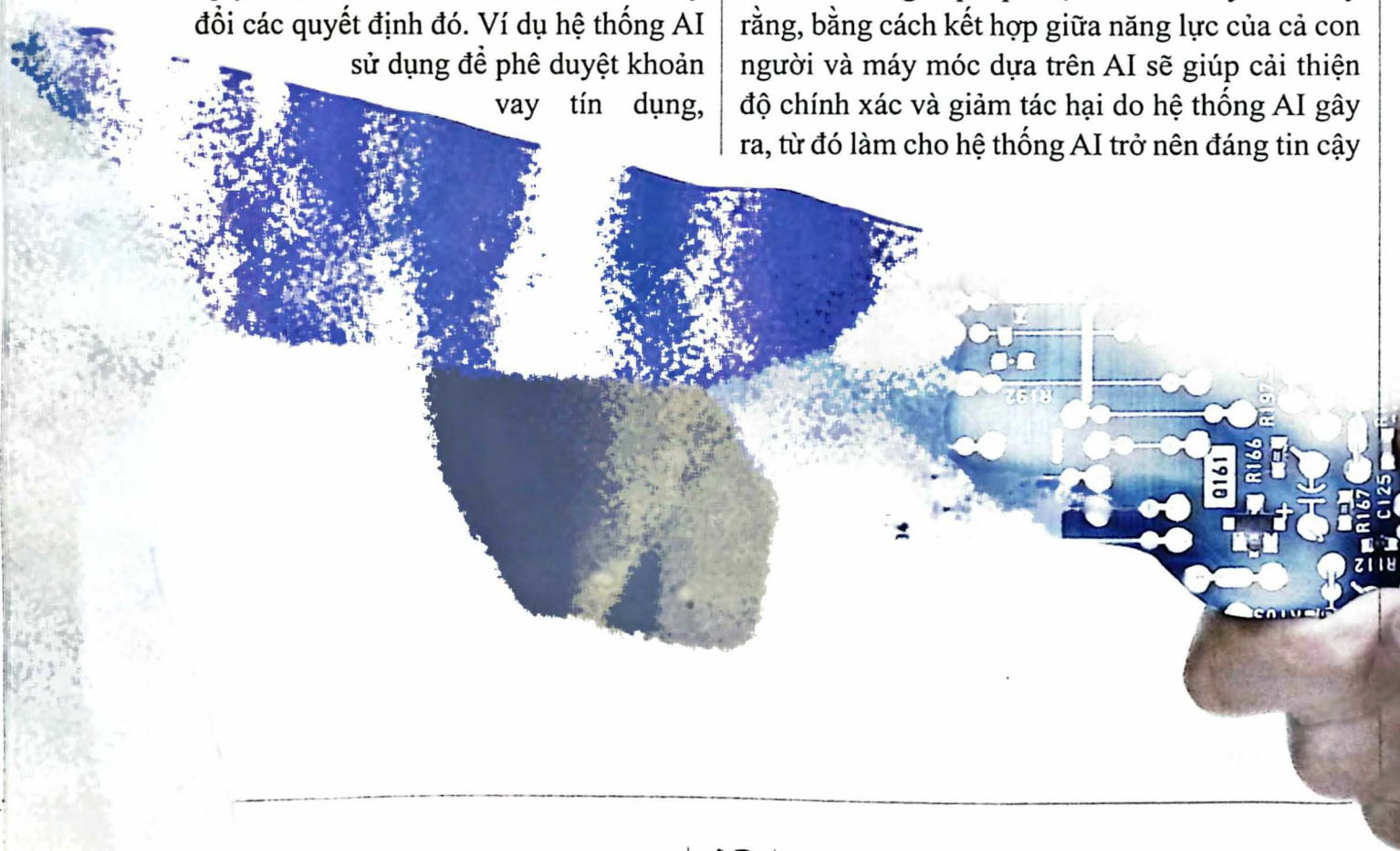
trọng, con người phải xem xét, cân nhắc, sửa đổi quyết định cuối cùng do AI đưa ra và được các hệ thống AI ghi nhận, học tập để dần cải thiện độ chính xác, tin cậy của chúng. Sự tham gia, can thiệp của con người là rất quan trọng để đảm bảo rằng máy móc hoạt động dựa vào AI đưa ra các quyết định có tính đạo đức và tuân thủ các hướng dẫn thực hành các quy tắc đạo đức được con người thiết lập cho nó.

Đối với các ứng dụng AI sử dụng trong lĩnh vực hoặc môi trường được đánh giá có mức độ rủi ro cao, các quyết định do hệ thống AI đưa ra chỉ có hiệu lực nếu được con người kiểm tra và xác thực. Ví dụ, hệ thống AI sử dụng trong y tế, nơi một quyết định sai lầm có thể dẫn đến hậu quả nguy hiểm đến sinh mạng con người, các bác sĩ nên xác thực quyết định của hệ thống AI dựa trên chuyên môn và kinh nghiệm của họ trước khi thực hiện chúng. Đối với các ứng dụng yêu cầu các quyết định của AI có hiệu lực ngay lập tức, cần có cách thức để con người có thể can thiệp, xem xét quyết định của AI, nếu cần họ có thể thay đổi các quyết định đó. Ví dụ hệ thống AI sử dụng để phê duyệt khoản vay tín dụng, nếu đơn đăng ký bị hệ thống AI từ chối, chuyên gia có thể xem xét lại đơn vay và thay đổi quyết định nếu cần.

Trong các hệ thống AI, con người phải được cung cấp năng lực giám sát hoạt động, giành quyền điều khiển, can thiệp ở mọi nơi, mọi lúc nếu như phát hiện rằng chúng hoạt động không phù hợp, hoặc các quyết định do hệ thống AI đưa ra không còn an toàn nữa. Ví dụ trong xe tự hành, nếu một số cảm biến bị lỗi hoặc xe hoạt động không bình thường, người lái xe có thể chiếm quyền điều khiển xe.

Một số phương pháp đã được đề xuất cho sự hợp tác giữa con người và máy móc để tăng cường độ tin cậy và độ chính xác của hệ thống AI. Những ví dụ điển hình về khung hợp tác tin cậy giữa người và AI sử dụng cho các ứng dụng như phát hiện xâm nhập, hệ thống sàng lọc giao dịch thẻ tín dụng, hệ thống phát hiện người dùng giả mạo trên mạng xã hội v.v.. đã tăng độ tin cậy hoạt động của các hệ thống AI lên nhiều lần.

Tất cả các giải pháp được đề xuất này cho thấy rằng, bằng cách kết hợp giữa năng lực của cả con người và máy móc dựa trên AI sẽ giúp cải thiện độ chính xác và giảm tác hại do hệ thống AI gây ra, từ đó làm cho hệ thống AI trở nên đáng tin cậy



hơn. Vì vậy, nguyên tắc này buộc các hệ thống AI phải trao quyền cho con người chứ không phải thay thế họ.

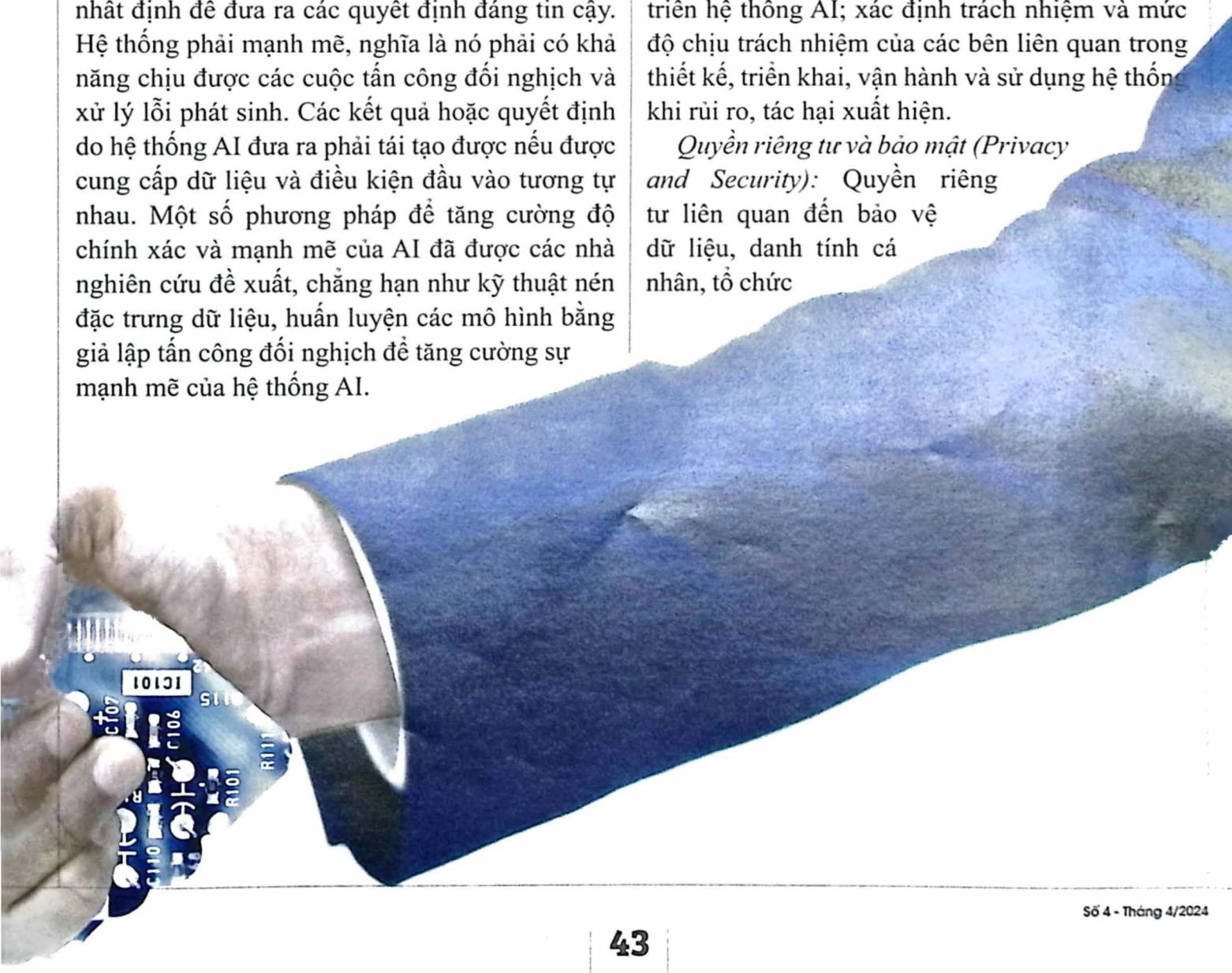
2. Nguyên tắc ngăn ngừa tác hại

Nguyên tắc này đảm bảo rằng hệ thống AI không được gây ra bất kỳ tổn hại vô ý hoặc cố ý nào cho con người và xã hội. Nó cũng đảm bảo rằng các hệ thống AI hoạt động trong môi trường an toàn và được bảo mật mà không gây ra bất kỳ tác hại cho bất kỳ ai. Hệ thống AI phải đáng tin cậy trong thực hiện các tác vụ ra quyết định. Để được như vậy, các yếu tố dưới đây cần được xem xét khi thiết kế và triển khai hệ thống AI:

Độ chính xác và mạnh mẽ (Accuracy and Robustness): Thuộc tính này đảm bảo rằng hệ thống AI phải có độ chính xác cao hơn ngưỡng nhất định để đưa ra các quyết định đáng tin cậy. Hệ thống phải mạnh mẽ, nghĩa là nó phải có khả năng chịu được các cuộc tấn công đối nghịch và xử lý lỗi phát sinh. Các kết quả hoặc quyết định do hệ thống AI đưa ra phải tái tạo được nếu được cung cấp dữ liệu và điều kiện đầu vào tương tự nhau. Một số phương pháp để tăng cường độ chính xác và mạnh mẽ của AI đã được các nhà nghiên cứu đề xuất, chẳng hạn như kỹ thuật nén đặc trưng dữ liệu, huấn luyện các mô hình bằng giả lập tấn công đối nghịch để tăng cường sự mạnh mẽ của hệ thống AI.

Trách nhiệm giải trình (Accountability): Thuộc tính này đề cập việc quy đối tượng chịu trách nhiệm đối với mọi quyết định (tốt hoặc xấu) do hệ thống AI đưa ra. Vì các thuật toán không thể chịu trách nhiệm về các quyết định của mình, nên các nhà thiết kế hệ thống AI hoặc bên liên quan khác phải chịu trách nhiệm về hoạt động của hệ thống bằng các khuôn khổ, quy trình kiểm tra, kiểm soát và các cơ chế giám sát thích hợp. Nó liên quan đến việc tuân thủ các quy định (chẳng hạn như văn bản quản lý của các cấp có thẩm quyền ban hành), các quy tắc để phát triển hệ thống AI có kiểm soát. Một số phương pháp kiểm tra, kiểm soát đã được đề xuất để ngăn ngừa tác hại như: khung nội bộ để kiểm soát hoạt động của các thuật toán trong suốt vòng đời phát triển hệ thống AI; xác định trách nhiệm và mức độ chịu trách nhiệm của các bên liên quan trong thiết kế, triển khai, vận hành và sử dụng hệ thống khi rủi ro, tác hại xuất hiện.

Quyền riêng tư và bảo mật (Privacy and Security): Quyền riêng tư liên quan đến bảo vệ dữ liệu, danh tính cá nhân, tổ chức



sở hữu dữ liệu dùng để huấn luyện hệ thống AI. Bảo mật hệ thống AI liên quan đến việc bảo vệ hệ thống khỏi các cuộc tấn công từ bên ngoài, can thiệp và làm xáo trộn hoạt động của hệ thống AI. Các phương pháp thực hiện chức năng này có thể là quy trình bảo vệ dữ liệu khi có hai hoặc nhiều bên liên quan tham gia phát triển hệ thống AI; phát hiện và ngăn chặn các loại hình tấn công; áp dụng và tuân thủ các quy định của luật hoặc các văn bản dưới luật về quyền riêng tư, bảo mật dữ liệu.

3. Nguyên tắc công bằng

Nguyên tắc công bằng đảm bảo rằng các quyết định do hệ thống AI đưa ra không được thiên vị và phân biệt đối xử. Hệ thống AI ứng dụng trong các lĩnh vực có thể có những hành vi phân biệt đối xử khiến cho các dự báo, quyết định đưa ra không công bằng, làm giảm niềm tin của người dùng. Nguyên tắc này đảm bảo rằng hệ thống AI phải đối xử bình đẳng với tất cả người dùng mà không thiên vị bất kỳ cá nhân, đối tượng hoặc nhóm người trong xã hội. Điều này có nghĩa là hệ thống AI phải có nghĩa vụ tuân thủ các giá trị đạo đức phổ quát, giá trị văn hóa mang tính bản sắc của quốc gia, dân tộc. Như vậy, nếu một hệ thống AI không được thiết kế, sử dụng tuân thủ các nguyên tắc về công bằng nêu trên sẽ là hệ thống có khả năng xuất hiện tính thiên vị và bất công.

Có nhiều nguyên nhân dẫn đến hệ thống AI hoạt động thiên vị và không công bằng, chẳng hạn như hệ thống AI được huấn luyện bởi dữ liệu sai lệch hoặc không trung thực, hoặc tính chất thống kê áp dụng cho dữ liệu không phản ánh đầy đủ và toàn diện các thuộc tính của nó đối với các đối tượng cần xử lý. Một ví dụ điển hình là bộ dữ liệu ImageNet được cộng đồng thị giác máy tính sử dụng rộng rãi hiện nay không phản ánh tính đa dạng về đặc điểm khuôn mặt theo chủng tộc, màu da hoặc theo vùng địa lý trên toàn cầu. Sự thiên vị có thể tạo ra bởi chính các thuật toán AI, ví dụ như khi thuật toán cố gắng tối đa hóa độ chính xác của nó đối với dữ liệu huấn luyện. Một số nguyên nhân khác tạo ra sự thiên vị có thể xuất hiện ở mọi khâu trong vòng đời của hệ thống, chẳng hạn như người thu thập dữ liệu, thiết kế hệ thống hoặc tương tác với hệ thống có thành kiến với một bộ phận nào đó trong xã hội. Vì vậy, thực hiện các biện pháp để làm cho hệ thống AI trở nên công bằng, không thiên vị là việc làm cần thiết.

Một số biện pháp giảm thiểu sự thiên vị, không công bằng được đề xuất như: kỹ thuật kiểm tra để phát hiện mọi tổ hợp thuộc tính dữ liệu đầu vào có khả năng phân biệt đối xử với bất kỳ cá nhân nào dựa trên giới tính, chủng tộc, sắc tộc; giải pháp hệ thống xếp hạng của bên thứ ba để phát hiện sự thiên vị; kiểm tra bằng cách sử dụng các bộ dữ liệu có và không có sự thiên vị.

4. Nguyên tắc giải thích

Nguyên tắc về khả năng giải thích liên quan đến việc hệ thống AI có khả năng đưa ra các giải thích, diễn giải cho các dự báo, quyết định mà nó đưa ra. Điều này xuất phát từ việc người ta lo ngại rằng độ phức tạp của các mô hình AI ngày càng cao khiến chúng đã trở thành những hộp đen khó hiểu và khó giải thích. Nguyên tắc giải thích đảm bảo rằng hoạt động của các hệ thống AI có thể được trao đổi cởi mở với các bên liên quan khác nhau và với những người có thể bị ảnh hưởng trực tiếp hoặc gián tiếp từ các quyết định của nó. Nguyên tắc này làm cho quá trình ra quyết định trở nên minh bạch, từ đó làm tăng sự tin tưởng của người dùng vào hệ thống. Nó cho phép người dùng hiểu chính xác lý do dẫn đến một quyết định cụ thể. Khả năng giải thích của hệ thống cũng giúp các nhà quản lý hiểu rõ hơn về hệ thống để đưa ra các luật, quy định phù hợp. Đồng thời nó giúp các nhà phát triển hệ thống phát hiện nguyên nhân sai sót và làm cho hệ thống hoạt động chính xác hơn.

Một số cách tiếp cận để làm cho hệ thống AI có khả năng giải thích được, chẳng hạn như tích hợp cơ chế giải thích vào vòng đời phát triển AI; phương pháp hậu kiểm, coi hệ thống AI như hộp đen để kiểm tra khả năng giải thích của nó; hoặc là tạo lập cơ chế giải thích toàn bộ hoạt động của mô hình, cơ chế cục bộ giải thích về một quyết định cụ thể do hệ thống đưa ra. Tất cả những cách tiếp cận này đều hướng tới mục tiêu làm cho hệ thống AI trở nên minh bạch và dễ hiểu đối với nhiều đối tượng người dùng khác nhau. Hình thức giải thích có thể thông qua các giao diện người dùng bằng đối thoại văn bản hoặc các mô hình mang tính trực quan hóa, dễ hiểu đối với người dùng.

Tuy nhiên, việc thực thi các cơ chế minh bạch cũng gặp nhiều thách thức, chẳng hạn như thiếu các yêu cầu và tiêu chuẩn rõ ràng trong lĩnh vực này; định nghĩa và khái niệm để cụ thể hóa các nguyên tắc còn chưa rõ ràng; chưa có sự thống nhất, thậm chí các nguyên tắc có thể xung đột với nhau. Ví dụ, tuân thủ nguyên tắc giải thích có thể xung đột với nguyên tắc ngăn ngừa tác hại, vì mô hình càng dễ hiểu và minh bạch thì càng dễ bị tấn công từ bên ngoài. Do đó, cần có sự cân bằng giữa các nguyên tắc dựa trên các yêu cầu ứng dụng và cần có những quy định nghiêm ngặt để chi phối hoạt động của hệ thống AI.

Một thách thức khác là một giải pháp có thể hiệu quả cho đối tượng này nhưng lại không hiệu quả cho đối tượng khác. Ví dụ cơ chế giải thích đối với nhà phát triển hệ thống có thể không có ý nghĩa đối với người dùng, vì họ không có kiến thức kỹ thuật. Do đó cần có những giải pháp cụ thể phù hợp với đối tượng và bối cảnh sử dụng. Một điều quan trọng nữa đó là tính ứng dụng đa dạng của AI cần có sự tham gia của đội ngũ chuyên gia đa ngành để phát triển chúng. Đảm bảo AI hoạt động đáng tin cậy là một lĩnh vực mới, cần rất nhiều nỗ lực nghiên cứu để làm cho hệ thống AI trở nên tin cậy.

Tất cả những nguyên tắc về AI đáng tin cậy trên nếu được tuân thủ đúng cách sẽ đảm bảo độ tin cậy của hệ thống AI, từ đó làm tăng độ tin cậy vào hệ thống. Với sự gia tăng áp dụng trí tuệ nhân tạo trong các lĩnh vực ứng dụng khác nhau, việc làm cho các hệ thống AI trở nên đáng tin cậy là đặc biệt quan trọng. Các phương pháp tiếp cận để tăng cường độ tin cậy cần được triển khai cho các hệ thống AI để chúng trở nên an toàn, chính xác, mạnh mẽ, công bằng, có thể giải thích được.

Thay lời kết

Đánh giá, thẩm định sự tin cậy của AI là vấn đề rất mới và phức tạp, nhưng có thể giải quyết và đã được minh chứng bằng thực tiễn triển khai ở một số quốc gia, khu vực. Quản lý nhà nước về hoạt động này phải có những quan điểm, phương pháp tiếp cận với nhận thức chung ở phạm vi toàn cầu. Hoạt động này cần bắt đầu từ việc xây dựng, hình thành khung pháp lý đưa ra các quy định, hướng dẫn thử nghiệm, các quy trình, công cụ thực thi; để tiến tới triển khai hoạt động này trong thực tiễn. THÔNG TIN & TRUYỀN THÔNG

Tài liệu tham khảo:

1. *Stastica, "Artificial Intelligence (AI) Worldwide - Statistics & Facts", 2/2024*
2. *Mike Thomas, "Six dangerous risks of Artificial Intelligence", Builtin. January 14, 2019*
3. *Sam Levin and Julia Carrie Wong, "Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian", The Guardian, 19 Mar 2018*
4. *Schlesinger, Ari, Kenton P. O'Hara, and Alex S. Taylor, "Let's talk about race: Identity, chatbots, and AI." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems", 2018*
5. *Angwin Julia, "Machine bias. ProPublica" (https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing), (2016)*
6. *Dastin, Jeffrey. "Amazon scraps secret AI recruiting tool that showed bias against women"; San Fransico, "CA: Reuters. Retrieved on October 9", (2018)*
7. *Universitat de Girona, "Artificial Intelligence, Ethics and Society", 2021*
8. *IIC, "The Industrial Internet of Things Trustworthiness Framework Foundations", 2021*