

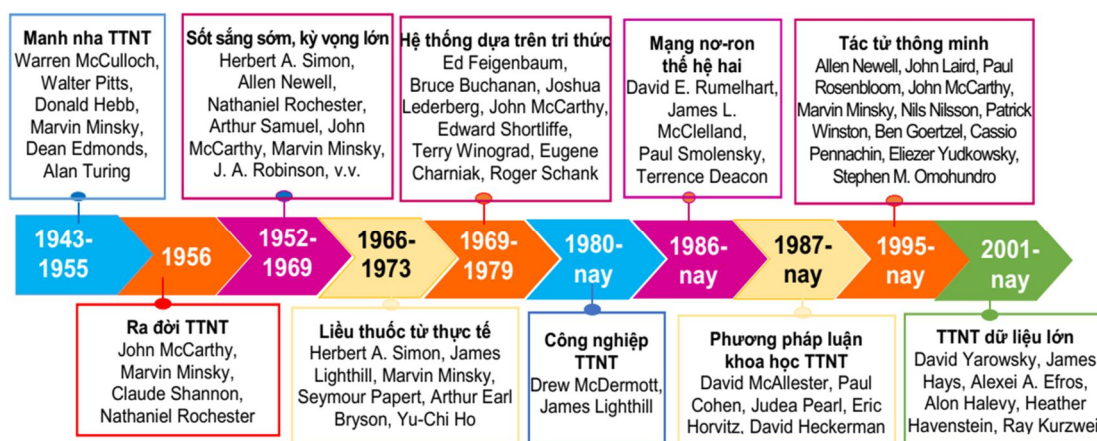
KHUYẾN NGHỊ ĐẠO ĐỨC TRÍ TUỆ NHÂN TẠO CỦA UNESCO VÀ LIÊN HỆ TỚI VIỆT NAM

Hà Quang Thụy, Phan Xuân Hiếu, Nguyễn Trí Thành, Trần Trọng Hiếu, Trần Mai Vũ,
Lê Đức Trọng, Lê Hoàng Quỳnh
Đại học Quốc gia Hà Nội

Trí tuệ nhân tạo (TTNT) đã được phát triển rực rỡ trong thời đại bùng nổ dữ liệu số, đóng vai trò của một công nghệ mục đích chung, đã thấm thấu vào hầu hết các công nghệ tiên tiến, tác động to lớn tới mọi hoạt động kinh tế, xã hội và đời sống con người, mang lại nhiều cơ hội phong phú và đa dạng song hành với những thách thức không nhỏ từ quá trình thiết kế, hình thành và sử dụng sản phẩm TTNT, theo đó đạo đức TTNT đã trở thành một trong những khía cạnh cốt lõi. Bài viết này cung cấp các nội dung khái quát về Khuyến nghị của UNESCO về đạo đức TTNT, phân tích thêm về hai nguyên tắc đạo đức TTNT là Ngăn ngừa thiên vị và công bằng và Khả năng giải thích được. Cuối cùng bài viết đưa ra một vài trao đổi liên quan tới đạo đức TTNT tại Việt Nam.

1. TRÍ TUỆ NHÂN TẠO: CƠ HỘI SONG HÀNH THÁCH THỨC

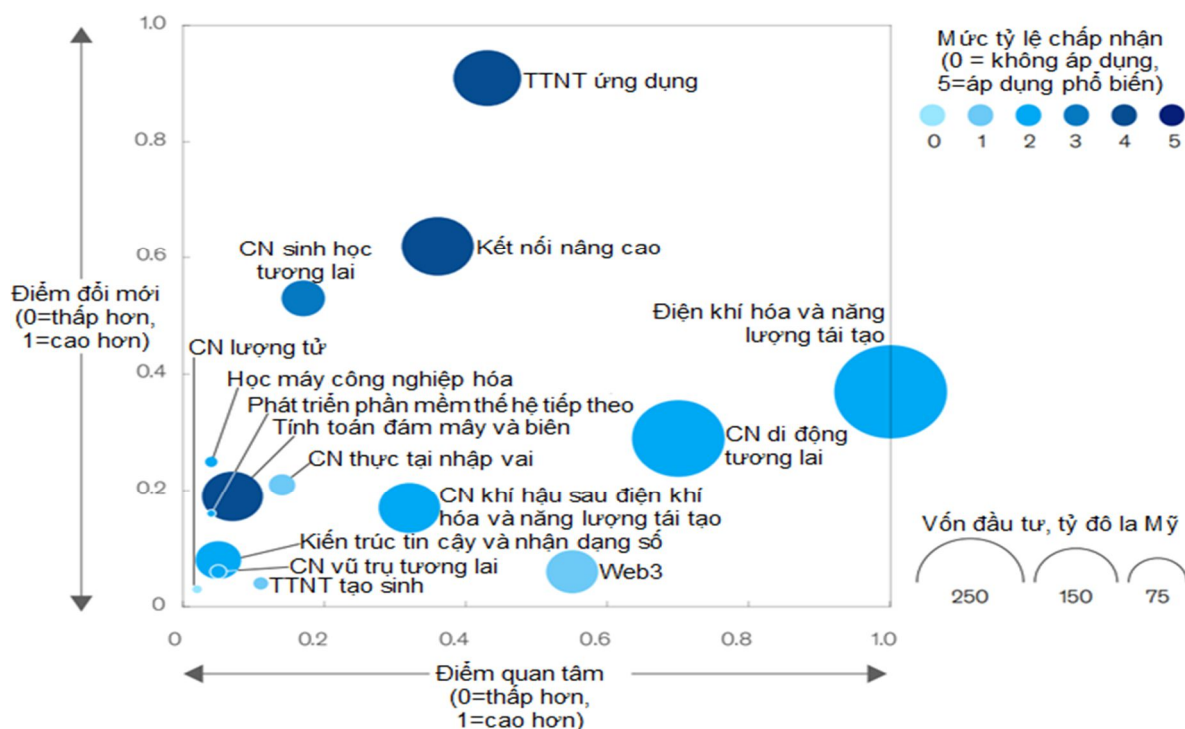
Thời đại bùng nổ dữ liệu số hiện nay, đặc biệt là dữ liệu số do người dùng tạo ra (User-Generated Content: UGC), đã tạo nên một mùa xuân phát triển rực rỡ của TTNT (TTNT dữ liệu lớn) sau một mùa đông thứ hai âm ỉ vào thập niên 1990 trong quá trình tiến hóa của nó. TTNT dữ liệu lớn tạo ra các cơ hội lớn theo bốn khu vực chính cho các



Hình 1. Tóm tắt quá trình tiến hóa của Trí tuệ nhân tạo; trong mỗi giai đoạn có danh sách các nhà khoa học TTNT tiêu biểu [1,2].

doanh nghiệp là Năng suất lao động, Cá nhân hóa, Tiết kiệm thời gian, Chất lượng cao và đóng góp vào kinh tế toàn cầu được dự báo lên tới 15700 tỷ đô la Mỹ vào năm 2030 [3]. Trong 15 khuynh hướng công nghệ nổi bật năm 2023 [4], TTNT có mức đổi mới (kết hợp bằng sáng chế và công bố khoa học) cao nhất và được phổ biến rộng rãi nhất (tỷ lệ

doanh nghiệp trực tiếp áp dụng TTNT đã tăng từ 20% vào năm 2017 lên trên 50% vào năm 2022; tuy nhiên, các vấn đề về tổ chức, kỹ thuật, đạo đức và quy định cần được giải quyết trước khi doanh nghiệp có thể nhận ra được toàn bộ tiềm năng của TTNT), đóng góp giá trị kinh tế của TTNT được ước tính là từ 17000 tới 26000 tỷ đô la Mỹ. Trong vài ba năm gần đây, TTNT tạo sinh (Generative AI) dựa trên mô hình học máy tạo sinh không giám sát học xấp xỉ hàm mật độ dữ liệu để sinh ra các dữ liệu mới không thể phân biệt được với dữ liệu thật. TTNT tạo sinh được kỳ vọng đóng góp từ



Hình 2. Mười lăm công nghệ xu hướng năm 2023 [4]

2600 tới 4400 tỷ đô la Mỹ vào nền kinh tế. Do đó, một phần ba số công ty trên toàn thế giới đang sử dụng TTNT tạo sinh (khoảng 40% lên kế hoạch mở rộng) với dự kiến đầu tư TTNT đạt khoảng 200 tỷ đô la Mỹ vào năm 2025, có khoảng 78,7% số Tiến sỹ TTNT mới năm 2021 là nam giới, tuy nhiên, 73% người được khảo sát trên phạm vi toàn thế giới cho biết cảm thấy lo ngại về rủi ro TTNT và 71% mong đợi các quy định ngăn ngừa rủi ro [5].

Nhà vật lý học lỗi lạc người Anh Stephen Hawking nhận định “*Những lợi ích tiềm năng của việc tạo ra trí thông minh là rất lớn. Nhưng nó cũng có thể là lần cuối cùng, trừ khi chúng ta học được cách tránh những rủi ro. ... sự trỗi dậy của TNTT mạnh mẽ sẽ là điều tốt nhất hoặc điều tồi tệ nhất từng xảy ra với nhân loại. Chúng ta vẫn chưa biết là cái nào.*” [6]. Năm nhận định nổi bật sau đây về ChatGPT, phần mềm TTNT tạo sinh nổi tiếng do OpenAI tạo ra vào cuối tháng 11/2022 [7]: (i) TTNT tạo sinh có thể nâng cao năng suất nhưng cũng có thể dẫn đến việc thay thế nhân viên con người; (ii) Việc giảng dạy, học tập và nghiên cứu học thuật

sẽ trải qua một số tác động mang tính biến đổi lớn nhất; (iii) Thiên vị, dữ liệu đào tạo lỗi thời, thiếu minh bạch và không đáng tin cậy là những mối lo ngại lớn; (iv) Tác động của TTNT tạo sinh đôi với việc tiếp thu kiến thức và chuyển đổi số cần được nghiên cứu; (v) Điều quan trọng là phải xác định và thực hiện các chính sách để bảo vệ khỏi việc sử dụng sai mục đích và lạm dụng TTNT tạo sinh.

Tổ chức Giáo dục, Khoa học và Văn hóa của Liên hợp quốc (UNESCO) nhận định là "Công nghệ TTNT mang lại lợi ích rất to lớn trong nhiều lĩnh vực, nhưng nếu không có các rào cản đạo đức, nó có nguy cơ tạo ra sự thiên vị và phân biệt đối xử trong thế giới thực, gây chia rẽ và đe dọa các quyền và tự do cơ bản của con người" [8].

2. KHUYẾN NGHỊ CỦA UNESCO VỀ ĐẠO ĐỨC TTNT

2.1. Sơ lược về khuyến nghị của UNESCO về đạo đức TTNT

Kỳ họp Đại hội đồng UNESCO lần thứ 40 từ ngày 12/11/2019 tới ngày 27/11/2019 tại Paris (Pháp) đã thông qua Quyết định 40 C/Resolution 37 về việc "Nghiên cứu sơ bộ về một công cụ thiết lập tiêu chuẩn khả thi về đạo đức của trí tuệ nhân tạo" nhằm xây dựng và ban hành một khuyến nghị của UNESCO về đạo đức TTNT đóng vai trò của một công cụ thiết yếu đảm bảo việc tăng cường xây dựng và thực thi luật pháp, chính sách và chiến lược quốc gia và quốc tế trong lĩnh vực TTNT, cũng như tăng cường hợp tác quốc tế về phát triển đạo đức và sử dụng TTNT trong hỗ trợ các Mục tiêu Phát triển Bền vững. Quyết định 40 C/Resolution 37 có ba điểm chính [9]:

- Quyết định rằng việc UNESCO chuẩn bị một công cụ thiết lập tiêu chuẩn quốc tế về đạo đức TTNT dưới hình thức khuyến nghị là kịp thời và phù hợp;
- Yêu cầu Tổng Giám đốc đảm bảo tổ chức đủ số lượng các cuộc tham vấn liên chính phủ hiệp thương (praesentia) về văn bản của khuyến nghị nói trên;
- Mời Tổng Giám đốc trình tại phiên họp thứ 41 văn bản dự thảo khuyến nghị về đạo đức TTNT theo Quy tắc về Thủ tục liên quan đến các khuyến nghị đối với các Quốc gia Thành viên và các công ước quốc tế quy định tại Điều IV, khoản 4 của Hiến chương.

Tại Kỳ họp Đại hội đồng UNESCO lần thứ 41 từ ngày 09/11/2021 tới ngày 24/11/2021 tại Paris (Pháp), Khuyến nghị về Đạo đức Trí tuệ Nhân tạo (Recommendation on the Ethics of Artificial Intelligence) của UNESCO đã được toàn bộ 193 quốc gia thành viên thông qua vào ngày 23/11/2021. Lưu ý rằng, Hoa Kỳ, một trong hai quốc gia được coi là siêu cường về TTNT, không là quốc gia thành viên của UNESCO tại thời điểm thông qua khuyến nghị; Hoa Kỳ rời khỏi UNESCO lần gần nhất vào ngày 31/12/2018 và mới trở lại UNESCO vào ngày 10/07/2023. Khuyến nghị ngầm định rằng, mọi thực thể xã hội (quốc gia, tổ chức/doanh nghiệp, cá nhân) đã có ý định tốt trong thiết kế, triển khai

và sử dụng TTNT song cần có hành vi đúng để mọi hoạt động đạt đúng mục đích tốt bởi vì có ý định tốt (moral) không luôn luôn đảm bảo đạt được hành vi đúng (ethical). Ngoài Lời mở đầu, văn bản khuyến nghị có tám chương với 141 điểm nội dung đề cập một cách toàn diện về: Phạm vi áp dụng, Mục đích và mục tiêu, Giá trị và nguyên tắc, Mười một Lĩnh vực hành động chính sách (đánh giá tác động đạo đức, quản trị và quản lý có đạo đức, chính sách dữ liệu, phát triển và hợp tác quốc tế, môi trường và hệ sinh thái, giới, văn hóa, giáo dục và nghiên cứu, truyền thông và thông tin, kinh tế và lao động, sức khỏe và phúc lợi xã hội), Giám sát và đánh giá, Việc sử dụng và khai thác Khuyến nghị hiện tại, Thúc đẩy Khuyến nghị hiện tại, Quy định thức [8]. Tiểu ban Khoa học Xã hội và Nhân văn của UNESCO đã giới thiệu các nội dung then chốt nhất của khuyến nghị [10] và các nội dung này sẽ được đề cập chi tiết hơn ở mục tiếp theo.

Khuyến nghị của UNESCO được coi là khuyến nghị toàn cầu đầu tiên trên thế giới về đạo đức TTNT và vẫn còn một chặng đường dài để cung cấp cho các Quốc gia Thành viên các nguồn lực có thể hành động nhằm đảm bảo việc thực hiện Khuyến nghị một cách hiệu quả. Để Khuyến nghị có khả năng áp dụng đặc biệt là các lĩnh vực hành động chính sách mở rộng, cho phép các nhà hoạch định chính sách biến các giá trị và nguyên tắc cốt lõi thành hành động liên quan đến quản trị dữ liệu, môi trường và hệ sinh thái, giới tính, giáo dục và nghiên cứu, sức khỏe và phúc lợi xã hội và nhiều lĩnh vực khác, nhiều hoạt động của UNESCO và các tổ chức quốc tế khác đã được tiến hành, bao gồm hai hội thảo quốc tế đã được tổ chức vào các năm 2022 [11] và 2024 [5]. Ngày 05/02/2024, tám công ty công nghệ hàng đầu thế giới (GSMA, INNIT, Lenovo Group, LG AI Research, Mastercard, Microsoft, Salesforce và Telefonica) đã ký một thỏa thuận mang tính đột phá nhằm xây dựng TTNT có đạo đức hơn, tích hợp các giá trị và nguyên tắc trong Khuyến nghị của UNESCO về Đạo đức của TTNT khi thiết kế và triển khai các hệ thống TTNT [5].

2.2. Những điểm chính từ Khuyến nghị của UNESCO về đạo đức TTNT

Nền tảng của Khuyến nghị là bảo vệ nhân quyền và nhân phẩm dựa trên việc thúc đẩy các nguyên tắc cơ bản như tính minh bạch và công bằng, luôn ghi nhớ tầm quan trọng của việc con người giám sát các hệ thống TTNT [8, 10].

Trọng tâm của Khuyến nghị là bốn giá trị cốt lõi đặt nền móng cho các hệ thống TTNT hoạt động vì lợi ích của nhân loại, cá nhân, xã hội và môi trường:

- Tôn trọng, bảo vệ và thúc đẩy nhân quyền, các quyền tự do cơ bản và phẩm giá con người,
- Sống trong xã hội hòa bình, công bằng và gắn kết với nhau,
- Đảm bảo tính đa dạng và toàn diện,

- Môi trường và hệ sinh thái hưng thịnh.

Cụ thể, mười nguyên tắc cốt lõi trong tiếp cận nhân quyền đối với TTNT trong Khuyến nghị như sau:

- Cân đối và không gây hại: Việc sử dụng hệ thống TTNT không được vượt quá những gì cần thiết để đạt được mục đích chính đáng. Đánh giá rủi ro nên được sử dụng để ngăn ngừa những tác hại có thể xảy ra do việc sử dụng đó.
- An toàn và bảo mật: Những tác hại không mong muốn (rủi ro an toàn) cũng như các lỗ hổng dễ bị tấn công (rủi ro bảo mật) cần được các tác nhân TTNT tránh và giải quyết.
- Quyền riêng tư và bảo vệ dữ liệu: Quyền riêng tư phải được bảo vệ và phát huy trong suốt vòng đời TTNT. Các khuôn khổ bảo vệ dữ liệu đầy đủ cũng cần được thiết lập.
- Quản trị và hợp tác đa bên và thích ứng: Luật pháp quốc tế và chủ quyền quốc gia phải được tôn trọng trong việc sử dụng dữ liệu, nghĩa là các quốc gia có thể quản lý dữ liệu được tạo ra trong hoặc đi qua lãnh thổ của họ. Ngoài ra, sự tham gia của các bên liên quan khác nhau là cần thiết cho các phương pháp tiếp cận toàn diện trong quản trị TTNT.
- Trách nhiệm và trách nhiệm giải trình: Hệ thống TTNT phải có thể kiểm tra được và theo dõi được. Cần có các cơ chế giám sát, đánh giá tác động, kiểm toán và thẩm định để tránh xung đột với các chuẩn mực nhân quyền và các mối đe dọa đối với phúc lợi môi trường.
- Minh bạch và giải thích được: Việc triển khai có đạo đức các hệ thống AI phụ thuộc vào tính minh bạch và khả năng giải thích của chúng. Ví dụ, mọi người nên được biết khi có quyết định được TTNT thông báo. Mức độ minh bạch và khả năng giải thích phải phù hợp với bối cảnh, vì có thể có những căng thẳng giữa tính minh bạch và khả năng giải thích cũng như các nguyên tắc khác như quyền riêng tư, an toàn và bảo mật.
- Sự giám sát và nỗ lực của con người: Các quốc gia thành viên nên đảm bảo rằng các hệ thống TTNT không thay thế trách nhiệm và trách nhiệm giải trình cuối cùng của con người.
- Tính bền vững: Các công nghệ TTNT cần được đánh giá dựa trên tác động của chúng đối với 'tính bền vững', được hiểu là một tập hợp các mục tiêu không ngừng phát triển, bao gồm cả những mục tiêu được đặt ra trong Mục tiêu Phát triển Bền vững của Liên hợp quốc.

- Nhận thức và học vấn: Sự hiểu biết của công chúng về TTNT và dữ liệu cần được thúc đẩy thông qua giáo dục mở và dễ tiếp cận, sự tham gia của công dân, kỹ năng kỹ thuật số và đào tạo đạo đức TTNT, hiểu biết về truyền thông và thông tin.
- Công bằng và không phân biệt đối xử: Các tác nhân TTNT nên thúc đẩy công bằng xã hội, sự công bằng và không phân biệt đối xử, đồng thời thực hiện cách tiếp cận toàn diện để đảm bảo tất cả mọi người đều có thể tiếp cận được lợi ích của TTNT.

Tóm tắt khuyến nghị nhận định rằng, mặc dù các giá trị và nguyên tắc là rất quan trọng để thiết lập nền tảng cho bất kỳ khung đạo đức TTNT nào, nhưng việc vượt ra ngoài các nguyên tắc cấp cao và hướng tới các chiến lược thực tế là hết sức cần thiết. Chính vì lý do đó, Khuyến nghị đặt ra mười một khu vực chính chính sách như được minh họa trong Hình 3.

Đồng thời, nhằm cung cấp nguồn lực bổ sung cho các quốc gia thành viên hành động đảm bảo thực hiện Khuyến nghị một cách hiệu quả, UNESCO phát triển hai phương pháp Đánh giá mức độ sẵn sàng (Readiness assessment methodology: RAM) và Đánh giá tác động đạo đức (Ethical impact assessment: EIA):



Hình 3. Mười một khu vực chính sách trong Khuyến nghị [12, 13]

- RAM được thiết kế để giúp đánh giá liệu các Quốc gia Thành viên có sẵn sàng thực hiện Khuyến nghị một cách hiệu quả hay không. RAM có thể giải quyết các câu hỏi như: (i) Có luật pháp nào để quản lý TTNT không? (ii) Cơ sở hạ tầng quốc gia hỗ trợ khả năng tiếp cận công nghệ TTNT không? RAM được xây dựng để giúp các Quốc gia Thành viên xác định tình trạng sẵn sàng của mình và tạo cơ sở để UNESCO điều chỉnh hỗ trợ xây dựng năng lực của mình. Nó cũng có thể được sử dụng để giúp các Quốc gia Thành viên xây dựng lộ trình hướng tới phát triển và sử dụng TTNT có đạo đức.

- EIA là một quy trình có cấu trúc nhằm thực hiện Khuyến nghị bằng cách giúp các nhóm dự án TTNT hợp tác với các cộng đồng bị ảnh hưởng để xác định và đánh giá các tác động mà hệ thống TTNT có thể gây ra. EIA có thể giải quyết các câu hỏi như: (i) Ai có khả năng bị ảnh hưởng xấu nhất bởi hệ thống TTNT này? (ii) Những tác động này sẽ diễn ra dưới hình thức nào? (iii) Có thể làm gì để ngăn chặn những tác hại này và phân bổ nguồn lực cho việc phòng ngừa tác hại này? Mục đích của EIA là tạo cơ hội để phản ánh những tác động tiềm tàng của một dự án TTNT và xác định các hành động phòng ngừa tác hại cần thiết.

Như đã được đề cập, đạo đức TTNT là chủ đề rất lớn và rất phức tạp, vì vậy, UNESCO đang triển khai chiến lược đa chiều với các chính phủ, khu vực tư nhân, các tổ chức học thuật và tổ chức xã hội dân sự để biến Khuyến nghị thành các chính sách và hành động [12]. Một số hoạt động điển hình trong chiến lược đa chiều nói trên là:

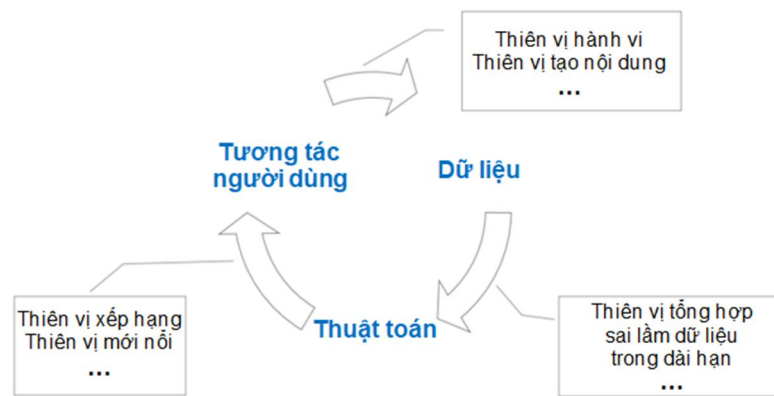
- Đài quan sát toàn cầu về đạo đức của TTNT (Global Observatory on the Ethics of AI), một nền tảng số đổi mới đóng vai trò là điểm tổng hợp cho những phân tích mới nhất về sự phát triển và sử dụng TTNT có đạo đức trên khắp thế giới cũng như kiến thức được tạo ra thông qua việc thực hiện Khuyến nghị ở các quốc gia khác nhau.
- Diễn đàn toàn cầu về đạo đức TTNT (Global Forum on Ethics of AI), một sự kiện cấp cao thường niên nhằm nâng cao kiến thức tiên tiến về những thách thức mà công nghệ TTNT đặt ra và tạo điều kiện thuận lợi cho việc học tập ngang hàng giữa các chính phủ và các bên liên quan khác.
- Mạng lưới Chuyên gia Đạo đức TTNT không biên giới (AI Ethics Experts Without Borders: AIEB), một cơ sở linh hoạt gồm các chuyên gia được triển khai tại các Quốc gia Thành viên theo nhu cầu để hỗ trợ thực hiện Khuyến nghị và áp dụng các công cụ xây dựng năng lực.
- Mạng lưới Phụ nữ vì Đạo đức TTNT (Women for AI Ethics: W4AIEthics), nền tảng dành cho các nhà lãnh đạo nữ có ảnh hưởng trong công nghiệp, chính phủ và xã hội dân sự, thúc đẩy những chuyển đổi theo hướng bình đẳng giới trong và thông qua TTNT.

2.3. Đạo đức TTNT về Thiên vị, công bằng và khả năng giải thích được

Mục này đưa ra một số bàn luận thêm về hai nguyên tắc quan trọng nhưng còn khá mới lạ về đạo đức TTNT là Thiên vị, công bằng và Khả năng giải thích được.

Nguyên tắc Thiên vị và công bằng có thể được phát biểu là “Không có bất kỳ định kiến hoặc/và thiên vị nào cho một cá nhân/một nhóm dựa trên các đặc điểm vốn có/có được của họ” và “Một hệ thống không công bằng là đưa ra các quyết định thiên về một

nhóm người cụ thể”. Ví dụ điển hình về phần mềm TTNT thiên vị như Phần mềm COMPAS thiên vị "dương giả" người Mỹ gốc Phi, Phần mềm thi sắc đẹp thiên vị da sẫm màu, Phần mềm nhận dạng khuôn mặt máy ảnh số dự đoán quá mức về trạng thái đang chớp mắt của người châu Á.

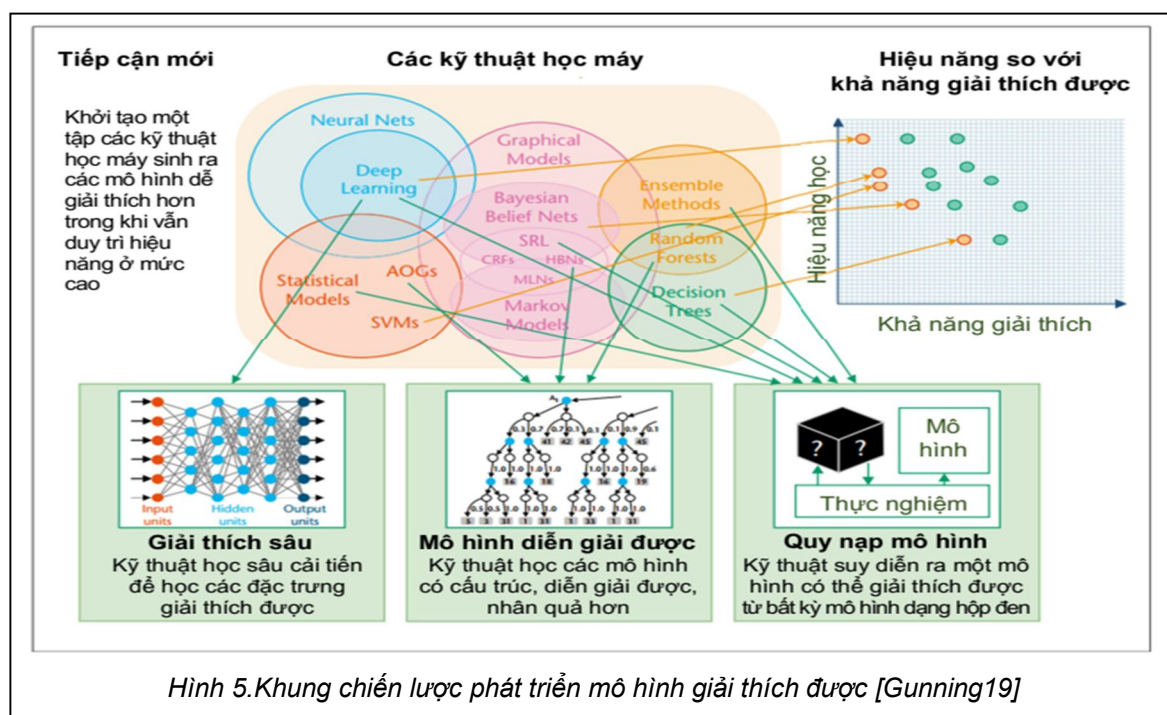


Hình 4. Vòng phản hồi thiên vị [13]

Thiên vị bị ẩn hoặc/và bị bỏ qua ở dữ liệu và/hoặc thuật toán. Tương tự, yếu tố không công bằng xuất hiện từ dữ liệu, từ thuật toán, từ tương tác người dùng. Hình 4 cho một khung về vòng phản hồi thiên vị trong các hệ thống TTNT. Một số giải pháp điển hình nhằm đảm bảo hệ thống TTNT đáp ứng nguyên tắc thiên vị và công bằng được trao đổi từ [13].

Nguyên tắc về Khả năng giải thích và minh bạch liên quan một thách thức không nhỏ của các ứng dụng TTNT là không thể giải thích được cho người dùng về các quyết định và hành động của chúng dù mang lại những lợi ích to lớn trong nhiều miền ứng dụng (giao thông vận tải, an ninh, y tế, tài chính, quốc phòng, v.v.) [14]. Hệ thống TTNT giải thích được khi nó có “khả năng bày tỏ được lý do tại sao một hệ thống TTNT đạt được một quyết định, đề xuất hoặc dự đoán cụ thể” [15], “các mô hình có thể giải thích được, khi kết hợp với các kỹ thuật giải thích hiệu quả, cho phép người dùng cuối hiểu, tin tưởng một cách thích hợp và quản lý hiệu quả hệ thống TTNT tạo sinh mới nổi” [14]. Hệ thống TTNT giải thích được tạo cơ hội thành công cho chuyên gia công nghệ, chuyên gia kinh doanh, chuyên gia pháp lý và rủi ro khi áp dụng TTNT và học máy vào thực tiễn [15]. Khi khả năng giải thích càng trở nên quan trọng thì cách thức để đạt được nó lại quan trọng hơn đáng kể trong bối cảnh hầu như các công cụ học máy tiên tiến thường vẫn là hộp đen. Giải pháp không chỉ đơn giản là tìm ra những cách tốt hơn để truyền đạt cách thức hoạt động của một hệ thống mà còn tạo ra các công cụ và quy trình nhằm giúp ngay cả các chuyên gia hiểu được sâu sắc kết quả và sau đó giải thích nó cho những người khác.

D. Gunning và D. Aha đề nghị một khung chiến lược phát triển các mô hình giải thích



được (Hình 5) [14]. L. Grennan và cộng sự [15] nhận định rằng các tổ chức xây dựng một khung khả năng giải thích và có được các công cụ hỗ trợ phù hợp sẽ có vị trí tốt hơn để nắm bắt toàn bộ giá trị của học sâu và những tiến bộ khác của TTNT và hai công việc chính cần được quan tâm thực hiện: (i) Thành lập một ủy ban quản trị TTNT để hướng dẫn các nhóm phát triển TTNT; (ii) Đầu tư tài năng, công nghệ có thể giải thích được, nghiên cứu và đào tạo một cách đúng đắn. Có thể coi quy trình xây dựng một hệ thống học máy mạnh, toàn diện để xây dựng ứng dụng TTNT giải thích được của A. Thampi [16] như một cụ thể hóa khung giải thích được. Quy trình này bao gồm ba giai đoạn là phát triển, kiểm tra và sản xuất, trong đó giai đoạn sản xuất được diễn ra khi ứng dụng TTNT đang hoạt động trong thực tiễn.

3. LIÊN HỆ VỚI VIỆT NAM

Việt Nam là một quốc gia thành viên có trách nhiệm trong UNESCO, nói riêng, Việt Nam rất có trách nhiệm trong hoạt động về đạo đức TTNT. Tại Hội thảo lần thứ hai [5], Việt Nam đã tham gia phiên họp Bộ trưởng về Trọng tâm là quản trị quốc gia về TTNT: Những hiểu biết sâu sắc từ các hoạt động Đánh giá mức độ sẵn sàng RAM của UNESCO và kinh nghiệm quốc gia. Tiếp thu kinh nghiệm quốc tế, chiến lược quốc gia về TTNT của Việt Nam [17] cần được nâng cấp, cải tiến, đồng thời, hệ thống các tiêu chí quốc gia của Việt Nam theo hai phương pháp Đánh giá mức độ sẵn sàng RAM và Đánh giá tác động đạo đức EIA cần được xây dựng. Trên cơ sở tham khảo kinh nghiệm của UNESCO và các tổ chức quốc tế khác (chẳng hạn, [18, 19, 20]), đặc biệt là Hướng dẫn về Quản trị và Đạo đức

TTNT của ASEAN [24], hệ thống các văn bản quy phạm pháp luật về TTNT và đạo đức TTNT của Việt Nam cần được hình thành từng bước.

Hạng	Giá trị thực 2016		Kỳ vọng 2030		Giá trị thực 2023		Kỳ vọng 2050	
	Quốc gia	Tỷ US\$	Quốc gia	Tỷ US\$	Quốc gia	Tỷ US\$	Quốc gia	Tỷ US\$
1	China	21269	China	38008	China	34643	China	58499
2	United States	18562	United States	23475	United States	27360	India	44128
3	India	8721	India	19511	India	14537	United States	34102
4	Japan	4932	Japan	5606	Russian Federation	6452	Indonesia	10502
5	Germany	3979	Indonesia	5424	Japan	6251	Brazil	7540
6	Russia	3745	Russia	4736	Germany	5857	Russia	7131
7	Brazil	3135	Germany	4707	Brazil	4454	Mexico	6863
8	Indonesia	3028	Brazil	4439	Indonesia	4333	Japan	6779
9	United Kingdom	2788	Mexico	3661	France	4169	Germany	6138
10	France	2737	United Kingdom	3638	United Kingdom	4026	United Kingdom	5369
11	Mexico	2307	France	3377	Turkiye	3767	Turkey	5184
12	Italy	2221	Turkey	2996	Italy	3452	France	4705
13	South Korea	1929	Saudi Arabia	2755	Mexico	3288	Saudi Arabia	4694
14	Turkey	1906	South Korea	2651	Korea, Rep.	2794	Nigeria	4348
15	Saudi Arabia	1731	Italy	2541	Spain	2553	Egypt	4333
16	Spain	1690	Iran	2354	Canada	2469	Pakistan	4236
17	Canada	1674	Spain	2159	Egypt, Arab Rep.	2120	Iran	3900
18	Iran	1459	Canada	2141	Saudi Arabia	2031	South Korea	3539
19	Australia	1189	Egypt	2049	Australia	1841	Philippines	3334
20	Thailand	1161	Pakistan	1868	Poland	1814	Vietnam	3176
21	Egypt	1105	Nigeria	1794	Thailand	1681	Italy	3115
22	Nigeria	1089	Thailand	1732	Iran, Islamic Rep.	1598	Canada	3100
23	Poland	1052	Australia	1663	Bangladesh	1567	Bangladesh	3064
24	Pakistan	988	Philippines	1615	Viet Nam	1502	Malaysia	2815
25	Argentina	879	Malaysia	1506	Pakistan	1493	Thailand	2782
26	Netherlands	866	Poland	1505	Nigeria	1414	Spain	2732
27	Malaysia	864	Argentina	1342	Netherlands	1398	South Africa	2570
28	Philippines	802	Bangladesh	1324	Argentina	1369	Australia	2564
29	South Africa	736	Vietnam	1303	Malaysia	1277	Argentina	2365
30	Colombia	690	South Africa	1148	Philippines	1262	Poland	2103
31	Bangladesh	628	Colombia	1111	Colombia	1122	Colombia	2074
32	Vietnam	595	Netherlands	1080	South Africa	957	Netherlands	1496

Hình 5. Kích thước kinh tế Việt Nam (GDP PPP) theo xu thế tăng năng suất [21, 22]

Như đã được đề cập, khu vực đóng góp quan trọng nhất của TTNT về kinh tế là tăng năng suất lao động. Một phần tư thế kỷ vừa qua, Việt Nam ở “làn nhanh” (“fast lane”) tăng năng suất lao động, đã rút ngắn được hơn tám năm trong một lộ trình mười lăm năm trở thành nền kinh tế hạng 25 thế giới về GDP PPP [21, 22]. Đây là thời điểm tăng cường đầu tư và đón đầu làn sóng năng suất tiếp theo để nước ta tăng tốc hội tụ mức năng suất của nền kinh tế tiên tiến [21]. Hai yếu tố đầu vào quyết định năng suất lao động là: lượng vốn trên mỗi lao động, mức tăng này còn được gọi là *tăng cường vốn* (capital deepening) và *vốn con người* (human capital) về trình độ học vấn, khả năng và kinh nghiệm tích lũy của người lao động.

Phát huy lợi thế về ổn định chính trị và về vị trí địa lý liền kề một siêu cường TTNT là Trung Quốc trong việc tạo vị thế là điểm đến quốc tế ưu việt thu hút vốn đầu tư nước ngoài và trong nước để *tăng cường vốn*, tăng cường đầu tư nguồn lực cho giáo dục (phổ thông và đại học) để *tăng cường vốn con người* nhằm đón đầu tốt làn sóng năng suất tiếp theo của thế giới.

Như đã được đề cập [25], Việt Nam cần tăng cường đầu tư cho giáo dục đại học, khoa học và công nghệ nhằm phát triển tài năng TTNT theo mục tiêu “Đến năm 2030, Việt Nam thuộc top dẫn đầu nhóm quốc gia có thu nhập trung bình cao về đào tạo tài năng và bồi dưỡng nhân tài TTNT theo mọi phương diện về nghiên cứu, sản xuất sản phẩm và triển khai ứng dụng TTNT”; các trường đại học/viện nghiên cứu trọng điểm về công nghệ và kinh tế cần hoàn thành nhiệm vụ chiến lược quốc gia này; Đội ngũ tài năng TTNT của đất nước là lực lượng chủ chốt tạo động lực tăng tốc trưởng thành số và TTNT của Việt Nam.

Tài liệu tham khảo

- [1] Stuart Russell, Peter Norvig. *Artificial Intelligence: A Modern Approach (Four edition)*. Prentice Hall, 2020.
- [2]. Nguyễn Thanh Thủy, Hà Quang Thụy, Phan Xuân Hiếu, Nguyễn Trí Thành. *Trí tuệ nhân tạo trong thời đại số: Bối cảnh thế giới và liên hệ với Việt Nam*. Tạp chí Công thương, trực tuyến ngày 21/08/2018.
- [3] Anand S. Rao, Gerard Verweij. *Sizing the prize: What's the real value of AI for your business and how can you capitalise*. PwC report, 2017.
- [4] McKinsey Technology Council. *Technology Trends Outlook 2023*. McKinsey & Company, July 2023.
- [5] *The 2nd Global Forum on the Ethics of AI: Changing the Landscape of AI Governance* organized by Slovenia, 5-6 February 2024, Kranj. <https://www.unesco.org/en/forum-ethics-ai>.
- [6] *Stephen Hawking's speech from the launch of the Leverhulme Centre for the Future of Intelligence on 19 October 2016*. <http://lcfi.ac.uk/resources/cfi-launch-stephen-hawking/>.
- [7] Yogesh K. Dwivedi, Nir Kshetri, Laurie Hughes et al (73 coauthors). *Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy*. Int. J. Inf. Manag. 71: 102642 (2023).
- [8] UNESCO. *Recommendation on the Ethics of Artificial Intelligence - Adopted on 23 November 2021*. UNESCO 2022.
- [9] UNESCO's Social and Human Sciences Sector. *Preliminary study on a possible standard-setting instrument on the ethics of artificial intelligence*. 40 C/Resolution 37, UNESCO 2020.
- [10] UNESCO's Social and Human Sciences Sector. *Key facts UNESCO's Recommendation on the Ethics of Artificial Intelligence*. SHS/2023/PI/H/1, UNESCO 2023.
- [11] The First Global Forum on the Ethics of AI, with the theme, "Ensuring inclusion in the AI world", will be hosted by the Czech Republic on 13 December 2022 in Prague. <https://www.unesco.org/en/articles/global-forum-ethics-artificial-intelligence-ensuring-inclusion-ai-world>.

- [12] www.unesco.org/artificial-intelligence
- [13] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, Aram Galstyan. *A Survey on Bias and Fairness in Machine Learning*. ACM Comput. Surv. 54(6): 115:1-115:35 (2022).
- [14] D. Gunning, D. Aha. *DARPA's Explainable Artificial Intelligence (XAI) Program*. AI Magazine, 40(2), 44-58, 2019.
- [15] Liz Grennan, Andreas Kremer, Alex Singla và Peter Zipparo. *Why businesses need explainable AI—and how to deliver it*. McKinsey's Article, September 29, 2022.
- [16] Ajay Thampi. *Interpretable AI: Building explainable machine learning systems*. Manning, 2022.
- [17] Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo đến năm 2030 được ban hành theo Quyết định số 127/QĐ-TTg của Thủ tướng Chính phủ.
- [18] Michael Chui, Chris Wigley, Simon London. *AI Ethics in today's world*. McKinsey Podcast, January 31, 2019.
- [19] Peter Zipparo, Rachel Dooley. *From Principles to Practice: Putting AI Ethics into Action*. McKinsey & Company, July 8, 2022.
- [20] Department of Industry, Science and Resources (Australian Government). *Australia's AI Ethics Principles*. <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>.
- [21] <https://www.pwc.com/gx/en/world-2050/assets/pwc-the-world-in-2050-full-report-feb-2017.pdf>.
- [22].https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.CD?most_recent_value_desc=true&locations=CN. June 03, 2024
- [23] Jan Mischke, Chris Bradley, Marc Canal, Olivia White, Sven Smit, Denitsa Georgieva. *Investing in productivity growth*. McKinsey Global Institute, March 27, 2024.
- [24] ASEAN Secretariat. *ASEAN Guide on AI Governance and Ethics*. The 4th ASEAN Digital Ministers' Meeting 2024 (ADGIM), February 2, 2024.
- [25] Hà Quang Thụy, Phạm Bảo Sơn, Phan Xuân Hiếu, Trần Trọng Hiếu, Trần Mai Vũ, Nguyễn Trí Thành, Phạm Văn Dũng, Trần Thế Hiệp. *Những yếu tố quan trọng trong Chiến lược quốc gia về trí tuệ nhân tạo*. Tạp chí Khoa học và Công nghệ Việt Nam 30/10/2020.