

# Công nghệ trí tuệ nhân tạo: Một số chính sách, luật định quốc tế và khuyến nghị cho Việt Nam<sup>1</sup>

Ngô Thị Mai Diên<sup>(\*)</sup>

**Tóm tắt:** Trí tuệ nhân tạo đang được xem là một trong những công nghệ đột phá nhất của thế kỷ XXI, có khả năng thay đổi mọi khía cạnh của cuộc sống con người. Ngày 26/01/2021, Thủ tướng Chính phủ đã ký Quyết định số 127/QĐ-TTg ban hành “Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030”. Chiến lược xác định sự cần thiết phải “xây dựng hệ thống văn bản quy phạm pháp luật và hành lang pháp lý liên quan đến trí tuệ nhân tạo”. Bài viết đề cập vấn đề nội dung các đạo luật, hướng dẫn, văn bản pháp lý của Liên minh châu Âu, Anh, Mỹ và Trung Quốc về trí tuệ nhân tạo mà Việt Nam có thể tham khảo để xây dựng chính sách, cơ chế, chiến lược quản lý và sử dụng công nghệ trí tuệ nhân tạo có trách nhiệm.

**Từ khóa:** Trí tuệ nhân tạo, Chính sách pháp luật, Liên minh châu Âu, Anh, Mỹ, Trung Quốc, Việt Nam

**Abstract:** Artificial intelligence is considered one of the most disruptive technologies of the 21<sup>st</sup> century, with potential to change every aspect of life. On January 26, 2021, the Prime Minister signed Decision No. 127/QĐ-TTg promulgating “The National Strategy on Research, Development and Application of Artificial Intelligence until 2030” which identifies the need to “build a system of legal documents and corridors related to artificial intelligence”. The article summarizes the content of laws, guidelines, and legal documents of the European Union, the United Kingdom, the United States, and China that Vietnam can refer to developing its policies, mechanisms, and strategies for managing and using artificial intelligence responsibly.

**Keywords:** Artificial Intelligence, Legal Policy, European Union, United Kingdom, United States, China, Vietnam

Ngày nhận bài: 30/7/2024; Ngày duyệt đăng: 01/9/2024

<sup>1</sup> Bài viết là kết quả nghiên cứu của Đề tài cấp cơ sở năm 2024 “Niên giám Thông tin Khoa học xã hội (Tập 20)”, do ThS. Ngô Thị Mai Diên, ThS. Trương Tuấn Anh đồng chủ nhiệm, Viện Thông tin Khoa học xã hội chủ trì.

<sup>(\*)</sup> ThS., NCVC, Viện Thông tin Khoa học xã hội, Viện Hàn lâm Khoa học xã hội Việt Nam;  
Email: maidiennis@gmail.com

## 1. Mở đầu

Trí tuệ nhân tạo (Artificial Intelligence - AI) là lĩnh vực khoa học máy tính (Computer science) tập trung vào việc nghiên cứu và tạo ra những phương pháp khoa học, công cụ thông minh có khả năng phân tích và thu thập nguồn dữ liệu khổng lồ giúp thực hiện mọi công việc theo yêu cầu của con người, thực hiện các nhiệm vụ thông minh thay thế con người. Công nghệ này đang ngày càng phát triển, được ứng dụng rộng rãi, hiệu quả trong nhiều lĩnh vực của đời sống, mang lại nhiều lợi ích cho con người. Tuy nhiên bên cạnh đó, AI cũng tiềm ẩn nhiều nguy cơ, rủi ro, đe dọa an ninh quốc gia, thách thức sự an nguy của loài người. Do vậy, các quyết định về việc phát triển và sử dụng AI cần được cân nhắc kỹ lưỡng. Nhiều quốc gia và tổ chức quốc tế đã và đang nỗ lực xây dựng các thể chế, chính sách, đạo luật, bộ nguyên tắc và quy định liên quan đến AI để đảm bảo sự phát triển và sử dụng công nghệ này có trách nhiệm, “vị nhân sinh”, bền vững, an toàn, đáng tin cậy.

## 2. Đạo luật Trí tuệ nhân tạo của Liên minh châu Âu

Ngày 13/3/2024, Nghị viện Liên minh châu Âu (EU) đã thông qua Đạo luật Trí tuệ nhân tạo (Artificial Intelligence Act - AIA) đầu tiên trên thế giới. Đạo luật gồm 13 chương, 113 điều và 13 phụ lục hướng đích mục tiêu kép: (i) bảo vệ các quyền cơ bản của con người, các giá trị dân chủ, pháp quyền cũng như sự bền vững của môi trường trước các thách thức do AI tạo ra; và (ii) thúc đẩy đổi mới sáng tạo, đưa châu Âu trở thành khu vực dẫn đầu trong lĩnh vực AI (Hùng Tuấn, 2024). Để hiện thực hóa mục tiêu này, các nhà lập pháp EU đã thiết kế cơ chế quản lý AI dựa trên việc đánh giá mức độ rủi ro của công nghệ này đối với

con người. Trong đó, có ba điểm đáng chú ý sau:

*Thứ nhất*, Đạo luật đặt ra ranh giới “cứng” cho sự phát triển của AI trong mối liên hệ với sự phát triển của loài người. Nói cách khác, AIA cấm một số ứng dụng AI đe dọa đến quyền công dân, bao gồm hệ thống nhận dạng sinh trắc học (Biometric Identification Systems - RBI) dựa trên các đặc điểm nhạy cảm và việc thu thập hình ảnh khuôn mặt không có chủ đích từ Internet hoặc đoạn phim camera giám sát an ninh để tạo cơ sở dữ liệu nhận dạng khuôn mặt. Các ứng dụng AI xâm phạm sự tồn tại, sự tự do, an toàn và phẩm giá của con người đều bị cấm. Chẳng hạn như các ứng dụng nhận diện cảm xúc, thao túng hành vi thông qua việc khai thác số tuổi, hiện trạng kinh tế, địa vị xã hội, tình trạng khuyết tật của con người; hay như RBI để suy đoán về chủng tộc, quan điểm chính trị, tôn giáo, xu hướng tính dục của con người. Một số trường hợp đặc biệt như tìm kiếm có chủ đích người mất tích, nạn nhân của hoạt động mua bán người, hoặc ngăn chặn một cuộc tấn công khủng bố... đều kèm theo các yêu cầu nghiêm ngặt và phải nhận được sự chấp thuận của cơ quan tư pháp hoặc cơ quan hành chính độc lập (Hùng Tuấn, 2024).

Các hệ thống AI mục đích chung (General Purpose AI) là một hệ thống AI có thể thực hiện các chức năng thường được áp dụng (như nhận dạng hình ảnh/giọng nói, tạo âm thanh/video, phát hiện mẫu, trả lời câu hỏi, dịch thuật...) như ứng dụng Chat GPT phải tuân thủ quy định về bản quyền, về tính minh bạch, phải nêu rõ cảnh báo đây là sản phẩm do máy móc tạo ra. Những nội dung hình ảnh, âm thanh hoặc video bị thao túng hoặc giả mạo (deepfakes) cần phải được dán nhãn rõ

ràng. Cùng với một số loại AI ứng dụng trong quản lý và vận hành các hệ thống cơ sở hạ tầng quan trọng như giao thông, điện, nước, khí đốt..., các nhà phát triển được yêu cầu hàng loạt nghĩa vụ liên quan đến việc đánh giá mức độ an toàn, tính minh bạch, đưa ra giải pháp quản lý dữ liệu, giám sát, kiểm soát rủi ro, năng lực giải quyết sự cố... (Hùng Tuấn, 2024).

*Thứ hai*, Đạo luật cung cấp cơ chế “mềm” để hỗ trợ quá trình đổi mới, phát triển AI. So với bản dự thảo năm 2021, các quy định về cơ chế thử nghiệm trong văn bản chính thức đã được bổ sung nhằm tạo môi trường pháp lý an toàn, ổn định cho các nhà phát triển, ứng dụng AI. Cơ chế này gồm hai loại: (i) Cơ chế quản lý thử nghiệm (Regulatory sandbox) cho phép các nhà phát triển AI đào tạo, thử nghiệm AI trong khoảng thời gian, không gian nhất định với sự kiểm soát của cơ quan nhà nước có thẩm quyền, trước khi đưa sản phẩm, dịch vụ ra thị trường. (ii) Cơ chế thử nghiệm AI có rủi ro cao trong thực tế (Real-world testing outside sandbox) cho phép các nhà cung cấp thử nghiệm AI ở thời điểm bất kỳ trước khi đưa dịch vụ, sản phẩm ra thị trường với điều kiện được sự chấp thuận và chịu giám sát từ cơ quan giám sát thị trường (Nguyễn Lan Phương, 2024).

*Thứ ba*, Đạo luật cung cấp cơ sở pháp lý hình thành các tổ chức đảm bảo hoặc hỗ trợ thực thi đạo luật. Ở cấp độ khu vực, EU thành lập 04 tổ chức: (i) Ủy ban Trí tuệ nhân tạo (Artificial Intelligence Board) để giám sát thực thi AIA; (ii) Văn phòng AI (Artificial Intelligence Office) để phụ trách phát triển chuyên môn trong lĩnh vực AI; (iii) Diễn đàn tư vấn (Advisory Forum) gồm các bên liên quan như doanh nghiệp, tổ chức xã hội và khối nghiên cứu hàn lâm...

để cung cấp tư vấn chuyên môn, kỹ thuật cho Hội đồng Châu Âu và Ủy ban Châu Âu (EC) trong các quyết định liên quan đến AI; (iv) Hội đồng khoa học (Scientific Panel) gồm các chuyên gia độc lập hỗ trợ các hoạt động thực thi đạo luật (Nguyễn Lan Phương, 2024).

Ở cấp độ quốc gia, EU yêu cầu mỗi quốc gia thành viên phải thành lập hoặc chỉ định 02 cơ quan: Cơ quan giám sát thị trường (Market Surveillance Authority) để đảm bảo các dịch vụ, sản phẩm AI được đưa vào sử dụng đều tuân thủ quy định pháp luật; Cơ quan thông báo (Notifying Authority) để đánh giá, chỉ định, thông báo và giám sát hoạt động của các tổ chức đánh giá sự phù hợp (Nguyễn Lan Phương, 2024).

Bên cạnh đó, các nhà lập pháp EU cũng khuyến khích các nhà phát triển AI, các hiệp hội nghề nghiệp, tổ chức xã hội, tổ chức nghiên cứu khoa học tham gia phát triển các Bộ quy tắc ứng xử (Codes of Conduct) đối với AI không thuộc loại rủi ro cao. Đồng thời, EC hướng tới xây dựng hướng dẫn thực thi Đạo luật, trong đó đặc biệt chú ý đến nhu cầu của các doanh nghiệp nhỏ và vừa, doanh nghiệp khởi nghiệp (Nguyễn Lan Phương, 2024).

### **3. Hướng dẫn phát triển hệ thống trí tuệ nhân tạo an toàn của Anh và Mỹ**

Ngày 27/11/2023, Trung tâm An ninh mạng Quốc gia (National Cyber Security Centre - NCSC) của Vương quốc Anh và Cơ quan An ninh Cơ sở hạ tầng và An ninh mạng (Cybersecurity and Infrastructure Security Agency - CISA) của Mỹ đã công bố tài liệu “Hướng dẫn phát triển hệ thống AI an toàn” (Guidelines for Secure Artificial Intelligence System Development)<sup>1</sup>. Đây là

<sup>1</sup> Xem: <https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf>

tài liệu hướng dẫn chi tiết đầu tiên về cách thức đảm bảo sử dụng an toàn công nghệ AI, nâng cao tiêu chuẩn an ninh mạng AI; hình thành sự hiểu biết chung, thống nhất, mang tính toàn cầu về các rủi ro liên quan đến AI cùng với chiến lược hiệu quả nhằm giảm thiểu rủi ro.

Hướng dẫn cụ thể bao gồm 04 lĩnh vực chính: thiết kế an toàn (secure design), phát triển an toàn (secure development), triển khai an toàn (secure deployment), vận hành và bảo trì an toàn (secure operation and maintenance). Theo nguyên tắc “An toàn ngay từ thiết kế”, các nhà cung cấp AI được yêu cầu phải đảm bảo sự an toàn, phải có các giải pháp bảo mật tích hợp ngay từ việc lựa chọn thông tin đầu vào, mẫu hình, mã nguồn, phần mềm, phần cứng... Trên cơ sở nguyên lý “bảo mật phải là yêu cầu cốt lõi mang tính liên tục và xuyên suốt”, các nhà cung cấp và phát triển AI được khuyến khích nhận trách nhiệm về tình trạng bảo mật trong suốt vòng đời của sản phẩm công nghệ, thay vì quy trách nhiệm cho khách hàng và người dùng. Bộ tài liệu “Hướng dẫn phát triển hệ thống AI an toàn” cũng coi trọng các thông tin báo cáo phát hiện rủi ro, đánh giá mức độ rủi ro và chia sẻ thông tin công khai để có thể nhanh chóng xử lý, khắc phục vấn đề một cách nhanh chóng và hiệu quả. Bộ tài liệu đã nhận được sự đồng thuận của 16 quốc gia trên thế giới gồm: Úc, Canada, Chile, Cộng hòa Czech, Estonia, Pháp, Đức, Israel, Ý, Nhật Bản, New Zealand, Nigeria, Na Uy, Ba Lan, Singapore, Hàn Quốc (Nguyễn Thu, 2023).

#### 4. Các văn bản pháp lý của Trung Quốc

Hiệp hội Trí tuệ nhân tạo Trung Quốc (中国人工智能学会, Chinese Association for Artificial Intelligence - CAAI) được thành lập từ năm 1981. Cho đến nay,

CAAI vẫn là tổ chức duy nhất của Trung Quốc được giao nhiệm vụ nghiên cứu AI và khoa học tình báo. Năm 1987, Đại học Thanh Hoa đã xuất bản ấn phẩm nghiên cứu đầu tiên của Trung Quốc về AI. Từ năm 1993, tự động hóa thông minh và trí tuệ thông minh đã là một phần trong kế hoạch công nghệ quốc gia của Trung Quốc. Đến những năm 2000, Trung Quốc tiếp tục mở rộng Quỹ Nghiên cứu và phát triển AI giúp số lượng các dự án nghiên cứu do Chính phủ tài trợ tăng lên đáng kể. Năm 2006, Trung Quốc công bố chính sách ưu tiên phát triển AI, đưa AI vào Kế hoạch trung hạn và dài hạn quốc gia về Phát triển khoa học và công nghệ (2006-2020).

Văn bản pháp lý đầu tiên của Trung Quốc liên quan đến AI là “Kế hoạch phát triển AI thế hệ mới” (新一代人工智能发展规划) được Quốc vụ viện Trung Quốc công bố ngày 20/7/2017. Đây là bản kế hoạch có nhiều tham vọng với thời gian biểu rõ ràng theo lộ trình: bắt kịp các nước phương Tây vào năm 2023, vượt qua các nước phương Tây vào năm 2025, dẫn đầu thế giới vào năm 2030. Trung Quốc cũng đặt mục tiêu đến năm 2025 sẽ thiết lập hệ thống quy phạm điều chỉnh AI bao gồm đạo luật về AI, bộ quy tắc đạo đức, hệ thống chính sách thông tin về AI để hình thành khả năng đánh giá và kiểm soát an ninh, rủi ro đối với AI (Dẫn theo: Nguyễn Ngọc Phương Hồng, Lưu Minh Sang, 2024).

Năm 2019, Trung Quốc thành lập Ủy ban chuyên môn Quản trị AI thế hệ mới trực thuộc Bộ Khoa học và Công nghệ Trung Quốc. Cơ quan này đã ban hành 02 văn bản hình thành nền móng quan trọng cho việc định hình khung pháp lý điều chỉnh AI: “Bộ nguyên tắc quản trị AI thế hệ

mới: Phát triển AI có trách nhiệm” (发展负责任的人工智能: 新一代人工智能治理原则发布)<sup>1</sup> ban hành tháng 6/2019, và “Chuẩn mực đạo đức cho AI thế hệ mới” (新一代人工智能伦理规范)<sup>2</sup> ban hành tháng 9/2021.

Ngày 13/7/2023, Cục Quản lý không gian mạng Trung Quốc phối hợp với Ủy ban Cải cách và Phát triển Quốc gia cùng Bộ Khoa học và Công nghệ Trung Quốc đã ban hành văn bản “Các biện pháp tạm thời quản lý các dịch vụ AI” (生成式人工智能服务管理暂行办法)<sup>3</sup>. Mục tiêu là nhằm thúc đẩy sự phát triển lành mạnh của AI thế hệ mới và các ứng dụng tiêu chuẩn liên quan, bảo vệ an ninh quốc gia và lợi ích chung của toàn xã hội, đồng thời bảo vệ các quyền và lợi ích hợp pháp của công dân và các tổ chức. Trung Quốc khuyến khích các ứng dụng sáng tạo của công nghệ AI tạo sinh (Generative Artificial Intelligence)<sup>4</sup> trong nhiều ngành và lĩnh vực nhằm mở ra cơ hội mới cho sự phát triển kinh tế và xã hội, hạn chế các tác động tiêu cực như lan truyền thông tin giả mạo, vi phạm bảo mật thông tin cá nhân và an toàn dữ liệu; đồng thời hỗ trợ các công ty, doanh nghiệp, các cơ sở giáo dục, nghiên cứu và các tổ chức liên quan khác thực hiện hợp tác trong lĩnh vực này. Trong khi đó, các cơ quan quản lý sẽ có trách nhiệm thực hiện phân loại và giám sát phân loại các dịch vụ AI tạo sinh; đảm bảo bảo mật thông tin trực tuyến, ngăn

chặn khả năng tiếp cận của người dùng dưới tuổi quy định (QT, 2023).

Ngày 18/3/2024, nhóm nghiên cứu gồm có 07 trường đại học<sup>5</sup> và Viện Nghiên cứu Công nghệ thông tin và Truyền thông Trung Quốc đã công bố bản “Dự thảo kiến nghị của chuyên gia” (Expert Proposal) nhằm phục vụ công tác góp ý, tham mưu cho dự thảo luật về AI của Trung Quốc. Bản dự thảo gồm 09 chương và 96 điều, bao gồm các nguyên tắc chung, các quy định về nghĩa vụ, trách nhiệm pháp lý của các nhà phát triển và cung cấp các sản phẩm và dịch vụ công nghệ AI. Trong bản dự thảo, nhóm nghiên cứu cũng nêu bật sự cần thiết phải phân bổ trách nhiệm rõ ràng hơn trong chuỗi giá trị AI nhằm ngăn chặn các nhà phát triển mô hình AI trốn tránh trách nhiệm thông qua ngôn ngữ hợp đồng hoặc với lý do bảo vệ bí mật kinh doanh; đồng thời đảm bảo sự chủ động của con người đối với AI thông qua giám sát trực tiếp và phương tiện kỹ thuật, kể cả khi AI hoạt động tự chủ. Nhóm nghiên cứu cũng đề xuất thành lập một cơ quan mới là Văn phòng AI Quốc gia (cơ quan đầu tiên trên toàn thế giới) để tập trung điều phối và giám sát việc quản lý công nghệ AI, tránh trường hợp phân tán giám sát hoặc chồng chéo quản lý và trách nhiệm hành chính (Dẫn theo: Nguyễn Ngọc Phương Hồng, Lưu Minh Sang, 2024).

Mới đây nhất, tháng 7/2024, Đại hội đồng Liên Hợp Quốc đã nhất trí thông qua Nghị quyết nâng cao năng lực AI do Trung

<sup>1</sup> Xem: [https://www.most.gov.cn/kjbgz/201906/t20190617\\_147107.html](https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html)

<sup>2</sup> Xem: [https://www.most.gov.cn/kjbgz/202109/t20210926\\_177063.html](https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html)

<sup>3</sup> Xem: [https://www.cac.gov.cn/2023-07/13/c\\_1690898327029107.htm](https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm)

<sup>4</sup> AI tạo ra nội dung mới trên cơ sở dữ liệu ban đầu, chứ không chỉ phân tích dữ liệu sẵn có.

<sup>5</sup> Đại học Khoa học và Công nghệ Trung Quốc, Đại học Khoa học và Công nghệ Bắc Kinh, Đại học Khoa học và Công nghệ Hoa Trung, Đại học Khoa học và Công nghệ Điện tử, Đại học Khoa học Kỹ thuật Điện tử Tây An, Đại học Khoa học và Công nghệ Nông lâm Tây Bắc, Đại học Công nghệ Quốc phòng.

Quốc dẫn đầu kêu gọi cộng đồng quốc tế tạo ra một môi trường kinh doanh “tự do, cởi mở, bao trùm và không phân biệt đối xử” giữa các quốc gia cho sự phát triển AI. Hơn 140 quốc gia đã bỏ phiếu thông qua Nghị quyết, với mong muốn đẩy mạnh hợp tác quốc tế và các hành động thực tiễn nhằm hỗ trợ các nước, đặc biệt là các nước đang phát triển, nâng cao năng lực AI, nâng cao tính đại diện và tiếng nói của các nước này trong quản trị AI toàn cầu (Cao Phong, 2024).

### 5. Khuyến nghị cho Việt Nam

Theo kết quả đánh giá và công bố trong báo cáo “Chỉ số sẵn sàng trí tuệ nhân tạo của chính phủ” do Oxford Insight thực hiện<sup>1</sup>, năm 2023, Việt Nam đứng thứ 59/193 quốc gia trên thế giới, đứng thứ 5/10 trong khối ASEAN về khai thác ứng dụng AI để vận hành và cung cấp dịch vụ, tăng một bậc so với năm 2022. Trong thời gian tới, với những tiềm lực sẵn có, chắc chắn AI sẽ tiếp tục phát triển rất nhanh. Để thúc đẩy triển khai hiệu quả hơn nữa “Chiến lược quốc gia về Nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030”, Việt Nam cần đẩy mạnh đầu tư vào công nghệ AI, không chỉ là ngân sách, nguồn nhân lực và vật lực, các phương tiện, trang thiết bị kỹ thuật hiện đại, các biện pháp, chính sách thúc đẩy sản xuất, tiêu dùng các sản phẩm, dịch vụ AI sản xuất trong nước; mà còn là xây dựng hệ thống pháp luật quản lý AI, trong đó quy định mức độ rủi ro, tính tin cậy, thực thi đạo đức và cách thức quản lý sản phẩm, ứng dụng AI.

Khung pháp lý về AI cần tập trung vào việc tìm kiếm điểm cân bằng giữa khuyến khích phát triển, thúc đẩy sáng tạo, đổi mới và quản lý, phòng ngừa rủi ro, đảm bảo an toàn trên cơ sở 04 nguyên tắc: *Một là* nguyên tắc lấy con người làm trung tâm, coi trọng các quyền con người và quyền dân sự cơ bản, lưu tâm đến quyền và lợi ích của các nhóm đặc thù, nhóm yếu thế như người già, người khuyết tật, đảm bảo rằng họ có quyền được giải thích hoặc quyền được từ chối sử dụng AI. *Hai là* nguyên tắc tuân thủ nghiêm ngặt các quy định pháp luật, tôn trọng quy chuẩn và đạo đức xã hội; nghiêm cấm tạo ra các nội dung kích động, gây nguy hại cho an ninh, lợi ích quốc gia, làm tổn hại đến hình ảnh đất nước; nghiêm cấm nội dung bạo lực, khiêu dâm, phân biệt đối xử, thông tin sai lệch trong cung cấp và sử dụng AI. *Ba là* nguyên tắc công khai, minh bạch, chính xác, đáng tin cậy trong việc xác định đối tượng sử dụng, cách thức sử dụng. *Bốn là* nguyên tắc tôn trọng quyền sở hữu trí tuệ và bảo mật thông tin người dùng trong quá trình thu thập và xử lý thông tin, chú thích dữ liệu (Xem thêm: Lưu Minh Sang, Trần Đức Thành, 2020).

Bên cạnh đó, Việt Nam cần tiếp tục nghiên cứu kinh nghiệm quốc tế để sửa đổi Luật Sở hữu trí tuệ (đã sửa đổi năm 2022) bao gồm các quy định về AI; nghiên cứu và xác định rõ tư cách pháp lý, bản chất pháp lý của AI và những thực thể mang AI, hướng đến việc xây dựng khung pháp lý điều chỉnh các mối quan hệ pháp luật liên quan đến AI như quan hệ về tài sản, quyền sở hữu, sở hữu trí tuệ, quan hệ lao động, bồi thường thiệt hại... Nếu như Mỹ và Úc chưa công nhận AI là đối tượng của bảo hộ quyền sở hữu trí tuệ, hay nói cách khác, những sản phẩm do AI tạo ra

<sup>1</sup> Xem: Oxford Insight (2023), Government AI Readiness Index 2023, <https://oxfordinsights.com/wp-content/uploads/2023/12/2023-Government-AI-Readiness-Index-1.pdf>

không được bảo hộ quyền sở hữu trí tuệ<sup>1</sup>, thì Vương quốc Anh hiện có một cách tiếp cận khá cởi mở, hướng đến việc ghi nhận quyền tác giả cho người đã tạo nên các chương trình máy tính, ghi nhận quyền sở hữu trí tuệ cho các tác phẩm được tạo ra bởi những “tác giả” không phải con người. Đạo luật Bản quyền, Kiểu dáng và Bằng sáng chế năm 1988 của Anh nêu rõ: “Trong trường hợp tác phẩm văn học, kịch, âm nhạc hay nghệ thuật được tạo ra từ máy tính, tác giả sẽ là người sắp xếp cần thiết cho việc tạo ra tác phẩm được thực hiện” (Xem: Văn Chiến, 2024).

Cùng với đó, Việt Nam cũng cần tiếp tục tăng cường hợp tác quốc tế, tham khảo kinh nghiệm chuyên gia để luật hóa công nghệ AI. Quyết định số 1290/QĐ-BKHCN về việc hướng dẫn một số nguyên tắc về nghiên cứu, phát triển các hệ thống AI có trách nhiệm đã được Bộ Khoa học và Công nghệ ban hành ngày 11/6/2024. Tài liệu đưa ra 09 nguyên tắc nghiên cứu, phát triển các hệ thống AI có trách nhiệm và hướng dẫn thực hiện<sup>2</sup>, trong đó có thể tóm lược một số nội dung của các nguyên tắc như: (i) *Tinh thần hợp tác, thúc đẩy đổi mới sáng tạo*: Tăng cường tính liên kết và khả năng tương tác giữa các hệ thống AI, tăng cường chia sẻ thông tin trong cộng đồng AI, ưu tiên phát triển các hệ thống AI phù hợp với các quy

chuẩn kỹ thuật, tiêu chuẩn quốc gia hoặc tiêu chuẩn quốc tế; (ii) *Tính minh bạch*: Kiểm soát đầu vào/đầu ra của hệ thống AI và khả năng giải thích các phân tích có liên quan để không gây ảnh hưởng đến tính mạng, thân thể, quyền riêng tư hoặc tài sản của người dùng hoặc bên thứ ba liên quan; (iii) *Khả năng kiểm soát*: Coi trọng năng lực kiểm soát hệ thống AI, đánh giá rủi ro và có các biện pháp ứng phó phù hợp; (iv) *An toàn*: Đảm bảo an toàn nội tại (giảm các yếu tố rủi ro như mức năng lượng của các thiết bị tạo ra sự kiện...) và an toàn chức năng (giảm thiểu rủi ro bằng cách sử dụng các thiết bị điều khiển bổ sung như tự động dừng khi có sự cố...); (v) *Bảo mật*: Lưu ý tính bảo mật thiết kế dựa trên đặc điểm của các công nghệ được áp dụng trong suốt quá trình phát triển hệ thống AI; (vi) *Quyền riêng tư*: Đảm bảo không vi phạm quyền riêng tư về không gian sống, về thông tin dữ liệu cá nhân và bí mật thông tin liên lạc của người dùng hoặc bên thứ ba; (vii) *Tôn trọng quyền và phẩm giá con người*: Đảm bảo tôn trọng nhân phẩm con người, không vi phạm các giá trị con người, đạo đức xã hội, không phân biệt đối xử, thiên vị, định kiến khi huấn luyện hệ thống AI; (viii) *Hỗ trợ người dùng*: tạo ra các giao diện, chức năng hỗ trợ, tạo điều kiện thuận lợi cho người dùng trong quá trình sử dụng; (ix) *Trách nhiệm giải trình*: Thực hiện trách nhiệm giải trình đối với các bên liên quan bao gồm cả người dùng hệ thống AI, cập nhật các vấn đề phát sinh để có giải pháp kịp thời.

## 6. Kết luận

Công nghệ AI dần trở thành “hạt nhân” của cuộc Cách mạng Công nghiệp lần thứ tư đang diễn ra mạnh mẽ trên toàn cầu. AI là một nền tảng đa năng, thông minh, được sử dụng rộng rãi trong nhiều lĩnh vực của đời sống con người, mang lại những sản

<sup>1</sup> Đạo luật Bản quyền năm 1968 của Úc quy định: “Quyền tác giả tồn tại trong một tác phẩm văn học, kịch, âm nhạc hoặc nghệ thuật gốc mà tác phẩm chưa được xuất bản và tác giả là công dân hoặc cư dân sinh sống tại Úc”. Cục Bản quyền Mỹ tuyên bố chỉ cho phép đăng ký quyền tác giả đối với các tác phẩm có được là do hoạt động sáng tạo của con người, và từ chối tất cả các đơn đăng ký quyền tác giả khi xác định được rằng các tác phẩm này không phải do con người sáng tạo (Văn Chiến, 2024).

<sup>2</sup> Xem: [https://sokhoahoc.sonla.gov.vn/?pageid=30963&p\\_steering=91177](https://sokhoahoc.sonla.gov.vn/?pageid=30963&p_steering=91177)

phẩm mới và dịch vụ mới, được kỳ vọng sẽ định hình lại nền kinh tế và các hoạt động xã hội, có tác động lớn đến năng suất, việc làm, phân phối thu nhập và toàn cầu hóa. Nhưng AI cũng khiến con người lo ngại về những rủi ro đạo đức và xã hội tiềm ẩn mà nó có thể gây ra. Việc luật hóa AI đã và đang được hầu hết các quốc gia trên thế giới thực thi và Việt Nam cũng không thể đứng ngoài xu thế đó. Các đạo luật, hướng dẫn, văn bản pháp lý của EU, Anh, Mỹ và Trung Quốc là những tham chiếu đáng tin cậy để Việt Nam nghiên cứu, tham khảo trong quá trình triển khai □

#### Tài liệu tham khảo

1. Văn Chiến (2024), “Công nghệ trí tuệ nhân tạo (AI) nhìn từ góc độ pháp luật sở hữu trí tuệ một số nước trên thế giới và Việt Nam”, Tạp chí điện tử *Pháp lý* ngày 08/3/2024, <https://phaply.net.vn/cong-nghe-tri-tue-nhan-tao-ai-nhin-tu-goc-do-phap-luat-so-huu-tri-tue-mot-so-nuoc-tren-the-gioi-va-viet-nam-a256438.html>
2. Đại học Fulbright Việt Nam (2024), *Giải mã những bí ẩn về ngành Trí tuệ nhân tạo (AI)*, <https://fulbright.edu.vn/vi/nganh-tri-tue-nhan-tao/>, ngày 04/6/2024.
3. Nguyễn Ngọc Phương Hồng, Lưu Minh Sang (2024), “Luật về trí tuệ nhân tạo và tầm nhìn chiến lược của Trung Quốc”, Tạp chí điện tử *Kinh tế Sài Gòn Online* ngày 05/5/2024, <https://thesaigontimes.vn/luat-ve-tri-tue-nhan-tao-va-tam-nhin-chien-luoc-cua-trung-quoc/>
4. Oxford Insight (2023), *Government AI Readiness Index 2023*, <https://oxfordinsights.com/wp-content/uploads/2023/12/2023-Government-AI-Readiness-Index-1.pdf>
5. Cao Phong (2024), “Nghị quyết do Trung Quốc dẫn đầu về AI được thông qua tại Liên Hợp Quốc”, Báo điện tử *Nhà báo & Công luận* ngày 02/7/2024, <https://congluan.vn/nghi-quyet-do-trung-quoc-dan-dau-ve-ai-duoc-thong-qua-tai-lien-hop-quoc-post301981.html>
6. Nguyễn Lan Phương (2024), “Đạo luật Trí tuệ nhân tạo đầu tiên trên thế giới”, Tạp chí điện tử *Kinh tế Sài Gòn Online* ngày 23/3/2024, <https://thesaigontimes.vn/dao-luat-tri-tue-nhan-tao-dau-tien-tren-the-gioi/>
7. Lưu Minh Sang, Trần Đức Thành (2020), “Trí tuệ nhân tạo và những thách thức pháp lý”, Tạp chí *Khoa học và Công nghệ Việt Nam* ngày 29/5/2020, <https://vjst.vn/vn/tin-tuc/3303/tri-tue-nhan-tao-va-nhung-thach-thuc-phap-ly.aspx>
8. QT (2023), “Trung Quốc chính thức ban hành quy tắc quản lý AI”, *Truyền hình Quốc hội Việt Nam* ngày 14/7/2023, <https://beta.quochoitv.vn/trung-quoc-chinh-thuc-ban-hanh-quy-tac-quan-ly-ai-182813.htm>
9. Nguyệt Thu (2023), “Mỹ và Anh phát triển các thỏa thuận hướng dẫn bảo mật mới về trí tuệ nhân tạo”, Tạp chí *An toàn thông tin* ngày 06/12/2023, <https://m.antoanthongtin.vn/cong-nghe-thong-tin/my-va-anh-phat-trien-cac-thoa-thuan-huong-dan-bao-mat-moi-ve-tri-tue-nhan-tao-109556>
10. Hùng Tuấn (2024), “Nghị viện châu Âu thông qua Đạo luật về trí tuệ nhân tạo”, Tạp chí *Khoa học & Công nghệ Việt Nam* ngày 17/4/2024, <http://vjst.vn/vn/tin-tuc/9292/nghi-vien-chau-au-thong-qua-dao-luat-ve-tri-tue-nhan-tao.aspx>