

TRÁCH NHIỆM PHÁP LÝ TRONG VỤ VIỆC LIÊN QUAN TRÍ TUỆ NHÂN TẠO

ĐỖ THỊ BẢO YẾN*

Trong bối cảnh hiện nay, sự phát triển nhanh chóng của trí tuệ nhân tạo (AI) đang làm thay đổi thế giới với tốc độ nhanh chóng. Kỹ thuật robot và AI giúp thay thế sức lao động của con người, tăng năng suất lao động và góp phần thúc đẩy phát triển nền kinh tế. Xu hướng này thể hiện sự phát triển tất yếu khách quan của nền văn minh nhân loại, mang lại đồng thời cả hệ quả tích cực và tiêu cực. Bài viết phân tích một số vấn đề về việc xác định trách nhiệm pháp lý trong vụ việc liên quan đến AI và có những gợi mở nhằm hoàn thiện quy định về AI ở Việt Nam trong giai đoạn tiếp theo.

Từ khóa: Trách nhiệm pháp lý; trí tuệ nhân tạo; hoạt động nghề nghiệp; thúc đẩy phát triển; hoàn thiện quy định về pháp lý.

In the current context, the rapid development of artificial intelligence (AI) is transforming the world at an unprecedented pace. Robotics and AI technologies are increasingly replacing human labor, boosting productivity, and contributing to economic growth. This trend represents an inevitable and objective stage of human civilization, bringing both positive and negative consequences. The article examines several key issues related to determining legal liability in AI-related cases. It provides recommendations for enhancing the legal framework governing AI in Vietnam over the coming period.

Keywords: Legal liability; artificial intelligence; professional activities; development promotion; improvement of legal regulations.

NGÀY NHẬN: 15/6/2025 NGÀY PHẢN BIỆN, ĐÁNH GIÁ: 12/10/2025 NGÀY DUYỆT: 14/11/2025

DOI: <https://doi.org/10.59394/qlnn.358.2025.1357>

1. Đặt vấn đề

Hiện nay, trên thế giới, khả năng tự học, bất chước và vượt qua con người về tốc độ, độ chính xác và khối lượng hoạt động tư duy của AI được công nhận và tận dụng trong nhiều lĩnh vực đời sống xã hội. Những vụ việc pháp lý liên quan đến AI, mỗi cá nhân, tổ chức đều phải chịu trách nhiệm về việc không tuân thủ quy định của pháp luật trong tạo ra và sử

dụng công cụ này cũng như những hậu quả bất hợp pháp; trong đó có yếu tố nguy hiểm về mặt xã hội do sử dụng AI gây ra. Để xác định đối tượng nào phải chịu trách nhiệm về tác hại do AI gây ra, cần thiết lập quá trình kiểm tra và đánh giá các đối tượng đó, trên cơ

*NCS của Trường Đại học Luật, Đại học Quốc gia Hà Nội

sở đó, xác định nguyên nhân hoạt động có hại của AI cũng như xác định trách nhiệm của những chủ thể tùy thuộc vào hình thức, yếu tố lỗi và hậu quả.

Cơ sở xác định trách nhiệm pháp lý xảy ra do sử dụng AI cần được làm rõ để tránh lỗ hổng, xung đột pháp lý và góp phần phá bỏ rào cản đối với sự phát triển của nền kinh tế khi ứng dụng công nghệ AI ngày càng lan rộng.

2. Cơ sở hình thành trách nhiệm pháp lý trong vụ việc liên quan đến AI

Ngày 12/7/2024, Đạo luật Trí tuệ nhân tạo (AIA) của Liên minh châu Âu (EU) đã chính thức được công bố, đánh dấu cột mốc quan trọng trong việc điều chỉnh việc sử dụng AI và bảo vệ quyền sở hữu trí tuệ. Cho đến nay, ở Việt Nam vẫn chưa có quy định cụ thể tương tự nhưng những thay đổi ở phạm vi quốc tế sẽ có tác động ảnh hưởng mạnh mẽ; đồng thời, đặt ra những thách thức và cơ hội cho doanh nghiệp và người sử dụng AI ở Việt Nam. Liên quan đến lĩnh vực này, nhiều quan hệ xã hội mới hình thành, đòi hỏi cần được pháp luật điều chỉnh.

Báo cáo về đạo đức của kỹ thuật robot của UNESCO đã liệt kê bốn đặc điểm chính của robot hiện đại là: khả năng di chuyển, khả năng tương tác, khả năng giao tiếp, khả năng tự chủ; qua đó, cho phép robot phân tích và đưa ra quyết định độc lập; đồng thời, thực hiện các hành động mà không cần sự can thiệp và kiểm soát của con người. Nghị quyết của Nghị viện châu Âu P8-TA-PROV (2017)0051: “Các chuẩn mực của pháp luật dân sự về robot”, AI được đề xuất là một hệ thống mạng vật lý tự chủ phi sinh học, có khả năng tương tác với các hệ thống tương tự khác, khả năng tự học, đưa ra quyết định, hành động theo các điều kiện tùy vào môi trường hoàn cảnh¹.

Tuy nhiên, việc sử dụng AI ở cấp độ nào thì việc chế tạo và sử dụng AI đều phụ thuộc vào con người. Do đó, cần phải xác định rõ ràng không chỉ mục tiêu và cách sử dụng AI mà còn cả những quy định giám sát hoạt động chế tạo

và sử dụng AI cũng như xác định trách nhiệm pháp lý trong những vụ việc liên quan đến AI. Khi một hành vi gây hại được thực hiện bởi một AI, trong khi AI đó không có khả năng chịu trách nhiệm pháp lý ngang bằng với một người. Trong trường hợp như vậy, cần phải kiểm tra và đánh giá hành vi (hành động hoặc không hành động) của chủ thể, như: lập trình viên, nhà sản xuất, người dùng hoặc bất kỳ đối tượng nào can thiệp bất hợp pháp nhằm xác định nguyên nhân gây ra hậu quả có hại. Theo đó, giả thiết xảy ra các trường hợp sau:

Thứ nhất, có khả năng chứng minh được nguyên nhân rõ ràng của sự kiện gây hại do AI. Khi đó dễ dàng xác định được trách nhiệm pháp lý cho các hành vi (hành động hoặc không hành động) của lập trình viên, nhà sản xuất, người dùng, sự hiện diện hoặc vắng mặt của mối quan hệ nhân quả giữa các hành vi của những chủ thể này và AI gây hậu quả tiêu cực.

Thứ hai, trường hợp phức tạp hơn khi có sự can thiệp bất hợp pháp của đối tượng khác nằm ngoài dự tính ban đầu, khiến AI đó phát triển thêm kỹ năng có hại dẫn đến hậu quả nguy hiểm cho xã hội. Điều này nằm ngoài mong muốn hay suy luận của nhà sản xuất trong quá trình AI tự học hỏi.

Thứ ba, trường hợp mà hầu hết các hành động nhằm xâm nhập và thay đổi hoạt động của AI được con người thực hiện một cách trực tiếp hoặc gián tiếp cho các mục đích bất hợp pháp nhưng dẫn tới hậu quả nặng nề hơn nằm ngoài ý định ban đầu. Khi đó, hành vi vi phạm pháp luật của các lập trình viên, nhà sản xuất và hầu hết người dùng AI có thể được thực hiện một cách vô ý, khi các chủ thể này không lường trước được khả năng xảy ra hậu quả có hại, có thể lường trước được hoặc đã lường trước được nhưng lại chủ quan cho rằng sẽ ngăn chặn được.

Như vậy, có những tính toán sai lầm được coi là hành vi do sơ suất mà trách nhiệm pháp lý tùy thuộc vào hậu quả được quy định về trách nhiệm dân sự, hành chính hoặc hình sự. Việc lập trình viên, nhà sản xuất, người

dùng đánh giá không đúng về mối nguy hiểm xã hội của hậu quả và mức độ nghiêm trọng của nó và (hoặc) đánh giá không đúng của những người này về mối quan hệ nhân quả giữa các hoạt động của họ trong lĩnh vực tạo ra và sử dụng AI để có kết quả của các hoạt động của AI.

3. Xây dựng và hoàn thiện quy định về trách nhiệm pháp lý trong vụ việc liên quan đến AI của Việt Nam từ kinh nghiệm một số nước

Việc hoàn thiện quy định về trách nhiệm pháp lý trong vụ việc liên quan đến AI là nhiệm vụ cấp thiết. Việt Nam có thể tham khảo vấn đề này để rút ra kinh nghiệm của một số nước trên thế giới.

Một là, cần từng bước xây dựng những quy định đầu tiên về nguyên tắc xây dựng và sử dụng AI đối với nhà sản xuất, lập trình viên và người dùng. Trọng tâm của các nguyên tắc này là bảo đảm an ninh và phát triển bền vững của xã hội, ưu tiên bảo đảm sự an toàn và quyền con người, thông qua việc đặt ra các quy tắc đối với nhà sản xuất và người sử dụng AI. Ví dụ, ở Liên bang Nga chưa có *Đạo luật AI*, tuy nhiên, nước này cũng đã ban hành hai văn bản là Nghị định số 490 ngày 10/10/2019 của Tổng thống Liên bang Nga (sửa đổi ngày 15/02/2024) về việc phát triển AI tại Liên bang Nga (cùng với “Chiến lược quốc gia về phát triển trí tuệ nhân tạo cho giai đoạn đến năm 2030”)² và Sắc lệnh số 2129-r ngày 19/8/2020 của Chính phủ Liên bang Nga về việc phê duyệt khái niệm phát triển quy định về quan hệ trong lĩnh vực AI và công nghệ rô-bốt đến năm 2024³. Nghị định số 490 đã thiết lập những nguyên tắc đầu tiên ở Liên bang Nga về an ninh và minh bạch trong hoạt động sáng tạo và sử dụng AI. Trong đó, quy định không được phép sử dụng AI vào mục đích bất hợp pháp; nhà sản xuất, lập trình viên chế tạo hoặc ứng dụng AI có nghĩa vụ thực hiện các biện pháp để ngăn ngừa và giảm thiểu rủi ro hậu quả tiêu cực do việc sử dụng nó; đồng thời, có nghĩa vụ bảo đảm năng lực giải thích hoạt động của AI và quá trình đạt được kết quả cũng như quyền truy

cập thông tin về các thuật toán được áp dụng của AI. Do đó, Việt Nam có thể tham khảo và dựa trên mức độ tuân thủ hoặc không tuân thủ các nguyên tắc này của người sáng tạo và người dùng, có thể xác định được sự hiện diện hay vắng mặt của trách nhiệm pháp lý.

Hai là, cần ưu tiên xây dựng quy định về AI đối với các lĩnh vực có phạm vi ảnh hưởng rộng và mức độ phổ biến cao. Ở Mỹ, sử dụng AI để hỗ trợ tòa án trong việc xét xử đang được quan tâm và phát triển. Tại Liên bang Nga, theo quy định hiện hành về AI, nếu nhà sản xuất, lập trình viên và người sử dụng AI không thể lường trước được những hậu quả tiêu cực từ hoạt động của AI thì vấn đề buộc chủ sở hữu AI phải chịu trách nhiệm dân sự mà không có lỗi theo Điều 1079 *Bộ luật Dân sự Liên bang Nga* về trách nhiệm đối với thiệt hại gây ra bởi nguồn nguy hiểm cao độ có thể được xem xét. Thực tiễn sử dụng AI trong lĩnh vực giao thông vận tải, y tế và các lĩnh vực thiết yếu khác cho thấy, tính tự động của AI vốn nằm ngoài tầm kiểm soát hoàn toàn của con người, có khả năng gây hại và nguy hiểm nghiêm trọng cho con người. Đặc biệt là các hệ thống robot thông minh phục vụ mục đích quân sự hoặc mục đích kép. Vì vậy, cần xây dựng và áp dụng các quy tắc và quy định đặc biệt cho các hệ thống để bảo đảm mức độ an toàn và trách nhiệm cao hơn trong quá trình tạo ra và sử dụng AI. Tuy nhiên, đối với Việt Nam, không áp dụng trong những lĩnh vực trên mà cần hoàn thiện quy định trong sử dụng AI vào những lĩnh vực, như: nhận dạng sinh trắc học, giáo dục - đào tạo, sử dụng các dịch vụ công...

Ba là, bổ sung quy định chi tiết về trường hợp khi các hành vi xâm nhập và thay đổi hoạt động của AI được thực hiện một cách bất hợp pháp và gây ra hậu quả nghiêm trọng hơn nằm ngoài mục đích ban đầu. Khi đó, hành vi vi phạm pháp luật của các lập trình viên, nhà sản xuất và hầu hết người dùng AI có thể được thực hiện một cách vô ý, khi các chủ thể này không lường trước được khả năng xảy ra hậu quả có hại nhưng đáng lẽ

phải lường trước và có thể lường trước được hoặc đã lường trước được nhưng lại chủ quan cho rằng sẽ ngăn chặn được.

Ngoài ra, nếu Việt Nam xây dựng một đạo luật về AI trong tương lai thì cần có những quy định rõ ràng về vấn đề cơ bản, gồm: (1) Bảo vệ thông tin cá nhân của công dân (bao gồm việc công bố thông tin về nguồn gốc dữ liệu, phương pháp thiết kế mô hình và các biện pháp đảm bảo sự công bằng, không phân biệt đối xử); (2) Thiết kế điều khoản sử dụng của nhà sản xuất; quy định về trách nhiệm và đạo đức của nhà sản xuất, lập trình viên; (3) Báo cáo vi phạm, chế tài đối với hành vi vi phạm.

Đối với việc sửa đổi, bổ sung các quy định hiện hành, trong đó có việc thêm vào những điều khoản về trách nhiệm pháp lý trong vụ việc có liên quan đến AI thì ngoài độ tuổi và năng lực pháp lý theo quy định của pháp luật, các chủ thể là lập trình viên, nhà sản xuất và người dùng với tư cách là chủ thể chịu trách nhiệm pháp lý cũng phải đáp ứng các yêu cầu chuyên môn liên quan. Ví dụ, ở Liên bang Nga, đối với vụ việc xét xử đối tượng tấn công trái phép hệ thống AI từ bên ngoài và thực hiện các thay đổi đối với hoạt động của hệ thống, Tòa án cần phải dựa trên cùng lúc các quy định về độ tuổi được quy định tại Điều 1074 *Bộ luật Dân sự Liên bang Nga*, Điều 2.3 *Bộ luật Xử lý vi phạm hành chính Liên bang Nga* và Điều 20 *Bộ luật Hình sự Liên bang Nga* và có năng lực pháp lý (Điều 21 *Bộ luật Dân sự Liên bang Nga*, Điều 25.3 *Bộ luật Xử lý vi phạm hành chính Liên bang Nga*, Điều 21 *Bộ luật Hình sự Liên bang Nga*⁴).

Bốn là, hoàn thiện quy định về trách nhiệm pháp lý trong vụ việc liên quan đến AI thì xác định yếu tố “lỗi” (cố ý, vô ý) trong quy phạm pháp luật là vấn đề đòi hỏi sự nghiên cứu kỹ lưỡng của nhà làm luật. Trong đó, “lỗi” có tính chất là tinh thần, thái độ của một người đối với hành vi bị pháp luật cấm do người đó thực hiện, như một yếu tố cần thiết của một hành vi phạm tội, có tầm quan trọng

đặc biệt trong việc xác định trách nhiệm pháp lý đối với các hành vi phạm tội trong lĩnh vực tạo ra và ứng dụng các hệ thống AI. Trong khi đó, yếu tố “lỗi” được đặc trưng bởi hai phương diện là trí tuệ và ý chí. Phương diện trí tuệ của “lỗi” bao hàm nhận thức về bản chất (nội dung và ý nghĩa) của hành vi và những hậu quả có thể xảy ra của nó. Ví dụ: hành vi cố ý phạm tội của lập trình viên, nhà sản xuất, người dùng và những người đã truy cập trái phép vào hệ thống AI. Còn phương diện ý chí bao gồm hoặc là mục đích của hành vi hoặc thái độ thiếu thận trọng, bất cẩn, cẩu thả hoặc quá mức chủ quan trong hành vi. Trong trường hợp xác định được lỗi là vô ý, trách nhiệm pháp lý được xác định linh hoạt tùy thuộc vào hậu quả đã xảy ra.

Năm là, Việt Nam cần tham khảo nguyên tắc quốc tế trong việc xây dựng quy định về bồi thường thiệt hại gây ra trong những vụ việc có liên quan đến AI. Ví dụ, hiện nay, ở một số nước, việc bồi thường thiệt hại gây ra trong những vụ việc có liên quan đến AI được thực hiện từ quỹ của một hệ thống bảo hiểm trách nhiệm dân sự bắt buộc được thiết lập đặc biệt dành cho các nhà phát triển và chủ sở hữu AI (đoạn 57 - 59 trong Nghị quyết ngày 16/2/2017 của Nghị viện châu Âu (EP) về các chuẩn mực của *Luật Dân sự về robot*).

Ngoài ra, *Đạo luật AI* ngày 12/7/2024 của EU là khuôn khổ pháp lý toàn diện đầu tiên trên thế giới về AI nhằm mục đích bảo đảm các hệ thống AI được phát triển và sử dụng một cách có trách nhiệm, minh bạch và công bằng. Trong đó, bảo đảm AI được phát triển và sử dụng ở châu Âu hoàn toàn phù hợp với các quyền và giá trị của EU, bao gồm: giám sát con người, an toàn, quyền riêng tư, minh bạch, không phân biệt đối xử, bảo đảm phúc lợi xã hội và môi trường. Đạo luật này đã thiết lập các tiêu chí phân loại các hệ thống AI dựa trên mức độ rủi ro mà có thể gây ra cho người dùng, gồm các cấp độ “rủi ro không thể chấp nhận được”, “rủi ro cao”, “rủi ro hạn chế” và “rủi ro thấp”.

Đối với việc hoàn thiện quy định liên quan đến AI, ở Việt Nam có thể tham khảo việc cấm một số ứng dụng AI nguy hiểm và quy định chi tiết những hành vi như thế nào thì được xác định là “rủi ro không thể chấp nhận được”, “rủi ro cao”, “rủi ro hạn chế” và “rủi ro thấp”. Ví dụ, cần cấm sử dụng các hệ thống AI khai thác điểm yếu của cá nhân, thu thập hình ảnh khuôn mặt từ internet hoặc CCTV (Closed-Circuit Television) mà không có mục tiêu để tạo cơ sở dữ liệu nhận dạng khuôn mặt. Trong *Đạo luật AI* của EU thì đây trường hợp hệ thống AI thuộc nhóm I - có nguy cơ tạo ra “rủi ro không thể chấp nhận được” và sẽ bị cấm đưa vào thị trường hoặc sử dụng tại EU, như thao túng hành vi nhận thức của con người hoặc các nhóm dễ bị tổn thương cụ thể (ví dụ: đồ chơi kích hoạt bằng giọng nói khuyến khích hành vi nguy hiểm ở trẻ em) hoặc loại AI phân loại con người dựa trên hành vi, tình trạng kinh tế - xã hội hoặc đặc điểm cá nhân.

Sáu là, cần quy định chi tiết về quy trình công khai các tác phẩm có bản quyền được sử dụng trong quá trình chế tạo AI. Ngoài ra, quy định chi tiết về việc người dùng được thông báo khi họ đang tương tác với AI, bao gồm các hệ thống AI có khả năng tạo ra hoặc xử lý nội dung hình ảnh, âm thanh hoặc video, đặc biệt là deepfakes. Để quá trình theo dõi, giám sát việc chế tạo và sử dụng AI đạt hiệu quả, cần xây dựng hệ thống giám sát thị trường cùng với thiết lập hàng rào pháp lý chặt chẽ. Ví dụ, tham khảo ý tưởng của các nước thuộc EU như các cơ quan chức năng giám sát việc tuân thủ đạo luật về AI và có quyền áp dụng các biện pháp chế tài đối với các vi phạm, mức phạt từ 35 triệu euro hoặc 7% doanh thu toàn cầu, tùy thuộc hành vi vi phạm và quy mô của các công ty⁵.

4. Kết luận

Vấn đề xác định trách nhiệm pháp lý đối với các hành vi phạm tội khác nhau trong lĩnh vực chế tạo và sử dụng AI vẫn còn nhiều tranh cãi cần tiếp tục nghiên cứu chuyên sâu

hơn. Đặc biệt là cần phân định rạch ròi giữa trách nhiệm pháp lý của các chủ thể và AI. Trong bối cảnh công nghệ khoa học ngày càng phát triển, ở bất kỳ quốc gia nào, đều có rủi ro về hành vi vi phạm pháp luật thông qua việc cố ý chế tạo và sử dụng AI trực tiếp cho các mục đích bất hợp pháp. Do đó, việc xác định trách nhiệm pháp lý của nhà sản xuất, lập trình viên hay người dùng càng cần được xác định chặt chẽ hơn. Đặc biệt, hành vi truy cập (xâm nhập) trái phép vào hệ thống AI từ bên ngoài và thay đổi hoạt động nghiêm trọng. Trường hợp cố ý sử dụng AI cho các mục đích bất hợp pháp thì AI sẽ được coi là công cụ để thực hiện hành vi phạm tội□

Chú thích:

1. *Нормы гражданского права о робототехнике, Резолюция Европарламента от 16 февраля 2017 года 2015/2013(INL) P8_TA-PROV(2017)0051*, включает текст Хартии робототехники. URL: <https://robotrends.ru>, truy cập ngày 09/6/2025.

2. Указ Президента РФ от 10/10/2019 N 490 (ред. от 15/2/2024) “О развитии искусственного интеллекта в Российской Федерации” (вместе с “Национальной стратегией развития искусственного интеллекта на период до 2030 года”).

3. Распоряжение Правительства РФ от 19/8/2020 N 2129-р “Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года”.

4. Апостолова Наталья Николаевна (2021), *Ответственность за вред, причиненный искусственным интеллектом*, Северо-Кавказский юридический вестник, № 1, 112-119.

5. EU Artificial Intelligence Act, *Article 99: Penalties*, <https://artificialintelligenceact.eu/article/99/>, truy cập ngày 09/6/2025.

Tài liệu tham khảo:

1. Dương Quỳnh Hoa (2020). *Trách nhiệm bồi thường thiệt hại trong kỷ nguyên trí tuệ nhân tạo*. Tạp chí Nghiên cứu Lập pháp, số 09 (409), tháng 5/2020.

2. *Doanh nghiệp sử dụng AI có trách nhiệm*. <https://nhandan.vn>, ngày 06/9/2024.