

KHUNG PHÁP LÝ ĐỐI VỚI ROBOT TRÍ TUỆ NHÂN TẠO TẠI VIỆT NAM: THỰC TRẠNG VÀ KIẾN NGHỊ HOÀN THIỆN

• LÊ THỊ THU HẰNG

Khoa Luật quốc tế, Học viện Ngoại giao

Email: thuhanglt@dav.edu.vn

TÓM TẮT:

Sự phát triển của robot tích hợp trí tuệ nhân tạo (AI) mở ra nhiều cơ hội nâng cao năng suất lao động và chất lượng dịch vụ, đồng thời đặt ra những thách thức mới về an toàn, trách nhiệm pháp lý và quản lý nhà nước. Trên cơ sở phân tích luật học kết hợp với nghiên cứu liên ngành và so sánh pháp luật, bài viết đánh giá khung pháp lý hiện hành của Việt Nam đối với robot trí tuệ nhân tạo theo cách tiếp cận quản lý vòng đời sản phẩm. Kết quả cho thấy, các quy định về trách nhiệm pháp lý, bảo vệ quyền lợi người tiêu dùng, tiêu chuẩn kỹ thuật và quản lý nhà nước đã bước đầu được hình thành nhưng còn phân tán, chưa đáp ứng yêu cầu quản lý đối với robot có mức độ tự chủ cao. Từ đó, bài viết đề xuất xây dựng bộ tiêu chuẩn, quy chuẩn kỹ thuật quốc gia; hoàn thiện cơ chế đánh giá sự phù hợp theo mức độ rủi ro; đồng thời ban hành tiêu chí pháp lý để phân biệt robot trí tuệ nhân tạo với robot thông minh thông thường nhằm xác định rõ trách nhiệm của các chủ thể liên quan.

Từ khóa: robot trí tuệ nhân tạo, an toàn robot, trách nhiệm pháp lý, tiêu chuẩn và quy chuẩn kỹ thuật, quản lý trí tuệ nhân tạo.

1. Đặt vấn đề

Sự phát triển của công nghệ tự động hóa, trí tuệ nhân tạo và công nghệ bán dẫn đang thúc đẩy việc nghiên cứu, phát triển và ứng dụng robot trong nhiều lĩnh vực của đời sống kinh tế - xã hội. Robot công nghiệp, robot dịch vụ, phương tiện tự hành và thiết bị bay không người lái ngày càng

được sử dụng rộng rãi, góp phần nâng cao năng suất lao động, cải thiện chất lượng dịch vụ và hỗ trợ con người trong những môi trường làm việc nguy hiểm. Tại Việt Nam, xu hướng này đã được xác định trong Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030, ban hành theo Quyết định số 127/QĐ-TTg của Thủ tướng Chính phủ năm 2021.

Tuy nhiên, khi được tích hợp trí tuệ nhân tạo, robot không còn chỉ thực hiện các thao tác được lập trình sẵn mà có thể học hỏi, thích ứng và đưa ra quyết định trong quá trình vận hành. Khả năng tự chủ cao hơn cũng làm gia tăng nguy cơ xuất hiện những hành vi ngoài dự kiến. Các nghiên cứu của OECD và Diễn đàn Kinh tế thế giới về quản trị trí tuệ nhân tạo đều cho thấy rủi ro của hệ thống AI không chỉ liên quan đến an toàn kỹ thuật mà còn đặt ra những vấn đề phức tạp về khả năng kiểm soát, quan hệ nhân quả và trách nhiệm của các chủ thể tham gia phát triển, sản xuất và vận hành hệ thống (OECD, 2024; World Economic Forum, 2024).

Ở Việt Nam, những vấn đề liên quan đến robot trí tuệ nhân tạo hiện được điều chỉnh thông qua nhiều nhóm quy định khác nhau. Luật Tiêu chuẩn và Quy chuẩn kỹ thuật năm 2006 (Quốc hội, 2006) và Luật Chất lượng sản phẩm, hàng hóa năm 2007 (Quốc hội, 2007) tạo cơ sở cho việc xây dựng yêu cầu kỹ thuật, đánh giá chất lượng và bảo đảm an toàn sản phẩm. Luật Bảo vệ quyền lợi người tiêu dùng năm 2023 (Quốc hội, 2023) quy định trách nhiệm đối với hàng hóa có khuyết tật, trong khi Luật Trí tuệ nhân tạo năm 2025 bước đầu đặt nền móng cho việc quản lý hoạt động nghiên cứu, phát triển và ứng dụng AI. Tuy vậy, các quy định này còn phân tán, chưa hình thành một khuôn khổ pháp lý thống nhất đối với robot trí tuệ nhân tạo, nhất là về phân loại theo mức độ tự chủ và rủi ro, đánh giá sự phù hợp trước khi đưa sản phẩm ra thị trường và phân định trách nhiệm giữa các chủ thể trong toàn bộ vòng đời của hệ thống.

Từ thực trạng đó, bài viết tập trung phân tích đặc điểm và rủi ro của robot trí tuệ nhân tạo, đánh giá mức độ đáp ứng của pháp luật hiện hành, đồng thời đề xuất hoàn thiện khung pháp lý theo hướng quản lý toàn bộ vòng đời sản phẩm và phân loại theo mức độ rủi ro. Đây cũng là cách tiếp cận được thể hiện rõ trong Đạo luật

Trí tuệ nhân tạo của Liên minh châu Âu (AI Act) và các nguyên tắc quản trị AI có trách nhiệm của OECD (European Parliament and Council of the European Union, 2024; OECD, 2024).

2. Phương pháp nghiên cứu

Bài viết sử dụng phương pháp phân tích luật học để hệ thống hóa, giải thích và đánh giá các quy định của pháp luật Việt Nam liên quan đến robot trí tuệ nhân tạo, bao gồm pháp luật về bảo vệ quyền lợi người tiêu dùng, trách nhiệm bồi thường thiệt hại, an toàn sản phẩm, tiêu chuẩn và quy chuẩn kỹ thuật, quản lý nhà nước và xử lý tội phạm công nghệ cao.

Bên cạnh đó, phương pháp phân tích, tổng hợp và nghiên cứu liên ngành được vận dụng nhằm làm rõ đặc điểm công nghệ, mức độ tự chủ và các rủi ro pháp lý của robot trí tuệ nhân tạo trong toàn bộ vòng đời của hệ thống, từ thiết kế, phát triển, thử nghiệm, sản xuất đến vận hành và xử lý sự cố. Trên cơ sở đó, bài viết đánh giá tính đầy đủ và khả năng áp dụng của các quy định hiện hành theo cách tiếp cận quản lý dựa trên vòng đời sản phẩm và mức độ rủi ro.

Ngoài ra, phương pháp so sánh pháp luật được sử dụng để đối chiếu khung pháp lý của Việt Nam với cách tiếp cận quản trị trí tuệ nhân tạo theo mức độ rủi ro của Liên minh châu Âu và các nguyên tắc quản trị AI của OECD, qua đó đề xuất những giải pháp phù hợp với điều kiện pháp lý và năng lực quản lý của Việt Nam.

3. Robot trí tuệ nhân tạo: Cơ hội, rủi ro và yêu cầu quản lý pháp lý

3.1. Sự phát triển của robot và bước ngoặt từ trí tuệ nhân tạo

Robot là hệ thống cơ điện tử được thiết kế để hỗ trợ hoặc thay thế con người trong những công việc nặng nhọc, lặp lại hoặc đòi hỏi độ chính xác cao. Về cấu tạo, robot thường gồm bộ phận cơ khí, cảm biến, bộ điều khiển, cơ cấu chấp hành

và phần mềm điều khiển. Đối với robot công nghiệp và hệ thống robot tích hợp, các yêu cầu về an toàn đã được quy định trong bộ tiêu chuẩn TCVN ISO 10218:2024 (Bộ Khoa học và Công nghệ, 2024a, 2024b).

Trong giai đoạn đầu, robot chủ yếu được sử dụng trong sản xuất công nghiệp dưới dạng tay máy, robot hàn, robot lắp ráp, dây chuyền tự động hoặc phương tiện dẫn đường tự động (AGV), qua đó góp phần nâng cao năng suất, tăng độ chính xác và giảm rủi ro cho người lao động. Cùng với sự phát triển của công nghệ số, phạm vi ứng dụng của robot ngày càng mở rộng sang y tế, dịch vụ, gia đình, phương tiện tự hành và một số hoạt động của khu vực công.

Điểm khác biệt quan trọng của robot tích hợp trí tuệ nhân tạo không nằm ở cấu tạo cơ khí mà ở năng lực xử lý thông tin. Khác với robot truyền thống vận hành theo chương trình được thiết lập trước, robot trí tuệ nhân tạo có thể thu thập dữ liệu, học hỏi, thích ứng và đưa ra quyết định với mức độ tự chủ cao hơn. Sự thay đổi này làm phát sinh những yêu cầu mới về an toàn, khả năng kiểm soát và trách nhiệm pháp lý, phù hợp với cách tiếp cận về khái niệm và quản lý hệ thống AI trong các tiêu chuẩn ISO/IEC hiện hành (ISO/IEC, 2022; ISO/IEC, 2023).

3.2. Khi robot trở nên khó lường

Việc tích hợp trí tuệ nhân tạo giúp robot hoạt động linh hoạt và thích ứng tốt hơn với môi trường, nhưng đồng thời cũng làm gia tăng các rủi ro về an toàn. Các nghiên cứu về quản trị AI đã cảnh báo robot phục vụ, robot hình người và phương tiện tự hành có thể phát sinh hành vi ngoài dự kiến khi hệ thống mất kiểm soát hoặc đưa ra quyết định không phù hợp với bối cảnh vận hành (World Economic Forum, 2024).

Khác với robot truyền thống, robot trí tuệ nhân tạo có thể thay đổi hành vi dựa trên dữ liệu và quá trình học hỏi. Rủi ro có thể xuất phát từ

lỗi cảm biến, sai lệch dữ liệu, thuật toán, bản cập nhật phần mềm hoặc sự tương tác giữa các thành phần trong hệ thống. Theo OECD, mức độ tự chủ càng cao thì việc xác định nguyên nhân sự cố, quan hệ nhân quả và trách nhiệm của các chủ thể liên quan càng trở nên phức tạp (OECD, 2024).

Vì vậy, việc quản lý robot trí tuệ nhân tạo không thể chỉ tập trung vào xử lý hậu quả sau khi sự cố xảy ra mà cần được thực hiện trong toàn bộ vòng đời của sản phẩm, từ thiết kế, phát triển, thử nghiệm và đánh giá sự phù hợp đến vận hành, cập nhật, giám sát và thu hồi khi cần thiết.

3.3. Tiêu chí nhận diện robot trí tuệ nhân tạo có mức độ rủi ro cao

Không phải mọi sản phẩm ứng dụng trí tuệ nhân tạo đều cần áp dụng cơ chế quản lý pháp lý đặc biệt. Dựa trên mức độ tự chủ và rủi ro, có thể phân thành 3 nhóm: (i) thiết bị cơ điện tử hoạt động theo chương trình cố định; (ii) robot thông minh có khả năng thích ứng trong phạm vi thiết kế; và (iii) robot tích hợp trí tuệ nhân tạo có khả năng học hỏi, tự điều chỉnh hành vi và đưa ra quyết định với mức độ tự chủ cao. Theo cách tiếp cận về khái niệm và đặc tính của hệ thống AI trong ISO/IEC 22989:2022, nhóm thứ ba là đối tượng đặt ra nhiều thách thức nhất đối với pháp luật.

Việc nhận diện robot cần áp dụng cơ chế quản lý đặc biệt nên dựa trên các tiêu chí như khả năng thu thập và xử lý dữ liệu; mức độ tự chủ trong quá trình ra quyết định; khả năng học hỏi hoặc cập nhật thuật toán; phạm vi ảnh hưởng đến con người, tài sản và môi trường; cũng như khả năng giám sát và can thiệp của con người. Đây cũng là những yếu tố phù hợp với cách tiếp cận quản trị AI dựa trên mức độ rủi ro được thể hiện trong AI Act (European Parliament and Council of the European Union, 2024).

Như vậy, phạm vi điều chỉnh của pháp luật không nên được xác định đơn thuần dựa trên việc sản phẩm có sử dụng trí tuệ nhân tạo hay không, mà cần căn cứ vào mức độ tự chủ và rủi ro của hệ thống. Cách tiếp cận này phù hợp với định hướng quản trị AI có trách nhiệm của OECD và mô hình phân loại rủi ro của Liên minh châu Âu, đồng thời tạo cơ sở để đánh giá và hoàn thiện khung pháp lý đối với robot trí tuệ nhân tạo tại Việt Nam (OECD, 2024; European Parliament and Council of the European Union, 2024).

4. Đánh giá khung pháp lý Việt Nam đối với robot trí tuệ nhân tạo

Không phải mọi robot hoặc sản phẩm ứng dụng trí tuệ nhân tạo đều cần áp dụng cơ chế quản lý pháp lý riêng. Đối tượng cần được ưu tiên điều chỉnh là các robot có mức độ tự chủ cao, có khả năng học hỏi, đưa ra quyết định và tiềm ẩn nguy cơ ảnh hưởng đến tính mạng, sức khỏe, tài sản hoặc môi trường.

Từ góc độ pháp lý, vấn đề cốt lõi không nằm ở tên gọi hay mức độ hiện đại của sản phẩm mà ở việc toàn bộ vòng đời của robot trí tuệ nhân tạo đã được đặt dưới một cơ chế quản lý đầy đủ hay chưa. Vì vậy, việc đánh giá khung pháp lý cần được thực hiện theo từng giai đoạn của vòng đời sản phẩm, từ nghiên cứu, thiết kế, phát triển và thử nghiệm đến sản xuất, đưa ra thị trường, vận hành, giám sát và xử lý sự cố. Đây cũng là cách tiếp cận được thể hiện trong AI Act và các khuyến nghị của OECD, qua đó giúp nhận diện những quy định đã có, những khoảng trống còn tồn tại và định hướng hoàn thiện pháp luật (European Parliament and Council of the European Union, 2024; OECD, 2024).

4.1. Không thừa nhận tư cách chủ thể pháp luật của robot trí tuệ nhân tạo

Một trong những vấn đề còn nhiều tranh luận là liệu robot trí tuệ nhân tạo có mức độ tự chủ cao có nên được thừa nhận là chủ thể của pháp

luật hay không. Nhiều nghiên cứu cho rằng mức độ tự chủ của hệ thống AI không đồng nhất với tư cách chủ thể pháp luật; khả năng học hỏi, thích ứng và đưa ra quyết định của robot chỉ phản ánh mức độ tự chủ về công nghệ, không đồng nghĩa với việc có năng lực pháp luật hoặc năng lực chịu trách nhiệm như cá nhân và pháp nhân (Bertolini, 2020; Pasquale, 2020). Việc hệ thống có thể hoạt động độc lập vì thế không làm phát sinh địa vị pháp lý của một chủ thể trong quan hệ pháp luật.

Mặt khác, việc trao tư cách chủ thể cho robot có thể làm phân tán trách nhiệm của nhà phát triển, nhà sản xuất, nhà cung cấp, chủ sở hữu và người vận hành. Điều này không chỉ gây khó khăn trong việc xác định trách nhiệm khi xảy ra thiệt hại mà còn làm giảm trách nhiệm phòng ngừa rủi ro ngay từ giai đoạn thiết kế, phát triển và triển khai sản phẩm.

Do đó, robot trí tuệ nhân tạo nên được xác định là đối tượng của quan hệ pháp luật dưới dạng sản phẩm hoặc phương tiện công nghệ đặc biệt. Mọi trách nhiệm phát sinh từ quá trình nghiên cứu, phát triển, sản xuất, cung cấp và vận hành phải được quy về các cá nhân, pháp nhân có liên quan. Cách tiếp cận này cũng phù hợp với các nguyên tắc về chủ thể và trách nhiệm dân sự được quy định trong Bộ luật Dân sự năm 2015 (Quốc hội, 2015a).

4.2. Hoàn thiện chế định trách nhiệm pháp lý đối với robot trí tuệ nhân tạo

4.2.1. Trách nhiệm đối với người tiêu dùng

Khi robot trí tuệ nhân tạo được đưa ra thị trường dưới dạng hàng hóa tiêu dùng, trách nhiệm của nhà sản xuất, nhà nhập khẩu và tổ chức, cá nhân kinh doanh được điều chỉnh trước hết bởi Luật Bảo vệ quyền lợi người tiêu dùng năm 2023 (Quốc hội, 2023). Theo đó, các chủ thể này phải chịu trách nhiệm đối với sản phẩm có khuyết tật, không bảo đảm an toàn hoặc không

cung cấp đầy đủ thông tin, cảnh báo và hướng dẫn sử dụng gây thiệt hại cho người tiêu dùng.

Tuy nhiên, các quy định hiện hành chủ yếu được xây dựng đối với hàng hóa truyền thống. Đối với robot trí tuệ nhân tạo, khuyết tật của sản phẩm có thể phát sinh không chỉ từ lỗi cơ khí mà còn từ thuật toán, dữ liệu huấn luyện, bản cập nhật phần mềm hoặc sự thay đổi hành vi của hệ thống trong quá trình vận hành. Điều này cho thấy, khái niệm “khuyết tật của sản phẩm” cần tiếp tục được cụ thể hóa trong pháp luật chuyên ngành để phù hợp với đặc thù của robot trí tuệ nhân tạo, đồng thời đặt ra yêu cầu hoàn thiện các quy định của Luật Chất lượng sản phẩm, hàng hóa năm 2007.

4.2.2. Trách nhiệm theo hợp đồng và ngoài hợp đồng

Thiệt hại do robot trí tuệ nhân tạo gây ra có thể xuất phát từ nhiều nguyên nhân như lỗi phần cứng, cảm biến, thuật toán, dữ liệu, bản cập nhật phần mềm hoặc sai sót trong quá trình vận hành. Vì vậy, trách nhiệm pháp lý không thể chỉ quy về một chủ thể mà cần được xác định tương ứng với vai trò của nhà phát triển, nhà sản xuất, nhà cung cấp phần mềm hoặc dữ liệu, nhà phân phối, chủ sở hữu, đơn vị vận hành hoặc người sử dụng theo quy định về trách nhiệm hợp đồng và bồi thường thiệt hại ngoài hợp đồng.

Các quy định của Bộ luật Dân sự năm 2015 về trách nhiệm hợp đồng và bồi thường thiệt hại ngoài hợp đồng vẫn có thể áp dụng đối với phần lớn trường hợp phát sinh. Tuy nhiên, đối với robot có mức độ tự chủ cao, việc xác định lỗi và quan hệ nhân quả sẽ phức tạp hơn, đòi hỏi phải bổ sung cơ chế lưu trữ dữ liệu, truy xuất và giải trình quá trình ra quyết định của hệ thống để làm căn cứ xác định trách nhiệm. Đây cũng là yêu cầu được nhấn mạnh trong AI Act với các hệ thống AI có mức độ rủi ro cao (European Parliament and Council of the European Union, 2024).

4.2.3. Trách nhiệm của nhà phát triển và nhà sản xuất

Đối với robot trí tuệ nhân tạo, trách nhiệm của nhà phát triển và nhà sản xuất cần được xác lập trong toàn bộ vòng đời của sản phẩm, từ thiết kế, đánh giá rủi ro, kiểm thử trước khi đưa vào sử dụng đến giám sát sau khi đưa ra thị trường, cập nhật phần mềm, khắc phục lỗi hỏng, cảnh báo và thu hồi sản phẩm khi phát hiện nguy cơ mất an toàn.

Đây cũng là định hướng được OECD và AI Act khuyến nghị đối với các hệ thống AI có mức độ rủi ro cao. Tuy nhiên, hiệu quả của cơ chế này không chỉ phụ thuộc vào quy định pháp luật mà còn đòi hỏi hệ thống đánh giá sự phù hợp, năng lực giám định kỹ thuật và cơ chế điều tra sự cố đủ khả năng hỗ trợ việc xác định trách nhiệm trong thực tiễn (OECD, 2024; European Parliament and Council of the European Union, 2024).

4.2.4. Nghiên cứu cơ chế trách nhiệm đối với robot trí tuệ nhân tạo có mức độ rủi ro cao

Bộ luật Dân sự đã quy định chế định bồi thường thiệt hại do nguồn nguy hiểm cao độ gây ra đối với một số loại tài sản có khả năng gây thiệt hại lớn. Đối với robot trí tuệ nhân tạo có mức độ tự chủ và rủi ro cao, có thể nghiên cứu áp dụng cơ chế trách nhiệm tương tự hoặc xây dựng cơ chế trách nhiệm nghiêm ngặt phù hợp với đặc thù của loại sản phẩm này.

Tiêu chí áp dụng không nên dựa vào việc sản phẩm có sử dụng trí tuệ nhân tạo hay không mà cần căn cứ vào mức độ tự chủ, khả năng gây thiệt hại và phạm vi tác động của hệ thống. Cách tiếp cận này vừa bảo đảm quản lý rủi ro hiệu quả, vừa tránh mở rộng trách nhiệm pháp lý đối với các sản phẩm có nguy cơ thấp, đồng thời tương đồng với xu hướng phân loại trách nhiệm theo mức độ rủi ro được đề cập trong một số nghiên cứu về trách nhiệm dân sự đối với robot tự chủ (Reille, 2021).

4.3. Tiêu chuẩn, quy chuẩn và cơ chế kiểm định đối với robot trí tuệ nhân tạo

4.3.1. Hệ thống tiêu chuẩn hiện hành

Tiêu chuẩn và quy chuẩn kỹ thuật là cơ sở bảo đảm an toàn và đánh giá sự phù hợp đối với robot trí tuệ nhân tạo trước khi đưa vào sử dụng. Việt Nam đã từng bước xây dựng hệ thống tiêu chuẩn quốc gia về robot và AI trên cơ sở hài hòa với bộ tiêu chuẩn ISO 10218 về robot công nghiệp và các tiêu chuẩn ISO/IEC về trí tuệ nhân tạo. Các tiêu chuẩn hiện hành chủ yếu điều chỉnh an toàn robot công nghiệp, robot hợp tác, thuật ngữ, phương pháp thử nghiệm và một số yêu cầu về quản lý vòng đời, chất lượng và tính đáng tin cậy của hệ thống AI (Bộ Khoa học và Công nghệ, 2024a, 2024b; ISO/IEC, 2023).

Mặc dù tạo nền tảng quan trọng cho hoạt động nghiên cứu, sản xuất và ứng dụng, các tiêu chuẩn này vẫn chủ yếu điều chỉnh từng lĩnh vực hoặc từng thành phần công nghệ riêng lẻ. Trong khi đó, robot trí tuệ nhân tạo là hệ thống tích hợp giữa cơ khí, điện tử, phần mềm và trí tuệ nhân tạo nên chưa có một khuôn khổ kỹ thuật thống nhất để đánh giá toàn bộ hệ thống.

4.3.2. Khoảng trống trong hệ thống tiêu chuẩn và cơ chế đánh giá sự phù hợp

Đặc điểm tích hợp giữa phần cứng, phần mềm, thuật toán, dữ liệu và kết nối mạng khiến mức độ an toàn của robot trí tuệ nhân tạo phụ thuộc vào sự vận hành đồng bộ của toàn bộ hệ thống, thay vì từng thành phần riêng lẻ.

Tuy nhiên, hệ thống tiêu chuẩn hiện hành vẫn chủ yếu điều chỉnh riêng đối với robot hoặc trí tuệ nhân tạo, chưa có các yêu cầu kỹ thuật tích hợp cũng như cơ chế đánh giá sự phù hợp trong toàn bộ vòng đời của robot trí tuệ nhân tạo. Mặc dù Việt Nam đã ban hành một số tiêu chuẩn quốc gia hài hòa với ISO về robot công nghiệp và quản lý hệ thống AI, các tiêu chuẩn này mới tập trung vào từng thành phần kỹ thuật hoặc từng lĩnh vực

ứng dụng, chưa hình thành khung đánh giá tổng thể đối với robot trí tuệ nhân tạo như một hệ thống tích hợp theo các tiêu chuẩn quốc gia và tiêu chuẩn quốc tế hiện hành (Bộ Khoa học và Công nghệ, 2024a, 2024b; ISO/IEC, 2023).

Như vậy, khoảng trống lớn nhất hiện nay không phải là thiếu tiêu chuẩn đối với robot hoặc trí tuệ nhân tạo, mà là thiếu cơ chế đánh giá tổng thể đối với một hệ thống tích hợp có khả năng tự học và thay đổi hành vi trong quá trình hoạt động.

4.3.3. Hoàn thiện bộ tiêu chuẩn, quy chuẩn kỹ thuật và cơ chế kiểm định theo mức độ rủi ro

Từ những khoảng trống của hệ thống pháp luật hiện hành, cần nghiên cứu xây dựng bộ tiêu chuẩn, quy chuẩn kỹ thuật quốc gia và cơ chế đánh giá sự phù hợp dành riêng cho robot trí tuệ nhân tạo theo hướng phân loại theo mức độ rủi ro.

Hệ thống tiêu chuẩn và quy chuẩn cần quy định các yêu cầu về mức độ tự chủ và rủi ro của robot; an toàn cơ khí, điện và chức năng; an toàn dữ liệu và an ninh mạng; khả năng giám sát, can thiệp và dừng khẩn cấp; khả năng truy xuất, giải trình quyết định của hệ thống; quản lý cập nhật phần mềm; thử nghiệm, giám sát sau khi đưa vào sử dụng và cơ chế cảnh báo, thu hồi khi phát hiện nguy cơ mất an toàn.

Bên cạnh đó, cần xây dựng cơ chế đánh giá sự phù hợp tương ứng với từng nhóm rủi ro. Theo đó, sản phẩm có rủi ro thấp có thể áp dụng cơ chế tự đánh giá hoặc tự công bố; nhóm rủi ro trung bình cần được đánh giá bởi tổ chức đánh giá sự phù hợp; còn robot có mức độ tự chủ hoặc rủi ro cao cần được kiểm định hoặc thẩm định trước khi đưa vào sử dụng và tiếp tục được giám sát trong quá trình vận hành. Cách tiếp cận này tương đồng với mô hình quản trị AI theo mức độ rủi ro của AI Act (European Parliament and Council of the European Union, 2024).

4.4. Cơ quan quản lý và năng lực thực thi

4.4.1. Xác định cơ quan đầu mối và cơ chế phối hợp

Hiệu quả của khung pháp lý đối với robot trí tuệ nhân tạo phụ thuộc không chỉ vào nội dung quy định mà còn vào cơ chế tổ chức thực thi. Luật Trí tuệ nhân tạo năm 2025 đã bước đầu xác lập cơ chế quản lý đối với hoạt động nghiên cứu, phát triển, cung cấp, triển khai và sử dụng hệ thống AI (Quốc hội, 2025). Tuy nhiên, do robot trí tuệ nhân tạo là hệ thống tích hợp nhiều lĩnh vực công nghệ, việc quản lý không thể chỉ giao cho một cơ quan duy nhất.

Do đó, cần xác định cơ quan đầu mối về quản lý trí tuệ nhân tạo, đồng thời thiết lập cơ chế phối hợp với cơ quan quản lý về tiêu chuẩn, chất lượng và các bộ quản lý chuyên ngành như giao thông, y tế, công nghiệp, quốc phòng và hàng không. Đối với robot có phạm vi ứng dụng liên ngành, cần quy định rõ trách nhiệm phối hợp, chia sẻ thông tin và xử lý sự cố nhằm bảo đảm tính thống nhất trong quản lý.

4.4.2. Xây dựng hệ thống đánh giá, kiểm định và phê duyệt theo mức độ rủi ro

Cùng với việc hoàn thiện hệ thống tiêu chuẩn, cần xây dựng cơ chế đánh giá sự phù hợp đối với robot trí tuệ nhân tạo dựa trên mức độ rủi ro. Theo đó, cần phân biệt rõ các hoạt động thử nghiệm, chứng nhận, kiểm định, thẩm định, cấp phép và giám sát sau khi sản phẩm được đưa vào sử dụng để áp dụng biện pháp quản lý phù hợp.

Không phải mọi robot trí tuệ nhân tạo đều cần được cấp phép trước khi đưa ra thị trường. Đối với sản phẩm có rủi ro thấp, có thể áp dụng cơ chế tự đánh giá hoặc công bố sự phù hợp; nhóm rủi ro trung bình cần được đánh giá bởi tổ chức độc lập; còn robot có mức độ tự chủ hoặc rủi ro cao cần được thẩm định hoặc cấp phép trước khi đưa vào sử dụng và chịu sự giám sát trong suốt quá trình vận hành.

Cách tiếp cận này vừa bảo đảm kiểm soát rủi ro, vừa hạn chế gánh nặng thủ tục hành chính và tạo điều kiện thúc đẩy đổi mới sáng tạo.

4.4.3. Nâng cao năng lực giám định và điều tra sự cố

Một trong những khó khăn lớn nhất khi xử lý sự cố liên quan đến robot trí tuệ nhân tạo là xác định nguyên nhân và chủ thể chịu trách nhiệm. Do robot là hệ thống tích hợp giữa phần cứng, phần mềm, dữ liệu, thuật toán và kết nối mạng, sự cố có thể phát sinh từ nhiều nguyên nhân đồng thời, làm cho việc xác định quan hệ nhân quả và trách nhiệm pháp lý trở nên phức tạp.

Vì vậy, cần xây dựng đội ngũ giám định viên có kiến thức liên ngành, đồng thời đầu tư phòng thử nghiệm, trung tâm kiểm định và hạ tầng mô phỏng phục vụ hoạt động giám định, tái hiện sự cố và phân tích nguyên nhân.

Bên cạnh đó, pháp luật cần quy định nghĩa vụ của nhà phát triển, nhà sản xuất và đơn vị vận hành trong việc lưu trữ nhật ký hoạt động, bảo toàn dữ liệu và chứng cứ điện tử nhằm phục vụ điều tra, giám định và giải quyết tranh chấp. OECD và AI Act cũng coi khả năng lưu giữ hồ sơ, truy xuất dữ liệu và giải trình hoạt động của hệ thống là những điều kiện quan trọng để bảo đảm trách nhiệm giải trình đối với AI có mức độ rủi ro cao (OECD, 2024; European Parliament and Council of the European Union, 2024).

5. Hoàn thiện pháp luật hình sự đối với hành vi liên quan đến robot trí tuệ nhân tạo

5.1. Hành vi can thiệp, xâm nhập và chiếm quyền điều khiển robot trí tuệ nhân tạo

Robot trí tuệ nhân tạo và các thiết bị kết nối mạng luôn tiềm ẩn nguy cơ bị xâm nhập hoặc chiếm quyền điều khiển trái phép. Khi sự cố xảy ra, nguyên nhân có thể xuất phát từ lỗi hồng bảo mật, lỗi phần mềm, sai sót trong vận hành hoặc hành vi tấn công có chủ đích, làm cho việc xác định trách nhiệm hình sự trở nên phức tạp.

Từ góc độ pháp luật hình sự, cần phân biệt 3 nhóm hành vi: (i) cố ý xâm nhập hoặc chiếm quyền điều khiển hệ thống để thực hiện hành vi phạm tội; (ii) vi phạm nghĩa vụ bảo đảm an toàn, an ninh mạng của nhà phát triển hoặc nhà sản xuất; và (iii) thiếu trách nhiệm trong quản lý, vận hành của chủ sở hữu hoặc đơn vị khai thác. Việc phân định rõ các nhóm hành vi này là cơ sở xác định trách nhiệm hình sự, đồng thời phân biệt với trách nhiệm dân sự và hành chính.

Mặc dù Bộ luật Hình sự năm 2015 (Quốc hội, 2015b), được sửa đổi, bổ sung năm 2017 (Quốc hội, 2017), đã có các quy định điều chỉnh tội phạm công nghệ cao, việc áp dụng đối với robot trí tuệ nhân tạo vẫn gặp khó khăn trong việc chứng minh quan hệ nhân quả giữa hành vi xâm nhập và hậu quả xảy ra. Vì vậy, ưu tiên hiện nay là hoàn thiện cơ chế thu thập chứng cứ số và giám định kỹ thuật hơn là bổ sung các tội danh mới.

5.2. Trách nhiệm pháp lý đối với hành vi che giấu hoặc phân loại sai robot trí tuệ nhân tạo

Một thách thức trong quản lý robot trí tuệ nhân tạo là việc nhà phát triển hoặc nhà sản xuất công bố không chính xác mức độ tự chủ hoặc chức năng của sản phẩm nhằm né tránh các nghĩa vụ về đánh giá sự phù hợp, kiểm định và giám sát.

Để khắc phục, cần xây dựng các tiêu chí pháp lý và kỹ thuật khách quan nhằm xác định mức độ tự chủ và mức độ rủi ro của robot, đồng thời thực hiện đánh giá thông qua cơ quan hoặc tổ chức đánh giá sự phù hợp có thẩm quyền. Đây cũng là cách tiếp cận được AI Act áp dụng đối với các hệ thống AI theo mức độ rủi ro (European Parliament and Council of the European Union, 2024).

Đối với hành vi cố ý cung cấp thông tin sai lệch hoặc che giấu rủi ro của sản phẩm, trước hết cần áp dụng các chế tài dân sự và hành chính theo pháp luật chuyên ngành. Chỉ khi hành vi có

lỗi cố ý, gây hậu quả nghiêm trọng hoặc đủ dấu hiệu cấu thành tội phạm mới đặt ra trách nhiệm hình sự.

5.3. Định hướng hoàn thiện pháp luật hình sự

Việc hoàn thiện pháp luật hình sự đối với robot trí tuệ nhân tạo cần kế thừa các quy định hiện hành, đồng thời đáp ứng yêu cầu của thực tiễn phát triển công nghệ.

Trong thời gian tới, cần tiếp tục rà soát khả năng áp dụng các tội danh hiện hành đối với hành vi xâm nhập, chiếm quyền điều khiển hoặc sử dụng robot trí tuệ nhân tạo làm công cụ thực hiện tội phạm; đồng thời hoàn thiện các quy định về lưu trữ dữ liệu, bảo toàn chứng cứ điện tử và nghĩa vụ bảo đảm an toàn của nhà phát triển, nhà sản xuất và đơn vị vận hành để tạo cơ sở xác định trách nhiệm khi xảy ra sự cố.

Việc hoàn thiện pháp luật cần dựa trên 3 nguyên tắc: (i) chỉ hình sự hóa các hành vi có tính nguy hiểm đáng kể cho xã hội; (ii) ưu tiên áp dụng các quy định hiện hành của Bộ luật Hình sự khi còn khả năng điều chỉnh; và (iii) bảo đảm sự thống nhất giữa pháp luật hình sự với pháp luật dân sự, hành chính và pháp luật chuyên ngành về trí tuệ nhân tạo. Cách tiếp cận này vừa bảo đảm hiệu quả phòng ngừa tội phạm, vừa tạo môi trường pháp lý ổn định cho hoạt động nghiên cứu, phát triển và ứng dụng robot trí tuệ nhân tạo. Định hướng này cũng phù hợp với xu hướng quản trị AI có trách nhiệm mà OECD và AI Act đang hướng tới (OECD, 2024; European Parliament and Council of the European Union, 2024).

6. Kết luận

Sự phát triển của robot trí tuệ nhân tạo với mức độ tự chủ ngày càng cao, khả năng học hỏi, thích ứng và tự đưa ra quyết định đang đặt ra những yêu cầu mới đối với bảo đảm an toàn, xác định trách nhiệm pháp lý và quản lý nhà nước.

So với robot truyền thống, đặc điểm này làm gia tăng tính khó dự báo trong quá trình vận hành, đồng thời phát sinh nhiều vấn đề pháp lý chưa được điều chỉnh đầy đủ.

Kết quả nghiên cứu cho thấy, Việt Nam đã bước đầu hình thành cơ sở pháp lý đối với robot trí tuệ nhân tạo thông qua các quy định về bảo vệ quyền lợi người tiêu dùng, trách nhiệm bồi thường thiệt hại, an toàn sản phẩm, tiêu chuẩn kỹ thuật và xử lý tội phạm công nghệ cao. Tuy nhiên, các quy định hiện hành còn phân tán và chưa hình thành một khung pháp lý thống nhất theo cách tiếp cận quản lý dựa trên mức độ rủi ro và toàn bộ vòng đời của sản phẩm.

Trên cơ sở đó, bài viết đề xuất hoàn thiện pháp luật theo hướng: (i) tiếp tục xác định trách nhiệm pháp lý đối với các cá nhân, pháp nhân

tham gia nghiên cứu, phát triển, sản xuất và vận hành robot trí tuệ nhân tạo; (ii) xây dựng hệ thống tiêu chuẩn, quy chuẩn kỹ thuật và cơ chế đánh giá sự phù hợp theo mức độ rủi ro; (iii) hoàn thiện cơ chế quản lý nhà nước, giám định và điều tra sự cố; và (iv) rà soát, hoàn thiện các quy định của pháp luật hình sự đối với các hành vi lợi dụng robot trí tuệ nhân tạo để thực hiện hành vi vi phạm pháp luật.

Việc hoàn thiện khung pháp lý không nhằm hạn chế đổi mới sáng tạo mà hướng tới xây dựng môi trường pháp lý an toàn, minh bạch và có khả năng kiểm soát rủi ro, qua đó tạo điều kiện cho robot trí tuệ nhân tạo được nghiên cứu, phát triển và ứng dụng một cách có trách nhiệm, phù hợp với yêu cầu phát triển kinh tế - xã hội của Việt Nam■

TÀI LIỆU THAM KHẢO:

- Bộ Khoa học và Công nghệ. (2024a). TCVN ISO 10218-1:2024 Robot và thiết bị robot - Yêu cầu an toàn đối với robot công nghiệp - Phần 1: Robot.
- Bộ Khoa học và Công nghệ. (2024b). TCVN ISO 10218-2:2024 Robot và thiết bị robot - Yêu cầu an toàn đối với robot công nghiệp - Phần 2: Hệ thống robot và tích hợp.
- Quốc hội. (2006). Luật Tiêu chuẩn và Quy chuẩn kỹ thuật số 68/2006/QH11 ngày 29/6/2006.
- Quốc hội. (2007). Luật Chất lượng sản phẩm, hàng hóa số 05/2007/QH12 ngày 21/11/2007.
- Quốc hội. (2015a). Bộ luật Dân sự số 91/2015/QH13 ngày 24/11/2015.
- Quốc hội. (2015b). Bộ luật Hình sự số 100/2015/QH13 ngày 27/11/2015.
- Quốc hội. (2017). Luật số 12/2017/QH14 ngày 20/6/2017 sửa đổi, bổ sung một số điều của Bộ luật Hình sự số 100/2015/QH13.
- Quốc hội. (2023). Luật Bảo vệ quyền lợi người tiêu dùng số 19/2023/QH15 ngày 20/6/2023.
- Quốc hội. (2025). Luật Trí tuệ nhân tạo số 134/2025/QH15 ngày 10/12/2025.
- Thủ tướng Chính phủ. (2021). Quyết định số 127/QĐ-TTg ngày 26/01/2021 phê duyệt Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030.
- Bertolini, A. (2020). Artificial Intelligence and Civil Liability. European Parliament.
- European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).
- ISO/IEC. (2022). ISO/IEC 22989:2022 Artificial intelligence - Concepts and terminology.
- ISO/IEC. (2023). ISO/IEC 42001:2023 Information technology - Artificial intelligence - Management system.
- OECD. (2024). Recommendation of the Council on Artificial Intelligence. Organisation for Economic Co-operation and Development.

Pasquale, F. (2020). *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Harvard University Press.

Reille, F. (2021). *Les robots autonomes et la responsabilité civile* (Mémoire de Master 2 recherche Droit privé général). Université Paris-Panthéon-Assas.

World Economic Forum. (2024). *Governance in the Age of Generative AI: A 360° Approach for Resilient Policy and Regulation*.

Ngày nhận bài: 5/5/2026

Ngày phản biện đánh giá và sửa chữa: 22/5/2026

Ngày chấp nhận đăng bài: 5/6/2026

LEGAL FRAMEWORK FOR ARTIFICIAL INTELLIGENCE ROBOTS IN VIETNAM: CURRENT STATUS AND RECOMMENDATIONS FOR IMPROVEMENT

● LE THI THU HANG

Faculty of International Law, Diplomatic Academy of Vietnam

Email: thuhanglt@dav.edu.vn

ABSTRACT:

The development of robots integrated with artificial intelligence (AI) offers significant opportunities to enhance labor productivity and service quality while also creating new challenges concerning safety, legal liability, and state administration. Using legal analysis combined with interdisciplinary research and comparative law, this paper evaluates Vietnam's current legal framework for artificial intelligence robots from a product lifecycle management perspective. The findings indicate that regulations on legal liability, consumer protection, technical standards, and state administration have initially been established but remain fragmented and insufficient to address the governance requirements of highly autonomous robots. Accordingly, the paper proposes developing national technical standards and regulations, improving risk-based conformity assessment mechanisms, and establishing legal criteria to distinguish artificial intelligence robots from conventional intelligent robots, thereby clearly determining the responsibilities of the relevant parties.

Keywords: artificial intelligence robots, robot safety, legal liability, technical standards and regulations, artificial intelligence governance.