

NGHĨA VỤ PHÁP LÝ CỦA NHÀ PHÁT TRIỂN VÀ NHÀ CUNG CẤP HỆ THỐNG TRÍ TUỆ NHÂN TẠO THEO PHÂN LOẠI RỦI RO

TRƯƠNG THỊ TƯỜNG VI*
ĐOÀN TRỌNG CHÍNH**

Bài viết phân tích các thách thức pháp lý trong việc xác định nghĩa vụ của nhà phát triển và nhà cung cấp hệ thống trí tuệ nhân tạo (AI). Trọng tâm của bài viết là làm rõ các nguyên tắc chung trong quản lý rủi ro AI; xung đột lợi ích và thiếu tính khách quan trong phân loại rủi ro ảnh hưởng đến việc xác định nghĩa vụ pháp lý của nhà phát triển, nhà cung cấp; trách nhiệm pháp lý chưa phù hợp với bối cảnh Việt Nam khi sử dụng các hệ thống AI từ nhà phát triển, nhà cung cấp nước ngoài. Trên cơ sở đó, đề xuất một số định hướng và giải pháp nhằm hoàn thiện khung pháp luật liên quan.

Từ khóa: Nghĩa vụ pháp lý; nhà phát triển; nhà cung cấp; AI; phân loại rủi ro.

This article analyzes legal challenges in determining the obligations of artificial intelligence (AI) system developers and providers. It clarifies general principles of AI risk management; the effects of conflicts of interest and a lack of objectivity in risk classification on determining developers' and providers' legal obligations; and the inadequacy of existing legal liability in the Vietnamese context when foreign entities develop or provide AI systems that are used. Accordingly, the article proposes several directions and solutions to improve the relevant legal framework.

Keywords: Legal obligations; developers; providers; AI; risk classification.

NGÀY NHẬN: 18/8/2026 NGÀY PHẢN BIỆN, ĐÁNH GIÁ: 01/10/2026 NGÀY DUYỆT ĐĂNG: 19/10/2026

Email: vittt@hcmute.edu.vn

DOI: <https://doi.org/10.59394/qlnn.371.2026.1687>

1. Đặt vấn đề

Sự phát triển mạnh mẽ của AI đã giải phóng sức lao động của con người và hiệu quả kinh tế cũng được nâng cao, tuy nhiên, nhiều thách thức liên quan đến việc kiểm soát rủi ro của AI cũng đã được đặt ra. Mặc dù Việt Nam đã ban hành Luật Trí tuệ nhân tạo năm 2025 (Quốc hội, 2025) và Nghị định số 142/2026/NĐ-CP ngày 30/4/2026 (Chính

phủ, 2026), nhưng vấn đề nghĩa vụ pháp lý của nhà phát triển và nhà cung cấp hệ thống AI theo cơ chế phân loại rủi ro đặt trong mối quan hệ với các chủ thể khác trong hệ sinh

* TS, Trường Đại học Công nghệ Kỹ thuật TP. Hồ Chí Minh

** TS, Trường Đại học Công nghệ Kỹ thuật TP. Hồ Chí Minh

thái AI cần được đánh giá lại để làm rõ trách nhiệm trong quản lý rủi ro, phân bổ trách nhiệm hợp lý trong hệ sinh thái AI và hoàn thiện khung pháp luật hiện hành, như: nguyên tắc chung trong quản lý rủi ro về hệ thống AI, xung đột lợi ích nếu có trong quá trình tự phân loại rủi ro và trách nhiệm pháp lý chưa phù hợp với bối cảnh Việt Nam khi sử dụng các hệ thống AI từ nhà phát triển, nhà cung cấp nước ngoài.

2. Cơ sở lý thuyết

2.1. Khái niệm

Điều 3 *Luật Trí tuệ nhân tạo* năm 2025 quy định AI là việc máy tính thực hiện các năng lực trí tuệ của con người, như: học tập, suy luận, nhận thức, phán đoán. Đây là khái niệm rộng và mang tính nền tảng nhưng không phải đối tượng bị điều chỉnh trực tiếp bằng nghĩa vụ pháp lý cụ thể.

Hệ thống AI là đối tượng quản lý trọng tâm của *Luật Trí tuệ nhân tạo* năm 2025. Theo khoản 2 Điều 3 của *Luật*, hệ thống AI là hệ thống dựa trên máy, được thiết kế để thực hiện các năng lực AI với các mức độ tự chủ khác nhau, có khả năng tự thích nghi sau khi được triển khai; căn cứ vào các mục tiêu được xác định rõ ràng hoặc được hình thành một cách ngầm định, hệ thống suy luận từ dữ liệu đầu vào để tạo ra các đầu ra như dự đoán, nội dung, khuyến nghị hoặc quyết định có thể gây ảnh hưởng tới môi trường vật lý hoặc môi trường số.

Theo Schuett (2025), cần phân biệt ba khái niệm: trí tuệ nhân tạo (AI), mô hình AI và hệ thống AI. AI là khái niệm công nghệ bao quát, bao gồm nhiều hướng tiếp cận như học máy, học sâu và robot. Mô hình AI là thành phần được thiết kế và huấn luyện trên dữ liệu để thực hiện các nhiệm vụ cụ thể như dự đoán, phân loại, nhận diện hình ảnh hoặc tạo sinh văn bản. Trong khi đó, hệ thống AI là một chỉnh thể được tích hợp từ mô hình AI và các thành phần liên quan, như: phần mềm, dữ liệu vận hành, giao diện và cơ chế kiểm soát nhằm thực hiện chức năng trong

môi trường thực tế. Chẳng hạn, Google Dịch, xe tự lái và hệ thống chấm công bằng nhận diện khuôn mặt đều là những ví dụ về hệ thống AI. Sự phân biệt này có ý nghĩa đối với quản lý nhà nước, bởi việc đánh giá và kiểm soát cần xem xét không chỉ mô hình AI mà còn toàn bộ hệ thống, mục đích sử dụng và tác động của nó trong thực tiễn.

2.2. Những đặc trưng rủi ro do hệ thống AI tạo ra

Về bản chất, những rủi ro từ sản phẩm AI là khó dự đoán và khó kiểm soát (Schuett, 2025), trong khi các lĩnh vực sản xuất khác, như: xây dựng, tạo sản phẩm, dịch vụ... thì dễ dự đoán hơn vì đầu vào có thể xác định được, có thể kiểm tra tiêu chuẩn thử nghiệm theo quy trình và đã xác định, có thể truy ngược lại lỗi trong quy trình. Các lĩnh vực kinh tế khác thường có tính ổn định tương đối và rủi ro thường không tự thay đổi theo thời gian nhưng một số loại hệ thống AI thì lại có khả năng tự học và đưa ra quyết định ngoài dự kiến của con người (Macrae, 2022).

Do vậy, hậu quả pháp lý của những đặc điểm này là không thể áp dụng nguyên tắc kiểm tra trước khi đưa ra thị trường một cách đầy đủ như các sản phẩm truyền thống. Do vậy, các rủi ro cần được phân loại và hỗ trợ quản lý thông qua các quy định pháp luật tương ứng từng cấp độ để có phương án xử lý cho từng rủi ro dự kiến phát sinh. Bên cạnh đó, đặc điểm này yêu cầu pháp luật cần có quy định nghĩa vụ giám sát hệ thống AI trong suốt vòng đời hoạt động, yêu cầu báo cáo định kỳ về hiệu suất và sự cố, thiết lập cơ quan chuyên trách có năng lực kỹ thuật để giám sát và quy định về hệ thống cảnh báo sớm khi phát hiện hành vi bất thường, cần xây dựng quy định có khả năng cập nhật nhanh chóng, tránh tính cứng nhắc của luật truyền thống.

2.3. Hai xu hướng quản lý rủi ro hệ thống AI hiện nay

Hệ thống AI vốn dĩ mang tính chất kỹ thuật - xã hội (Macrae, 2022), có nghĩa là

chúng bị ảnh hưởng bởi các quy tắc kỹ thuật và hành vi xã hội của con người nên rủi ro và lợi ích của AI có thể phát sinh từ sự tương tác giữa các khía cạnh kỹ thuật kết hợp với các yếu tố xã hội liên quan đến cách sử dụng AI. Điểm (2) Điều 3 *Luật Trí tuệ nhân tạo* của Liên minh châu Âu (EU) định nghĩa “rủi ro” có nghĩa là sự kết hợp giữa xác suất xảy ra thiệt hại và mức độ nghiêm trọng của thiệt hại đó. *Luật Trí tuệ nhân tạo* năm 2025 của Việt Nam đã nêu nội dung phân cấp quản lý hệ thống AI theo mức độ rủi ro nhưng chưa đưa ra được định nghĩa cụ thể về rủi ro trong AI.

Phân cấp rủi ro trong quản lý AI là phương pháp phân loại AI thành các mức độ rủi ro khác nhau dựa trên mức độ tiềm năng gây hại cho con người, xã hội, quyền cơ bản hoặc an ninh. Theo định hướng pháp luật về AI của Việt Nam, các sản phẩm và hệ thống AI được phân loại thành ba cấp độ rủi ro khác nhau là rủi ro cao, rủi ro trung bình và rủi ro thấp (khoản 1 Điều 9 *Luật Trí tuệ nhân tạo* năm 2025). Mục đích chính của việc phân cấp này là áp dụng các yêu cầu quản lý phù hợp và tương xứng với mức độ rủi ro của từng hệ thống AI - mức độ rủi ro càng cao thì các yêu cầu về kiểm soát, minh bạch và trách nhiệm càng nghiêm ngặt. Với sự tiếp cận dựa trên rủi ro này, nguồn lực đã được tập trung vào những hệ thống AI có khả năng gây ảnh hưởng đáng kể đến quyền con người, an toàn, sức khỏe hoặc các lợi ích xã hội quan trọng bởi cơ quan quản lý. Đồng thời, nhờ phương thức quản lý theo mức độ rủi ro, sự cân bằng giữa mục tiêu bảo vệ xã hội trước các tác động tiêu cực tiềm ẩn của AI và nhu cầu thúc đẩy phát triển công nghệ cũng được bảo đảm.

Các quốc gia quản lý AI dựa trên đánh giá mức độ rủi ro như EU (Regulation - EU, 2024), Việt Nam sử dụng phương pháp tiếp cận dựa trên rủi ro để điều chỉnh các hệ thống AI gắn liền với các khung pháp lý hiện hành của các quốc gia đó. Phương pháp tiếp cận dựa trên rủi ro được tập trung vào việc

xác định và giải quyết trước các mối đe dọa nghiêm trọng nhất đối với người dùng, thay vì tất cả các rủi ro đều được xử lý với mức độ cần thiết như nhau. Với phương pháp này thì các rủi ro cao được ưu tiên giải quyết trước, sau đó sẽ chuyển sang xử lý các vấn đề có mức độ ưu tiên thấp hơn. Trong khi đó, Anh đưa ra các nguyên tắc để kiểm soát rủi ro chứ không phải là các quy định cụ thể (UK Government, 2023). Năm 2022, Hoa Kỳ đưa ra 5 nguyên tắc bảo vệ người dân trước hệ thống AI là an toàn, không phân biệt, bảo vệ dữ liệu riêng tư, minh bạch - giải thích được và có lựa chọn thay thế do con người kiểm soát (White House Office of Science and Technology Policy - OSTP, 2022). Pháp luật Hoa Kỳ sự thiếu vắng một đạo luật điều chỉnh AI trên toàn liên bang nhưng có các quy định trong luật của từng tiểu bang, nghĩa là các quy tắc điều chỉnh AI không được nêu cụ thể trong văn bản luật nào mà chỉ phác thảo định hướng chiến lược trong các sắc lệnh hành pháp thay vì một đạo luật của cơ quan lập pháp bao quát, thống nhất. Cách tiếp cận linh hoạt này có thể giúp thúc đẩy đổi mới sáng tạo nhưng cũng có thể tạo ra sự can thiệp khác nhau giữa các cơ quan khiến việc người dùng khó dự đoán hướng xử lý về pháp lý.

3. Các vấn đề pháp lý và giải pháp

Một là, xác định nguyên tắc chung trong quản trị hệ thống AI với nghĩa vụ pháp lý của nhà phát triển và nhà cung cấp.

Cách tiếp cận dựa trên rủi ro được xem là một cơ chế quan trọng trong pháp luật về AI của Việt Nam. Theo cách tiếp cận này, mức độ điều chỉnh pháp lý được thiết kế tương ứng với mức độ rủi ro mà hệ thống AI có thể gây ra đối với xã hội, qua đó, tạo ra sự cân bằng giữa yêu cầu quản lý nhà nước và nhu cầu thúc đẩy đổi mới sáng tạo. Tuy nhiên, đối với các hệ thống AI được xếp vào nhóm có rủi ro cao thì thường phức tạp hơn do phải đáp ứng các yêu cầu pháp lý chặt chẽ về đánh giá, giám sát và minh bạch. Trên cơ sở phân loại ba cấp độ rủi ro khác nhau, các

quyền và nghĩa vụ pháp lý của nhà phát triển, nhà cung cấp hệ thống AI cần được xác định tương ứng với mức độ rủi ro mà nó có thể gây ra. Nói cách khác, mức độ rủi ro càng cao thì yêu cầu về nghĩa vụ pháp lý đối với chủ thể sáng tạo AI càng nghiêm ngặt, đặc biệt trong các lĩnh vực, như: giám sát, đánh giá rủi ro, bảo đảm tính minh bạch, cũng như trách nhiệm giải trình trong quá trình vận hành hệ thống.

Về phương diện lý luận pháp lý, cách tiếp cận này phản ánh sự vận dụng nguyên tắc cân xứng (Comella, 2025) - một nguyên tắc nền tảng của nhà nước pháp quyền vào lĩnh vực quản lý các công nghệ mới. Theo đó, các biện pháp quản lý cần được thiết kế sao cho tương xứng với mức độ rủi ro mà đối tượng điều chỉnh có thể gây ra. Mục tiêu của nguyên tắc này là thiết lập sự cân bằng giữa hai yêu cầu quan trọng: *một mặt*, bảo vệ những giá trị thiết yếu của xã hội, như: an toàn, sức khỏe cộng đồng, quyền con người và an ninh, *mặt khác*, bảo đảm các quy định pháp luật không trở thành rào cản quá mức đối với sự phát triển của khoa học, công nghệ và hoạt động đổi mới sáng tạo.

Từ góc độ quản trị rủi ro, pháp luật không thể và cũng không nên tìm cách loại bỏ hoàn toàn mọi rủi ro phát sinh từ việc phát triển và sử dụng AI. Thay vào đó, pháp luật cần thiết lập một cơ chế quản trị rủi ro hiệu quả trong đó gánh nặng tuân thủ bao gồm: chi phí pháp lý, chi phí kỹ thuật và chi phí vận hành tổ chức được phân bổ tương ứng với mức độ rủi ro tiềm ẩn của hệ thống AI.

Hai là, xung đột lợi ích và thiếu tính khách quan trong phân loại rủi ro ảnh hưởng đến việc xác định nghĩa vụ pháp lý của nhà phát triển, bên cung cấp.

Theo quy định của khoản 1 Điều 10 Luật Trí tuệ nhân tạo năm 2025, nhà cung cấp tự phân loại hệ thống AI của chính mình thì có thể dẫn đến tình trạng đánh giá thấp rủi ro để tránh các nghĩa vụ pháp lý phức tạp hơn, như: thông báo, kiểm tra định kỳ. Điều này có

nguy cơ làm giảm hiệu lực của toàn bộ khung phân loại (Sandra Wachter, 2024) ngay từ đầu. Như vậy các hệ thống AI thực tế có rủi ro cao có thể bị cho là rủi ro thấp hoặc trung bình dẫn đến việc giám sát lỏng lẻo và tiềm ẩn nguy cơ gây hại. Giả định công ty A phân loại hệ thống giao dịch tự động của họ là rủi ro trung bình để tránh thủ tục kiểm tra định kỳ phức tạp, thấy vậy công ty B dù có hệ thống tương tự nhưng kém an toàn hơn cũng sẽ phân loại là trung bình để không bị mất lợi thế cạnh tranh.

Đối với tình huống này, cần đánh giá theo hướng sau:

Thứ nhất, Quyết định số 33/2026/QĐ-TTg liệt kê 46 loại hệ thống trí tuệ nhân tạo rủi ro cao thuộc 6 lĩnh vực (Thủ tướng Chính phủ, 2026) và Điều 5 và 6 Nghị định số 142/2026/NĐ-CP khẳng định nhà cung cấp phải tự thực hiện phân loại hệ thống AI trước khi đưa vào sử dụng và chịu trách nhiệm trước pháp luật về kết quả phân loại. Đây là cơ chế kết hợp giữa danh mục cứng và đánh giá theo bối cảnh cụ thể của từng hệ thống. Bên cạnh đó, pháp luật còn quy định nghĩa vụ lập hồ sơ kỹ thuật, đánh giá sự phù hợp đối với hệ thống rủi ro cao, thông báo kết quả phân loại qua cổng thông tin điện tử và cơ sở dữ liệu quốc gia về AI, cũng như cơ chế kiểm tra định kỳ hoặc kiểm tra chọn mẫu, giám sát thông qua cơ sở dữ liệu quốc gia về hệ thống AI. Về mặt thiết kế, các công cụ này nhằm tạo ra sự cân bằng giữa tính linh hoạt của tự đánh giá và sự giám sát của Nhà nước tránh áp dụng máy móc mọi hệ thống trong cùng lĩnh vực đều bị xếp vào nhóm rủi ro cao. Tuy nhiên, chính sự kết hợp này cũng tạo ra điểm hạn chế trong quá trình thực hiện sẽ phụ thuộc vào sự tự nguyện tuân thủ của nhà cung cấp chính chủ thể được lợi từ việc phân loại thấp.

Thứ hai, Quyết định số 33/2026/QĐ-TTg và Nghị định số 142/2026/NĐ-CP tạo khoảng trống cho việc phân loại rủi ro thấp: Hệ thống chỉ bị coi là rủi ro cao nếu (1) Thuộc danh

mục; (2) Có khả năng gây thiệt hại đáng kể; (3) Không thuộc trường hợp được loại trừ theo khoản 2 Điều 8 Nghị định số 142/2026/NĐ-CP. Do vậy, nếu nhà cung cấp có thể viện dẫn một trong các trường hợp loại trừ để tự phân loại hệ thống của mình xuống mức trung bình dù hệ thống đó về bản chất thuộc danh mục rủi ro cao. Tuy nhiên, hiệu quả thực tế của cơ chế này vẫn phụ thuộc vào năng lực phát hiện và xử lý hành vi phân loại sai của cơ quan quản lý. Nếu nguồn lực kiểm tra, giám sát còn hạn chế hoặc việc phát hiện chưa được thực hiện một cách hệ thống và kịp thời thì rủi ro nhà cung cấp đánh giá thấp mức độ rủi ro để giảm nghĩa vụ tuân thủ vẫn có thể tồn tại. Do đó, dù khung pháp lý đã được xây dựng với các quy định phòng ngừa bảo vệ, vấn đề bảo đảm tính khách quan trong phân loại rủi ro vẫn cần được tiếp tục hoàn thiện thông qua việc tăng cường năng lực giám sát độc lập, nâng cao chế tài và hoàn thiện cơ chế báo cáo, kiểm tra sau triển khai.

Thứ ba, khoản 5 và khoản 6 Điều 10 Luật Trí tuệ nhân tạo năm 2025 quy định cơ quan quản lý nhà nước thực hiện kiểm tra, giám sát định kỳ hoặc đột xuất theo mức độ rủi ro. Tuy nhiên, vấn đề trọng tâm là khả năng phát hiện hành vi phân loại sai. Ngoài ra, nghiên cứu đã chỉ ra rằng, cách tiếp cận phân loại rủi ro theo cấp độ bộc lộ nhiều hạn chế khi rủi ro của AI thay đổi liên tục, phi tuyến tính (Hứa Thị Thùy Anh, 2026), theo dữ liệu, bối cảnh triển khai và khả năng tự học của hệ thống nên phân loại tĩnh có thể dẫn đến nghĩa vụ điều tiết không tương xứng hoặc không nắm bắt được rủi ro mới nổi và tích lũy. Do vậy, nếu cơ quan quản lý không có nguồn lực hoặc phương thức để kiểm tra tính chính xác của việc tự phân loại một cách hệ thống, chế tài dù nghiêm khắc cũng khó phát huy tác dụng răn đe.

Đối với những nội dung trên, cần quy định rõ tiêu chí khả năng gây thiệt hại đáng kể và các trường hợp loại trừ tại khoản 2 Điều 8 Nghị định số 142/2026/NĐ-CP theo hướng

tránh việc viện dẫn loại trừ để hạ mức rủi ro; Yêu cầu nhà cung cấp công bố công khai tóm tắt hồ sơ phân loại rủi ro (không bao gồm bí mật thương mại) trên Cổng thông tin điện tử một cửa và Cơ sở dữ liệu quốc gia về AI, tạo điều kiện giám sát xã hội.

Ba là, trách nhiệm pháp lý chưa phù hợp với bối cảnh Việt Nam sử dụng các hệ thống trí tuệ nhân tạo từ nhà phát triển, nhà cung cấp nước ngoài.

Khoản 2 Điều 29 Luật Trí tuệ nhân tạo năm 2025 quy định: sau khi bồi thường, bên triển khai yêu cầu nhà cung cấp, nhà phát triển hoặc các bên liên quan hoàn trả khoản tiền bồi thường nếu có thỏa thuận giữa các bên. Cách diễn đạt này có nghĩa là quyền truy đòi của bên triển khai chỉ tồn tại khi có thỏa thuận trước đó chứ không phải là quyền mặc định theo Luật. Đây là điểm bất cập lớn vì trong thực tế, các hợp đồng mua bán công nghệ AI thường được soạn thảo bởi nhà cung cấp lớn từ nước ngoài với điều khoản giới hạn trách nhiệm chặt chẽ. Trong bối cảnh đó, bên triển khai Việt Nam thường ở vị thế yếu hơn trong đàm phán sẽ khó có thể đạt được thỏa thuận truy đòi có lợi.

Ngay cả khi có thỏa thuận, việc thực thi quyền truy đòi rất phức tạp khi nhà phát triển hoặc nhà cung cấp ở nước ngoài không có hiện diện pháp lý tại Việt Nam. Cơ chế pháp lý hiện hành chưa có quy định rõ ràng về thẩm quyền tài phán và công nhận phán quyết nước ngoài trong các tranh chấp truy đòi liên quan đến ứng dụng AI. Khoản 2 Điều 29 không quy định thứ tự ưu tiên truy đòi giữa nhà cung cấp và nhà phát triển, cũng không quy định tỷ lệ phân bổ trách nhiệm khi có nhiều bên cùng có lỗi. Điều này dẫn đến tình trạng nhiều trường hợp không thể xác định ai phải hoàn trả bao nhiêu, đặc biệt trong chuỗi cung ứng AI phức tạp với nhiều nhà phát triển mô hình nền tảng, bên tích hợp và bên tùy chỉnh. Do vậy, cần:

(1) Bổ sung quy định về bảo hiểm trách nhiệm bắt buộc đối với hệ thống AI rủi ro cao;

đồng thời, khuyến khích phát triển thị trường bảo hiểm AI (Faure & cộng sự 2022) thông qua các ưu đãi thuế và hỗ trợ kỹ thuật từ nhà nước nhằm bảo đảm nguồn lực bồi thường khi rủi ro phát sinh ngoài dự kiến.

(2) Thiết lập quyền truy đòi mặc định theo luật cho bên triển khai đối với nhà cung cấp và nhà phát triển khi thiệt hại phát sinh do lỗi kỹ thuật của mô hình AI, không phụ thuộc vào thỏa thuận giữa các bên.

(3) Quy định giới hạn trách nhiệm tối đa cho nhà cung cấp và nhà phát triển có phân biệt theo mức độ rủi ro và quy mô doanh nghiệp.

4. Kết luận

Sự phát triển nhanh của AI mang lại nhiều lợi ích nhưng cũng tạo ra rủi ro phức tạp, khó dự đoán và khó kiểm soát. Bài viết chỉ ra rằng, nghĩa vụ pháp lý của nhà phát triển và nhà cung cấp hệ thống AI không thể chỉ dựa vào việc gắn nhãn rủi ro cao, trung bình hay thấp. Hiệu quả của cơ chế này phụ thuộc vào tính khách quan khi phân loại rủi ro, năng lực giám sát độc lập của Nhà nước và cơ chế phân bổ trách nhiệm hợp lý trong chuỗi cung ứng AI, đặc biệt với nhà phát triển, nhà cung cấp nước ngoài.

Bài viết đề xuất làm rõ tiêu chí và trường hợp loại trừ để hạn chế hạ thấp mức rủi ro; tăng cường công khai thông tin phân loại; thiết lập quyền truy đòi mặc định; quy định bảo hiểm trách nhiệm bắt buộc với hệ thống rủi ro cao; giới hạn trách nhiệm phù hợp mức rủi ro và quy mô doanh nghiệp. Việc hoàn thiện theo hướng này giúp cân bằng kiểm soát rủi ro với thúc đẩy đổi mới sáng tạo; đồng thời, nâng cao hiệu lực thực thi pháp luật AI tại Việt Nam □

Tài liệu tham khảo:

- * Bài viết là sản phẩm công trình nghiên cứu cấp trường, mã số T2026-49, Trường Đại học Công nghệ Kỹ thuật TP. Hồ Chí Minh.
- Chính phủ (2026). *Nghị định số 142/2026/NĐ-CP ngày 30/4/2026 quy định chi tiết một số điều và biện pháp thi hành Luật Trí tuệ nhân tạo.*

- Comella, V. F. (2025). *Beyond the principle of proportionality*. In Comparative constitutional theory (pp. 312-330). Edward Elgar Publishing.
- Faure, M. G., & Li, S. (2022). Artificial intelligence and (compulsory) insurance. *Journal of European Tort Law*, 13(1), 1-24. <https://doi.org/10.1515/jetl-2022-0001> https://cris.maastrichtuniversity.nl/ws/portalfiles/portal/185498922/Faure-2022-Artificial-Intelligence-and-_Compulsory_-Insurance.pdf
- Hứa Thị Thùy Anh (21/9/2026). *Bất cập của phân loại rủi ro theo cấp độ trong điều tiết AI: Kinh nghiệm quốc tế và hàm ý cho Việt Nam*. *Tạp chí Công Thương*, <https://nckh.tapchicongthuong.vn/nckhct-media/26/8/10/12--so-19-2025.pdf?md5=Wq6TnB4mkfVIDgonUOoLkg&expires=1791439040#1#1>
- Macrae, C. (2022). *Learning from the Failure of Autonomous and Intelligent Systems: Accidents, Safety, and Sociotechnical Sources of Risk*. *Risk Analysis*, 42: 1999-2025. <https://doi.org/10.1111/risa.13850>
- Quốc hội (2025). *Luật Trí tuệ nhân tạo năm 2025*.
- Regulation (EU) (2024). *1689 of the European Parliament and of the Council of 13 June 2024*.
- Schuett, J. (2025). *Frontier AI developers need an internal audit function*. *Risk Analysis*, 45, 1332-1352. <https://doi.org/10.1111/risa.17665>
- Sandra Wachter (2024). *Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond*, *Yale Journal of Law & Technology*, Volume 26, Issue 3, Available at SSRN: <https://ssrn.com/abstract=4924553> or <http://dx.doi.org/10.2139/ssrn.4924553>
- Thủ tướng Chính phủ (2026). *Quyết định số 33/2026/QĐ-TTg về danh mục hệ thống trí tuệ nhân tạo rủi ro cao*.
- UK Government (2023). *Implementing the UK's AI regulatory principles: initial guidance for regulators*. <https://www.gov.uk/government/publications/implementing-the-uks-ai-regulatory-principles-initial-guidance-for-regulators/implementing-the-uks-ai-regulatory-principles-initial-guidance-for-regulators>.
- White House Office of Science and Technology Policy (OSTP) (2022). *Blueprint for an AI Bill of Rights*. Oct. 2022. <https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/>