BẤT CẬP CỦA CÁCH TIẾP CẬN PHÂN LOẠI RỦI RO THEO CẤP ĐỘ TRONG ĐIỀU TIẾT AI: KINH NGHIỆM QUỐC TẾ VÀ HÀM Ý CHO VIỆT NAM

HỨA THỊ THÙY ANH
Trường Đại học Kinh tế - Luật - Đại học Quốc gia TP. Hồ Chí Minh
University of Economics and Law, Vietnam National University, Ho Chi Minh City, Vietnam
Email: httanhhcmulaw@gmail.com

Ngày nhận bài: 18/5/2026
Ngày sửa bài: 8/6/2026
Ngày duyệt đăng bài: 19/6/2026

Tóm tắt:

Sự phát triển nhanh chóng của trí tuệ nhân tạo (AI) đang làm thay đổi cách thức cung ứng dịch vụ trong nền kinh tế số, đồng thời đặt ra nhiều thách thức đối với các mô hình điều tiết truyền thống. Hiện nay, nhiều quốc gia lựa chọn cách tiếp cận quản trị AI dựa trên phân loại rủi ro theo cấp độ, trong đó tiêu biểu là Đạo luật AI của Liên minh châu Âu (EU AI Act). Tuy nhiên, cách tiếp cận này bộc lộ nhiều hạn chế khi rủi ro của AI thay đổi liên tục, phi tuyến tính và thay đổi liên tục theo dữ liệu, bối cảnh triển khai và khả năng tự học của hệ thống. Bài viết phân tích những bất cập của mô hình phân loại rủi ro theo cấp độ dẫn đến sự thiếu tương xứng trong điều tiết. Trên cơ sở đó, bài viết làm rõ xu hướng quốc tế chuyển sang quản trị AI dựa trên cấu trúc rủi ro động, đồng thời đề xuất một số kiến nghị hoàn thiện pháp luật cho Việt Nam.

Limitations of the tiered risk-classification approach in AI regulation: International experience and implications for Vietnam

Abstract:

The rapid development of artificial intelligence is transforming service delivery in the digital economy while challenging traditional regulatory models. Many jurisdictions have adopted tiered risk-classification approaches to AI governance, most notably the European Union's AI Act. However, this model has significant limitations because AI risks are dynamic, nonlinear, and continuously influenced by data, deployment contexts, system interactions, and adaptive learning. Static classification may therefore result in disproportionate regulatory obligations or fail to capture emerging and cumulative risks. This study examines these limitations and analyzes the international shift toward dynamic, context-based, and lifecycle-oriented AI governance. Based on international experience, it proposes recommendations for Vietnam to introduce continuous risk assessment, strengthen post-deployment monitoring, clarify accountability, and develop adaptive regulatory mechanisms suitable for evolving AI systems.

Keywords: *artificial intelligence, AI governance, dynamic risk assessment, lifecycle governance, AI regulation.*

1. Đặt vấn đề

Trí tuệ nhân tạo (AI) đang trở thành nền tảng công nghệ cốt lõi của nền kinh tế số và thương mại dịch vụ hiện đại. Không chỉ được sử dụng trong các lĩnh vực công nghệ truyền thống, AI hiện diện ngày càng sâu rộng trong y tế, tài chính, giáo dục, logistics, giao thông, thương mại điện tử, quảng cáo số và quản trị công. Các hệ thống AI không chỉ hỗ trợ con người xử lý dữ liệu mà còn có khả năng tự học, tự thích nghi và đưa ra các quyết định có ảnh hưởng trực tiếp đến quyền, lợi ích và cơ hội tiếp cận dịch vụ của cá nhân, tổ chức.

Sự mở rộng nhanh chóng của AI đồng thời làm gia tăng áp lực đối với các mô hình quản lý nhà nước truyền thống. Tại Việt Nam, Luật Trí tuệ nhân tạo 2025 và các văn bản hướng dẫn thi hành cũng cho thấy xu hướng tiếp cận tương tự EU thông qua việc phân loại hệ thống AI theo mức độ rủi ro và áp dụng các cơ chế kiểm soát tương ứng. Tuy nhiên, cách tiếp cận quản trị dựa trên phân loại rủi ro cố định đang bộc lộ nhiều giới hạn khi áp dụng đối với AI do các yếu tố rủi ro đặc thù của công nghệ này.

Thực tế cho thấy, nhiều quốc gia và tổ chức quốc tế bắt đầu dịch chuyển từ mô hình quản trị rủi ro cố định sang các mô hình quản trị linh hoạt để thích ứng, linh hoạt và dựa trên cấu trúc rủi ro cụ thể của trí tuệ nhân tạo. Xu hướng này nhấn mạnh việc đánh giá AI theo bối cảnh sử dụng thực tế, quản trị xuyên suốt vòng đời hệ thống và liên tục cập nhật nghĩa vụ pháp lý theo sự thay đổi của rủi ro. Bài viết tập trung phân tích những giới hạn của mô hình phân loại rủi ro theo cấp độ trong điều tiết AI và đề xuất một số hàm ý chính sách cho Việt Nam.

2. Tổng quan nghiên cứu và cơ sở lý thuyết

2.1. Cách tiếp cận quản trị AI dựa trên cấp độ rủi ro theo quy định pháp luật Việt Nam

Cùng với xu hướng phát triển khung quản trị AI trên thế giới, Việt Nam đang từng bước chuyển từ mô hình điều chỉnh phân tán sang mô hình quản trị AI chuyên biệt dựa trên rủi ro thông qua Luật Trí tuệ nhân tạo năm 2025 và Nghị định số 142/2026/NĐ-CP. Theo đó, pháp luật Việt Nam thiết lập một cấu trúc phân tầng rủi ro đối với hệ thống AI. Ở cấp độ cao nhất, một số hành vi ứng dụng AI bị nghiêm cấm tuyệt đối, như hành vi thao túng nhận thức, lợi dụng nhóm yếu thế, tạo nội dung giả mạo (deepfake) nhằm đe dọa an ninh quốc gia hoặc xâm hại nghiêm trọng đến quyền và lợi ích hợp pháp của cá nhân[1]. Đối với các hệ thống AI còn lại, pháp luật phân chia thành 3 nhóm gồm AI rủi ro cao, AI rủi ro trung bình và AI rủi ro thấp dựa trên các tiêu chí như mức độ ảnh hưởng đến quyền con người, an toàn xã hội, lĩnh vực triển khai, phạm vi tác động và khả năng duy trì sự giám sát của con người[2].

Đồng thời, pháp luật Việt Nam cũng bắt đầu tiếp cận theo hướng điều tiết linh hoạt hơn khi thừa nhận rằng không phải mọi hệ thống AI trong cùng một lĩnh vực đều tạo ra mức độ rủi ro tương đương nhau. Ví dụ, một số hệ thống có thể được loại trừ khỏi nhóm rủi ro cao nếu chỉ phục vụ mục đích nội bộ hoặc vẫn duy trì cơ chế đánh giá và kiểm soát thực chất của con người trước khi đưa ra quyết định cuối cùng[3].

Trên cơ sở phân loại rủi ro, pháp luật Việt Nam thiết lập các nghĩa vụ quản trị tương ứng và phân định trách nhiệm theo vai trò của từng chủ thể trong chuỗi cung ứng AI. Có thể thấy, cách tiếp cận của Việt Nam về cơ bản đang tiếp thu xu hướng quản trị AI dựa trên rủi ro tương tự nhiều hệ thống pháp luật hiện đại, đặc biệt là mô hình của EU. Tuy nhiên, việc phân loại hệ thống AI theo cấp cố định chưa phản ánh đầy đủ sự đa dạng về nguồn phát sinh rủi ro của từng hệ thống AI cụ thể vì hai hệ thống AI cùng được xếp vào nhóm "rủi ro cao" có thể phát sinh rủi ro từ những nguyên nhân hoàn toàn khác nhau[4].

2.2. Nguyên nhân các quốc gia lựa chọn mô hình điều tiết mang tính cố định

Trước hết, mô hình điều tiết cố định xuất phát từ nhu cầu bảo đảm tính chắc chắn và khả năng dự báo của pháp luật[5]. Theo quan niệm truyền thống về nhà nước pháp quyền, pháp luật cần tạo ra các chuẩn mực ổn định, rõ ràng và có khả năng tiên liệu để các chủ thể có thể xác định trước hành vi hợp pháp của mình[6].

Bên cạnh đó, xu hướng điều tiết cố định còn phản ánh tư duy phòng ngừa trước sự bất định của công nghệ AI[7]. Trong bối cảnh hậu quả của AI có thể ảnh hưởng trực tiếp đến lợi ích công cộng nhưng mức độ rủi ro thực tế lại chưa thể được đo lường đầy đủ, nhiều quốc gia có xu hướng áp dụng nguyên tắc phòng ngừa nhằm can thiệp sớm trước khi rủi ro phát sinh trên thực tế[8].

Trong nhiều trường hợp, sự lựa chọn mô hình điều tiết cố định phản ánh không chỉ yêu cầu bảo đảm an toàn mà còn là kết quả của thiên hướng duy trì nguyên trạng trong hoạt động quản trị nhà nước[9]. Đồng thời, mô hình này cho phép Nhà nước duy trì khả năng can thiệp trực tiếp và tức thời đối với thị trường[10].

3. Từ bản chất rủi ro đặc thù của AI đến những bất cập trong cách tiếp cận quản trị AI dựa trên cấp độ rủi ro

3.1. Về bản chất rủi ro đặc thù của AI

Bản thân các đặc tính của AI không tạo ra rủi ro mà rủi ro chỉ phát sinh khi các đặc tính này cộng hưởng và đặt trong bối cảnh vận hành cụ thể. Điều này khiến rủi ro của AI không còn là trạng thái cố định, cụ thể:

Một là, AI có khả năng tiếp tục thích nghi và thay đổi hành vi sau khi đã được triển khai trên thực tế[11]. Trong các hệ thống học máy và học sâu, kết quả đầu ra không hoàn toàn được xác định trước bởi các quy tắc cứng như phần mềm truyền thống mà được hình thành thông qua quá trình điều chỉnh liên tục các tham số nội bộ dựa trên dữ liệu đầu vào và môi trường vận hành.

Hai là, AI còn vận hành với mức độ phi minh bạch cao hơn đáng kể so với các công nghệ truyền thống. Trong các mô hình học sâu, dữ liệu được xử lý thông qua hàng triệu hoặc hàng tỷ tham số liên kết phức tạp, hình thành hiện tượng "hộp đen" [12], nơi ngay cả nhà phát triển cũng khó có thể giải thích đầy đủ logic dẫn đến một quyết định cụ thể. Sự kết hợp giữa khả năng tự học, tính thích nghi hậu triển khai và tính khó giải thích khiến cấu trúc rủi ro của AI trở nên khó dự báo hơn nhiều so với các công cụ kỹ thuật truyền thống.

Ba là, rủi ro của AI còn mang tính hệ thống và phi tuyến tính[13]. Khác với các công cụ truyền thống thường tạo ra các rủi ro cục bộ, sai lệch trong AI có thể nhanh chóng bị khuếch đại thông qua khả năng tự động hóa và triển khai trên quy mô lớn. Nhiều hệ thống AI cũng không vận hành độc lập mà liên tục tương tác với các thuật toán, nền tảng dữ liệu và hệ sinh thái số khác, từ đó làm phát sinh rủi ro và hệ thống và các hiệu ứng lan truyền ngoài dự kiến[14].

Thứ tư, là rủi ro của AI phụ thuộc mạnh mẽ vào bối cảnh triển khai thực tế. Cùng một mô hình AI nhưng khi được sử dụng trong các lĩnh vực khác nhau có thể tạo ra mức độ tác động hoàn toàn khác nhau[15].

3.2. Những bất cập trong cách tiếp cận quản trị AI dựa trên cấp độ rủi ro

Từ các đặc tính trên, mô hình phân loại rủi ro cố định bắt đầu bộc lộ nhiều giới hạn trong hoạt động điều tiết AI. Một mặt, việc gắn nhãn "rủi ro cao" cho toàn bộ các hệ thống AI trong cùng một lĩnh vực có thể dẫn đến hiện tượng điều tiết quá mức, khi các hệ thống có mức độ tác động hạn chế vẫn phải chịu cùng một cường độ nghĩa vụ tuân thủ với các hệ thống có khả năng ảnh hưởng trực tiếp đến quyền và lợi ích của cá nhân[16]. Mặt khác, mô hình phân loại cố định cũng có nguy cơ bỏ sót nhiều dạng rủi ro quan trọng, đặc biệt đối với các hệ thống AI tuy không được xếp vào nhóm "rủi ro cao" nhưng lại có khả năng tạo ra tác động xã hội lớn thông qua hiệu ứng tích lũy, thao túng hành vi hoặc khuếch đại thông tin trên quy mô rộng[17].

Ngoài ra, mô hình phân loại cố định còn chưa phản ánh đầy đủ hiện tượng dữ liệu đầu ra của hệ thống tiếp tục được sử dụng làm dữ liệu đầu vào cho các vòng huấn luyện tiếp theo, các sai lệch hoặc thiên kiến ban đầu có thể bị khuếch đại theo thời gian[18].

Những giới hạn trên cho thấy, mô hình quản trị AI dựa trên phân loại rủi ro theo cấp độ tuy có ý nghĩa quan trọng trong việc thiết lập khung quản trị ban đầu, nhưng chưa đủ để phản ánh đầy đủ tính động và cấu trúc rủi ro phức hợp của AI hiện đại.

4. Xu hướng chuyển sang quản trị AI thay đổi theo cấu trúc rủi ro của AI trên thế giới và hàm ý cho Việt Nam

4.1. Xu hướng chuyển sang quản trị AI thay đổi theo cấu trúc rủi ro của AI trên thế giới

Một trong những xu hướng nổi bật của quản trị AI hiện nay là chuyển từ phân loại rủi ro cứng nhắc sang đánh giá rủi ro dựa trên bối cảnh sử dụng thực tế của hệ thống và sự thay đổi rủi ro liên tục của AI[19]. Cách tiếp cận này giúp chuyển trọng tâm quản trị từ "AI thuộc nhóm rủi ro nào" sang "AI tạo ra rủi ro như thế nào trong bối cảnh cụ thể". Xu hướng dịch chuyển này hiện đã hình thành ở quy mô toàn cầu. OECD và G20 hiện ghi nhận khoảng 50 quốc gia cam kết thực hiện các nguyên tắc AI linh hoạt và có khả năng thích ứng cao. Đồng thời, theo dữ liệu của Ngân hàng Thế giới, đã có 73 cơ sở thử nghiệm pháp lý (regulatory sandboxes) được triển khai tại 57 quốc gia và khu vực pháp lý trên thế giới nhằm hỗ trợ cơ chế quản trị công nghệ mang tính thử nghiệm và thích ứng. Bên cạnh đó, Đài quan sát chính sách AI của OECD hiện ghi nhận hơn 700 sáng kiến chính sách AI đến từ khoảng 60 quốc gia và Liên minh châu Âu (EU), trong đó phần lớn đều nhấn mạnh việc quản trị AI dựa trên bối cảnh, quản trị theo vòng đời và cơ chế đánh giá rủi ro động[20] thay vì xác lập nghĩa vụ pháp lý dựa trên một phân cấp rủi ro cố định hoặc lĩnh vực sử dụng rộng lớn, các mô hình quản trị hiện đại ngày càng nhấn mạnh việc đánh giá cấu trúc phát sinh rủi ro thực tế của từng hệ thống AI cụ thể.

Nhận thức này được phản ánh rõ nét trong các khuôn khổ quản trị AI hàng đầu thế giới. OECD hiện là một trong những tổ chức đi đầu trong việc thúc đẩy mô hình đánh giá AI dựa trên bối cảnh và tính động của rủi ro. Khung phân loại AI của OECD từ chối cách tiếp cận phân loại "rủi ro cao - rủi ro thấp", thay vào đó yêu cầu đánh giá hệ thống AI thông qua nhiều tiêu chí kết hợp đan xen như dữ liệu đầu vào, mô hình AI, bối cảnh kinh tế - xã hội, nhiệm vụ đầu ra và tác động đối với con người[21]. Đáng chú ý, OECD nhấn mạnh rằng vị trí rủi ro của một hệ thống AI có thể thay đổi liên tục khi hệ thống mở rộng quy mô triển khai, tiếp thu dữ liệu mới hoặc thay đổi hành vi thông qua cơ chế học máy.

Tại Hoa Kỳ, Khung quản lý rủi ro AI của Viện Tiêu chuẩn và Công nghệ Quốc gia (NIST AI RMF)[22] cũng chuyển trọng tâm từ việc cấm đoán theo lĩnh vực sang quản lý rủi ro theo bối cảnh vận hành và toàn bộ vòng đời của hệ thống. Thay vì áp dụng các danh mục rủi ro cố định, NIST xây dựng các "hồ sơ rủi ro" cho phép điều chỉnh nghĩa vụ quản trị phù hợp với từng môi trường triển khai cụ thể. Trong khi đó, Vương quốc Anh lựa chọn cách tiếp cận linh hoạt và thúc đẩy đổi mới thông qua việc sử dụng các cơ sở thử nghiệm pháp lý, cho phép cơ quan quản lý và doanh nghiệp đánh giá trực tiếp các rủi ro thực tế của thuật toán trong môi trường kiểm soát trước khi áp dụng các nghĩa vụ pháp lý chính thức[23].

Trong mô hình này, nghĩa vụ pháp lý không còn được xác lập dựa trên "nhãn rủi ro" cố định của hệ thống AI, mà phụ thuộc vào cấu trúc dữ liệu, mức độ tự chủ, quy mô triển khai và các tác động thực tế mà hệ thống tạo ra đối với xã hội và quyền con người. Điều này là phù hợp với những bản chất rủi ro đặc thù của AI như đã nêu trên.

4.2. Hàm ý cho Việt Nam

Từ những phân tích trên, pháp luật cần chuyển sang mô hình đánh giá rủi ro linh hoạt dựa trên cấu trúc vận hành và bối cảnh triển khai thực tế của từng hệ thống AI cụ thể. Điều này đặc biệt quan trọng bởi cùng phân loại AI có thể tạo ra các mức độ rủi ro hoàn toàn khác nhau tùy thuộc vào dữ liệu đầu vào, quy mô triển khai, mức độ tự chủ và nhóm chủ thể chịu tác động. Bên cạnh đó, kinh nghiệm quốc tế cũng cho thấy nghĩa vụ pháp lý cần được thiết kế theo hướng tương xứng với cấu trúc rủi ro thực tế của hệ thống AI thay vì áp dụng đồng loạt cho toàn bộ một lĩnh vực. Điều này không chỉ giúp nâng cao hiệu quả quản trị rủi ro mà còn hạn chế nguy cơ hình thành các rào cản gia nhập thị trường không cần thiết đối với dịch vụ AI dưới góc độ của GATS

5. Kết luận

Mô hình quản trị AI dựa trên phân loại rủi ro theo cấp độ hiện nay đang bộc lộ nhiều hạn chế khi áp dụng đối với các hệ thống AI có khả năng tự học, thích nghi và liên tục thay đổi trong quá trình vận hành. Rủi ro của AI không còn mang tính cố định mà ngày càng thể hiện đặc tính động, phi tuyến tính và phụ thuộc mạnh mẽ vào bối cảnh triển khai thực tế. Điều này làm xuất hiện nhiều dạng rủi ro mới như rủi ro mạng lưới, tác động tích lũy, vòng lặp phản hồi và sự tiến hóa của thuật toán theo thời gian. Trong bối cảnh đó, xu hướng quản trị AI trên thế giới đang chuyển dịch mạnh mẽ từ mô hình phân loại rủi ro cố định sang cơ chế đánh giá rủi ro theo bối cảnh cụ thể, quản trị theo vòng đời và hiệu chỉnh nghĩa vụ pháp lý liên tục theo cấu trúc rủi ro thực tế của hệ thống. Đây cũng là kinh nghiệm cần thiết để hoàn thiện pháp luật Việt Nam[n]■

TÀI LIỆU TRÍCH DẪN

[1]Điều 7 Luật Trí tuệ nhân tạo năm 2025

[2]Điều 9 Luật Trí tuệ nhân tạo năm 2025

[3]Điều 8 Nghị định số 142/2026/NĐ-CP

[4]Bullock, J. B., Chen, Y.-C., et al. (2024). *The Oxford Handbook of AI Governance*. Oxford University Press, pp. 441–460.

[5]Barfield, W. (2020). *The Cambridge Handbook of the Law of Algorithms*. Cambridge University Press, p. 224.

[6]Corrales, M., Fenwick, M., & Forgó, N. (2018). *Robot, AI and Future Law*. Springer, p. 89.

[7]Chesterman, S. (2021). *We, the Robots? Regulating Artificial Intelligence and the Limits of the Law*. Cambridge University Press, p. 183.

[8]Nikolinakos, N. T. (2023). *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies: The AI Act*. Springer, p. 275.

[9]Mathis, K., & Tor, A. (2023). *Law and Economics of the Digital Transformation*. Springer, p. 364.

[10]Kovač, M. (2024). *Generative Artificial Intelligence: A Law and Economics Approach to Optimal Regulation and Governance*. Springer, pp. 113-114.

[11]Nikolinakos, N. T. (2023). *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies: The AI Act*. Springer, pp. 352-353.

[12]Nikolinakos, N. T. (2023). *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies: The AI Act*. Springer, pp. 43-44.

[13]Antunes, H. S., Freitas, P. M., Oliveira, A. L., Pereira, C. M., Sequeira, E. V., & Xavier, L. B. (2024). *Multidisciplinary Perspectives on Artificial Intelligence and the Law*. Springer, pp. 419-421.

[14]Smuha, N. A. (2025). *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*. Cambridge University Press, pp. 243–244.

[15]Bullock, J. B., Chen, Y.-C., et al. (2024). *The Oxford Handbook of AI Governance*. Oxford University Press, pp. 183-197.

[16]Nikolinakos, N. T. (2023). *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies: The AI Act*. Springer, pp. 435-436.

[17]Nikolinakos, N. T. (2023). *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies: The AI Act*. Springer, pp. 508-509.

[18]Antunes, H. S., Freitas, P. M., Oliveira, A. L., Pereira, C. M., Sequeira, E. V., & Xavier, L. B. (2024). *Multidisciplinary Perspectives on Artificial Intelligence and the Law*. Springer, p. 421.

[19]Organisation for Economic Co-operation and Development (OECD). (2022). *OECD Framework for the Classification of AI Systems*.

[20]Bullock, J. B., Chen, Y.-C., et al. (2024). *The Oxford Handbook of AI Governance*. Oxford University Press, pp. 253-275.

[21]Organisation for Economic Co-operation and Development (OECD). (2022). *OECD Framework for the Classification of AI Systems*.

[22]Singh, T. (2024). *Artificial Intelligence and Ethics: A Field Guide for Stakeholders*. CRC Press, pp. 189-190.

[23]Corrales, M., Fenwick, M., & Forgó, N. (2018). *Robot, AI and Future Law*. Springer, pp. 89-90.

TÀI LIỆU THAM KHẢO:

Antunes, H. S., Freitas, P. M., Oliveira, A. L., Pereira, C. M., Sequeira, E. V., & Xavier, L. B. (2024). *Multidisciplinary perspectives on artificial intelligence and the law*. Springer.

Barfield, W. (2020). *The Cambridge handbook of the law of algorithms*. Cambridge University Press.

Bullock, J. B., Chen, Y.-C., et al. (2024). *The Oxford handbook of AI governance*. Oxford University Press.

Chesterman, S. (2021). *We, the robots? Regulating artificial intelligence and the limits of the law*. Cambridge University Press.

Corrales, M., Fenwick, M., & Forgó, N. (2018). *Robot, AI and future law*. Springer.

European Union. (2025). *Artificial Intelligence Act*.

Kellmeyer, P., Mueller, O., Burgard, W., & Voeneky, S. (Eds.). (2022). *The Cambridge handbook of responsible artificial intelligence: Interdisciplinary perspectives*. Cambridge University Press.

Kovač, M. (2024). *Generative artificial intelligence: A law and economics approach to optimal regulation and governance*. Springer.

Mathis, K., & Tor, A. (2023). *Law and economics of the digital transformation*. Springer.

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*.

Nikolinakos, N. T. (2023). *EU policy and legal framework for artificial intelligence, robotics and related technologies: The AI Act*. Springer.

Organisation for Economic Co-operation and Development. (2022). *OECD framework for the classification of AI systems*. OECD Publishing.

Singh, T. (2024). *Artificial intelligence and ethics: A field guide for stakeholders*. CRC Press.

Smuha, N. A. (2025). *The Cambridge handbook of the law, ethics and policy of artificial intelligence*. Cambridge University Press.