

XÁC ĐỊNH CHỦ THỂ CHỊU TRÁCH NHIỆM HÌNH SỰ ĐỐI VỚI HÀNH VI XÂM PHẠM DỮ LIỆU TRONG BỐI CẢNH TRÍ TUỆ NHÂN TẠO

TRẦN HOÀNG THÙY VY¹ - LÊ HỒ TRUNG HIẾU^{1*} - TRẦN ANH KHOA²

¹Khoa Luật, Trường Đại học Văn Lang

Faculty of Law, Van Lang University

²Cử nhân Luật, Trường Đại học Luật Thành phố Hồ Chí Minh

Bachelor of Laws, Ho Chi Minh City University of Law

*Tác giả liên hệ - Email: hieu.lht@vlu.edu.vn

Ngày nhận bài: 10/7/2026

Ngày sửa bài: 27/7/2026

Ngày duyệt đăng bài: 14/8/2026

Tóm tắt:

Trong bối cảnh trí tuệ nhân tạo (AI) ngày càng được ứng dụng rộng rãi trong thu thập, xử lý và phân tích dữ liệu, các hành vi xâm phạm dữ liệu có sự tham gia của AI đặt ra những vấn đề mới trong việc xác định chủ thể chịu trách nhiệm hình sự. Bài viết phân tích cơ sở pháp lý và những vấn đề đặt ra đối với việc xác định chủ thể chịu trách nhiệm hình sự trong các hành vi xâm nhập trái phép hệ thống, sử dụng trái phép thông tin, khai thác dữ liệu cá nhân, đầu độc dữ liệu huấn luyện và một số hình thức tấn công được thực hiện với sự hỗ trợ của AI. Trên cơ sở đó, bài viết đề xuất một số giải pháp hoàn thiện pháp luật hình sự Việt Nam nhằm xác định rõ chủ thể và căn cứ chịu trách nhiệm hình sự đối với hành vi xâm phạm dữ liệu trong bối cảnh phát triển trí tuệ nhân tạo.

Từ khóa: trí tuệ nhân tạo, xâm phạm dữ liệu, trách nhiệm hình sự, chủ thể của tội phạm, pháp nhân thương mại

Identifying subjects bearing criminal liability for data infringements in the context of artificial intelligence

Abstract:

As artificial intelligence (AI) is increasingly applied to data collection, processing, and analysis, data infringements involving AI raise new issues in identifying subjects bearing criminal liability. This study analyzes the legal basis and emerging challenges in determining criminal liability for unauthorized system intrusion, unlawful use of information, exploitation of personal data, poisoning of training data, and certain forms of attack carried out with AI assistance. On this basis, the study proposes several solutions for improving Vietnamese criminal law in order to clarify liable subjects and the grounds for criminal liability for data infringements in the context of artificial intelligence development.

Keywords: artificial intelligence, data infringement, criminal liability, criminal subject, commercial legal entity

1. Đặt vấn đề

AI là một hệ thống được thiết kế để hoạt động một cách tự chủ, có khả năng tự thích ứng và phát triển, tự dự đoán, tự suy luận và đề xuất sau khi được nhập dữ liệu đầu vào (khoản 1 Điều 3 Regulation (EU) 2024/1689). Hiểu đơn giản, AI có khả năng thực hiện đa dạng các tác vụ dựa trên dữ liệu mà không cần đến sự can thiệp, điều khiển từ con người.

Tại Việt Nam hiện nay, ngoài “Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030” được ban hành theo Quyết định số 127/QĐ-TTg năm 2021, khung pháp lý cho dữ liệu và AI đã và đang được xây dựng chuyên biệt hơn với Luật Dữ liệu năm 2024 có hiệu lực từ ngày 01/7/2025, Luật Bảo vệ dữ liệu cá nhân năm 2025 có hiệu lực từ ngày 01/01/2026, Luật Trí tuệ nhân tạo năm 2025 có hiệu lực từ ngày 01/3/2026, cùng Nghị định số 165/2025/NĐ-CP và Nghị định số 356/2025/NĐ-CP để hướng dẫn thi hành. Bên cạnh đó, Luật An ninh mạng 2018 vẫn là đạo luật an ninh mạng đang có hiệu lực tại thời điểm hiện tại và sẽ được thay thế bằng Luật An ninh mạng năm 2025, có hiệu lực từ ngày 01/7/2026. Mặc dù khung pháp lý cho dữ liệu và AI được xây dựng rất phong phú, nhưng các nguyên tắc nền tảng của Bộ luật Hình sự (BLHS) vẫn không có sự thay đổi. Các nghĩa vụ, tiêu chuẩn hành xử và giới hạn xử lý dữ liệu có thể được xác lập bằng các quy phạm pháp luật về dữ liệu và AI. Tuy nhiên, nếu muốn truy cứu TNHS thì vẫn phải căn cứ vào BLHS, nghĩa là dựa vào 4 yếu tố cấu thành tội phạm.

Trước đây, xâm phạm dữ liệu được nhận diện qua các hành vi cụ thể như xâm nhập trái phép vào hệ thống, sao chép dữ liệu, phát tán phần mềm độc hại... thì trong bối cảnh trí tuệ nhân tạo hiện nay, các hành vi này có thể được thực hiện một cách tự động, không cần phải có sự can thiệp từ con người. Vấn đề đặt ra là trách nhiệm hình sự phải được xác định như thế nào, đối với ai, trên nền tảng chứng minh nào, và pháp luật Việt Nam hiện hành đang gặp những khoảng trống gì trong bối cảnh AI.

2. Cơ sở pháp lý hiện hành về chủ thể của tội phạm

Theo khoa học Luật Hình sự, có 4 yếu tố cấu thành tội phạm bao gồm: khách thể, chủ thể, mặt khách quan và mặt chủ quan của tội phạm. Điều 8 BLHS quy định “Tội phạm là hành vi nguy hiểm cho xã hội được quy định trong BLHS, do người có năng lực trách nhiệm hình sự hoặc pháp nhân thương mại thực hiện một cách cố ý hoặc vô ý, xâm phạm các quan hệ xã hội được luật hình sự bảo vệ”. Nghĩa là, chỉ người nào thỏa mãn điều kiện về tuổi chịu TNHS theo quy định tại Điều 12 BLHS và không thuộc trường hợp quy định tại Điều 21 BLHS hoặc pháp nhân thương mại thỏa mãn Điều 75 và Điều 76 BLHS mới phải chịu TNHS. Theo pháp luật hình sự hiện hành, không có bất cứ quy định nào cho phép công nhận một phần mềm hay một hệ thống trí tuệ nhân tạo nào là chủ thể độc lập. Do đó, mặc dù AI có thể tự phát triển, tự ra quyết định gây hại hoặc làm trung gian cho các hành vi xâm phạm dữ liệu thì pháp luật hình sự Việt Nam chỉ truy cứu TNHS con người cụ thể hoặc pháp nhân thương mại thỏa mãn điều kiện được quy định tại Điều 75 và Điều 76 BLHS.

Hiện nay, trong bối cảnh AI, các tội phạm trong công nghệ thông tin, mạng viễn thông là các tội gần nhất với các hành vi xâm phạm dữ liệu có thể được áp dụng nếu như AI là công cụ để con người thực hiện các hành vi phạm tội. Tuy nhiên, trong phạm vi chịu TNHS của pháp nhân thương mại được quy định tại Điều 76 BLHS chưa bao gồm nhóm tội này. Như vậy, có thể thấy hiện nay, dù hành vi xâm phạm dữ liệu của AI thông qua quyết định của tổ chức, vì lợi ích của tổ chức thì TNHS vẫn được quy về cho cá nhân. Đây là một điểm bất cập quan trọng vì trên thực tế, nhiều hành vi xâm phạm dữ liệu trong bối cảnh AI được thực hiện bởi các tổ chức và mang lợi ích về cho doanh nghiệp.

3. AI không phải là chủ thể độc lập

Khi nói về công nghệ, những cụm từ rất phổ biến như “AI quyết định”, “AI phân biệt”, “AI làm lộ dữ liệu” thường xuất hiện, mặc dù đứng ở vai trò chủ ngữ, AI không được coi là chủ thể thực hiện hành vi phạm tội trong Luật Hình sự mà chỉ là những hiện tượng được mô tả về kỹ thuật¹. Như đã đề cập, BLHS hiện hành quy định, chủ thể của tội phạm phải là con người hoặc pháp nhân thương mại.

Simon Chesterman chỉ ra, việc tạo ra nhân cách pháp lý cho AI hoàn toàn có thể nhưng không nên làm như vậy. AI dù tự học, tự phát triển nhưng vẫn dựa trên các thuật toán và dữ liệu đầu vào, hoàn toàn không có khả năng hiểu được đúng - sai, hợp pháp - bất hợp pháp để lựa chọn hành vi, do không có ý chí (will) và nhận thức (normative understanding)². Bên cạnh đó, việc coi AI là một chủ thể độc lập có thể tạo ra “lỗ hổng pháp lý” cho các cá nhân và pháp nhân đứng sau AI, mọi trách nhiệm sẽ được đẩy sang AI.

Từ góc độ pháp luật hình sự Việt Nam, mục đích của hình phạt được quy định cụ thể tại Điều 31 BLHS, là “nhằm trừng trị người, pháp nhân thương mại phạm tội và giáo dục họ ý thức tuân theo pháp luật và các quy tắc của cuộc sống, ngăn ngừa họ phạm tội mới; giáo dục người, pháp nhân thương mại khác tôn trọng pháp luật, phòng ngừa và đấu tranh chống tội phạm”. Nếu chủ thể của tội phạm là AI, thì việc trừng trị và giáo dục không có ý nghĩa gì, vì AI không phải là con người, không có khả năng tiếp thu, nhìn nhận và sửa chữa sai lầm. Do đó, về khía cạnh hình sự, việc xác định TNHS và truy cứu TNHS phải quy về con người hoặc tổ chức đứng sau AI, còn AI chỉ có thể được coi như công cụ, phương tiện phạm tội³.

Trong bối cảnh AI, vấn đề không nằm ở việc cần có thêm một chủ thể mới để quy TNHS và trừng phạt, mà cần phải làm rõ hơn cơ chế xác định TNHS trong “vòng đời” của AI. Khi cơ chế tự học, tự phát triển và tự quyết định của AI dẫn đến hành vi xâm phạm dữ liệu thì vấn đề đặt ra đó là việc xác định chủ thể pháp luật nào trong các giai đoạn thiết kế, huấn luyện, triển khai, sử dụng AI phải chịu TNHS.

4. Xác định chủ thể chịu TNHS trong vòng đời của AI

Chủ thể của tội phạm xâm phạm dữ liệu trong bối cảnh trí tuệ nhân tạo là những cá nhân hiện diện trong vòng đời của AI, tùy theo từng tình huống cụ thể trên thực tế mà chủ thể có thể là người thu thập dữ liệu, người gán nhãn dữ liệu, người huấn luyện mô hình, người lập trình, người kiểm thử, người triển khai, người quản lý sản phẩm, người vận hành, người dùng, nhà cung cấp cloud, bên thứ ba tích hợp API... Hệ thống trí tuệ nhân tạo hoàn toàn có thể tạo ra những hành vi mà con người không trực tiếp ra lệnh, đây được Matthias gọi là “khoảng trống trách nhiệm”⁴. Mặc dù vậy, “khoảng trống trách nhiệm” không đồng nghĩa với việc không truy cứu TNHS. Việc truy cứu TNHS chỉ được chuyển từ “người trực tiếp thực hiện” sang “người có nghĩa vụ quyết định và kiểm soát” hoạt động của AI.

Để xác định cá nhân nào trong vòng đời của AI có thể là chủ thể của tội phạm, có thể căn cứ vào 5 tiêu chí sau:

Thứ nhất, người đó có nghĩa vụ giám sát, hạn chế hoặc ngăn chặn không cho AI gây ra các hậu quả nguy hiểm hay không. Các nghĩa vụ này có thể là nghĩa vụ pháp lý, nghề nghiệp, chức vụ, nhiệm vụ, được phát sinh từ các vị trí quản lý, quy trình nội bộ, hợp đồng dịch vụ, tiêu chuẩn nghề nghiệp hoặc từ quy định pháp luật về bảo vệ dữ liệu và quản trị AI.

Thứ hai, người đó biết hoặc buộc phải biết những rủi ro xảy ra, có thể dựa vào báo cáo kiểm tra, kết quả kiểm toán thuật toán, cảnh báo nội bộ hoặc phản ánh của người dùng.

Thứ ba, người đó có quyền và có khả năng ngăn chặn hay không. Quyền để ngăn chặn rủi ro xảy ra có thể là quyền thu hồi mô hình, tắt hệ thống, hạn chế quyền truy cập, vô hiệu hóa plugin...

Thứ tư, có mối quan hệ nhân quả giữa hành vi và hậu quả hay không. Trong bối cảnh AI, các hành vi không thực hiện hoặc thực hiện sai nghĩa vụ dẫn đến hậu quả nghiêm trọng có thể bị ẩn bởi nhiều tầng kỹ thuật như dữ liệu đầu vào, mô hình, API... Do đó, việc xác định và chứng minh mối quan hệ nhân quả này cần phải dựa trên chuỗi hoạt động kỹ thuật hoặc chuỗi quản trị hệ thống.

Và cuối cùng, hành vi của người này phải được quy định trong tội danh cụ thể theo BLHS. Điều 2 BLHS có quy định rõ “Chỉ người nào/pháp nhân thương mại nào phạm một tội đã được BLHS quy định mới phải chịu trách nhiệm hình sự”. Nếu một người vi phạm nghĩa vụ về bảo vệ dữ liệu hoặc quản trị AI, gây ra hậu quả nghiêm trọng nhưng hành vi này chưa được quy định trong BLHS thì không bị truy cứu TNHS. Điều kiện mang tính giới hạn này nhằm phân hóa rõ các trách nhiệm pháp lý, tránh hình sự hóa những hành vi không cần thiết.

5. Một số giải pháp hoàn thiện

Thứ nhất, công nhận AI là công cụ, phương tiện phạm tội hoặc là yếu tố trung gian để thực hiện tội phạm. Để tránh việc tạo ra lá chắn trách nhiệm cho các cá nhân và tổ chức đứng sau AI, pháp luật hình sự Việt Nam cần tiếp tục giữ vững lập trường AI không phải là chủ thể của tội phạm. TNHS cần được quy cho con người hoặc pháp nhân thương mại, và AI là công cụ, là môi trường để thực hiện hành vi phạm tội.

Thứ hai, pháp luật hình sự Việt Nam cần bổ sung các tiêu chí để xác định chủ thể chịu TNHS trong vòng đời của AI. AI không hoạt động đơn lẻ, do một người trực tiếp thực hiện mà được hình thành từ nhiều giai đoạn khác nhau. Để tránh quy TNHS một cách hình thức, cần có các tiêu chí cụ thể để cơ quan tiến hành tố tụng có thể xác định ai là người có hành vi, quyền và nghĩa vụ liên quan đến việc kiểm soát và quản trị AI cũng như mối quan hệ nhân quả với hậu quả dữ liệu bị xâm phạm.

Thứ ba, phạm vi chịu TNHS của pháp nhân thương mại được quy định tại Điều 76 BLHS cần được mở rộng. Việc các tội phạm về công nghệ thông tin và dữ liệu không được liệt kê trong Điều 76 BLHS tạo ra khoảng trống khá lớn khi hành vi và lợi ích của tổ chức nhưng TNHS lại quy về cho cá nhân. Không ít trường hợp các lập trình viên làm việc theo chỉ đạo của tổ chức, doanh nghiệp. Do đó, việc quy TNHS về cho cá nhân là thiếu tính công bằng, tạo sức ép lên cá nhân cấp dưới và có thể là một trong các lý do khiến chảy máu chất xám tại Việt Nam.

Thứ tư, cần có sự kết nối giữa Luật Hình sự với các pháp luật về dữ liệu để mở rộng phạm vi truy cứu TNHS. Luật Bảo vệ dữ liệu cá nhân năm 2025 và Nghị định số 356/2025/NĐ-CP đặt ra nhiều nguyên tắc và nghĩa vụ đối với hoạt động xử lý dữ liệu cá nhân để đảm bảo quyền của chủ thể dữ liệu và trách nhiệm của bên kiểm soát hoặc bên xử lý dữ liệu. Tuy nhiên, việc xử lý vi phạm hành chính trong một số trường hợp vẫn là quá nhẹ so với hậu quả nghiêm trọng của hành vi gây ra. Căn cứ vào những yếu tố, như: mục đích vụ lợi, mục đích gây hại của hành vi; quy mô, tính chuyên nghiệp của hành vi; tính chất của dữ liệu bị xâm hại; và hậu quả nghiêm trọng của hành vi để có thể hình sự hóa những hành vi xâm phạm dữ liệu, tránh trường hợp xử lý quá nhẹ những hành vi thật sự nguy hiểm cho xã hội.

6. Kết luận

Mặc dù, hành vi xâm phạm dữ liệu trong bối cảnh AI khác hoàn toàn với các hành vi truyền thống, nguyên tắc nền tảng của Luật Hình sự Việt Nam vẫn chưa thay đổi. AI không phải là chủ thể độc lập của tội phạm, TNHS chỉ đặt ra đối với cá nhân hoặc pháp nhân thương mại. Hiện nay, AI chỉ được xem là công cụ hoặc bối cảnh phạm tội, điều này không làm TNHS biến mất mà chỉ làm quá trình xác định chủ

thể của tội phạm trở nên phức tạp hơn. Các cơ quan tiến hành tố tụng cần làm rõ vai trò và nghĩa vụ của những người xuất hiện trong vòng đời của AI cũng như mối quan hệ nhân quả giữa hành vi của họ và hậu quả xảy ra trên thực tế. Đối với pháp nhân thương mại, khoảng trống về phạm vi chịu TNHS cho các hành vi xâm phạm dữ liệu cần được sửa đổi, bổ sung để đảm bảo xử lý đúng người, đúng tội, tránh để AI trở thành công cụ né tránh TNHS ■

TÀI LIỆU TRÍCH DẪN:

¹Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 177.

²Chesterman, S. (2020). Artificial intelligence and the limits of legal personality. *International and Comparative Law Quarterly*, 69(4), 842-843.

³Simmler, M. (2024). Ensuring accountability for robots and AI under criminal law. W. Barfield, Y.-H. Weng & U. Pagallo (Eds.), *The Cambridge Handbook of the Law, Policy, and Regulation for Human - Robot Interaction*, Cambridge University Press, 798-812.

⁴Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 177.

TÀI LIỆU THAM KHẢO:

Chính phủ. (2021). Quyết định số 127/QĐ-TTg ngày 26/01/2021 ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030.

Chính phủ. (2025). Nghị định số 165/2025/NĐ-CP ngày 30/06/2025 quy định chi tiết một số điều và biện pháp thi hành Luật Dữ liệu.

Chính phủ. (2025). Nghị định số 356/2025/NĐ-CP ngày 31/12/2025 quy định chi tiết một số điều và biện pháp thi hành Luật Bảo vệ dữ liệu cá nhân.

Quốc hội. (2018). Luật số 24/2018/QH14: Luật An ninh mạng, ban hành ngày 12/06/2018.

Quốc hội. (2024). Luật số 60/2024/QH15: Luật Dữ liệu, ban hành ngày 30/11/2024.

Quốc hội. (2025). Luật số 91/2025/QH15: Luật Bảo vệ dữ liệu cá nhân, ban hành ngày 26/06/2025.

Quốc hội. (2025). Luật số 116/2025/QH15: Luật An ninh mạng, ban hành ngày 10/12/2025.

Quốc hội. (2025). Luật số 134/2025/QH15: Luật Trí tuệ nhân tạo, ban hành ngày 10/12/2025.

Văn phòng Quốc hội. (2025). Văn bản hợp nhất số 135/VBHN-VPQH ngày 05/09/2025: Bộ luật Hình sự.

Chesterman, S. (2020). Artificial intelligence and the limits of legal personality. *International and Comparative Law Quarterly*, 69(4), 819-844. <https://doi.org/10.1017/S0020589320000366>.

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). *Official Journal of the European Union*.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175-183. <https://doi.org/10.1007/s10676-004-3422-1>.

Simmler, M. (2024). Ensuring accountability for robots and AI under criminal law. In W. Barfield, Y.-H. Weng, & U. Pagallo (Eds.), *The Cambridge Handbook of the Law, Policy, and Regulation for Human-Robot Interaction* (798-812). Cambridge University Press. <https://doi.org/10.1017/9781009386708.050>.