

HIỆN TƯỢNG “ẢO GIÁC TRÍ TUỆ NHÂN TẠO” VÀ NGUY CƠ VI PHẠM SỞ HỮU TRÍ TUỆ - TRÁCH NHIỆM PHÁP LÝ TRONG MÔI TRƯỜNG SỐ VÀ MỘT SỐ KIẾN NGHỊ HOÀN THIỆN QUY ĐỊNH PHÁP LUẬT

TS NGUYỄN PHƯƠNG THẢO

Khoa Luật Dân sự, Trường Đại học Luật Tp. Hồ Chí Minh,
Faculty of Civil Law, Ho Chi Minh City University of Law,

Email: npthao@hcmulaw.edu.vn

NGUYỄN THỊ KIM HOÀNG

Phòng pháp chế Công ty CP xây dựng thương mại Chính Bắc,
Legal Department, Chinh Bac Construction and Trading Joint Stock Company,

Email: 2153801012087@email.hcmulaw.edu.vn

Tóm tắt

Hiện tượng “AI hallucinations” hay còn gọi là ảo giác AI, xảy ra khi trí tuệ nhân tạo (AI) tạo ra những thông tin sai lệch hoặc bịa đặt, dẫn đến những nguy cơ rủi ro vi phạm về sở hữu trí tuệ (SHTT). Điều này đặt ra câu hỏi phân định về trách nhiệm pháp lý của các bên liên quan, bao gồm cả nhà phát triển AI, bên trung gian cung cấp dịch vụ và người sử dụng các nền tảng công nghệ AI. Trên cơ sở đó, bài viết phân tích quy định pháp luật hiện hành về SHTT của Việt Nam, nhận diện những khoảng trống pháp lý và tham khảo kinh nghiệm của một số quốc gia. Từ đó, bài viết đưa ra những kiến nghị hoàn thiện quy định pháp luật, đảm bảo cân bằng hài hòa giữa phát triển công nghệ và bảo vệ quyền SHTT trong kỷ nguyên số.

Từ khóa: hiện tượng “AI Hallucinations”, sở hữu trí tuệ, trách nhiệm pháp lý, kinh nghiệm của một số quốc gia

Abstract

The phenomenon of “AI hallucinations” occurs when artificial intelligence (AI) generates false or fabricated information, leading to potential risks of intellectual property (IP) infringement. This issue raises critical questions regarding the legal liability of stakeholders, including AI developers, intermediary AI service providers, and end-users. Based on this context, the article analyzes Vietnam's current intellectual property laws, identifies existing legal gaps, and consults the international experiences of several countries. Consequently, the paper proposes recommendations to refine legal regulations, ensuring a harmonious balance between technological advancement and the protection of IP rights in the digital era.

Keywords: AI hallucinations, intellectual property, legal liability, legal experience of several countries

DOI: <https://doi.org/10.70236/khplvn.737>

Ngày nhận bài: 13/10/2025

Ngày duyệt đăng: 02/04/2026

Trước sự phát triển mạnh mẽ của AI, đặc biệt là các mô hình AI tạo sinh (generative AI)¹ đã cho thấy khả năng ứng dụng rộng rãi của khoa học công nghệ trong nhiều lĩnh vực như giáo dục, y tế, khoa học kỹ thuật, đặc biệt là dịch vụ pháp lý... Tuy nhiên, cùng với khả năng tự động học sâu (deep learning), học máy (machine learning) và việc tự tạo ra các nội dung, những hệ thống AI hiện đại cũng phát sinh những vấn đề đáng quan tâm là hiện tượng

¹ Generative AI (AI tạo sinh) là một lĩnh vực tiên tiến trong trí tuệ nhân tạo với khả năng tạo ra dữ liệu mới như văn bản, hình ảnh, âm thanh, video, dựa trên các mẫu dữ liệu có sẵn. Khác với các hệ thống AI truyền thống chỉ thực hiện phân loại hoặc dự đoán, Generative AI có thể sáng tạo và tạo ra nội dung mới gần giống với dữ liệu thực tế. Một số ví dụ phổ biến về AI tạo sinh có thể kể đến như: ChatGPT, Dall.E, Bard, Midjourney, Copilot, Gemini, Grok...

ảo giác AI (*AI Hallucinations*). Đây chính là việc AI tự tổng hợp và tạo ra những thông tin sai lệch, mang tính bịa đặt, không có thật nhưng được trình bày như thể chính xác và đáng tin cậy. Dưới góc độ pháp lý, hiện tượng ảo giác AI đặt ra nhiều vấn đề đáng quan ngại đối với mỗi thông tin mà người dùng tiếp cận, đặc biệt là trong lĩnh vực sở hữu trí tuệ (SHTT). Việc AI tự động tạo ra các nội dung chứa yếu tố vi phạm bản quyền hoặc tái hiện lại các tác phẩm đã được bảo hộ và trích dẫn sai lệch nguồn tài liệu có thể dẫn đến hành vi vi phạm quyền SHTT.

Bên cạnh đó, về phương diện lập pháp, nhiều quốc gia có hệ thống pháp luật phát triển đã chủ động xây dựng khuôn khổ pháp lý điều chỉnh hoạt động của AI và các hệ quả pháp lý phát sinh từ công nghệ này. Điển hình như Liên minh châu Âu (*European Union*, EU) với Đạo luật về AI cùng các điều khoản cụ thể về nghĩa vụ và trách nhiệm pháp lý của nhà cung cấp dịch vụ AI nhằm bảo đảm an toàn, minh bạch và trách nhiệm giải trình trong quá trình triển khai và khai thác công nghệ. Tại Hoa Kỳ, mặc dù chưa có đạo luật riêng và thống nhất để điều chỉnh về AI, các vụ việc thực tiễn đã được Tòa án giải quyết trên cơ sở các quy định hiện hành và các vụ việc điển hình, đã gián tiếp hình thành khung trách nhiệm pháp lý đối với hành vi sử dụng AI. Từ những kết quả nghiên cứu này, nhóm tác giả tập trung nghiên cứu hiện tượng ảo giác AI dưới góc nhìn SHTT, phân tích và đánh giá trách nhiệm pháp lý thông qua các vụ việc thực tiễn, đồng thời đưa ra các kiến nghị hoàn thiện quy định pháp luật nhằm bảo vệ quyền và lợi ích hợp pháp của các bên liên quan, bảo đảm khai thác hiệu quả quyền SHTT trong kỷ nguyên số.

1. Hiện tượng ảo giác AI dưới góc nhìn pháp lý

1.1. Khái niệm và cơ chế kỹ thuật của hiện tượng ảo giác AI

Thuật ngữ ảo giác AI được hiểu là hiện tượng các mô hình ngôn ngữ lớn (*Large Language Models*, LLM), thường thể hiện dưới dạng chatbot AI tạo sinh hoặc công cụ thị giác máy tính, tự tạo ra thông tin không đúng sự thật, nội dung thiếu căn cứ hoặc sai lệch với thực tế, nhưng lại được trình bày bằng cấu trúc ngôn ngữ logic, mạch lạc, dễ gây nhầm lẫn về tính xác thực. Hiện tượng này dẫn đến việc mô hình cung cấp đầu ra sai nhưng thể hiện với độ tin cậy cao về mặt ngữ nghĩa và hình thức, tiềm ẩn nguy cơ tác động tiêu cực đối với quyền lợi ích hợp pháp của tổ chức, cá nhân trong môi trường số.² Theo đó, nhóm tác giả nhận định rằng thuật ngữ của hiện tượng ảo giác AI có thể được diễn giải theo cách hiểu thông thường là hiện tượng mô hình AI tạo sinh đưa ra kết quả không trùng khớp với dữ liệu được đào tạo hoặc không tồn tại trên thực tế. Khác với những lỗi kỹ thuật thông thường, hiện tượng này không hoàn toàn xuất phát từ lỗi hệ thống mà đây là một đặc điểm vốn có của các mô hình xác suất trong học sâu, trong đó AI không truy xuất dữ liệu theo cách logic có kiểm chứng mà dự đoán từ ngữ tiếp theo dựa trên xác suất thống kê từ tập dữ liệu được đào tạo. Điều này có thể dẫn đến tình trạng “ảo hóa về tính

2 Adele Athiyaman, “Artificial intelligence in community economic development: The invisible danger of AI hallucinations”, *Research Brief*, Vol. 7(7), 2025; Chrystal R. China, “What are AI hallucinations?”, *IBM*, 2023, <https://www.ibm.com/think/topics/ai-hallucinations>, truy cập ngày 21/04/2025.

xác thực”, đặc biệt nguy hiểm trong các lĩnh vực giáo dục, y tế, tài chính và pháp luật khi mà thông tin sai lệch có thể gây hậu quả pháp lý nghiêm trọng và ảnh hưởng đến các vấn đề xã hội. Về mặt pháp lý, hệ quả của ảo giác AI chính là việc tạo ra các nội dung không chính xác, không kiểm chứng được nguồn gốc, từ đó làm phát sinh nguy cơ xâm phạm quyền SHTT, điển hình như: trích dẫn sai, tạo mới nội dung vi phạm quyền tác giả hoặc thu thập tổng hợp và sử dụng trái phép các sản phẩm có bản quyền. Bởi do các mô hình AI hiện nay được đào tạo từ các kho dữ liệu trực tuyến khổng lồ nên có nguy cơ tái sử dụng các đoạn văn, hình ảnh và nội dung gốc mà không có cơ chế kiểm tra bản quyền, gây ra tranh chấp về quyền SHTT.

1.2. Tính chất pháp lý của nội dung do AI tạo ra

Theo quy định tại Điều 6 và Điều 13 Luật SHTT năm 2005 (sửa đổi, bổ sung năm 2009, 2019, 2022 và 2025)³ (viết tắt là Luật SHTT năm 2005) quyền tác giả phát sinh kể từ thời điểm tác phẩm được sáng tạo và được thể hiện dưới hình thức vật chất nhất định, bởi tác giả là cá nhân tức một chủ thể có tư cách pháp lý dân sự theo quy định của Bộ luật Dân sự năm 2015 (BLDS năm 2015). Bên cạnh đó, tại Điều 13 khoản 1 còn quy định “Tổ chức cá nhân có tác phẩm được bảo hộ quyền tác giả gồm người trực tiếp sáng tạo ra tác phẩm và chủ sở hữu quyền theo các Điều 37 đến Điều 42”,⁴ có thể thấy pháp luật chỉ công nhận hai nhóm chủ thể được hưởng quyền tác giả: (i) Người trực tiếp sáng tạo ra tác phẩm, tức cá nhân có hoạt động sáng tạo mang tính chủ quan và (ii) Chủ sở hữu quyền tác giả theo các trường hợp được chuyển giao, thỏa thuận, giao kết hợp đồng hoặc theo quy định pháp luật, chẳng hạn như tác phẩm được tạo ra theo nhiệm vụ công vụ hoặc hợp đồng lao động. Và do không đáp ứng tiêu chí “chủ thể là con người” theo quy định pháp luật nên AI không thể được xem là chủ thể có năng lực hành vi dân sự để thực hiện hoạt động sáng tạo một cách có ý thức.

Trước đây pháp luật SHTT Việt Nam chưa có những quy định cụ thể điều chỉnh những vấn đề về nội dung do AI tạo ra. Tuy nhiên, với Nghị định 100/2026/NĐ-CP,⁵ Luật Sở hữu trí tuệ năm 2005 (sửa đổi, bổ sung năm 2009, 2019, 2022 và 2025 – Gọi là Luật SHTT hiện hành)⁶ cùng các văn bản khác có liên quan đã có những điều khoản điều chỉnh những vấn đề liên quan đến AI. Cụ thể, tại Điều 15 của Nghị định 100/2026/NĐ-CP đã có những bổ sung Điều 10a vào sau Điều 10 của Nghị định 65/2023/NĐ-CP. Theo đó tại “Điều 10a. Quyền đối với đối tượng quyền sở hữu công nghiệp được tạo ra có sử dụng hệ thống trí tuệ nhân tạo” đã thể hiện tư duy lập pháp nhạy bén với xu hướng của khoa học công nghệ hiện đại bằng cách khẳng định vai trò trung tâm của con người trong quá trình sáng tạo. Theo khoản 1, hệ thống AI

3 Hiện nay được tổng hợp theo Văn bản hợp nhất số 155/VBHN-VPQH ngày 09/9/2025.

4 Điều 13 Luật Sở hữu trí tuệ năm 2005 (sửa đổi, bổ sung năm 2009, 2019 và 2022).

5 Nghị định số 100/2026/NĐ-CP ngày 31/3/2026 sửa đổi, bổ sung một số điều của Nghị định số 65/2023/NĐ-CP ngày 23/8/2023 của Chính phủ quy định chi tiết một số điều và biện pháp thi hành Luật Sở hữu trí tuệ về sở hữu công nghiệp, bảo vệ quyền sở hữu công nghiệp, quyền đối với giống cây trồng và quản lý nhà nước về sở hữu trí tuệ được sửa đổi, bổ sung bởi Nghị định số 15/2026/NĐ-CP ngày 14/1/2026 và Nghị định số 33/2026/NĐ-CP ngày 21/1/2026 của Chính phủ (Có hiệu lực thi hành ngày 01/4/2026).

6 Hiện nay được tổng hợp theo Văn bản hợp nhất số 155/VBHN-VPQH ngày 09/9/2025.

chỉ được xem là công cụ hỗ trợ và quyền sở hữu công nghiệp đối với sáng chế, kiểu dáng công nghiệp chỉ được xác lập khi con người có “đóng góp đáng kể” vào việc tạo ra đối tượng đó. Cách tiếp cận này giúp bảo vệ triết lý cốt lõi của luật SHTT là tôn vinh sự sáng tạo của con người, đồng thời ngăn chặn việc AI được công nhận là một chủ thể quyền độc lập vốn là một vấn đề pháp lý phức tạp và đầy tranh cãi trên phạm vi toàn cầu hiện nay. Việc cụ thể hóa tiêu chuẩn “đóng góp đáng kể” tại khoản 2 là một quy định thiết thực và tất yếu trong kỷ nguyên số, giúp quy định này có tính khả thi cao. Thay vì để khái niệm này mang tính định tính trừu tượng, Điều 10a đã liệt kê rõ ràng các quy trình sáng tạo, thiết kế từ khâu xác định ý tưởng, lựa chọn dữ liệu đầu vào, đến việc đánh giá và tinh chỉnh kết quả để làm thay đổi cấu trúc hoặc chức năng cốt lõi của sản phẩm sáng tạo.⁷ Với quy định này, người sử dụng AI buộc phải tham gia thực chất vào quá trình tư duy và phải có những đóng góp nhất định trong quá trình sáng tạo, tránh tình trạng vi phạm liên chính khoa học và “trục lợi” bảo hộ đối với những kết quả thuần túy do máy móc tạo ra mà không có giá trị sáng tạo mới từ trí tuệ của con người. Bên cạnh đó, khoản 3 và khoản 4 đã đưa ra các giải pháp dung hòa lợi ích một cách linh hoạt. Đối với những sản phẩm hỗ trợ bởi AI không đủ tiêu chuẩn bảo hộ quyền sở hữu công nghiệp, pháp luật vẫn cho phép quyền sử dụng và chuyển giao, giúp tận dụng tối đa lợi ích từ công nghệ cho phát triển xã hội mà không gây lãng phí nguồn lực. Từ những phân tích trên, có thể kết luận rằng AI không thể trở thành chủ thể của quyền tác giả, bởi AI không có năng lực pháp luật và hành vi dân sự, do đó AI không thể là tác giả về mặt pháp lý. Vì vậy, các nội dung do AI tạo ra nếu không có sự can thiệp sáng tạo đáng kể từ con người sẽ không thể phát sinh quyền tác giả theo pháp luật SHTT Việt Nam hiện hành.

Bên cạnh đó, trong thực tiễn đã đặt ra nhiều vướng mắc, nhất là trong các trường hợp AI tái tạo hoặc mô phỏng lại tác phẩm đã được bảo hộ, dẫn đến nguy cơ xâm phạm quyền tác giả. Điển hình như nếu một hệ thống AI ảo giác tạo thông tin sai bằng cách tái tạo lại phần nội dung có bản quyền mà không có sự cho phép của chủ thể quyền, thì câu hỏi đặt ra: chủ thể nào sẽ phải chịu trách nhiệm pháp lý trong môi trường số? Người sử dụng hay người cung cấp dịch vụ AI. Đó là vấn đề còn nhiều bỏ ngỏ trong chế tài xử lý đối với những vi phạm bản quyền có sự can thiệp của AI. Theo đó, Văn phòng Bản quyền Hoa Kỳ (*US Copyright Office*), các nội dung không có yếu tố sáng tạo của con người sẽ không được cấp bản quyền. Điển hình như qua vụ việc của *Zarya of the Dawn* (2023), đơn đăng ký bản quyền cho tác phẩm truyện tranh có hình ảnh tạo từ AI Midjourney đã bị từ chối một phần vì lý do thiếu yếu tố con người.⁸ Ngược lại, trong trường hợp hệ thống AI khai thác và tái tạo các nội dung đã được đăng ký bản quyền, chẳng hạn như trích dẫn lại đoạn văn từ sách, báo hoặc tài liệu học thuật, thì nguy cơ xâm phạm quyền tác giả trở nên rõ ràng và dễ nhận diện hơn. Một trong những vụ kiện nổi bật gần đây là *The New*

7 Khoản 2 của Điều 10a Nghị định số 100/2026/NĐ-CP.

8 Tony Analla, “Zarya of the dawn: How AI is changing the landscape of copyright protection”, *Harvard JOLT Digest*, 2023, <https://jolt.law.harvard.edu/digest/zarya-of-the-dawn-how-ai-is-changing-the-landscape-of-copyright-protection>, truy cập ngày 21/04/2025.

York Times kiện Microsoft và OpenAI (2023), trong đó nguyên đơn là New York Times, cùng với Trung tâm Nghiên cứu Điều tra (*Center for Investigative Research Inc.*) cáo buộc ChatGPT của OpenAI đã vi phạm bản quyền làm suy yếu thương hiệu, vi phạm Đạo luật Bản quyền Thiên niên kỷ Kỹ thuật số (*Digital Millennium Copyright Act, DMCA*) và cạnh tranh không lành mạnh liên quan đến chiếm đoạt trái phép. Các tổ chức báo chí cáo buộc các bị cáo huấn luyện các LLM của họ bằng cách sử dụng nội dung thu thập từ internet, bao gồm cả nội dung có bản quyền từ các trang web của nguyên đơn.⁹ Như vậy có thể thấy những nội dung do AI tạo ra và những phát sinh từ AI cũng là những vấn đề đáng quan tâm trong học thuật và khoa học pháp lý, đặc biệt là lĩnh vực SHTT khi mà AI được tích hợp nhiều khả năng vượt trội có thể xâm phạm đến quyền SHTT của các chủ thể.

1.3. Quy định pháp luật của một số quốc gia trong việc quy định về trách nhiệm của các bên đối với nội dung do AI tạo ra

Tại Hoa Kỳ, hiện nay chưa có đạo luật riêng và thống nhất ở cấp Liên bang điều chỉnh về AI, tuy nhiên, các vấn đề liên quan đến nội dung do AI tạo ra được điều chỉnh chủ yếu qua các quy định pháp lý hiện hành, tùy theo từng lĩnh vực như bảo vệ người tiêu dùng, trách nhiệm dân sự, quyền riêng tư hay SHTT. Luật Bản quyền của Hoa Kỳ (the U.S. Copyright Act) (17 U.S.C. § 101 et seq.),¹⁰ chỉ công nhận bảo hộ đối với những tác phẩm có sự “sáng tạo của con người” theo nguyên tắc tác giả là con người (*human authorship*). Do đó, nếu AI tạo ra một tác phẩm mà không có sự sáng tạo đáng kể từ con người, tác phẩm đó sẽ không đủ điều kiện để được bảo hộ quyền tác giả. Hiểu một cách đơn giản hơn, để được bảo hộ, tác phẩm đó phải là kết quả của quá trình “lao động sáng tạo thực chất”.¹¹ Bên cạnh đó, trách nhiệm pháp lý đối với việc vi phạm bản quyền sẽ thuộc về người sử dụng AI, cụ thể là những người vận hành, phát triển hoặc sử dụng công nghệ AI mà không tuân thủ các quy định về quyền SHTT. Điều này có thể dẫn đến tình huống các cá nhân hoặc tổ chức chịu trách nhiệm nếu AI của họ tạo ra nội dung vi phạm quyền tác giả.¹²

Tại EU, ngày 01/08/2024 Đạo luật về AI chính thức có hiệu lực và được xem là một trong những bước đi đầu tiên trong việc xây dựng khung pháp lý toàn diện cho AI. Đạo luật này yêu cầu các nhà cung cấp AI phải tuân thủ các nghĩa vụ kiểm soát rủi ro, đặc biệt đối với các hệ thống AI có tác động lớn đến các quyền cơ bản của con người (Điều 9, Điều 16, Điều 50, Điều 53 và Điều 55 cùng các điều khoản liên quan).¹³ Trường hợp hệ thống AI không tuân thủ nghĩa vụ kiểm soát rủi ro, nhà cung cấp có thể bị truy cứu trách nhiệm, áp dụng chế tài xử phạt và họ phải chịu trách nhiệm đối với các hành vi vi phạm

9 Blake Brittain, “Judge explains order for New York Times in OpenAI copyright case”, *Reuters*, 2025, <https://www.reuters.com/legal/litigation/judge-explains-order-new-york-times-openai-copyright-case-2025-04-04/>, truy cập ngày 21/04/2025.

10 U.S. Copyright Act of 1976, 17 U.S.C. §§ 101 et seq.

11 Nguyễn Phương Thảo, Nguyễn Thị Xuân Hoa, Nguyễn Thị Kim Hoàng, “Bảo hộ tác phẩm tạo ra từ trí tuệ nhân tạo theo pháp luật Hoa Kỳ và gợi mở hoàn thiện pháp luật Việt Nam”, *Tạp chí Dân luật và Pháp luật*, số 432(7), 2025, tr. 75.

12 Copyright Law of the United States (Title 17).

13 EU Artificial Intelligence Act, <https://www.artificial-intelligence-act.com/>, truy cập ngày 21/04/2025.

của AI, bao gồm cả lỗi ảo giác AI trong trường hợp tạo ra thông tin sai lệch hoặc xâm phạm quyền tác giả. Quy định này khẳng định trách nhiệm pháp lý của nhà cung cấp trong trường hợp vi phạm do AI gây ra, điều này phản ánh tầm quan trọng của trách nhiệm các bên trong bối cảnh số hóa khi mà việc sử dụng AI trong các lĩnh vực như SHTT, tài chính và y tế¹⁴ ngày càng phổ biến. Bên cạnh đó, cũng như nhóm tác giả đã đề cập, tại Việt Nam pháp luật chưa có quy định riêng điều chỉnh các nội dung do AI tạo ra, đặc biệt là trường hợp ảo giác AI. Luật SHTT hiện hành chỉ tập trung vào việc bảo vệ quyền tác giả đối với các tác phẩm có sự sáng tạo của con người. Cụ thể, tại Điều 13 của Luật SHTT hiện hành quy định chỉ có người sáng tạo và chủ sở hữu quyền tác giả được công nhận là chủ thể quyền tác giả, không có quy định đối với tác phẩm tạo ra bởi AI. Điều này tạo ra khoảng trống pháp lý đối với việc điều chỉnh trách nhiệm pháp lý khi AI tạo ra nội dung vi phạm quyền SHTT. Do đó, cần có sự điều chỉnh bổ sung trong các quy định pháp lý để đảm bảo việc bảo vệ quyền và lợi ích hợp pháp của các bên liên quan trong môi trường số.

Từ phân tích trên có thể thấy, Hoa Kỳ và EU tuy khác nhau về quy định pháp luật điều chỉnh nhưng đều có những điểm chung trong việc quy định trách nhiệm đối với nội dung do AI tạo ra. Hoa Kỳ vẫn dựa chủ yếu vào hệ thống pháp luật hiện hành, điển hình là Luật Bản quyền, chỉ bảo hộ những tác phẩm có sự sáng tạo của con người và quy trách nhiệm chính cho người sử dụng, vận hành hoặc phát triển AI khi nội dung đầu ra xâm phạm quyền SHTT. Trong khi đó, EU với Đạo luật AI đã xây dựng khung pháp lý chủ động hơn, đặt ra nghĩa vụ kiểm soát rủi ro cho nhà cung cấp hệ thống AI và cho phép truy cứu trách nhiệm của họ khi hệ thống gây ra hậu quả, bao gồm cả trường hợp ảo giác AI dẫn đến thông tin sai lệch hoặc xâm phạm quyền tác giả. Cách tiếp cận của Hoa Kỳ dựa trên những vụ việc thực tiễn và án lệ còn với EU, cách quy định pháp luật có phần mở rộng và chi tiết hơn. Việc nghiên cứu quy định của Hoa Kỳ và EU về trách nhiệm đối với nội dung do AI tạo ra có ý nghĩa quan trọng đối với Việt Nam trong việc nhận diện khoảng trống pháp lý hiện nay và định hướng hoàn thiện khung pháp lý về phân định trách nhiệm giữa các bên liên quan khi AI tạo ra nội dung xâm phạm SHTT, đặc biệt trong trường hợp ảo giác AI.

2. Thực tiễn và hệ quả pháp lý từ hiện tượng ảo giác AI

*Vụ Mata v. Avianca (2023) của Tòa án quận Hoa Kỳ, SD New York: Luật sư bị phạt vì trích dẫn án lệ không có thật do ChatGPT tạo ra.*¹⁵ Vào năm 2022, tại Tòa án quận Hoa Kỳ, SD New York, ông Roberto Mata (Nguyên đơn) đã kiện hãng hàng không Avianca, Inc (Bị đơn) vì bị xe đẩy trên máy bay gây thương tích trong chuyến bay đến sân bay quốc tế. Vấn đề phát sinh khi luật sư đại

14 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance).

15 United States District Court, S.D. New York., *MATA v. AVIANCA INC*, No. 1:22-cv-01426 (S.D.N.Y. 2023), <https://caselaw.findlaw.com/court/us-dis-crt-sd-new-yor/2335142.html>, truy cập ngày 22/04/2025.

diện nguyên đơn, ông Steven Schwartz – Công ty Luật Levidow đệ trình bản luận cứ pháp lý có trích dẫn sáu án lệ nhằm bác bỏ đề nghị của bị đơn. Tuy nhiên, khi bị đơn có đơn phản bác, Tòa án phát hiện toàn bộ sáu án lệ không có thật, do luật sư sử dụng ChatGPT để hỗ trợ mà không xác minh lại. Luật sư Schwartz thừa nhận rằng ông không quen thuộc với công nghệ AI và tin tưởng rằng ChatGPT cung cấp thông tin chính xác. Ông còn đính kèm ảnh chụp màn hình cho thấy ChatGPT xác nhận các án lệ là có thật, trong khi đó chỉ là “ảo giác AI”. Tòa án xử phạt hành chính 5.000 USD, yêu cầu luật sư và văn phòng đại diện thông báo quyết định kỷ luật cho các cơ quan có liên quan, đồng thời phải tham gia khóa đào tạo bắt buộc về đạo đức nghề nghiệp và công nghệ. Tương tự như vụ việc trên, vụ việc *Wadsworth v. Walmart Inc.* (2025) của Tòa án quận Hoa Kỳ, D. Wyoming¹⁶ cũng là một ví dụ điển hình. Vụ việc liên quan đến các luật sư từ Công ty Morgan & Morgan sử dụng AI (ChatGPT) để tạo ra các trích dẫn án lệ không có thật trong một vụ kiện liên quan đến sản phẩm lỗi. Nguyên đơn là Stephanie Wadsworth và Matthew Wadsworth (người giám hộ cho các con) khởi kiện Walmart Inc. và Jetson Electric Bikes, LLC với yêu cầu bồi thường thiệt hại liên quan đến sản phẩm hoverboard (một chiếc ván trượt điện) bị lỗi dẫn đến bốc cháy và thiêu rụi nhà của Nguyên đơn. Trong quá trình chuẩn bị tài liệu pháp lý, luật sư Rudwin Ayala đã sử dụng AI để tạo ra các trích dẫn án lệ, nhưng những trích dẫn này không tồn tại, do hệ thống AI bịa đặt, không có cơ sở pháp lý thực tế. Tòa án liên bang Wyoming đã ra phán quyết xử phạt các luật sư từ Morgan & Morgan vì không kiểm tra tính chính xác của các tài liệu pháp lý trước khi đệ trình lên tòa. Tòa án đã xử phạt tiền đối với các luật sư và cũng đã chỉ trích các luật sư vì đã không thực hiện các bước kiểm tra tính xác thực của các trích dẫn pháp lý trước khi ký tên vào tài liệu giao nộp.¹⁷ Có thể thấy hành vi trích dẫn các án lệ không có thật từ AI không thể coi là một sự cố thông thường bởi điều này cho thấy sự chủ quan của các luật sư trong việc sử dụng công nghệ mới trong các ngành nghề yêu cầu tính chính xác cao như pháp luật.

*Vụ Robby Starbuck kiện Meta (2024): Công nghệ AI của Meta đã lan truyền thông tin sai lệch cho người dùng:*¹⁸ Trong vụ việc này, ông Robby Starbuck khởi kiện Meta Platforms, Inc. tại Tòa án Tối cao bang Delaware, yêu cầu bồi thường 5 triệu USD với cáo buộc hệ thống AI do Meta phát triển đã tạo lập và lan truyền thông tin sai lệch liên quan đến việc ông tham gia vụ bạo loạn Điện Capitol ngày 06/01/2021 và có liên hệ với các tổ chức cực đoan như QAnon và Holocaust denial. Mặc dù ông Starbuck nhiều lần yêu cầu gỡ bỏ nội dung,

16 Bernie Pazanowski, “Morgan & Morgan lawyers fined for ‘Hallucinated’ AI Citations”, *Bloomberg Law*, 2025, <https://news.bloomberglaw.com/litigation/morgan-morgan-lawyers-fined-for-hallucinated-ai-citations?>, truy cập ngày 22/04/2025.

17 Thẩm phán Tòa án quận Hoa Kỳ Kelly Rankin tại Wyoming cho biết “Khi các luật sư chuyển sang thế giới AI, nhiệm vụ kiểm tra nguồn tin và thực hiện cuộc điều tra hợp lý về luật hiện hành vẫn không thay đổi” (tạm dịch). Xem Sara Merken, “Judge fines lawyers in Walmart lawsuit over fake, AI-generated cases”, 2025, *Reuters*, <https://www.reuters.com/legal/government/judge-fines-lawyers-walmart-lawsuit-over-fake-ai-generated-cases-2025-02-25/?>, truy cập ngày 30/04/2025.

18 David Spector, “Robby Starbuck, Meta settle lawsuit over AI chatbot defamation claim”, 2025, <https://www.foxbusiness.com/media/robby-starbuck-meta-settle-lawsuit-after-ai-chatbot-defamed-him>, truy cập ngày 14/8/2026.

AI của Meta vẫn tiếp tục phát tán thông tin, gây ảnh hưởng nghiêm trọng đến danh dự và an toàn của ông và gia đình. Đây là một trong những vụ kiện đầu tiên tại Hoa Kỳ nhằm trực tiếp truy cứu trách nhiệm của nhà cung cấp AI, phản ánh sự dịch chuyển pháp lý từ việc xử lý người dùng sang yêu cầu trách nhiệm quản trị, kiểm soát nội dung và cơ chế bồi thường của doanh nghiệp công nghệ. Dù chưa có phán quyết cuối cùng, song vụ kiện trên cho thấy đây là vụ việc điển hình như một tiền lệ quan trọng, đặt ra yêu cầu cấp thiết đối với các doanh nghiệp công nghệ phải kiểm soát nội dung do AI tạo ra, bảo đảm cơ chế minh bạch, trách nhiệm và bồi thường khi xảy ra thiệt hại.

Vụ Getty Images kiện Stability AI năm 2023: Đây là một trong những vụ tranh chấp pháp lý nổi bật liên quan đến hiện tượng ảo giác AI và vi phạm SHTT trong AI.¹⁹ Theo đó, Getty Images cáo buộc Stability AI đã sử dụng trái phép hơn 12 triệu hình ảnh có bản quyền từ cơ sở dữ liệu của mình để đào tạo mô hình AI Stable Diffusion mà không có sự cho phép. Vụ kiện này đang được xử lý tại Tòa án Liên bang Delaware (Hoa Kỳ) và Tòa án Tối cao Anh với yêu cầu bồi thường lên đến 1,7 tỷ USD cùng mức phạt 150.000 USD cho mỗi hình ảnh bị vi phạm. Bên cạnh đó, đầu năm 2025 và trong thời gian gần đây, trào lưu “Ghibli hóa”²⁰ bằng AI đã bùng nổ và lan rộng trên khắp các nền tảng số đã làm phát sinh nhiều tranh luận pháp lý về ranh giới giữa sự sáng tạo thực chất và hành vi xâm phạm quyền SHTT. Trào lưu này đã làm nổi lên những vấn đề pháp lý đáng quan ngại về đạo đức sử dụng AI,²¹ đặt ra nhiều câu hỏi bản khoăn về tương lai của những nghệ sĩ cũng như giá trị sáng tạo thực chất của con người trước vấn nạn sao chép phong cách nghệ thuật. Có quan điểm cho rằng AI chỉ đơn thuần là một công cụ hỗ trợ phổ biến nghệ thuật đến với công chúng một cách rõ nét. Ngược lại, nhiều quan điểm khác cũng cho rằng AI đã vận hành dựa trên các thuật toán học từ hàng triệu bức tranh sẵn có, trong đó có phong cách Ghibli nhưng không có sự xin phép hay ghi nhận nguồn thông tin của họa sĩ gốc.²² Theo nguyên tắc cơ bản của các hệ thống pháp luật trên thế giới về bảo hộ quyền tác giả, để được công nhận bảo hộ thì chính tác phẩm phải đạt được mức độ sáng tạo tối thiểu đủ để chứng minh rằng tác phẩm đó nên được khai thác độc quyền bởi tác giả hoặc chủ sở hữu quyền của tác phẩm đó.²³ Với các tác phẩm được tạo ra từ dữ liệu đào tạo AI mà không có sự xin phép từ chính tác giả thì trường hợp này chính là vi phạm về SHTT. Từ góc

19 Kevin Madigan, “Top takeaways from order in the Andersen v. Stability AI copyright case”, *Copyright Alliance*, 2024, <https://copyrightalliance.org/andersen-v-stability-ai-copyright-case/>, truy cập ngày 30/04/2025.

20 Ghibli hóa là trào lưu sử dụng AI để chuyển đổi các hình ảnh thông thường thành phong cách hoạt hình đặc trưng của hãng hoạt hình nổi tiếng Nhật Bản Studio Ghibli. Phong cách Ghibli nguyên bản nổi tiếng với cách vẽ tay tỉ mỉ và lối kể chuyện đầy chất thơ được Hayao Miyazaki phác họa.

21 Hiện tại phía Studio Ghibli và Bandai Namco yêu cầu OpenAI ngừng sử dụng sản phẩm của mình để huấn luyện AI và cho rằng: “hành vi sao chép trong quá trình học máy có thể cấu thành vi phạm bản quyền”. Xem Kim, “Studio Ghibli và Bandai Namco yêu cầu OpenAI ngừng sử dụng sản phẩm của mình để huấn luyện AI”, *GenK*, 2025, <https://genk.vn/studio-ghibli-va-bandai-namco-yeu-cau-openai-ngung-su-dung-san-pham-cua-minh-de-huan-luyen-ai-165251108140506082.chn>, truy cập ngày 05/11/2025.

22 Hà Trang, “Còn sót AI phong cách Ghibli: Tồn vinh hay cú tát vào ngành sáng tạo?”, *Báo Tiền Phong*, 2025, <https://tienphong.vn/con-sot-ai-phong-cach-ghibli-ton-vinh-hay-cu-tat-vao-nganh-sang-tao-post1730550.tpo>, truy cập ngày 30/04/2025.

23 Marcos Wachowics, Lukas Ruthes Goncalves, *Artificial intelligence and creativity: New concepts in intellectual property*, Nxb. Gedai, 2019, tr. 25-26.

độ pháp lý, đây là minh chứng cho khoảng trống pháp lý trong việc xác định hành vi xâm phạm quyền SHTT khi AI sao chép phong cách nghệ thuật của tác giả, qua đó đặt ra yêu cầu hoàn thiện cơ chế bảo hộ đối với sáng tạo mang tính phong cách trong kỷ nguyên số.

Qua các tình huống pháp lý đã phân tích, có thể nhận định rằng hiện tượng ảo giác AI không chỉ là sai sót mang tính kỹ thuật mà còn là nguồn phát sinh nghĩa vụ pháp lý đối với các chủ thể tham gia vào quá trình phát triển và khai thác hệ thống AI, đặc biệt trong các lĩnh vực có yêu cầu cao về độ chính xác như pháp luật. Tuy nhiên, pháp luật Việt Nam cũng như nhiều quốc gia khác vẫn chưa có cơ chế xác định cụ thể trách nhiệm pháp lý giữa người sử dụng, nhà phát triển và nhà cung cấp dịch vụ AI, dẫn đến nguy cơ phát sinh tranh chấp, xâm phạm quyền SHTT khi AI tự động sao chép, tạo lập và cung cấp thông tin sai lệch. Do đó, việc hoàn thiện khung pháp lý nhằm xác lập rõ mối quan hệ giữa hành vi vi phạm, thiệt hại và chủ thể phải chịu trách nhiệm là yêu cầu cần thiết để bảo đảm tính minh bạch, an toàn pháp lý và trách nhiệm giải trình trong hệ sinh thái AI.

3. Một số kiến nghị hoàn thiện quy định pháp luật Việt Nam

Thứ nhất, cần thiết phải xây dựng khung pháp lý riêng về AI và quy định về trách nhiệm pháp lý giữa người sử dụng và các nhà cung cấp dịch vụ AI.²⁴ Trong ngày 23/11/2024, Quốc hội đã thảo luận tại tổ về dự án Luật Công nghiệp công nghệ số, trong quá trình thảo luận, nhiều đại biểu đồng tình với việc cho rằng cần dành một chương riêng và cần nghiên cứu toàn diện để xây dựng một đạo luật riêng về AI của Việt Nam.²⁵ Nhóm tác giả cũng đồng tình với quan điểm trên là cần ban hành một đạo luật riêng về AI để điều chỉnh toàn diện những vấn đề phát sinh từ công nghệ này, bao gồm cả việc bảo hộ quyền tác giả, tác phẩm... và các vấn đề liên quan đến hiện tượng ảo giác AI. Thời gian qua, Việt Nam đã thông qua Luật Công nghiệp công nghệ số (CNCNS)²⁶ và Luật Trí tuệ nhân tạo.²⁷ Đây được xem là bước đột phá thể chế cho kỷ nguyên số và là bước đi chiến lược kiến tạo nền tảng pháp lý cho thời đại số ở Việt Nam. Giống như Hoa Kỳ và trong hướng dẫn của Văn phòng Bản quyền Hoa Kỳ (*United States Copyright Office, USCO*), Việt Nam cần bày tỏ quan điểm rõ ràng trong việc xác định rõ phạm vi tác phẩm được bảo hộ khi có sự tham gia của AI dựa trên nguyên tắc “Con người là trung tâm và AI chỉ là công cụ hỗ trợ”.²⁸ Có quan điểm cho rằng cần bổ sung hướng dẫn về điều kiện bảo hộ quyền tác giả đối với

24 Tương ứng Điều 12, 14, 15, 19, 20 và một số điều tại Dự thảo Luật Trí tuệ nhân tạo.

25 Cao Nguyên, “Cần thiết có đạo luật riêng của Việt Nam về AI”, *Báo Lao Động*, 2024, <https://laodong.vn/thoi-su/can-thiet-co-dao-luat-rieng-cua-viet-nam-ve-ai-1425452.ldo>, truy cập ngày 30/04/2025.

26 Ngày 14/6/2025 tại Kỳ họp thứ 9, Quốc hội khóa XV đã chính thức thông qua Luật Công nghiệp công nghệ số (CNCNS) với 441/445 tán thành (tỷ lệ tán thành 92,26%). Luật này có hiệu lực từ ngày 01/01/2026.

27 Hiện nay, Việt Nam đã thông qua Luật CNCNS ngày 14/6/2025 (có hiệu lực vào ngày 01/01/2026) và đã có một chương riêng quy định về trí tuệ nhân tạo. Tuy nhiên, các quy định tại Luật CNCNS mới chỉ dừng lại ở các nguyên tắc, khái quát chung chưa đủ sức tạo cơ chế xử lý và hành lang pháp lý toàn diện, đồng bộ để điều chỉnh, phát triển AI. Theo đó, Dự thảo gồm: 09 Chương, 70 điều. Xem thêm tại: Minh Hiền, “Dự thảo Luật Trí tuệ nhân tạo: Quản lý AI dựa trên 4 cấp độ rủi ro”, *Báo điện tử Chính phủ*, 08/10/2025, <https://baochinhphu.vn/du-thao-luat-tri-tue-nhan-tao-quan-ly-ai-dua-tren-4-cap-do-rui-ro-102251006165316259.htm>, truy cập ngày 09/10/2025.

28 Đây cũng là quy định tại khoản 1 Điều 4 Dự thảo Luật Trí tuệ nhân tạo.

nội dung do AI tạo ra theo hướng: (i) chỉ bảo hộ khi có sự thể hiện rõ yếu tố sáng tạo của con người, đặc biệt trong việc lựa chọn, sắp xếp và xử lý tài liệu do AI hỗ trợ; (ii) phạm vi bảo hộ chỉ giới hạn ở phần đóng góp sáng tạo độc lập của con người, không bao gồm nội dung do AI tự tạo ra.²⁹ Nhóm tác giả cũng đồng tình với quan điểm trên theo hướng (i) nhưng vẫn phải tuân theo nguyên tắc cốt lõi rằng con người là “trung tâm” và AI chỉ là “công cụ hỗ trợ”, không thể thay thế con người. Điều này không chỉ nhằm tôn vinh giá trị sáng tạo thực chất của con người và hướng đến mục tiêu phát triển bền vững chính là sử dụng AI an toàn, minh bạch và hợp lý, tránh và ngăn ngừa những hành vi vi phạm SHTT cũng như ảnh hưởng tiêu cực đến các thông tin của người sử dụng dịch vụ AI.

Bên cạnh đó, đối với vấn đề trách nhiệm pháp lý của các nhà cung cấp dịch vụ AI, có thể thấy trong nhiều trường hợp, những nhà cung cấp dịch vụ AI không trực tiếp thực hiện hành vi vi phạm mà chỉ cung cấp nền tảng hoặc công cụ hỗ trợ cho người sử dụng tạo ra nội dung đầu ra. Tuy nhiên, nếu các công cụ này được vận hành mà không có biện pháp kiểm soát chặt chẽ và hợp lý, nhà cung cấp vẫn có thể bị quy trách nhiệm theo cơ chế “biết hoặc phải biết”. Do đó, pháp luật cần đặt ra cơ chế trách nhiệm rõ ràng, xác định các tiêu chí như: mức độ kiểm soát của nhà cung cấp đối với dữ liệu đầu vào và đầu ra, khả năng nhận biết hành vi vi phạm. Điều này không chỉ có giá trị trong việc định lượng trách nhiệm, mà còn tạo áp lực buộc các nhà cung cấp AI phải chủ động xây dựng chính sách vận hành minh bạch, hệ thống kiểm soát nội bộ chặt chẽ. Song song đó, cũng cần ban hành một khung pháp lý giới hạn trách nhiệm phù hợp cho các nhà cung cấp dịch vụ AI. Bởi các đặc tính tự học và vận hành theo thuật toán xác suất, đầu ra của AI có thể vượt ngoài khả năng kiểm soát trực tiếp của con người, khiến các nhà cung cấp dịch vụ AI có nguy cơ bị truy cứu trách nhiệm pháp lý ngay cả khi lỗi không xuất phát từ chủ thể vận hành. Việc quy định giới hạn trách nhiệm không nhằm mục tiêu miễn trừ hoàn toàn nghĩa vụ của nhà cung cấp, mà là nhằm xây dựng một khuôn khổ pháp lý minh định, xác định rõ ràng các điều kiện và ngưỡng trách nhiệm có thể chấp nhận được khi xảy ra các vụ việc trên thực tiễn. Tuy nhiên, cơ chế này chỉ được áp dụng khi nhà cung cấp dịch vụ AI đã thực hiện đầy đủ biện pháp kiểm soát hợp lý, bao gồm công bố chính sách sử dụng minh bạch, triển khai cơ chế “thông báo và gỡ bỏ” với nội dung vi phạm SHTT và hợp tác với cơ quan quản lý khi xử lý vi phạm.

Thứ hai, đề xuất xây dựng cơ chế truy vết và công khai dữ liệu huấn luyện của mô hình AI. Việc này nhằm đảm bảo minh bạch nguồn gốc nội dung AI tạo ra, phục vụ cho việc xác định hành vi xâm phạm quyền tác giả hoặc đạo văn. Cơ chế này có thể được vận hành thông qua quy định bắt buộc về khả năng truy xuất nguồn gốc AI (AI traceability) một nguyên tắc đang được khuyến nghị trong các tài liệu của WIPO (2023) và OECD AI Principles (2019). Việc nội luật hóa cơ chế này không chỉ tăng cường trách nhiệm của nhà phát triển AI, mà còn giúp cơ quan thực thi pháp luật, tòa án và các chủ thể quyền có cơ sở xác minh nguồn gốc,

29 Vũ Thị Hải Yến, “Tác động của trí tuệ nhân tạo đến việc bảo hộ quyền tác giả – Thực tiễn pháp lý và liên hệ với Việt Nam”, *Tạp chí Luật học*, số đặc biệt, 2024, tr. 112.

đánh giá mức độ vi phạm và phân định rõ trách nhiệm pháp lý giữa các bên liên quan trong quá trình xử lý tranh chấp SHTT do AI tạo ra.

*Thứ ba, cần tăng cường kiểm soát đầu ra của hệ thống AI thông qua cơ chế tiền kiểm và hậu kiểm.*³⁰ Hầu hết các nghiên cứu đã chỉ ra rằng “tiền kiểm” và “hậu kiểm” phải được kết hợp song song bởi “tiền kiểm” giúp kiểm soát dữ liệu đầu vào và đánh giá rủi ro ban đầu, nhưng không thể dự liệu hết hành vi của AI. Do đó, quy trình “hậu kiểm” đóng vai trò giám sát, xử lý vi phạm và bảo đảm trách nhiệm. Vì vậy, cần xây dựng cơ chế kết hợp hai giai đoạn để vừa thúc đẩy đổi mới vừa bảo vệ hiệu quả quyền SHTT trong bối cảnh AI phát triển mạnh mẽ. Đặc biệt đối với các lĩnh vực nhạy cảm và đòi hỏi tính chính xác cao như y tế, báo chí, pháp luật và giáo dục, mọi thông tin sai lệch do ảo giác AI gây ra có thể dẫn đến hậu quả pháp lý nghiêm trọng. Việc kiểm soát không nhằm cản trở sáng tạo công nghệ, mà hướng đến bảo đảm các nội dung do AI tạo ra không vi phạm pháp luật hiện hành, không gây phương hại đến quyền SHTT hay an toàn thông tin. Việt Nam có thể học hỏi mô hình từ Đạo luật AI của EU khi quy định rõ nghĩa vụ kiểm định, đánh giá và công bố rủi ro đối với các hệ thống AI được xếp loại “rủi ro cao”. Đồng thời, cần rà soát, sửa đổi một số quy định pháp luật hiện hành như Luật An ninh mạng, BLHS, BLDS về chế định bồi thường thiệt hại ngoài hợp đồng... và các văn bản Luật khác, để bổ sung các quy định cụ thể về hành vi sử dụng AI nhằm tạo, lan truyền hoặc khai thác thông tin sai lệch, qua đó xác lập cơ chế xử lý vi phạm và trách nhiệm pháp lý rõ ràng đối với cả người sử dụng và nhà cung cấp hệ thống AI khi xảy ra hậu quả pháp lý do ảo giác AI gây ra.

Thứ tư, Việt Nam cần đẩy mạnh hợp tác quốc tế trong việc điều chỉnh pháp luật liên quan đến AI và SHTT, đặc biệt trong bối cảnh các rủi ro từ ảo giác AI đang ngày càng vượt khỏi phạm vi một quốc gia. Việc cập nhật kịp thời các hướng dẫn, khuyến nghị từ Tổ chức Sở hữu trí tuệ thế giới (World Intellectual Property Organization, WIPO), Cơ quan Sáng chế và nhãn hiệu Hoa Kỳ (United States Patent and Trademark Office, USPTO), cũng như Ủy ban châu Âu (European Commission, EC) về quyền SHTT đối với tác phẩm do AI tạo ra là rất cần thiết. Đồng thời, cần tích cực tham gia vào các diễn đàn quốc tế và các nhóm xây dựng tiêu chuẩn để tiếp cận các mô hình pháp lý tiên tiến, đặc biệt trong lĩnh vực giám sát nội dung đầu ra của AI, truy vết dữ liệu và xây dựng vùng trách nhiệm pháp lý.

Ngoài ra, cần nâng cao nhận thức, đào tạo chuyên sâu về công nghệ AI và pháp luật AI thông qua các chương trình đào tạo và tập huấn. Điển hình như để tạo nền tảng pháp lý vững chắc cho việc quản lý và giám sát AI hiệu quả trên thực tế, cần có đội ngũ nhân lực các ngành nghề tư pháp, hỗ trợ tư pháp, nhà nghiên cứu các lĩnh vực và các chuyên gia pháp lý có kiến thức sâu rộng về AI. Thông qua các chương trình đào tạo chuyên sâu sẽ giúp họ hiểu rõ về những đặc điểm, thách thức và cơ hội mà AI mang lại trong các lĩnh vực sáng tạo, sản xuất cũng như ứng dụng hiệu quả và có trách nhiệm trong việc bảo vệ quyền SHTT. Hơn hết cần ưu tiên chú trọng phát triển năng lực nghiên cứu

30 Trương Trọng Hiếu, “Kinh nghiệm của Châu Âu trong điều chỉnh pháp lý về trí tuệ nhân tạo”, Kỷ yếu Hội thảo khoa học “Tác động của trí tuệ nhân tạo và các công nghệ số mới nổi khác đối với sự phát triển của luật sư”, Nxb Tư Pháp, 2023, tr. 62.

và ứng dụng pháp luật AI. Hiện nay đã có nhiều quyết định của Thủ tướng Chính phủ về việc nghiên cứu, phát triển và ứng dụng AI như: Quyết định số 127/QĐ-TTg ngày 26/01/2021 của Thủ tướng Chính phủ về việc ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030 và những quyết định khác. Điều này cho thấy sự quan tâm của Đảng và Nhà nước trong việc phát triển năng lực nghiên cứu ứng dụng AI vào đời sống của hầu hết các lĩnh vực. Không chỉ vậy, việc tăng cường nhận thức xã hội về trách nhiệm và đạo đức trong phát triển AI cũng là vấn đề đáng quan tâm. Cũng giống như tính liêm chính trong học thuật hay các tiêu chuẩn được xem là đạo văn, Việt Nam cần tăng cường nhận thức cho mỗi người dùng AI trong việc nhận thức về giá trị, đạo đức của sáng tạo thực chất và trách nhiệm của việc sử dụng AI. Những vấn đề này cần được ban hành dưới hình thức Bộ Quy tắc, Hướng dẫn sử dụng hoặc Quy định hướng dẫn về AI. Việc triển khai đồng thời các giải pháp trên sẽ là nền tảng vững chắc cho các cá nhân và tổ chức những kiến thức cần thiết về đạo đức và trách nhiệm xã hội khi phát triển và ứng dụng AI, mỗi cá nhân sẽ có ý thức, trách nhiệm trong việc sử dụng AI minh bạch, an toàn và hiệu quả trong kỷ nguyên số.

Kết luận

Hiện tượng ảo giác AI không chỉ là biểu hiện sai lệch về thuật toán của công nghệ mà là thách thức pháp lý mang tính hệ thống, đòi hỏi sự điều chỉnh kịp thời của pháp luật. Vì vậy, cần sớm xây dựng một khung pháp lý toàn diện, minh bạch và khả thi, xác định rõ trách nhiệm của các chủ thể tham gia vào chuỗi phát triển và vận hành AI, đồng thời bảo đảm cơ chế truy vết, kiểm soát và phòng ngừa rủi ro hiệu quả. Đây là điều kiện tiên quyết để bảo vệ quyền SHTT, duy trì niềm tin pháp lý trong môi trường số và hướng tới một nền công nghệ AI phát triển có trách nhiệm, an toàn và bền vững trong kỷ nguyên số. ●

Tài liệu tham khảo

- [1] Adee Athiyaman, “Artificial intelligence in community economic development: The invisible danger of AI hallucinations”, *Research Brief*, Vol. 7(7), 2025
- [2] Trương Trọng Hiếu, “Kinh nghiệm của Châu Âu trong điều chỉnh pháp lý về trí tuệ nhân tạo”, trong *Kỷ yếu Hội thảo khoa học “Tác động của trí tuệ nhân tạo và các công nghệ số mới nổi khác đối với sự phát triển của luật tư”*, Nxb. Tư Pháp [trans: Truong Trong Hieu, “European experience in legal regulation of artificial intelligence”, in *Proceedings of the Scientific Conference “Impact of artificial intelligence and other emerging digital technologies on the development of private law”*, Judicial Publishing House]
- [3] Nguyễn Phương Thảo, Nguyễn Thị Xuân Hoa, Nguyễn Thị Kim Hoàng, “Bảo hộ tác phẩm tạo ra từ trí tuệ nhân tạo theo pháp luật Hoa Kỳ và gợi mở hoàn thiện pháp luật Việt Nam”, *Tạp chí Dân chủ và Pháp luật*, số 432(7), 2025 [trans: Nguyen Phuong Thao, Nguyen Thi Xuan Hoa, Nguyen Thi Kim Hoang, “Protection of works created by artificial intelligence under US law and recommendations for improving Vietnamese law”, *Journal of Democracy and Law*, Vol. 7(432), 2025]
- [4] Marcos Wachowics, Lukas Ruthes Goncalves, *Artificial intelligence and creativity: new concepts in intellectual property*, Nxb. Gedai, 2019
- [5] Vũ Thị Hải Yến, “Tác động của trí tuệ nhân tạo đến việc bảo hộ quyền tác giả – Thực tiễn pháp lý và liên hệ với Việt Nam”, *Tạp chí Luật học*, số đặc biệt, 2024 [trans: Vu Thi Hai Yen, “Impact of artificial intelligence on copyright protection – Legal practice and implications for Vietnam”, *Journal of Law*, Special Issue, 2024]