ĐỊNH HƯỚNG XÂY DỰNG HÀNH LANG PHÁP LÝ CHO HOẠT ĐỘNG ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG NGHIÊN CỨU KHOA HỌC XÃ HỘI VÀ NHÂN VĂN Ở VIỆT NAM

PHẠM THỊ THÚY NGA*

Tóm tắt: *Sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) đang đặt ra nhiều thách thức pháp lý mới cho hoạt động nghiên cứu khoa học xã hội và nhân văn (KHXH&NV), đặc biệt trong bối cảnh dữ liệu định tính và ngữ cảnh văn hóa - xã hội ngày càng được xử lý thông qua công cụ công nghệ. Bài viết phân tích đặc điểm đặc thù của KHXH&NV dưới góc nhìn pháp lý - đạo đức, nhận diện các rủi ro pháp lý điển hình trong quá trình ứng dụng AI vào nghiên cứu học thuật, từ đó đề xuất các định hướng xây dựng hành lang pháp lý phù hợp. Trên cơ sở tiếp cận liên ngành, bài viết gợi mở ba nhóm giải pháp chính sách: (1) Bổ sung quy định về đạo đức và trách nhiệm pháp lý trong các văn bản pháp luật hiện hành; (2) Phát triển bộ quy tắc đạo đức nghiên cứu áp dụng riêng cho lĩnh vực KHXH&NV có sử dụng AI; (3) Xây dựng cơ chế sandbox pháp lý để kiểm nghiệm chính sách trong thực tiễn. Các đề xuất này nhằm bảo vệ giá trị nhân văn trong nghiên cứu, tăng cường tính chính danh học thuật, đồng thời thúc đẩy đổi mới sáng tạo có trách nhiệm trong thời đại số.*

Từ khóa: *Trí tuệ nhân tạo; khoa học xã hội và nhân văn; hành lang pháp lý; chính sách pháp luật; đạo đức nghiên cứu; sandbox pháp lý.*

Abstract: *The rapid development of artificial intelligence (AI) presents new legal challenges for research in social sciences and humanities (SSH), especially as qualitative data and socio-cultural contexts are increasingly processed through technological tools. This article analyzes the distinctive characteristics of SSH from legal and ethical perspectives, identifies typical legal risks arising from the application of AI in academic research, and proposes orientations for developing an appropriate legal framework. Adopting an interdisciplinary approach, the article suggests three main policy solutions: (1) supplementing current legislation with provisions on ethics and legal responsibility; (2) developing a dedicated code of research ethics for SSH involving AI applications; and (3) establishing a legal sandbox mechanism to pilot policies in practice. These proposals aim to protect humanistic values in research, enhance academic legitimacy, and promote responsible innovation in the digital age.*

Keywords: *Artificial intelligence; social sciences and humanities; legal framework; legal policy; research ethics; legal sandbox.*

Ngày nhận bài: 12/3/2025; Ngày sửa bài: 20/4/2025; Ngày duyệt đăng bài: 20/5/2025.

1. Đặt vấn đề

Trong bối cảnh AI đang chuyển hóa nhanh chóng cách thức tạo lập và xử lý tri thức khoa học, lĩnh vực KHXH&NV đang đối mặt với một loạt các thách thức pháp lý, đạo đức và học thuật chưa từng có tiền lệ. Khác với khoa học kỹ thuật - nơi dữ liệu chủ yếu là các chỉ số định lượng, KHXH&NV vận hành trên nền tảng các hệ quy chiếu văn hóa, biểu tượng, giá trị xã hội - vốn khó

* TS., Viện Nhà nước và Pháp luật, Viện Hàn lâm KHXH Việt Nam; Email: ngapham@isl.gov.vn

lượng hóa và dễ bị giản lược bởi các thuật toán tự động. AI có thể hỗ trợ phân tích văn bản, nhận diện mẫu hình hành vi hoặc tạo sinh nội dung, nhưng đồng thời cũng làm nảy sinh các vấn đề pháp lý phức tạp như: quyền riêng tư trong phỏng vấn định tính, trách nhiệm pháp lý khi AI sinh ra nội dung sai lệch, hay xung đột quyền tác giả trong nghiên cứu có sự tham gia của máy học.

Tại Việt Nam, Chiến lược quốc gia về AI đến năm 2030 đã xác định rõ vai trò trung tâm của con người và yêu cầu phát triển hành lang pháp lý đồng bộ. Tuy nhiên, các văn bản pháp luật hiện hành vẫn còn thiếu vắng các quy định cụ thể điều chỉnh hoạt động nghiên cứu có ứng dụng AI, đặc biệt là trong lĩnh vực KHXH&NV. Việc phát triển chính sách pháp luật trong lĩnh vực này không thể chỉ dựa vào khung kỹ thuật công nghệ, mà cần tiếp cận từ chính đặc thù học thuật của KHXH&NV, trong đó đặt con người - với tư cách là chủ thể kiến tạo tri thức - làm trung tâm. Đây là cơ sở lý luận và thực tiễn để thúc đẩy việc xây dựng một hành lang pháp lý đa tầng, linh hoạt nhưng hiệu lực, nhằm vừa bảo vệ giá trị nhân văn trong nghiên cứu, vừa thúc đẩy ứng dụng công nghệ một cách có trách nhiệm.

2. Hoạt động ứng dụng trí tuệ nhân tạo trong khoa học xã hội và nhân văn

Trong bối cảnh AI phát triển nhanh chóng và tác động sâu rộng đến các lĩnh vực khoa học, bao gồm cả KHXH&NV, việc điều chỉnh pháp luật đối với công nghệ mới không thể chỉ dựa vào các công cụ pháp lý truyền thống. Thay vào đó, cần đặt trong một nền tảng lý luận đa chiều, phản ánh sự thay đổi bản chất của quan hệ xã hội dưới tác động của công nghệ. Pháp luật cần đóng vai trò như một công cụ nhận diện, phòng ngừa và quản lý các rủi ro mới, đặc biệt là những rủi ro chưa chắc chắn và khó lường phát sinh từ việc ứng dụng AI vào hoạt động học thuật. Trong trường hợp của KHXH&NV, các rủi ro như thiên kiến thuật toán, tổn hại đến quyền riêng tư, sai lệch diễn ngôn hay xói mòn đạo đức nghiên cứu cần được nhận diện từ sớm, với các công cụ điều chỉnh thích hợp.

Đặc điểm đặc thù của KHXH&NV

So với các ngành khoa học kỹ thuật - công nghệ, KHXH&NV có những đặc điểm nhận thức luận và phương pháp luận riêng biệt, tạo nên sự khác biệt căn bản trong cách thức thiết kế nghiên cứu, xử lý dữ liệu, cũng như đánh giá độ tin cậy của kết quả. Trước hết, KHXH&NV đặt trọng tâm vào việc hiểu biết con người như một chủ thể mang tính lịch sử, văn hóa và ý thức, thay vì chỉ quan sát các hiện tượng vật lý hoặc thiết lập quan hệ nhân quả định lượng như trong các ngành khoa học tự nhiên. Các đối tượng nghiên cứu trong KHXH&NV - như giá trị xã hội, hành vi tập thể, ý thức cộng đồng, thực hành văn hóa - không tồn tại một cách "khách quan" độc lập, mà thường mang tính diễn ngôn, gắn với ngữ cảnh xã hội và được kiến tạo qua lịch sử. Điều này dẫn đến việc các phương pháp nghiên cứu định tính, phân tích văn bản, phân tích diễn ngôn hoặc tiếp cận phê phán thường được sử dụng rộng rãi, đòi hỏi sự thấu cảm, lý giải và diễn giải thay vì đo lường đơn thuần.

Ứng dụng AI trong KHXH&NV

Năm nhóm rủi ro chính liên quan đến việc ứng dụng AI trong nghiên cứu KHXH&NV được chỉ ra, bao gồm: (1) Vi phạm quyền riêng tư cá nhân trong khảo sát và phỏng vấn; (2) Mơ hồ trong xác định quyền tác giả khi sử dụng AI sinh nội dung; (3) Sai lệch do dữ liệu huấn luyện thiếu đại

diện; (4) Nguy cơ mất tính minh bạch trong diễn giải; và (5) Xói mòn chuẩn mực đạo đức học thuật do lạm dụng công cụ công nghệ. Do đó, cần thiết lập một hành lang pháp lý mềm linh hoạt, kết hợp giữa đạo đức nghiên cứu và quy định pháp lý để kiểm soát tốt hơn (Ford et al., 2024)[1]. Việc ứng dụng AI - vốn được xây dựng dựa trên các mô hình học máy, xác suất thống kê và xử lý dữ liệu lớn - vào nghiên cứu KHXH&NV cần được tiếp cận với sự thận trọng và cân nhắc cao độ. Mặc dù AI có thể xử lý khối lượng lớn dữ liệu văn bản và phát hiện các mẫu hình phổ biến, song lại gặp nhiều hạn chế trong việc nắm bắt ngữ nghĩa sâu, tính đa tầng của biểu tượng, cũng như khả năng kiến tạo tri thức đặc thù của các lĩnh vực xã hội học, nhân học hay triết học. Theo Mokander và Schroeder (2024)[2], các hệ thống AI hiện nay thiếu ba năng lực cốt lõi để có thể phát triển lý thuyết xã hội một cách thực chất: khả năng biểu diễn khái niệm phức tạp bằng ngôn ngữ hình thức (semanticization), khả năng chuyển giao kiến thức giữa các bối cảnh (transferability), và khả năng tự sinh tạo mô hình lý thuyết (generativity). Những năng lực này vốn gắn liền với phân tích phản tư - một phẩm chất khó có thể thay thế bằng các thuật toán học máy hiện hành. Hơn nữa, ngôn ngữ trong nghiên cứu KHXH&NV không chỉ là công cụ truyền đạt thông tin mà còn là phương tiện biểu đạt bản sắc, quyền lực và cấu trúc văn hóa. Tuy nhiên, như Browning và LeCun (2022)[3] đã chỉ ra, các hệ thống AI hiện nay - chủ yếu được huấn luyện trên chuỗi ký tự phi ngữ cảnh - không thể nắm bắt đầy đủ các biến thể văn hóa hoặc chiều sâu ngữ nghĩa cần thiết cho việc phân tích hiện tượng xã hội phức tạp. Điều này càng thể hiện rõ khi phân tích các hiện tượng mang tính hệ hình, nơi mà kiến thức được kiến tạo thông qua lịch sử, tương tác xã hội và quyền lực biểu tượng. Thực tế cũng cho thấy, các hệ thống AI thường được thiết kế mà không xét đến tính xã hội của dữ liệu, dẫn đến việc "tự động hóa sự mơ hồ" và tái sản xuất các thiên lệch vốn có trong cấu trúc xã hội (Birhane, 2022)[4]. Điều này gây rủi ro nghiêm trọng khi AI được sử dụng như một công cụ trung lập trong phân tích các hiện tượng mang tính phức hợp cao - ví dụ như định kiến chủng tộc, bất bình đẳng giới, hoặc chuyển đổi chính trị - mà không được kiểm soát bằng khung đạo đức hoặc tiêu chuẩn học thuật tương thích. Nhiều mô hình AI có thể tạo ra "ảo tưởng sự thật" (hallucination) - nội dung trông có vẻ hợp lý nhưng sai sự thật hoặc gây hiểu nhầm - ảnh hưởng đến độ tin cậy của nghiên cứu KHXH&NV. Trong bối cảnh nghiên cứu KHXH&NV thường làm việc với dữ liệu định tính, các hiện tượng như sai trích dẫn, dựng ngữ cảnh hoặc tạo ra bằng chứng giả mạo không chỉ ảnh hưởng đến chất lượng khoa học mà còn tiềm ẩn hệ lụy đạo đức nghiêm trọng (Madanchian & Taherdoost, 2025)[5]. Bên cạnh đó, các nghiên cứu định tính - đặc biệt là những nghiên cứu dựa trên phân tích nội dung, phân tích diễn ngôn hoặc lý thuyết nền tảng - yêu cầu khả năng gắn kết nội dung với ngữ cảnh

[1] Ford, H., Bentall, C., & Jain, S. (2024), Ethical AI in Social Sciences Research: Are We Gatekeepers or Revolutionaries?, Societies, 15(3), 62. https://www.mdpi.com/2075-4698/15/3/62

[2] Mokander, J., & Schroeder, R. (2024), *AI and Social Theory*, arXiv preprint. https://arxiv.org/abs/2407.06233

[3] Browning, J., & LeCun, Y. (2022), *AI and the limits of language*. Noema Magazine, https://www.noemamag.com/ai-and-the-limits-of-language/

[4] Birhane, A. (2022), *Automating ambiguity: Challenges and pitfalls of artificial intelligence*, arXiv preprint. https://arxiv.org/abs/2206.04179

[5] Mitra Madanchian & Hamed Taherdoost (2025), *The impact of artificial intelligence on research efficiency*, Results in Engineering, 26, 104743. DOI: 10.1016/j.rineng.2025.104743

xã hội cụ thể. Tuy nhiên, các mô hình học máy hiện tại không thể tái tạo được thao tác lý giải đặc thù của nhà nghiên cứu xã hội, nhất là khi phải xử lý sự nhập nhằng ngữ nghĩa, tầng lớp biểu tượng, hoặc các hiện tượng liên ngành gắn với quyền lực và văn hóa (Delve, 2024)[6]. Nếu thiếu cơ chế giám sát và tiêu chuẩn kiểm định phù hợp, việc sử dụng AI trong nghiên cứu có thể dẫn đến "suy diễn giản lược" hoặc ngụy tạo tính khách quan thông qua các kết luận có vẻ hợp lý nhưng thực chất không phản ánh đúng bản chất xã hội.

Đặc biệt, các hệ quy chiếu đạo đức trong nghiên cứu KHXH&NV gắn liền với trách nhiệm đối với cộng đồng được nghiên cứu - bao gồm quyền được diễn giải đúng bối cảnh, quyền riêng tư và quyền không bị định kiến tái tạo từ mô hình AI. Trong khi đó, hệ thống pháp luật hiện hành vẫn chủ yếu dựa trên khung kỹ thuật trung lập, chưa đủ năng lực bao quát tính chất phức tạp của mối quan hệ giữa người - cộng đồng - diễn ngôn trong lĩnh vực nghiên cứu xã hội và nhân văn. Những tình huống như AI "phân tích diễn ngôn chính trị" nhưng vô tình gắn nhãn lệch lạc cho nhóm dân cư hoặc AI tổng hợp phỏng vấn nhưng bỏ qua tầng nghĩa ngữ cảnh - là những ví dụ điển hình cho thấy sự cần thiết của một khung pháp lý chuyên biệt về đạo đức và trách nhiệm sử dụng AI trong nghiên cứu có yếu tố xã hội - văn hóa.

Do đó, nếu không làm rõ tính đặc thù của KHXH&NV so với các lĩnh vực kỹ thuật - công nghệ, thì yêu cầu xây dựng khung pháp lý riêng cho việc ứng dụng AI vào nghiên cứu học thuật có thể bị xem là thiếu căn cứ khoa học. Việc xác lập một hành lang pháp lý phù hợp không nên chỉ dựa trên bản chất kỹ thuật của AI, mà cần xuất phát từ đặc điểm nhận thức luận và chuẩn mực học thuật nội tại của các ngành KHXH&NV. Khung pháp lý đó cần hướng đến việc bảo vệ tính nhân văn, quyền được tái hiện đúng thực tại xã hội và quyền của cộng đồng trong việc không bị "giản lược hóa" thành dữ liệu - qua đó duy trì tính chính danh của tri thức và đạo đức khoa học trong kỷ nguyên số.

3. Xây dựng hành lang pháp lý cho ứng dụng AI trong nghiên cứu khoa học xã hội và nhân văn

3.1. Về hành lang pháp lý cho ứng dụng AI trong nghiên cứu KHXH&NV

Trong bối cảnh nghiên cứu khoa học xã hội và nhân văn, khái niệm "hành lang pháp lý" cần được hiểu theo nghĩa rộng, không chỉ bao gồm các quy định mang tính bắt buộc (luật, nghị định), mà còn bao gồm cả các nguyên tắc hướng dẫn đạo đức nghiên cứu, cơ chế tự giám sát trong cộng đồng học thuật và các quy chuẩn nghề nghiệp do các tổ chức học thuật, cơ sở nghiên cứu ban hành. Điều này xuất phát từ đặc trưng của KHXH&NV là không chỉ tạo ra tri thức, mà còn kiến tạo giá trị xã hội - văn hóa, đòi hỏi một cơ chế pháp lý - đạo đức đa tầng, mềm dẻo nhưng đủ ràng buộc để kiểm soát rủi ro và duy trì niềm tin học thuật.

Một trong những bước tiến mang tính nền tảng trong định hình phạm vi pháp lý toàn cầu về AI là việc Hội đồng Châu Âu thông qua Công ước khung về Trí tuệ nhân tạo, quyền con người, dân chủ và pháp quyền vào năm 2024. Đây là hiệp ước quốc tế đầu tiên đặt ra các nghĩa vụ pháp lý ràng buộc giữa các quốc gia về việc phát triển, triển khai và sử dụng AI, đặc biệt trong các lĩnh vực có nguy cơ cao ảnh hưởng đến quyền con người và giá trị dân chủ, bao gồm cả lĩnh vực học thuật và

[6] Delve (2024), *AI in qualitative data analysis*. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2949882125000283

nghiên cứu xã hội. Công ước yêu cầu các quốc gia thành viên phải bảo đảm rằng việc sử dụng AI trong các hoạt động như nghiên cứu KHXH&NV phải tuân thủ các nguyên tắc cơ bản về minh bạch, trách nhiệm giải trình và không phân biệt đối xử; đồng thời phải thiết lập cơ chế đánh giá tác động xã hội - pháp lý trong suốt vòng đời của hệ thống AI (Council of Europe, 2024)[7]. Báo cáo của Hội đồng Châu Âu (2021) về tác động của AI đối với quyền con người, dân chủ và pháp quyền đã mở rộng phạm vi điều chỉnh pháp luật từ lĩnh vực công nghệ thuần túy sang các bối cảnh xã hội - học thuật. Báo cáo chỉ ra rằng, khi AI được sử dụng trong các lĩnh vực như giáo dục, nghiên cứu hoặc xuất bản học thuật, các nguy cơ về vi phạm quyền riêng tư, định kiến thuật toán và thiếu minh bạch ngày càng trở nên phức tạp, khó kiểm soát. Do đó, cần phải thiết kế một khung pháp lý đa tầng, kết hợp luật cứng với các nguyên tắc đạo đức học thuật và các quy tắc hành nghề nội ngành để giám sát quá trình ứng dụng AI trong KHXH&NV (Council of Europe, 2021)[8]. Yeung & Lodge đặc biệt nhấn mạnh đến các vấn đề phát sinh trong bối cảnh ứng dụng AI vào hoạt động học thuật, như khó khăn trong xác định chủ thể chịu trách nhiệm khi kết quả nghiên cứu bị sai lệch do lỗi thuật toán, hoặc khi hệ thống "hộp đen" không thể giải thích được quy trình xử lý dữ liệu. Tác giả khẳng định rằng lĩnh vực nghiên cứu KHXH&NV cần có các quy định đặc thù hơn về quyền dữ liệu, trách nhiệm pháp lý và nghĩa vụ minh bạch hóa trong hoạt động phân tích xã hội dựa trên AI (Yeung & Lodge, 2020)9. Đến nay có khoảng 33 quốc gia đã xây dựng dự thảo pháp lý về AI. Tuy nhiên, thế giới chưa có bộ quy chuẩn mang tính tổng thể, chưa có tiếng nói chung trên toàn cầu[10]. Nhìn chung, các nhóm vấn đề pháp lý trọng yếu đặt ra khi ứng dụng AI trong KHXH&NV bao gồm: (1) Vấn đề về quyền riêng tư và dữ liệu cá nhân trong nghiên cứu định tính, phân tích văn bản xã hội, hoặc khảo sát nhóm yếu thế; (2) Vấn đề về trách nhiệm pháp lý khi có sai lệch do thuật toán gây ra trong quá trình trích xuất, xử lý hoặc diễn giải dữ liệu nghiên cứu; (3) Vấn đề về quyền tác giả và sở hữu trí tuệ, nhất là trong các sản phẩm học thuật có sự tham gia của AI trong sáng tạo nội dung; (4) Vấn đề về minh bạch và khả năng truy nguyên khi sử dụng các mô hình AI khép kín, thiếu khả năng giải thích (black box); (5) Vấn đề về đạo đức nghiên cứu và công bằng học thuật, khi AI bị lạm dụng trong viết bài, trích dẫn hoặc tạo dữ liệu giả.

3.2. Thực trạng hành lang pháp lý và gợi mở định hướng xây dựng hành lang pháp lý cho ứng dụng AI trong nghiên cứu khoa học xã hội và nhân văn ở Việt Nam

Tại Việt Nam, nhận thức về vai trò và tác động của AI trong đời sống khoa học và quản lý đã có những chuyển biến rõ nét trong vài năm gần đây. Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030 được Thủ tướng Chính phủ phê duyệt tại Quyết định số 127/QĐ-TTg ngày 26/1/2021 đã xác định

[7] Council of Europe (2024), *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, Retrieved from https://en.wikipedia.org/wiki/Framework_Convention_on_Artificial_Intelligence

[8] Council of Europe (2021), *Artificial intelligence, human rights, democracy and the rule of law*, Retrieved from https://arxiv.org/abs/2202.02776

[9] Yeung, K. & Lodge, M. (2020), "Legal and human rights issues of AI: Gaps, challenges and vulnerabilities", *European Journal of Risk Regulation*,11(2),273-300. https://www.sciencedirect.com/science/article/pii/S2666659620300056

[10] Lê Phúc & Trần Dương (2025), "Sự cần thiết xây dựng khung pháp lý về AI: Kinh nghiệm của các nước và một số đề xuất cho Việt Nam", *Tạp chí Pháp lý điện tử*. https://phaply.net.vn/su-can-thiet-xay-dung-khung-phap-ly-ve-ai-kinh-nghiem-cua-cac-nuoc-va-mot-so-de-xuat-cho-viet-nam-a258209.html

rõ mục tiêu xây dựng hành lang pháp lý và cơ chế quản lý thông thoáng đáp ứng yêu cầu thúc đẩy nghiên cứu, phát triển công nghệ AI một cách an toàn, hiệu quả và bền vững; Phát triển và ứng dụng AI lấy con người và doanh nghiệp làm trung tâm, tránh lạm dụng công nghệ và xâm phạm quyền, lợi ích hợp pháp của tổ chức, cá nhân[11]. Thời gian qua, Thủ tướng Chính phủ đã ban hành 3 quyết định quan trọng liên quan đến chuyển đổi số đất nước, gồm Quyết định 749 phê duyệt Chương trình chuyển đổi số quốc gia đến năm 2025, Quyết định 942 phê duyệt Chiến lược phát triển chính phủ điện tử hướng tới chính phủ số giai đoạn 2021 - 2025, Quyết định 411 phê duyệt Chiến lược phát triển kinh tế số và xã hội số đến năm 2025. Các quyết định này đều nhấn mạnh đến ứng dụng AI trong các lĩnh vực chính phủ số, kinh tế số và xã hội số.

Tuy nhiên, khi đối chiếu với thực tiễn điều chỉnh pháp luật hiện hành, có thể nhận thấy rằng hành lang pháp lý cho ứng dụng AI trong lĩnh vực KHXH&NV vẫn đang ở giai đoạn khởi đầu, thiếu tính chuyên biệt và định hướng dài hạn. Hiện nay, các văn bản pháp luật chủ yếu điều chỉnh công nghệ AI tại Việt Nam (như Luật An toàn thông tin mạng năm 2015, Luật Giao dịch điện tử năm 2023, các quy định về dữ liệu cá nhân và an ninh mạng) chủ yếu tiếp cận từ góc độ kỹ thuật - công nghệ - quản lý hành chính, trong khi chưa có văn bản nào tập trung điều chỉnh hoạt động ứng dụng AI trong môi trường học thuật và nghiên cứu khoa học. Đặc biệt, trong lĩnh vực KHXH&NV - nơi dữ liệu nghiên cứu thường là định tính, mang tính văn hóa - biểu tượng và gắn với quyền con người thì các rủi ro pháp lý như vi phạm quyền riêng tư, sai lệch diễn ngôn, xung đột quyền tác giả, hoặc tái tạo định kiến xã hội vẫn chưa được nhận diện và điều chỉnh đầy đủ trong hệ thống pháp luật hiện hành.

Bên cạnh đó, các cơ chế đánh giá rủi ro đạo đức, trách nhiệm pháp lý của nhà nghiên cứu khi sử dụng AI, hay nghĩa vụ minh bạch trong công bố kết quả nghiên cứu được hỗ trợ bởi AI hiện vẫn thiếu vắng trong Luật Khoa học và Công nghệ năm 2013 và các văn bản hướng dẫn thi hành. Việc thiếu các chuẩn mực đạo đức nghiên cứu đối với ứng dụng AI trong KHXH&NV cũng khiến hội đồng đạo đức hoặc tổ chức xét duyệt đề tài gặp khó khăn trong việc đánh giá tác động của AI một cách toàn diện.

Có thể nói, hành lang pháp lý cho ứng dụng AI trong KHXH&NV ở Việt Nam hiện nay đang tồn tại khoảng trống kép: vừa thiếu vắng quy định pháp luật cứng (hard law), vừa chưa có các chuẩn mực mềm (soft norms) như quy tắc đạo đức, hướng dẫn nghiên cứu hay cơ chế thử nghiệm chính sách[12]. Khoảng trống này đặt ra yêu cầu cấp thiết đối với việc xây dựng các quy phạm mới, mang tính đặc thù và thích ứng cao, nhằm bảo vệ tính chính danh học thuật, đảm bảo quyền con người và thúc đẩy phát triển công nghệ có trách nhiệm trong nghiên cứu khoa học xã hội và nhân văn.

Tại Việt Nam, để xây dựng hành lang pháp lý cho ứng dụng AI trong nghiên cứu

[11] Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030 (ban hành theo Quyết định số 127/QĐ-TTg ngày 26/1/2021 Thủ tướng Chính phủ).

[12] Xem Nguyễn Giang Trường, "Thực trạng về vấn đề bảo vệ dữ liệu cá nhân, kinh nghiệm một số quốc gia, khu vực và đề xuất hoàn thiện pháp luật ở Việt Nam hiện nay", *Tạp chí Công Thương*, (12)/2024, tr.70-75; Lê Thị Minh, "Sự sáng tạo của trí tuệ nhân tạo trong mối liên hệ với pháp luật về quyền tác giả", Tạp chí *Luật học*, (5)/2024; Nguyễn Thị Quế Anh, "Hoàn thiện pháp luật sở hữu trí tuệ Việt Nam trong bối cảnh phát triển trí tuệ nhân tạo", Tạp chí *Luật học (Đại học Quốc gia Hà Nội)*, (3)/2022.

KHXH&NV, có thể cân nhắc ba định hướng cơ bản:

Thứ nhất, bổ sung nội dung về đạo đức và trách nhiệm pháp lý trong các văn bản điều chỉnh hoạt động khoa học - công nghệ, đặc biệt trong Luật KH&CN và Luật An toàn thông tin. Việc bổ sung nội dung về AI, bao gồm chuẩn mực đạo đức, trách nhiệm giải trình, cơ chế bảo vệ dữ liệu cá nhân trong nghiên cứu KHXH&NV là cần thiết nhằm lấp đầy khoảng trống điều chỉnh pháp luật. Các điều khoản bổ sung nên làm rõ trách nhiệm của nhà nghiên cứu khi sử dụng AI, quyền của người tham gia nghiên cứu khi dữ liệu của họ được xử lý tự động, cũng như nghĩa vụ đánh giá rủi ro đạo đức trước khi triển khai đề tài. Đây là nền tảng để từng bước hình thành một khung pháp lý thống nhất, có khả năng tích hợp vào hệ thống pháp luật hiện hành, mà không cần ban hành luật mới trong ngắn hạn.

Thứ hai, phát triển bộ quy tắc đạo đức nghiên cứu áp dụng riêng cho các lĩnh vực có sử dụng công nghệ AI, do Bộ Khoa học và Công nghệ phối hợp với các tổ chức học thuật chủ trì. Các quy định pháp luật mang tính cưỡng chế, cần thiết phải xây dựng một bộ quy tắc đạo đức nghiên cứu chuyên biệt, áp dụng cho các ngành có sử dụng AI trong nghiên cứu KHXH&NV. Bộ quy tắc này nên được xây dựng theo hướng liên ngành, trong đó Bộ Khoa học và Công nghệ đóng vai trò chủ trì, phối hợp với các viện nghiên cứu chuyên ngành, hội đồng đạo đức quốc gia và các tổ chức xuất bản khoa học. Nội dung bộ quy tắc nên bao gồm: (i) Nghĩa vụ công khai mức độ sử dụng AI trong nghiên cứu; (ii) Yêu cầu đánh giá thiên lệch thuật toán và khả năng truy nguyên kết quả; (iii) Nguyên tắc về bảo vệ quyền riêng tư, đặc biệt trong các nghiên cứu định tính hoặc liên quan đến cộng đồng dễ bị tổn thương; và (iv) Trách nhiệm giải trình của nhà nghiên cứu khi sử dụng AI trong xử lý dữ liệu và diễn giải kết quả. Việc này không chỉ củng cố lòng tin vào nghiên cứu học thuật mà còn tạo điều kiện để hệ thống tạp chí, cơ sở đào tạo và tổ chức nghiên cứu nội bộ áp dụng nhất quán các chuẩn mực đạo đức học thuật trong bối cảnh chuyển đổi số.

Thứ ba, xây dựng cơ chế thí điểm (regulatory sandbox) cho các đề tài nghiên cứu có sử dụng AI để đánh giá rủi ro, kiểm định chuẩn mực và thử nghiệm khung pháp lý phù hợp với bối cảnh xã hội - văn hóa Việt Nam. Theo Báo cáo của OECD (2022)[13], sandbox pháp lý được định nghĩa là môi trường thử nghiệm có kiểm soát, nơi các tổ chức có thể triển khai công nghệ mới dưới sự giám sát của cơ quan quản lý mà không phải tuân thủ đầy đủ các quy định pháp luật hiện hành. Điều này cho phép đánh giá rủi ro và tác động xã hội của AI trước khi áp dụng rộng rãi, đặc biệt hữu ích trong các lĩnh vực như KHXH&NV, nơi mà các chuẩn mực đạo đức và quyền con người đóng vai trò then chốt. Trong bối cảnh công nghệ phát triển nhanh hơn tốc độ lập pháp, việc áp dụng mô hình sandbox pháp lý là một giải pháp linh hoạt và có tính khả thi cao tại Việt Nam. Đây là cách tiếp cận đang được nhiều quốc gia áp dụng để thử nghiệm khung quản lý trong môi trường rủi ro thấp, có kiểm soát và giám sát chặt chẽ. Đối với lĩnh vực nghiên cứu KHXH&NV, một cơ chế sandbox sẽ cho phép các nhóm nghiên cứu được phép triển khai đề tài sử dụng AI trong môi trường

[13] OECD (2022), *Regulatory sandboxes in artificial intelligence*, Organisation for Economic Cooperation and Development, https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence_8f80a0e6-en.html

pháp lý "mềm" với điều kiện bắt buộc phải tuân thủ các yêu cầu về đánh giá tác động, bảo vệ dữ liệu cá nhân và báo cáo định kỳ về hiệu quả - rủi ro đạo đức. Trong bối cảnh nghiên cứu KHXH&NV, sandbox đặc biệt quan trọng khi AI được sử dụng để phân tích dữ liệu nhạy cảm hoặc ảnh hưởng đến các nhóm dễ bị tổn thương[14]. Việc tổ chức sandbox không chỉ giúp kiểm định thực tiễn các chuẩn mực pháp lý - đạo đức đang xây dựng, mà còn cung cấp dữ liệu chính sách để điều chỉnh kịp thời các quy phạm pháp luật chính thức. Mô hình này cũng có thể kết nối với các chương trình tài trợ nghiên cứu đổi mới sáng tạo, qua đó khuyến khích những đề tài chất lượng cao nhưng có yếu tố công nghệ chưa ổn định về mặt pháp lý, tạo động lực cho chuyển đổi số học thuật gắn với trách nhiệm xã hội và nhân văn. Tại Việt Nam, việc áp dụng cơ chế sandbox pháp lý trong nghiên cứu KHXH&NV có thể được triển khai thông qua các dự án thí điểm do Bộ Khoa học và Công nghệ phối hợp với các viện nghiên cứu và trường đại học chủ trì. Các sandbox này nên tập trung vào việc đánh giá tác động xã hội của AI, thiết lập các tiêu chuẩn đạo đức trong nghiên cứu, và phát triển các khung pháp lý linh hoạt phù hợp với bối cảnh văn hóa và xã hội Việt Nam.

4. Kết luận

Sự gia tăng ứng dụng AI trong nghiên cứu khoa học xã hội và nhân văn (KHXH&NV) mang lại nhiều tiềm năng đổi mới nhưng đồng thời cũng đặt ra những thách thức pháp lý đáng kể, đặc biệt liên quan đến quyền riêng tư, đạo đức học thuật và trách nhiệm pháp lý. Trong khi AI được kỳ vọng hỗ trợ phân tích và tạo sinh tri thức, thì chính đặc thù của KHXH&NV - với dữ liệu mang tính biểu tượng, ngữ cảnh và nhân văn sâu sắc - lại đòi hỏi một hệ thống pháp luật có khả năng nhận diện và điều chỉnh các rủi ro phi kỹ thuật mà AI có thể gây ra. Tại Việt Nam, hành lang pháp lý cho lĩnh vực này còn chưa đầy đủ. Các quy định hiện hành mới chỉ tập trung vào khía cạnh công nghệ - kỹ thuật, trong khi thiếu các quy phạm chuyên biệt về đạo đức nghiên cứu, minh bạch thuật toán, hay trách nhiệm của các chủ thể trong ứng dụng AI vào nghiên cứu KHXH&NV. Để lấp đầy khoảng trống pháp lý hiện nay, cần bổ sung các quy định liên quan đến đạo đức và trách nhiệm pháp lý trong các văn bản pháp luật hiện hành, đồng thời phát triển một bộ quy tắc đạo đức nghiên cứu chuyên biệt cho các lĩnh vực có sử dụng AI, phù hợp với đặc thù của KHXH&NV. Bên cạnh đó, việc thiết lập một cơ chế thử nghiệm chính sách có kiểm soát cũng được coi là giải pháp khả thi nhằm đánh giá và hoàn thiện các quy phạm điều chỉnh trong điều kiện thực nghiệm. Những định hướng này, nếu được triển khai đồng bộ, sẽ góp phần hình thành một hành lang pháp lý vừa linh hoạt, vừa bảo đảm chuẩn mực, qua đó giúp việc ứng dụng AI trong KHXH&NV phát huy được tiềm năng hỗ trợ nghiên cứu mà vẫn không làm tổn hại đến các giá trị nhân văn, đạo đức học thuật và quyền con người - những nền tảng không thể thiếu trong đời sống khoa học đương đại.

TÀI LIỆU THAM KHẢO

1. Birhane, A (2022), *Automating ambiguity: Challenges and pitfalls*

[14] Truby et al (2021), Truby, J., Brown, C., & Dahdal, A. (2021), *A sandbox approach to regulating high-risk artificial intelligence applications*, University of Twente Research Information System. https://ris.utwente.nl/ws/files/304762207/a_sandbox_approach_to_regulating_high_risk_artificial_intelligence_applications.pdf

of artificial intelligence. arXiv preprint, https://arxiv.org/abs/2206.04179

2. Browning, J., & LeCun, Y(2022), *AI and the limits of language,* Noema Magazine, https://www.noemamag.com/ai-and-the-limits-of-language/

3. Council of Europe. (2021), *Artificial intelligence, human rights, democracy and the rule of law,* https://arxiv.org/abs/2202.02776

4. Council of Europe (2024), *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law,* https://en.wikipedia.org/wiki/Framework_Convention_on_Artificial_Intelligence

5. Delve (2024), *AI in qualitative data analysis,* ScienceDirect, https://www.sciencedirect.com/science/article/pii/S2949882125000283

6. Ford, H., Bentall, C., & Jain, S, Ethical AI in social sciences research: Are we gatekeepers or revolutionaries? *Societies,* 15(3)/2024, https://www.mdpi.com/2075-4698/15/3/62

7. Lê Phúc & Trần Dương (2025), "Sự cần thiết xây dựng khung pháp lý về AI: Kinh nghiệm của các nước và một số đề xuất cho Việt Nam", Tạp chí *Pháp lý điện tử,* https://phaply.net.vn/su-can-thiet-xay-dung-khung-phap-ly-ve-ai-kinh-nghiem-cua-cac-nuoc-va-mot-so-de-xuat-cho-viet-nam-a258209.html

8. Madanchian, M., & Taherdoost, H, "The impact of artificial intelligence on research efficiency", *Results in Engineering,* 26/2025, 104743, https://doi.org/10.1016/j.rineng.2025.104743

9. Mokander, J., & Schroeder, R (2024), *AI and social theory.* arXiv, https://arxiv.org/abs/2407.06233

10. Nguyễn Giang Trường (2024), "Thực trạng về vấn đề bảo vệ dữ liệu cá nhân, kinh nghiệm một số quốc gia, khu vực và đề xuất hoàn thiện pháp luật ở Việt Nam hiện nay", Tạp chí *Công Thương,* (12)/2024.

11. Nguyễn Thị Quế Anh, "Hoàn thiện pháp luật sở hữu trí tuệ Việt Nam trong bối cảnh phát triển trí tuệ nhân tạo", Tạp chí *Luật học,* (3)/2022.

12. OECD (2022), *Regulatory sandboxes in artificial intelligence,* Organisation for Economic Cooperation and Development, https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence_8f80a0e6-en.html

13. Lê Thị Minh, "Sự sáng tạo của trí tuệ nhân tạo trong mối liên hệ với pháp luật về quyền tác giả", Tạp chí *Luật học,* số 5/2024.

14. Truby, J., Brown, C., & Dahdal, A (2021), *A sandbox approach to regulating high-risk artificial intelligence applications,* University of Twente Research Information System, https://ris.utwente.nl/ws/files/304762207/a_sandbox_approach_to_regulating_high_risk_artificial_intelligence_applications.pdf

15. Yeung, K., & Lodge, M. (2020), "Legal and human rights issues of AI: Gaps, challenges and vulnerabilities", *European Journal of Risk Regulation,* 11(2)/2020, https://www.sciencedirect.com/science/article/pii/S2666659620300056

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