Tạp chí Khoa học Nông nghiệp Việt Nam 2026, 24(4): 528-536 https://doi.org/10.31817/tckhnnvn.2026.24.04.10
vie.vjas.vn

SỰ SẴN SÀNG CHO PHÁT TRIỂN NHÂN LỰC TRÍ TUỆ NHÂN TẠO CÓ ĐẠO ĐỨC TẠI VIỆT NAM

Nguyễn Minh Đức*, Nguyễn Đăng Nam

Khoa Kinh tế Chính trị, Trường Đại học Kinh tế, Đại học Quốc gia Hà Nội

**Tác giả liên hệ: ducnm.ueb@vnu.edu.vn*

Ngày nhận bài: 06.10.2025 Ngày chấp nhận đăng: 19.03.2026

TÓM TẮT

Trong bối cảnh Việt Nam đang đẩy mạnh phát triển trí tuệ nhân tạo (AI) như một trụ cột chiến lược trong quá trình chuyển đổi số, câu hỏi về mức độ sẵn sàng về nhân lực cho nghiên cứu, phát triển và ứng dụng AI một cách có đạo đức đang trở nên cấp thiết. Nghiên cứu này nhằm đánh giá về sự sẵn sàng của Việt Nam trong việc phát triển nguồn nhân lực AI theo khuyến nghị của UNESCO về Đạo đức AI. Bài viết tiếp cận theo 5 trụ cột: (1) thể chế đạo đức; (2) nhân lực và kỹ năng; (3) dữ liệu và hạ tầng; (4) công bằng và bao trùm; (5) giám sát và thực thi; và áp dụng phương pháp phân tích nội dung các chiến lược quốc gia, chính sách hiện hành về AI và dữ liệu từ các tài liệu đã công bố của các báo cáo trong nước và quốc tế. Kết quả nghiên cứu chỉ ra rằng mức độ sẵn sàng của Việt Nam còn hạn chế, không đồng đều giữa các trụ cột. Dựa trên các phát hiện, nghiên cứu này đề xuất một số khuyến nghị chính sách nhằm củng cố nền tảng thể chế, tích hợp đào tạo đạo đức vào chương trình nghiên cứu, phát triển AI, xây dựng năng lực dữ liệu và thúc đẩy mô hình hợp tác mở.

Từ khóa: Khung RAM của UNESCO, rủi ro đạo đức AI, đào tạo nhân lực AI, chính sách phát triển AI.

Vietnam's Readiness for Ethical Artificial Intelligence Workforce Development

ABSTRACT

In the context of Vietnam promoting artificial intelligence (AI) as a strategic pillar of digital transformation, the question of workforce readiness for the ethical development and application of AI has become increasingly urgent. This study aims to assess Vietnam's readiness for AI human resource development in alignment with UNESCO's Recommendation on the Ethics of Artificial Intelligence. The paper applies five pillars: (1) ethical governance; (2) human resources and skills; (3) data and infrastructure; (4) fairness and inclusion; and (5) monitoring and enforcement and uses content analysis of national strategies, current policies, and data attracted from published reports and journal articiles. The findings reveal that Vietnam's level of readiness remains limited, uneven across the pillars. Based on the findings, this study proposes a set of policy recommendations aimed at strengthening institutional foundations, integrating ethics training into AI research and development, enhancing data capacity, and fostering open collaboration models.

Keywords: UNESCO RAM framework, AI ethical risks, AI human resource development, AI development policy.

1. ĐẶT VẤN ĐỀ

Trong bối cảnh trí tuệ nhân tạo (AI) đang nhanh chóng tái định hình cách thức xã hội vận hành, các quốc gia ngày càng đối diện với yêu cầu cấp thiết: phát triển AI một cách có đạo đức, công bằng, minh bạch và đáng tin cậy. AI không còn là một công nghệ trung lập, mà là kết quả của tập hợp các quyết định chính trị, kỹ thuật và xã hội, tiềm ẩn khả năng tái tạo hoặc làm trầm trọng thêm bất bình đẳng, thiên lệch và mất kiểm soát xã hội (Floridi & Cowls, 2019; Jobin & cs., 2019).

Nhận thức được điều đó, vào năm 2021, UNESCO đã ban hành Khuyến nghị toàn cầu về đạo đức AI, trong đó lần đầu tiên giới thiệu một khung đánh giá chính thức mang tên RAM - Readiness Assessment Methodology. Khung RAM được thiết kế nhằm hỗ trợ các quốc gia xây dựng năng lực phát triển và sử dụng AI một cách đạo đức.

Nhiều nghiên cứu gần đây chỉ ra rằng phần lớn thất bại trong triển khai AI có trách nhiệm không bắt nguồn từ việc thiếu nguyên tắc đạo đức, mà từ sự thiếu hụt năng lực đạo đức của đội ngũ thiết kế, triển khai và vận hành hệ thống AI (Floridi & Cowls, 2019; Mittelstadt, 2019). Vấn đề đạo đức AI đã thu hút sự quan tâm trong nghiên cứu quốc tế, nhưng các đánh giá về mức độ sẵn sàng của nhân lực AI có đạo đức còn hạn chế. Khoảng trống này cũng xuất hiện tại Việt Nam, nơi phát triển AI có trách nhiệm đã được xác lập như một định hướng chiến lược trong dài hạn. Quyết định số 127/QĐ-TTg năm 2021 đã đề ra các mục tiêu rõ ràng về xây dựng trung tâm nghiên cứu AI, phát triển hạ tầng dữ liệu và nâng cao chất lượng nguồn nhân lực AI đến năm 2030. Vận dụng các khuyến nghị về đạo đức AI của UNESCO (2021), bài nghiên cứu này đặt ra hai mục tiêu chính. Thứ nhất, đánh giá sự sẵn sàng cho phát triển nhân lực AI có đạo đức tại Việt Nam. Thứ hai, đề xuất các khuyến nghị chính sách có thể tích hợp đạo đức AI trong các chiến lược phát triển AI theo định hướng lớn về phát triển khoa học - công nghệ và đổi mới sáng tạo.

2. PHƯƠNG PHÁP NGHIÊN CỨU

Nghiên cứu được thực hiện đánh giá theo hai bước chính. Thứ nhất, nghiên cứu này tiến hành tổng hợp và phân tích các văn bản chính sách, tài liệu học thuật bao gồm các văn kiện từ UNESCO, OECD, Oxford Insights, cùng các báo cáo về thị trường lao động trong lĩnh vực công nghệ thông tin, và các nghiên cứu học thuật được xuất bản trên các tạp chí khoa học. Thứ hai, để đánh giá toàn diện mức độ sẵn sàng của hệ thống nhân lực trong việc phát triển và vận hành AI có đạo đức tại Việt Nam, nghiên cứu này vận dụng các khuyến nghị căn bản về đạo đức AI do UNESCO phát triển (UNESCO, 2021). Nghiên cứu sử dụng phương pháp nghiên cứu tài liệu (desk study), dựa trên dữ liệu từ văn bản chính sách, báo cáo và nghiên cứu học thuật. Việc lựa chọn RAM của UNESCO làm khung phân tích chính do RAM là khung tích hợp đồng thời thể chế - nhân lực - đạo đức - và điều kiện thực thi. Các khung khác như OECD AI Principles, hay UNDP AILA chủ yếu định hướng chính sách hoặc đánh giá hệ sinh thái AI ở mức tổng thể, không sát với đánh giá năng lực nhân lực AI có đạo đức.

Các khuyến nghị của UNESCO giúp việc đo lường mức độ hoàn thiện thể chế, và đặc biệt nhấn mạnh đến yếu tố con người, bao gồm năng lực kỹ thuật, hiểu biết đạo đức, tư duy phản biện và điều kiện thực hành trong toàn bộ hệ sinh thái nghiên cứu, phát triển và ứng dụng AI. Vận dụng khung RAM của UNESCO, nghiên cứu đã xây dựng 5 trụ cột chính gắn với phát triển nguồn nhân lực AI sẵn sàng cho đạo đức AI phù hợp với định hướng của Việt Nam (Bộ Thông tin và Truyền thông, 2024), bao gồm: (1) Thể chế đạo đức, (2) Nguồn nhân lực và kỹ năng, (3) Dữ liệu và hạ tầng, (4) Công bằng và bao trùm và (5) Giám sát và thực thi.

Thêm vào đó, nghiên cứu cũng áp dụng phương pháp so sánh nhằm đặt kết quả đánh giá của Việt Nam trong một hệ tham chiếu rộng hơn. Việc lựa chọn các quốc gia được thực hiện có chủ đích: EU có chuẩn pháp lý cao nhất về quản trị AI, Ấn Độ đại diện cho chiến lược đào tạo kỹ năng AI bao trùm, Singapore là nước tiên phong về quản trị AI và đánh giá thực hành; Thái Lan có mức độ tương đồng về quản trị AI trong khu vực ASEAN.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Thể chế về đạo đức AI

Trong những năm gần đây, Việt Nam đã thể hiện sự chủ động nhất định trong việc xây dựng nền tảng thể chế cho phát triển trí tuệ nhân tạo. Cột mốc quan trọng là Quyết định số 127/QĐ-TTg (2021) phê duyệt Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng AI đến năm 2030, trong đó nhấn mạnh phát triển hạ tầng nghiên cứu, dữ liệu và khung pháp lý cho AI. Gần đây hơn, Bộ Thông tin và Truyền thông (2024) ban hành Hướng dẫn phát triển AI có trách nhiệm, dựa trên các nguyên tắc phổ quát như minh bạch, bảo vệ quyền riêng tư, trách nhiệm giải trình và giám sát của con người, Quyết định số 1002/QĐ-TTg của Thủ tướng Chính phủ ngày 24/5/2025 về "Phê duyệt

Đề án đào tạo nguồn nhân lực phục vụ phát triển công nghệ cao giai đoạn 2025-2035 và định hướng tới năm 2045" đã lần đầu đặt đào tạo nhân lực AI trong đề án đào tạo công nghệ cao dài hạn. Một số sáng kiến phi nhà nước, như Ủy ban Đạo đức AI của VINASA hay Sổ tay AI Việt Nam (Private Compliance, 2023), cũng phản ánh sự gia tăng nhận thức về quản trị đạo đức AI trong cộng đồng chuyên môn.

Tuy nhiên, các nỗ lực này chủ yếu dừng ở mức định hướng và khuyến nghị, chưa được thể chế hóa thành các quy định pháp lý có tính ràng buộc. Đặc biệt, chưa có cơ chế chính thức gắn đạo đức AI với đào tạo, cấp chứng chỉ nghề nghiệp hoặc chuẩn hành nghề của nhân lực AI. Điều này khiến các nguyên tắc đạo đức khó chuyển hóa thành chuẩn mực thực hành trong nghiên cứu, phát triển và triển khai AI.

Đối chiếu với khuyến nghị về đạo đức AI của UNESCO, khoảng cách thể chế của Việt Nam trở nên rõ nét hơn. Báo cáo RAM của UNESCO (2025) cho thấy Việt Nam mới đạt mức "đang phát triển" ở hầu hết các trụ cột, đặc biệt yếu ở khía cạnh pháp lý - quản trị và cơ chế giám sát. Hiện chưa có quy định bắt buộc về phân loại rủi ro AI, đánh giá tác động đạo đức hoặc trách nhiệm giải trình thuật toán, trong khi đây là các yếu tố cốt lõi để bảo đảm nhân lực AI thực hành có trách nhiệm.

Các đánh giá độc lập cũng củng cố nhận định này. Theo chỉ số sẵn sàng AI của Chính phủ do Oxford Insights (2024) công bố, Việt Nam xếp thứ 76 toàn cầu và đứng thứ 5 trong ASEAN, phản ánh mức độ sẵn sàng thể chế ở mức trung bình. Báo cáo AILA của UNDP (2025) tiếp tục chỉ ra việc thiếu khung tiêu chuẩn đạo đức AI, thiếu quy trình quản lý rủi ro và không có yêu cầu đánh giá tác động đạo đức như những điểm nghẽn chính.

So với định hướng chung của ASEAN - nơi đạo đức AI được xem là vấn đề thể chế hóa thông qua cơ quan điều phối, chuẩn trách nhiệm nghề nghiệp và cơ chế giám sát độc lập - Việt Nam hiện vẫn tiếp cận theo hướng thử nghiệm và tự nguyện (ASEAN, 2024). Việt Nam thuộc nhóm đa số (cùng với Thái Lan, Indonesia, Malaysia, Philippines) đã ban hành Chiến lược quốc gia, Hướng dẫn Đạo đức AI, nhưng chưa hình thành cơ quan kiểm toán AI chuyên biệt.

Cụ thể, Thái Lan chưa thiết lập cơ quan kiểm toán AI độc lập mang tính pháp lý, nhưng đã hình thành các nhóm chuyên gia tư vấn và trung tâm chuyên trách về quản trị AI nhằm hỗ trợ xây dựng chính sách và hướng dẫn đạo đức cho các lĩnh vực then chốt. Các cơ chế này cho thấy bước chuyển từ định hướng sang hỗ trợ triển khai AI có trách nhiệm. Singapore đi xa hơn với thành lập Tổ chức Xác minh (AI Verify Foundation) và Trung tâm Đạo đức và Quản trị AI (Ethics and Governance Hub) – các thiết chế phối hợp công - tư để kiểm toán và hướng dẫn quản trị AI, dù vẫn dựa trên nguyên tắc tự nguyện (Bảng 1).

Sự khác biệt này làm nổi bật khoảng cách thể chế giữa Việt Nam và Singapore, đặc biệt ở khâu thực thi, nơi các nguyên tắc đạo đức được chuyển hóa thành tiêu chuẩn kỹ thuật, quy trình đánh giá rủi ro và trách nhiệm giải trình. Việc thiếu các cơ chế này trực tiếp hạn chế khả năng đào tạo và triển khai nhân lực AI có nền tảng đạo đức vững chắc tại Việt Nam (UNESCO, 2025).

Bảng 1. Khung chính sách AI quốc gia

Quốc gia	Chính sách AI quốc gia	Cơ chế giám sát đạo đức	Tổ chức kiểm toán AI độc lập
Singapore	Khung quản trị AI (AI Governance Framework, 2019)	Khung kỹ thuật xác minh AI (AI Verify)	Trung tâm Đạo đức và Quản trị AI Singapore (Ethics & Governance Hub)
Thái Lan	Hướng dẫn quản trị AI (AI Governance Guidelines, 2022)	Có nhóm tư vấn chuyên gia	Chưa có cơ quan kiểm toán AI độc lập có tính pháp lý
Việt Nam	Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo đến năm 2030 (2021)	Mới dừng ở cấp định hướng, khuyến nghị	Chưa có kiểm toán hoặc đánh giá rủi ro bắt buộc

Nguồn: IMDA (2024); ETDA (2022); Thủ Tướng Chính phủ (2021).

3.2. Nguồn nhân lực và kỹ năng

Nguồn nhân lực không chỉ là lực lượng thực thi mà còn là chủ thể kiến tạo, phản biện và kiểm soát đạo đức công nghệ. Trụ cột thứ hai cho sự sẵn sàng về đạo đức AI này xem con người là yếu tố quyết định để đảm bảo rằng các hệ thống AI không gây tổn hại xã hội và các giá trị nhân bản (Floridi & Cowls, 2019; Jobin & cs., 2019).

Tại Việt Nam, lực lượng lao động ICT tăng nhanh nhưng nhân lực AI chuyên sâu vẫn rất hạn chế. Theo Bộ Thông tin và Truyền thông (2024), Việt Nam có khoảng 1,3 triệu lao động ICT, song chỉ khoảng 5.000-6.000 người có chuyên môn thực sự về AI (Dang & Dao, 2021). UNESCO (2025) cũng ghi nhận năng lực AI chuyên sâu, đặc biệt trong các lĩnh vực liên quan đến đạo đức AI, còn ở mức thấp so với yêu cầu phát triển và quản trị AI có trách nhiệm.

- Khoảng cách giữa kỹ năng kỹ thuật và năng lực đạo đức

Đào tạo AI tại Việt Nam hiện vẫn thiên về thuật toán và hiệu năng, trong khi các nội dung như đánh giá thiên lệch dữ liệu, công bằng thuật toán hay tác động xã hội mới chỉ được đề cập hạn chế và hiếm khi là học phần bắt buộc (UNESCO, 2025). Khoảng cách này khiến kỹ sư AI thiếu năng lực phản biện đạo đức, ngay cả khi trình độ kỹ thuật cao - một rủi ro đã được Floridi & Cowls (2019) cảnh báo trong triển khai AI có trách nhiệm.

- Hệ sinh thái đào tạo còn phân mảnh và thiếu thực hành đạo đức

Hoạt động đào tạo nhân lực AI tại Việt Nam hiện chủ yếu tập trung ở một số cơ sở giáo dục hàng đầu như Đại học Bách khoa Hà Nội, Đại học Công nghệ (Đại học Quốc gia Hà Nội), FPT, VinUni và Đại học Quốc gia TP. Hồ Chí Minh. Tuy nhiên, hệ sinh thái đào tạo này vẫn thiếu các điều kiện đảm bảo tích hợp thực hành đạo đức. Cụ thể, các chương trình đào tạo AI còn thiếu mạng lưới mentor có kinh nghiệm triển khai AI có trách nhiệm, hạn chế về hạ tầng tính toán và cơ sở dữ liệu phục vụ thử nghiệm các mô hình về công bằng, minh bạch và thiếu các sandbox cho phép sinh viên thử nghiệm mô hình AI trong bối cảnh thực tế có sự giám sát xã hội (UNDP, 2025). Một số sáng kiến gần đây về đào tạo đạo đức AI, như Sổ tay AI Việt Nam (Private Compliance, 2023) hay các khóa học, hội thảo đầu tiên về đạo đức AI được tổ chức tại Việt Nam, cho thấy sự gia tăng nhận thức về tầm quan trọng của nhân lực AI có trách nhiệm. Tuy nhiên, các sáng kiến này chủ yếu mang tính thử nghiệm, chưa được chuẩn hóa trong chương trình đào tạo chính quy, chưa gắn với chuẩn đầu ra hay cơ chế đánh giá năng lực đạo đức của người học.

Hệ quả là người học thường thành thạo các chỉ số hiệu năng như sự chính xác, hiệu suất của mô hình, nhưng không quen thuộc với khái niệm kiểm định đạo đức hay đánh giá tác động xã hội của mô hình. UNESCO (2025) cũng cho thấy Việt Nam đạt mức điểm thấp ở tiêu chí năng lực đánh giá tác động xã hội của AI. Theo OECD (2023), sự thiếu hụt này làm trầm trọng thêm khoảng cách giữa khả năng lập trình và khả năng hành động có trách nhiệm trong thực tiễn.

- Thiếu khung năng lực quốc gia về đạo đức AI

Hiện nay, Việt Nam chưa ban hành hệ thống đánh giá năng lực AI cấp quốc gia hay khung chuẩn kỹ năng đạo đức trong lĩnh vực này, trong khi đây là thành phần then chốt để bảo đảm nguồn nhân lực đủ năng lực vận hành AI có trách nhiệm (UNESCO, 2025). Điều này dẫn đến một số hệ quả như: các trường đại học chưa được yêu cầu bắt buộc tích hợp đào tạo đạo đức AI trong chương trình chuẩn; chưa có tiêu chí kiểm định chất lượng đầu ra liên quan đến đạo đức trong đào tạo kỹ sư AI; doanh nghiệp khi tuyển dụng kỹ sư AI cũng không có tiêu chuẩn rõ ràng về đạo đức nghề nghiệp hoặc trách nhiệm xã hội.

- So sánh quốc tế

So sánh với hai quốc gia châu Á là Singapore và Ấn Độ có thể thấy rõ những hạn chế trong mô hình phát triển nguồn nhân lực AI của Việt Nam. Việt Nam chưa xây dựng được một khung kỹ năng quốc gia về AI, chưa có cơ chế kiểm định đầu ra liên quan đến năng lực

đạo đức. Trong khi đó, Singapore đã đi trước một bước với hệ thống đào tạo được chuẩn hóa theo lộ trình năng lực, trong đó yếu tố đạo đức không bị tách rời mà trở thành một tiêu chí xuyên suốt. Singapore đã đưa chuẩn đạo đức AI vào khung kỹ năng quốc gia từ năm 2020 và đã xây dựng lộ trình phát triển nhân lực AI với yếu tố đạo đức được tích hợp xuyên suốt cả ba cấp trong đó kỹ năng AI: cơ bản, ứng dụng và lãnh đạo (The Government of Singapore, 2023; Allen & cs., 2024). Hơn nữa, nước này còn thiết lập quan hệ đối tác chặt chẽ giữa nhà nước - đại học - doanh nghiệp, thúc đẩy hình thành một hệ sinh thái AI có trách nhiệm.

Ấn Độ cũng đã triển khai nhiều chương trình đào tạo AI miễn phí qua nền tảng MOOC, trong đó một số khóa học có lồng ghép nội dung về đạo đức và trách nhiệm AI. Mô hình này giúp mở rộng khả năng tiếp cận kiến thức AI theo hướng đạo đức cho công chúng.

3.3. Dữ liệu và hạ tầng

Theo UNESCO (2021), khả năng truy cập, xử lý và sử dụng hạ tầng dữ liệu, tính toán không chỉ là điều kiện kỹ thuật, mà còn là yếu tố cốt lõi bảo đảm rằng nhân lực AI có thể phát triển các hệ thống minh bạch, công bằng và có trách nhiệm. Trong bối cảnh AI ngày càng can thiệp vào các quyết định mang tính xã hội, dữ liệu trở thành một thực thể đạo đức cần được kiểm định. Việc đào tạo nhân lực AI có đạo đức vì vậy không thể tách rời khỏi năng lực tiếp cận dữ liệu kiểm định và môi trường hạ tầng đạt chuẩn để thực hành kỹ thuật đạo đức.

- Dữ liệu huấn luyện AI

Việt Nam chưa có cơ sở dữ liệu đào tạo AI chuẩn hóa, thiếu cơ chế đánh giá rủi ro đạo đức của dữ liệu và năng lực hạ tầng tính toán vẫn ở mức thấp (UNESCO, 2025). Theo Sổ tay AI Việt Nam các tổ chức phát triển AI trong nước gặp nhiều khó khăn trong việc tiếp cận dữ liệu đa dạng, có truy xuất nguồn gốc và có cơ chế đánh giá thiên lệch, những yếu tố then chốt để đào tạo và kiểm thử mô hình AI có trách nhiệm. Thực tế, các nguồn dữ liệu công chủ yếu nằm rải rác trong các bộ, ngành mà không thống nhất về chuẩn siêu dữ liệu, không có khả năng truy xuất nguồn gốc và chưa có cơ chế đánh giá đạo đức. Sinh viên và kỹ sư AI vì vậy không có cơ hội tiếp cận dữ liệu thực tế có kiểm định độ công bằng mô hình hoặc xác định các yếu tố có thể vi phạm quyền riêng tư. Việc đào tạo đạo đức AI do đó vẫn thiên về lý thuyết, thiếu nền tảng thực hành. Việc thiếu dữ liệu mở chất lượng cao là một trong những cản trở lớn nhất đối với năng lực phát triển và đánh giá AI có trách nhiệm của Việt Nam (UNDP, 2025; UNESCO, 2025).

- Hạ tầng tính toán

Việc huấn luyện các mô hình AI hiện đại, đặc biệt là các mô hình có chức năng đánh giá đạo đức như khả năng giải thích, tính công bằng hay trách nhiệm giải trình, đòi hỏi năng lực tính toán cao, bao gồm hệ thống GPU chuyên dụng hoặc trung tâm tính toán hiệu năng cao. Tuy nhiên, năng lực tính toán của Việt Nam ở mức "còn hạn chế", đặc biệt trong bối cảnh đào tạo kỹ năng đạo đức AI đòi hỏi hạ tầng để kiểm thử các mô hình phức tạp (UNESCO, 2025). Tài nguyên tính toán chủ yếu tập trung tại một số doanh nghiệp lớn như Viettel, Vingroup, FPT. Trong khi đó, các trường đại học và viện nghiên cứu nhỏ, cũng như các startup AI, gần như không thể tiếp cận các nguồn lực này.

Bảng 2. Mô hình đào tạo nhân lực AI

Quốc gia	Mô hình đào tạo nhân lực AI	Đặc điểm nổi bật
Singapore	Lộ trình phát triển nhân tài AI	Chuẩn kỹ năng AI 3 cấp, tích hợp đạo đức ở mọi tầng
Ấn Độ	Chương trình quốc gia về AI	MOOC đạo đức AI miễn phí, liên kết Coursera, edX, IITs
Việt Nam	Phân tán tại các trường đại học lớn, chưa đồng bộ	Thiếu chuẩn quốc gia, chưa có hệ thống đánh giá kỹ năng

Nguồn: Niti Aayog (2018); The Government of Singapore (2023).

- Chưa có hệ thống kiểm định đạo đức dữ liệu, mô hình

Một điểm yếu nghiêm trọng trong hệ sinh thái đạo đức AI tại Việt Nam là thiếu vắng cơ chế kiểm định đạo đức cho dữ liệu và mô hình. Trong khi EU và Singapore đều có các hướng dẫn kỹ thuật và công cụ cụ thể để kiểm thử thiên vị, công bằng hoặc khả năng truy xuất mô hình, Việt Nam vẫn chưa có bất kỳ quy chuẩn nào thực hiện các chức năng này.

Hệ quả là các kỹ sư và sinh viên AI không được đào tạo về cách kiểm định dữ liệu từ góc độ đạo đức, không có hướng dẫn phát hiện sai lệch cấu trúc và không có công cụ xác định mức độ đại diện xã hội của tập dữ liệu. Điều này dẫn đến việc các mô hình AI được phát triển trong nước có nguy cơ cao thiếu trách nhiệm đạo đức (Floridi & Cowls, 2019; Mittelstadt, 2019).

- So sánh quốc tế

Phân tích trên cho thấy rõ khoảng cách giữa Việt Nam với các quốc gia khác. Về hạ tầng hỗ trợ AI, Singapore vận hành cổng dữ liệu quốc gia Data.gov.sg kết hợp với Khung chia sẻ dữ liệu tin cậy, đồng thời cung cấp tài nguyên tính toán hiệu năng cao thông qua Trung tâm Siêu máy tính quốc gia (NSCC) để phục vụ nghiên cứu và doanh nghiệp. Trong khi đó, Ấn Độ phát huy hiệu quả của nền tảng dữ liệu chính phủ mở và mạng lưới siêu máy tính thuộc Chương trình Siêu máy tính quốc gia (NSM) kết nối các viện công nghệ hàng đầu (IITs). Hạ tầng dữ liệu mở và năng lực tính toán này tạo nền tảng thiết yếu để các cơ sở đào tạo tại Ấn Độ triển khai các chương trình thực hành chuyên sâu về đạo đức AI, bao gồm các bài toán về tính công bằng, quyền riêng tư và khả năng giải thích dựa trên các tập dữ liệu thực tế đã được chuẩn hóa. Việt Nam hiện mới bắt đầu xây dựng hạ tầng và dữ liệu, dẫn đến hệ sinh thái nhân lực AI thiếu công cụ thực hành và kiểm thử đạo đức trong môi trường dữ liệu thực. Điều này ảnh hưởng trực tiếp đến chất lượng sẵn sàng của nhân lực triển khai AI có trách nhiệm.

3.4. Công bằng và bao trùm

Công bằng và tính bao trùm là một trụ cột cốt lõi trong khung đánh giá mức độ sẵn sàng của UNESCO, phản ánh khả năng bảo đảm rằng các nhóm xã hội khác nhau có cơ hội tham gia vào quá trình phát triển, triển khai và giám sát AI (UNESCO, 2021). Trong tiếp cận này, bao trùm không chỉ là vấn đề phân phối lợi ích, mà còn là điều kiện tiên quyết để hình thành nguồn nhân lực AI có khả năng nhận diện và giảm thiểu các thiên lệch xã hội trong hệ thống AI.

Tại Việt Nam, mặc dù đã có các chính sách thúc đẩy đào tạo công nghệ, mức độ bao trùm trong phát triển nhân lực AI vẫn còn hạn chế. Tỷ lệ nữ giới tham gia các ngành liên quan đến AI chỉ dao động khoảng 15-18%, thấp hơn mức trung bình ASEAN và đáng kể so với Singapore (Dang & Dao, 2021; ASEAN, 2024). Sự thiếu đại diện này làm gia tăng nguy cơ thiên lệch giới trong thiết kế và vận hành hệ thống AI, đồng thời làm suy giảm năng lực phản biện đạo đức từ bên trong đội ngũ phát triển công nghệ.

Bảng 3. Dữ liệu và hạ tầng cho đào tạo nhân lực AI

Quốc gia	Hệ sinh thái dữ liệu & hạ tầng	Gắn kết với đào tạo nhân lực đạo đức
Singapore	Cổng dữ liệu mở Data.gov.sg và Khung chia sẻ dữ liệu tin cậy; Trung tâm Siêu máy tính quốc gia (NSCC)	Cung cấp sandbox cho đại học; hướng dẫn kiểm thử công bằng
Ấn Độ	Nền tảng dữ liệu chính phủ mở; Chương trình Siêu máy tính quốc gia (NSM)	Hợp tác với IITs để xây dựng học phần đạo đức AI trên dữ liệu thực tế đã được làm sạch
Việt Nam	Dữ liệu phân mảnh, chưa chuẩn hóa; chưa có Trung tâm Sêu tính toán AI quốc gia	Không có sandbox; chưa tích hợp pháp lý dữ liệu vào đào tạo đạo đức

Nguồn: Niti Aayog (2018); The Government of Singapore (2023); UNESCO (2025); UNDP (2025).

Bảng 4. Chính sách nhân lực AI bao trùm

Quốc gia	Chính sách nhân lực AI bao trùm	Tỷ lệ tham gia nữ giới/ nhóm yếu thế	Hạ tầng hỗ trợ tiếp cận đào tạo AI
Singapore	AI4E (AI cho mọi người) - đào tạo toàn dân miễn phí	28% sinh viên ngành AI là nữ	Nền tảng học AI trực tuyến mở; sandbox đại học
Ấn Độ	AI cho tất cả (AI for all) - chương trình AI hóa giáo dục phổ thông	Chương trình đa ngôn ngữ, miễn phí	Kết hợp MOOC, dữ liệu mở, tài nguyên cho trường công
Việt Nam	Chưa thực thi chính sách phổ cập AI	15-18% nữ; thiếu dữ liệu vùng sâu	Thiếu nền tảng phổ cập; chưa có cơ chế học AI mở

Nguồn: Niti Aayog (2018); Dang & Dao (2021); The Government of Singapore (2023); UNESCO (2025); UNDP (2025).

Bên cạnh đó, hệ sinh thái nhân lực AI tại Việt Nam hiện tập trung chủ yếu ở Hà Nội và TP.HCM, chiếm đến 82% tổng lực lượng lao động trong ngành công nghệ thông tin (Topdev, 2025; Vietnamworks, 2025), trong khi các vùng Trung bộ, Tây Bắc và Tây Nguyên gần như không có chương trình đào tạo AI chuyên sâu. Sự mất cân đối vùng miền này khiến nhiều cộng đồng không có tiếng nói trong quá trình thiết kế và giám sát AI, làm gia tăng rủi ro áp dụng công nghệ không phù hợp với bối cảnh xã hội địa phương.

So với các quốc gia tiên phong trong khu vực, Việt Nam chưa có chính sách đào tạo AI mang tính phổ cập gắn với các chuẩn mực đạo đức (UNDP, 2025). Trong khi Singapore triển khai chương trình "AI cho mọi người" và Ấn Độ thúc đẩy sáng kiến "AI cho tất cả" nhằm mở rộng tiếp cận AI có trách nhiệm cho toàn dân, Việt Nam vẫn thiếu các cơ chế tương tự ở cấp quốc gia. Điều này khiến đào tạo nhân lực AI có nguy cơ trở thành đặc quyền của một số nhóm xã hội, làm suy giảm tính đại diện và năng lực đạo đức của hệ sinh thái AI trong dài hạn.

3.5. Giám sát và thực thi

Giám sát và thực thi là trụ cột then chốt nhằm bảo đảm các nguyên tắc đạo đức AI được chuyển hóa từ tuyên ngôn thành cơ chế vận hành và trách nhiệm ràng buộc trong thực tiễn (UNESCO, 2021). Đây cũng là nơi giao thoa giữa năng lực kỹ thuật và chuẩn mực đạo đức xã hội, tạo điều kiện để nhân lực AI hình thành năng lực phản biện, giải trình và tự điều chỉnh theo các giá trị đạo đức chung.

- Thiếu khung kiểm toán và đánh giá tác động đạo đức mô hình AI

Hiện nay, Việt Nam mới dừng ở các hướng dẫn phát triển AI có trách nhiệm (Bộ Thông tin và Truyền thông, 2024), trong khi chưa có quy định ràng buộc về phân loại rủi ro đạo đức, kiểm toán mô hình hay đánh giá tác động đạo đức trước khi triển khai AI (UNESCO, 2025). Khoảng trống này tạo ra sự lệch pha đáng kể giữa chính sách và thực hành, đặc biệt trong đào tạo và phát triển nhân lực AI. Ngược lại, EU yêu cầu các hệ thống AI rủi ro cao phải chịu kiểm toán độc lập theo EU AI Act (2023), còn Singapore triển khai bộ công cụ Xác minh AI nhằm hỗ trợ kiểm định tính công bằng và khả năng giải thích. Việc thiếu các cơ chế tương tự khiến kỹ sư AI tại Việt Nam khó tiếp cận chuẩn mực quốc tế về trách nhiệm giải trình, một thành tố cốt lõi của đạo đức nghề nghiệp trong AI.

- Không có kênh phản hồi, khiếu nại hay điều chỉnh xã hội

Việt Nam chưa có cổng thông tin nào công khai danh sách mô hình AI đang được triển khai, chưa có cơ chế chính thức để phản hồi tác động xã hội từ hệ thống AI, chưa có quy trình điều tra độc lập từ các mô hình sai lệch. Báo cáo của UNESCO (2025) cho rằng Việt Nam chưa có cơ chế minh bạch hóa và phản hồi từ người dân, làm suy giảm khả năng giám sát xã hội đối với các hệ thống AI đang vận hành. Theo Jobin & cs. (2019), thiếu vắng cơ chế phản hồi là dấu hiệu rõ nhất của đạo đức hình thức, nơi nguyên tắc đạo đức được tuyên bố nhưng không tạo ra hành vi đạo đức cụ thể trong thực tiễn triển khai.

Bảng 5. Cơ chế kiểm toán mô hình AI

Quốc gia	Cơ chế kiểm toán mô hình AI	Báo cáo đạo đức bắt buộc	Cổng/kênh phản hồi công khai
EU	Kiểm toán bắt buộc đối với mô hình rủi ro cao theo AI Act (2023); các tổ chức phải đăng ký mô hình AI vào hệ thống kiểm soát quốc gia	Có yêu cầu báo cáo EIA (Ethical Impact Assessment) với mô hình rủi ro cao	Có Cổng tiếp nhận khiếu nại AI từ cá nhân và tổ chức
Singapore	AI Verify Toolkit - hướng dẫn kỹ thuật + đánh giá độc lập; áp dụng tự nguyện nhưng khuyến nghị mạnh với mô hình triển khai công	Có yêu cầu tự đánh giá đạo đức và chuẩn bị sẵn sàng báo cáo giải trình	Có Cổng phản hồi về AI tích hợp với dịch vụ công số
Việt Nam	Chưa có bất kỳ cơ chế kiểm toán chính thức hoặc phân loại rủi ro AI	Không có yêu cầu báo cáo đạo đức hay đánh giá EIA	Không tồn tại kênh chính thức phản hồi, khiếu nại về AI

Nguồn: European Parliament (2024); Allen & cs. (2024); The Government of Singapore (2023); UNESCO (2025).

4. KHUYẾN NGHỊ CHÍNH SÁCH

Dựa trên các phát hiện đã trình bày ở trên, chúng tôi đề xuất một số định hướng chính sách gợi mở nhằm tăng cường mức độ sẵn sàng về đạo đức cho nhân lực AI tại Việt Nam. Các khuyến nghị này hướng đến việc tham khảo và thảo luận thêm bởi các cơ quan hoạch định chính sách, cơ sở đào tạo và các bên liên quan trong hệ sinh thái AI.

Trước hết, cần xem xét khả năng xây dựng một khung năng lực quốc gia về đạo đức AI, tích hợp các năng lực kỹ thuật như phát hiện thiên lệch, truy xuất nguồn dữ liệu và giải thích mô hình, cùng với các năng lực xã hội như đánh giá tác động xã hội và trách nhiệm giải trình. Khung này là cơ sở để các cơ sở giáo dục, viện nghiên cứu và doanh nghiệp tham chiếu trong thiết kế chương trình đào tạo, bồi dưỡng và đánh giá năng lực nhân lực AI.

Bên cạnh đó, việc phát triển một hệ sinh thái dữ liệu và hạ tầng tính toán mở - bao gồm các kho dữ liệu huấn luyện đã được kiểm định đạo đức, hệ thống thử nghiệm AI (AI sandbox) và cụm GPU hỗ trợ nghiên cứu các mô hình công bằng - là điều kiện quan trọng để đảm bảo rằng đạo đức không chỉ là lý thuyết, mà có thể được thực hành thực tế. Việt Nam có thể tham khảo kinh nghiệm từ Singapore và Ấn Độ về dữ liệu mở có kiểm định và các trung tâm tính toán chung để từng bước hình thành môi trường thực hành đạo đức AI.

Thêm vào đó, chính sách AI cũng cần quan tâm đến tính bao trùm xã hội, nhằm bảo đảm rằng mọi nhóm dân cư đều có khả năng tham gia vào hệ sinh thái AI. Việc mở rộng các chương trình phổ cập AI cho cộng đồng, đặc biệt là phụ nữ, nhóm dân tộc thiểu số, thanh niên và lao động phổ thông, nên được cân nhắc như một hướng đi để nâng cao sự đa dạng và giảm thiểu thiên lệch xã hội trong thiết kế và vận hành hệ thống AI.

Cuối cùng, cần thiết lập hoặc củng cố các cơ chế giám sát và phản hồi để đảm bảo rằng các hệ thống AI vận hành trong khu vực công và trong các dịch vụ nhạy cảm đều được công bố minh bạch, có đánh giá rủi ro và được kiểm toán đạo đức bởi các tổ chức độc lập.

5. KẾT LUẬN

Kết quả nghiên cứu cho thấy Việt Nam đã thể hiện cam kết chiến lược về phát triển AI, song mức độ thể chế hóa các tiêu chuẩn đạo đức, cơ chế giám sát, cũng như năng lực triển khai trong hệ sinh thái phát triển nhân lực vẫn còn hạn chế. Đặc biệt, sự thiếu vắng khung năng lực đạo đức AI, hệ thống dữ liệu được kiểm định, hạ tầng tính toán mở và cơ chế kiểm toán - phản hồi xã hội đã tạo ra những khoảng trống đáng kể giữa năng lực kỹ thuật và khả năng nhận diện - phản biện - giải trình đạo đức của lực lượng nhân lực AI. Từ các phát hiện trên, các khuyến nghị chính sách được đề xuất nhằm củng cố nền tảng thể chế, tích hợp đào tạo đạo đức AI vào phát triển nguồn nhân lực.

Như vậy, về phương diện học thuật, bài viết đóng góp ba điểm nổi bật. Thứ nhất, bài viết xây dựng một khung phân tích tích hợp dựa trên

UNESCO RAM, cho phép đánh giá không chỉ năng lực kỹ thuật mà cả năng lực xã hội, đạo đức của hệ sinh thái nhân lực AI, một hướng tiếp cận còn mới. Thứ hai, nghiên cứu đã hệ thống hóa các bằng chứng thực tiễn dựa trên các kết quả đánh giá tài liệu chính sách, các xuất bản từ nhiều nguồn uy tín, nhờ đó tạo ra một bức tranh nhất quán về sự sẵn sàng cho phát triển nhân lực AI có trách nhiệm của Việt Nam. Thứ ba, bài viết bổ sung góc nhìn so sánh quốc tế, qua đó nhận diện rõ hơn những yếu tố đang ảnh hưởng đến sự phát triển nhân lực AI có đạo đức tại Việt Nam.

TÀI LIỆU THAM KHẢO

Allen J.G., Loo J. & Campoverde J.L.L. (2024). Governing intelligence: Singapore's evolving AI governance framework. Cambridge Forum on AI: Law and governance. 1(e12): 1-20.

ASEAN S. (2024). ASEAN guide on AI governance and ethics. ASEAN. Retrieved from https://asean.org/book/asean-guide-on-ai-governance-and-ethics on Dec 25, 2025.

Bộ Thông tin và Truyền thông (2024). Báo cáo Tổng kết công tác năm 2024, Phương hướng, nhiệm vụ năm 2025. Truy cập từ https://mic.mediacdn.vn/639352410187198464/2024/12/28/3-bao-cao-tom-tat-17353979879001550546803.pdf ngày 5/5/2025.

Dang Q.A. & Dao Q.V. (2021). Human resources development readiness in ASEAN: Vietnam country report. Coventry University.

ETDA - Electronic Transactions Development Agency (2022). Thailand's AI Governance Guideline. Retrieved from https://www.etda.or.th/getattachment/Our-Service/AIGC/Research-and-Recommendation/Thailand%E2%80%99s-AI-Governance-Guideline-for-Executive_2023.pdf.aspx?lang=th-TH on Sep 8, 2025.

European Parliament (2024). Artificial Intelligence Act. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689 on Aug 8, 2025.

Floridi L. & Cowls J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review. 4(1).

IMDA - Infocomm Media Development Authority (2024). AI Verify Foundation and AI Verify Framework. Retrieved from https://www.imda.gov.sg/about-imda/emerging-technologies-and-research/artificial-intelligence on Jul 6, 2025.

Jobin A., Ienca M. & Vayena E. (2019). The global landscape of AI ethics guidelines. Nature machine intelligence. 1(9): 389-399.

Mittelstadt B. (2019). Principles alone cannot guarantee ethical AI. Nature machine intelligence. 1(11): 501-507.

Niti Aayog (2018). AI for All: India's National AI Strategy. Retrieved from https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf ngày on Jul 6, 2025.

OECD (2023). Artificial Intelligence and the Labour Market: What Do We Know So Far? Retrieved from https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/01/the-impact-of-artificial-intelligence-on-the-labour-market_a4b9cac2/7c895724-en.pdf on May 5, 2025.

Oxford Insights (2024). Government AI Readiness Index. Retrieved from https://oxfordinsights.com/ai-readiness/ai-readiness-index/ on Sep 9, 2025.

Private Compliance (2023). Vietnam AI handbook: A practical guide to responsible artificial intelligence in Vietnam. Private Compliance Asia.

The Government of Singapore (2023). Singapore's National AI Strategy 2.0 (NAIS 2.0). Retrieved from https://www.smartnation.gov.sg/ initiatives/national-ai-strategy on Sep 9, 2025.

Thủ Tướng Chính Phủ (2021). Quyết định Ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo đến năm 2023. Truy cập từ https://chinhphu.vn/?pageid=27160&docid=202565&tagid=6&type=1 ngày 5/5/2024.

Thủ Tướng Chính Phủ (2025). Quyết định số 1002/QĐ-TTg của Thủ tướng CP về "Phê duyệt Đề án đào tạo nguồn nhân lực phục vụ phát triển công nghệ cao giai đoạn 2025-2035 và định hướng tới năm 2045".

Topdev (2025). Báo Cáo Thị trường IT Việt Nam 2024-2025: Toàn cảnh Nhân tài CNTT và Công nghệ Việt Nam. Truy cập từ https://topdev.vn/blog/bao-cao-thi-truong-it-viet-nam-2024/ ngày 5/5/2025.

UNDP (2025). Artificial Intelligence Landscape Assessment (AILA): Vietnam 2025. Vietnam: United Nations Development Programme.

UNESCO (2021). Recommendation on the ethics of artificial intelligence. Retrieved from https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence on Mar 5, 2024.

UNESCO (2025). Viet Nam: Artificial Intelligence Readiness Assessment Report. Viet Nam: United Nations Educational, Scientific and Cultural Organization.

Vietnamworks (2025). Vietnamworks inTECH: Báo cáo Thực trạng nhân sự và tuyển dụng ngành CNTT trong làn sóng trí tuệ nhân tạo giai đoạn 2024-2025. Truy cập từ https://www.vietnamworks.com/tram-sac/edocuments/it-report-2024 ngày 7/5/2025.