

ĐO LƯỜNG NHẬN THỨC ĐẠO ĐỨC AI CỦA SINH VIÊN THEO 7 NGUYÊN TẮC UNESCO (2021): THỰC TRẠNG VÀ HÀM Ý

Bùi Thanh Thủy¹, Nguyễn Thu Hà¹, Phạm Thúy Hà¹,
Hoàng Thúy An¹, Nguyễn Thị Hải¹

Tóm tắt: Nghiên cứu này vận dụng Khuyến nghị về đạo đức trí tuệ nhân tạo của UNESCO (2021) để xây dựng và áp dụng một thang đo dựa trên các nguyên tắc đạo đức AI nhằm đánh giá mức độ nhận thức đạo đức AI của sinh viên đại học Việt Nam. Khảo sát được thực hiện với 307 sinh viên tại các cơ sở giáo dục đại học. Kết quả cho thấy nhận thức đạo đức AI của sinh viên đạt mức tương đối cao, đặc biệt ở các khía cạnh độ tin cậy kỹ thuật, quyền riêng tư và minh bạch. Phân tích so sánh cho thấy giới tính, khối ngành và tần suất sử dụng AI không tạo ra sự khác biệt có ý nghĩa thống kê về mức độ nhận thức đạo đức AI. Trong khi đó, sự khác biệt theo năm học được ghi nhận, với sinh viên năm nhất có mức nhận thức cao hơn sinh viên năm hai. Trên cơ sở đó, nghiên cứu đề xuất tăng cường giáo dục đạo đức AI trong chương trình đào tạo đại học, đặc biệt ở giai đoạn đầu, nhằm hình thành thói quen sử dụng AI có trách nhiệm cho sinh viên.

Từ khóa: đạo đức trí tuệ nhân tạo, nhận thức, sinh viên, giáo dục đại học.

1. MỞ ĐẦU

Trong bối cảnh chuyển đổi số và cách mạng công nghiệp 4.0, việc ứng dụng trí tuệ nhân tạo (AI) đã ngày càng trở nên rộng khắp và tác động tới mọi mặt của đời sống xã hội (Huang và cộng sự, 2022) [1]. Trong lĩnh vực giáo dục, các công cụ AI đã giúp cải thiện hiệu suất học tập, tăng cường sáng tạo và hỗ trợ việc học tự chủ của sinh viên. Khả năng cá nhân hóa và trở thành trợ lý học tập thông minh đã tạo ra những thay đổi mới trong việc học, sáng tạo và nghiên cứu của sinh viên (Singh & Aziz, 2025; Chanpetch & Songserm, 2023) [2,3]. Tuy nhiên, bên cạnh những cơ hội, việc ứng dụng AI trong giáo dục cũng kéo theo những rủi ro đạo đức đáng kể như phụ thuộc công nghệ, mất khả năng tư duy phản biện, và mơ hồ về tính xác thực của sản phẩm học tập do AI hỗ trợ (Zulfa et al., 2025) [4]. Vì vậy, trong bối cảnh giáo dục đại học tại Việt Nam, nơi mà sinh viên ngày càng sử dụng các công cụ AI trong học tập, sáng tạo và nghiên cứu, việc đánh giá nhận thức đạo đức về AI trở nên hết sức cần thiết.

Mặc dù đã có nhiều công trình khảo sát về ứng dụng AI trong giáo dục và nhận thức về AI tại các quốc gia khác, nhưng tại Việt Nam vẫn chưa có nhiều nghiên cứu liên quan tới nhận thức đạo đức về AI của sinh viên đại học. Các thang đo hiện có thường được phát triển ở các quốc gia khác và chưa chắc đã phù hợp với đặc thù văn hóa, giáo dục và

¹ Trường Đại học Khoa học Xã hội và Nhân văn, ĐHQGHN

hành vi sử dụng AI của sinh viên Việt Nam. Vì vậy, nghiên cứu này vận dụng Khuyến nghị về đạo đức trí tuệ nhân tạo của UNESCO (2021) để đo lường nhận thức đạo đức AI của sinh viên đại học Việt Nam. Nghiên cứu nhằm: (1) làm rõ thực trạng nhận thức của sinh viên theo các nguyên tắc đạo đức AI của UNESCO; (2) xem xét sự khác biệt giữa các nhóm sinh viên nhằm nhận diện các nguyên tắc cần ưu tiên trong giáo dục đạo đức AI; và (3) đề xuất hàm ý giáo dục nhằm thúc đẩy việc sử dụng AI có trách nhiệm trong giáo dục đại học.

2. NỘI DUNG NGHIÊN CỨU

2.1. Cơ sở lý thuyết

2.1.1. Đạo đức trí tuệ nhân tạo

Đạo đức AI là một lĩnh vực liên ngành, nghiên cứu các vấn đề đạo đức nảy sinh trong quá trình thiết kế, phát triển, triển khai và sử dụng các hệ thống trí tuệ nhân tạo. Theo Siau & Wang (2020) [5], đạo đức AI bao gồm hai khía cạnh: đạo đức của AI (ethics of AI) - đề cập đến các nguyên tắc, quy định và chính sách định hướng việc phát triển công nghệ, và AI có đạo đức (ethical AI) – liên quan đến việc xây dựng hệ thống có khả năng hành xử theo các tiêu chuẩn đạo đức đã xác lập.

Theo Kazim & Koshiyama (2021) [6], đạo đức AI là một lĩnh vực nghiên cứu có phạm vi rộng, bao gồm tác động tâm lý học (tính tự chủ tinh thần, nhận biết khi tương tác với thực thể phi con người), xã hội học (công bằng, thiên lệch, tác động môi trường), và chính trị học (ảnh hưởng tới dân chủ, cấu trúc kinh tế – quyền lực). Trong khi đó, Hanna và cộng sự (2025) [7] nhấn mạnh rằng đạo đức AI cần được lồng ghép vào toàn bộ vòng đời phát triển công nghệ, từ thu thập dữ liệu, xây dựng thuật toán, vận hành cho đến giám sát liên tục, nhằm ngăn ngừa tổn hại và bảo đảm kết quả công bằng. Weidener & Fisher (2024) [8], định nghĩa đạo đức AI là một tập hợp các giá trị, nguyên lý và kỹ thuật nhằm hướng dẫn hành vi đạo đức trong việc sử dụng AI dựa trên các chuẩn mực đúng – sai được xã hội công nhận. Tuy vậy, Mittelstadt (2019) [9] cảnh báo rằng các nguyên tắc này vẫn còn mang tính trừu tượng, thiếu cơ chế thực thi rõ ràng trong bối cảnh phát triển AI hiện nay. Trong giáo dục, Leta & Vancea (2023) [10] cho rằng việc triển khai một khung đạo đức toàn diện – bao gồm chính sách minh bạch, kiểm soát thuật toán và ngăn chặn thiên lệch – là điều kiện tiên quyết để đảm bảo một môi trường giáo dục công bằng và có trách nhiệm.

Từ những phân tích trên, có thể khẳng định rằng đạo đức AI không chỉ là vấn đề kỹ thuật mà còn là một năng lực cần thiết trong thời đại số. Năng lực này đặc biệt quan trọng trong giáo dục đại học, nơi mà sinh viên cần được trang bị không chỉ kiến thức công nghệ mà còn là khả năng đánh giá và ra quyết định một cách có đạo đức trong thế giới ngày càng phụ thuộc vào AI.

2.1.2. Khuyến nghị về đạo đức trí tuệ nhân tạo của UNESCO

Năm 2021, UNESCO đã ban hành Khuyến nghị về đạo đức AI [11]. Tài liệu này đã đưa ra các nguyên tắc nhằm bảo đảm rằng AI được thiết kế, triển khai và sử dụng theo hướng an toàn, có trách nhiệm và tôn trọng quyền con người. Trong tài liệu này có 10 nguyên tắc nhưng nhóm tác giả của nghiên cứu này chỉ lựa chọn bảy nguyên tắc phù hợp nhất với cấp độ nhận thức của cá nhân để xây dựng thang đo đánh giá sinh viên. Ba nguyên tắc còn lại liên quan nhiều hơn tới quản trị cấp vĩ mô, chính sách quốc gia và cơ chế tổ chức, không phù hợp để đo lường dưới góc độ cá nhân sinh viên. Do đó, nhóm nghiên cứu không đưa các nguyên tắc này vào để nghiên cứu.

Nguyên tắc thứ nhất là “Chủ quyền của con người và giám sát của con người” – con người phải giữ vai trò trung tâm, duy trì quyền tự chủ và tham gia giám sát trong toàn bộ quá trình vận hành của hệ thống AI. Nguyên tắc thứ hai là “Độ tin cậy kỹ thuật và an toàn” - các hệ thống AI phải ổn định, đáng tin cậy, được thử nghiệm đầy đủ và có cơ chế ứng phó khi xảy ra lỗi hoặc rủi ro kỹ thuật. Nguyên tắc thứ ba là “Quyền riêng tư và quản trị dữ liệu” – đề cập đến việc bảo vệ dữ liệu cá nhân, đảm bảo minh bạch trong quá trình thu thập, xử lý và sử dụng dữ liệu, đồng thời duy trì quyền kiểm soát của người dùng đối với thông tin của họ. Nguyên tắc thứ tư là “Minh bạch và khả năng giải thích” – hệ thống AI phải hoạt động một cách minh bạch, cho phép người dùng hiểu được căn cứ, dữ liệu và quy trình đưa ra quyết định hoặc gợi ý. Nguyên tắc thứ năm là “Công bằng và không phân biệt đối xử” – nhấn mạnh việc ngăn chặn mọi hình thức thiên vị do dữ liệu hoặc thuật toán gây ra, nhằm đảm bảo rằng AI không tạo ra bất công hoặc làm tổn hại bất kỳ nhóm xã hội nào. Nguyên tắc thứ sáu là “Phúc lợi xã hội và môi trường” – xem xét các tác động rộng hơn của AI đối với môi trường, văn hóa, đời sống cộng đồng và phát triển bền vững, bảo đảm rằng AI phục vụ lợi ích chung của xã hội. Và nguyên tắc cuối cùng là “Trách nhiệm và giải trình” – yêu cầu xác định rõ trách nhiệm giữa các bên tham gia thiết kế, phát triển và sử dụng AI, đồng thời bảo đảm khả năng truy vết và giải trình khi hệ thống gây ra hậu quả tiêu cực.

Bảy nguyên tắc trên tạo thành cơ sở lý thuyết quan trọng cho nghiên cứu này. Chúng cũng là nền tảng cho việc xây dựng thang đo nhằm đánh giá mức độ nhận thức đạo đức về AI của sinh viên trong bối cảnh giáo dục đại học Việt Nam.

2.1.3. Các nghiên cứu liên quan đến đạo đức AI

Nhận thức đạo đức AI trong sinh viên đang trở thành chủ đề được quan tâm trong giáo dục đại học trên toàn thế giới. Tại Nhật Bản, Ghotbi & Ho (2021) [12] ghi nhận sinh viên lo ngại nhiều nhất về thất nghiệp do AI, trong khi các yếu tố như cảm xúc, phân biệt đối xử và kiểm soát xã hội cũng được đề cập. Tại Đông Nam Á, Valerio (2024) [13] cho thấy sinh viên Philippines có mức nhận thức cao về AI nhưng cũng bày tỏ lo ngại về quyền riêng tư và việc làm. Ferhataj et al. (2025) [14] tại Albania khẳng định vai trò của minh bạch và trách nhiệm pháp lý, nhấn mạnh nhu cầu tăng cường giáo dục đạo đức AI trong trường đại học.

Về công cụ đo lường, các thang đo như AEPS (Saatci, 2025) [15] và AIERS (Wang et al., 2025) [16] đã được phát triển để đánh giá nhận thức đạo đức AI qua các yếu tố như minh bạch, trách nhiệm, công bằng, và nhận thức phê phán. Các nghiên cứu cũng chỉ ra sự khác biệt theo giới tính và tần suất sử dụng AI. Ngoài ra, can thiệp giáo dục như mô-đun phản biện trực tuyến (Usher & Barak, 2024) [17] đã chứng minh hiệu quả trong việc nâng cao khả năng nhận diện và giải quyết tình huống đạo đức của sinh viên.

Tuy nhiên, phần lớn các nghiên cứu hiện nay mới chỉ tiếp cận nhận thức đạo đức AI của sinh viên thông qua các khía cạnh đạo đức tổng hợp hoặc các thang đo được phát triển trong những bối cảnh giáo dục cụ thể. Chưa có nhiều nghiên cứu đặt trọng tâm vào việc đo lường nhận thức theo các nguyên tắc đạo đức chuẩn hóa ở cấp độ quốc tế như Khuyến nghị của UNESCO (2021). Đặc biệt, các nghiên cứu xem xét cách các nguyên tắc đạo đức AI của UNESCO được sinh viên nhận thức và ưu tiên trong bối cảnh giáo dục đại học còn thiếu. Đây chính là khoảng trống mà nghiên cứu này hướng tới, nhằm cung cấp cơ sở khoa học cho việc thiết kế giáo dục đạo đức AI trong các cơ sở giáo dục đại học Việt Nam.

2.2. Phương pháp nghiên cứu

2.2.1. Đối tượng và mẫu nghiên cứu

Đối tượng nghiên cứu là sinh viên đại học đang theo học tại 8 cơ sở giáo dục đại học tại Việt Nam, thuộc hai khối ngành chính: khoa học tự nhiên và kỹ thuật, và khoa học xã hội và nhân văn. Nghiên cứu sử dụng phương pháp chọn mẫu thuận tiện để tiếp cận người tham gia. Mẫu nghiên cứu được cân đối về giới tính, năm học và ngành đào tạo nhằm đảm bảo tính đại diện cho nhiều nhóm đối tượng và mức độ tiếp xúc khác nhau với công nghệ AI.

2.2.2. Công cụ thu thập dữ liệu

Nghiên cứu sử dụng bảng hỏi khảo sát để thu thập dữ liệu. Bảng hỏi gồm hai phần là thông tin cá nhân và thang đo nhận thức đạo đức AI. Phần thông tin cá nhân thu thập các thông tin về giới tính, năm học, khối ngành đào tạo và tần suất sử dụng AI trong học tập hoặc công việc hàng ngày.

Các biến này được dùng để phân tích sự khác biệt trong nhận thức đạo đức AI giữa các nhóm sinh viên. Phần Thang đo nhận thức đạo đức AI đo lường mức độ nhận thức đạo đức của sinh viên dựa trên thang Likert 5 mức (1 = Hoàn toàn không đồng ý đến 5 = Hoàn toàn đồng ý). Thang đo gồm 21 mục hỏi và chia thành 7 khía cạnh được xây dựng dựa trên bảy nguyên tắc đạo đức của UNESCO (2021) như đã nêu ở mục 2.2. Các khía cạnh bao gồm: (1) Chủ quyền và giám sát của con người, (2) Độ tin cậy kỹ thuật và an toàn, (3) Quyền riêng tư và quản trị dữ liệu, (4) Minh bạch và khả năng giải thích, (5) Công bằng và không thiên vị, (6) Phúc lợi xã hội và môi trường, và (7) Trách nhiệm và giải trình. Mỗi nguyên tắc được đo bằng 3 mục hỏi độc lập.

2.2.3. Quy trình thu thập và xử lý dữ liệu

Dữ liệu được thu thập thông qua bảng hỏi trực tuyến được thiết kế trên nền tảng Google Forms và phân phối trong khoảng thời gian khảo sát tháng 10 – 11/2025. Người tham gia được mời tham gia khảo sát trên cơ sở tự nguyện và không nhận bất kỳ hình thức khuyến khích vật chất nào.

Dữ liệu được xử lý bằng phần mềm SPSS, với các phân tích gồm: kiểm định độ tin cậy thang đo, phân tích mô tả theo các nguyên tắc đạo đức AI của UNESCO (2021), và phân tích so sánh nhóm (t-test, ANOVA) nhằm xác định các ưu tiên trong giáo dục đạo đức AI cho sinh viên.

2.2.4. Đạo đức nghiên cứu

Nghiên cứu được thực hiện trên nguyên tắc tự nguyện và ẩn danh. Tất cả người tham gia được thông tin đầy đủ về mục đích, nội dung và quyền rút lui khỏi khảo sát bất kỳ lúc nào. Dữ liệu thu thập được bảo mật và chỉ phục vụ cho mục đích nghiên cứu học thuật. Để giảm thiểu thiên lệch xã hội, bảng hỏi được thiết kế ẩn danh, không có đáp án đúng – sai, và người tham gia được nhấn mạnh rằng khảo sát không nhằm đánh giá cá nhân hay kết quả học tập.

2.3. Kết quả nghiên cứu

2.3.1. Thông tin mô tả mẫu

Nghiên cứu đã khảo sát tổng số 307 sinh viên, bao gồm đại diện từ các nhóm giới tính, ngành học, năm học và mức độ sử dụng AI khác nhau. Dữ liệu mô tả được trình bày trong Bảng 1 dưới đây:

Bảng 1. Mức độ sử dụng AI các nhóm

Biến	Nhóm	Số phiếu	Tỷ lệ
Giới tính	Nam	151	49,2 %
	Nữ	156	50,8%
Khối ngành	Khoa học tự nhiên và kỹ thuật	151	49,2 %
	Khoa học xã hội và nhân văn	156	50,8%
Năm học	Năm nhất	78	25,4 %
	Năm hai	76	24,8 %
	Năm ba	75	24,4%
	Năm tư	78	25,4 %
Tần suất sử dụng	Rất thường xuyên	91	29,6%

Biến	Nhóm	Số phiếu	Tỷ lệ
	Thường xuyên	155	50,5%
	Thỉnh thoảng	61	19,9%

Kết quả cho thấy tỷ lệ giới tính và khối ngành trong mẫu được phân bố gần như đồng đều. Về năm học, bốn nhóm sinh viên từ năm nhất đến năm tư có tỷ lệ tương đối cân bằng, dao động từ 24,4% đến 25,4%. Đặc biệt, phần lớn sinh viên cho biết sử dụng AI ở mức thường xuyên hoặc rất thường xuyên (80,1%). Dữ liệu này cho phép phân tích các nhóm một cách khách quan và đáng tin cậy, đồng thời hỗ trợ kiểm định các giả thuyết so sánh sự khác biệt trong nhận thức đạo đức AI theo đặc điểm cá nhân và hành vi sử dụng công nghệ.

2.3.2. Kiểm định độ tin cậy thang đo

Để đảm bảo độ tin cậy và giá trị sử dụng của thang đo nhận thức đạo đức về AI, nghiên cứu đã tiến hành kiểm định hệ số Cronbach's alpha và phân tích nhân tố khám phá. Kết quả tổng hợp được trình bày trong Bảng 2 cho thấy tất cả các thang đo đều đạt độ tin cậy cao với hệ số Cronbach's alpha lớn hơn 0.80. Các hệ số tương quan biến – tổng đều lớn hơn 0.60. Điều này cho thấy các biến quan sát có tương quan chặt chẽ với tổng thang đo. Chỉ số KMO của từng nhóm đều lớn hơn 0.6, kiểm định Bartlett's Test đều có ý nghĩa thống kê ($p < 0.001$), giá trị tải nhân tố nhỏ nhất đều vượt ngưỡng 0.8 trong các nhóm thang đo, tổng phương sai trích (TVE) dao động từ 73.5% đến 84.1%. Điều này chỉ ra rằng cấu trúc các thang đo là phù hợp, các biến quan sát hội tụ tốt vào từng nhóm yếu tố và thang đo có giá trị cấu trúc vững chắc.

Bảng 2. Kết quả kiểm định thang đo

Khía cạnh	Cronbach's alpha	Hệ số tương quan biến - tổng nhỏ nhất	KMO	p-value	Hệ số tải nhân tố nhỏ nhất	TVE (%)
Chủ quyền và giám sát của con người	.842	.645	.676	<0.001	.834	76.411
Độ tin cậy kỹ thuật và an toàn	.905	.792	.753	<0.001	.907	84.082
Quyền riêng tư và quản trị dữ liệu	.860	.656	.704	<0.001	.833	78.583

Khía cạnh	Cronbach's alpha	Hệ số tương quan biến - tổng nhỏ nhất	KMO	p-value	Hệ số tải nhân tố nhỏ nhất	TVE (%)
Minh bạch và khả năng giải thích	.860	.720	.731	<0.001	.876	78.261
Công bằng và không thiên vị	.818	.653	.717	<0.001	.844	73.533
Phúc lợi xã hội và môi trường	.878	.725	.733	<0.001	.874	80.374
Trách nhiệm và giải trình	.854	.649	.707	<0.001	.831	77.531
Tổng	.970	.674	.951	<0.001	.658	70.228

Note: KMO: Kaiser- Meyer- Olkin index; TVE: total variance explained

2.3.3. Mức độ nhận thức đạo đức về AI

Kết quả phân tích tại Bảng 3 cho thấy mức độ nhận thức đạo đức về AI của sinh viên đạt mức tương đối cao với ĐTB = 4,16 và DLC = 0.841. Điều này phản ánh rằng sinh viên không chỉ nhận biết sự tồn tại của các vấn đề đạo đức trong quá trình sử dụng AI mà còn đồng thuận ở mức cao với các yêu cầu đạo đức cốt lõi. Trong số các yếu tố thành phần, “Chủ quyền và giám sát của con người”, “Độ tin cậy kỹ thuật và an toàn” và “Quyền riêng tư và quản trị dữ liệu” có điểm trung bình cao nhất (ĐTB = 4.20). Đây đều là các vấn đề nổi bật trong bối cảnh AI tạo sinh được sử dụng phổ biến và người dùng thường xuyên đối diện với nguy cơ lỗi thuật toán, sai lệch dữ liệu hoặc rủi ro rò rỉ thông tin cá nhân. Điều này cho thấy sinh viên đã có sự nhạy cảm nhất định đối với các rủi ro về an toàn kỹ thuật và bảo mật thông tin – các kỹ năng quan trọng trong năng lực số và năng lực sử dụng AI có trách nhiệm.

Các yếu tố như “Phúc lợi xã hội và môi trường”, “Trách nhiệm và giải trình” và “Minh bạch và khả năng giải thích” cũng được đánh giá cao với ĐTB đều lớn hơn 4.1. Kết quả này cho thấy sinh viên không chỉ nhận thức ở các khía cạnh kỹ thuật cá nhân, mà còn nhận thức được những yêu cầu ở tầm hệ thống như minh bạch thuật toán hoặc trách nhiệm của tổ chức khi triển khai AI. Mặc dù các yếu tố “Công bằng và không thiên vị” và “Chủ quyền và giám sát của con người” có điểm trung bình thấp hơn so với các nhóm

còn lại, nhưng vẫn duy trì mức trên 4.0. Điều này cho thấy sinh viên có nhận thức khá vững về vai trò trung tâm của con người trong giám sát AI cũng như khả năng nhận diện vấn đề thiên lệch thuật toán. Đây vốn là một chủ đề phức tạp thường đòi hỏi sinh viên có kiến thức chuyên sâu hơn. Tóm lại, kết quả này chỉ ra rằng sinh viên đại học tại Việt Nam đã có một mức nhận thức tích cực và có hiểu biết cơ bản về đạo đức AI. Đây là tín hiệu đáng khích lệ trong bối cảnh các công cụ AI đang ngày càng được ứng dụng sâu vào hoạt động học tập, nghiên cứu và sáng tạo của người học.

Bảng 3. Thực trạng mức độ nhận thức đạo đức về AI

Khía cạnh	ĐTB	ĐLC
Chủ quyền và giám sát của con người	4.20	.935
Độ tin cậy kỹ thuật và an toàn	4.20	.975
Quyền riêng tư và quản trị dữ liệu	4.20	.934
Minh bạch và khả năng giải thích	4.19	.866
Công bằng và không thiên vị	4.03	.952
Phúc lợi xã hội và môi trường	4.12	.961
Trách nhiệm và giải trình	4.12	.950
Tổng	4.16	.841

2.3.4. Mức độ nhận thức đạo đức về AI theo các nhóm

Kết quả phân tích tại Bảng 4 cho thấy không có sự khác biệt có ý nghĩa thống kê giữa nam và nữ trong tất cả các khía cạnh đạo đức đã khảo sát ($p > 0.05$). Mặc dù có khác biệt nhỏ về điểm trung bình giữa hai nhóm – trong đó sinh viên nữ có xu hướng đạt điểm cao hơn ở hầu hết các khía cạnh – nhưng giá trị p từ kiểm định t đều vượt ngưỡng ý nghĩa thống kê. Phân tích so sánh giữa hai nhóm sinh viên theo khối ngành đào tạo – khoa học tự nhiên và kỹ thuật và khoa học xã hội và nhân văn cũng cho thấy sự khác biệt không có ý nghĩa thống kê giữa hai nhóm ở tất cả các khía cạnh đạo đức được khảo sát ($p > 0.05$). Điều này cho thấy rằng sự nhận thức đạo đức về AI của sinh viên không bị chi phối đáng kể bởi yếu tố giới tính hay khối ngành đào tạo. Nó phản ánh mức độ tiếp cận công nghệ và thông tin tương đồng giữa các nhóm trong bối cảnh AI đang được phổ biến rộng rãi trong giáo dục đại học.

Bảng 4. Mức độ nhận thức đạo đức về AI theo giới tính và khối ngành

Biến	Nhóm	ĐTB	ĐLC	Giá trị p (Levene's test)	Giá trị p (t – test)
Giới tính	Nam	4.12	.972	.003	.433
	Nữ	4.20	.691		
Khối ngành	Khoa học tự nhiên và kỹ thuật	4.11	1.012	.001	.283
	Khoa học xã hội và nhân văn	4.21	.630		

Kết quả tại Bảng 5 cho thấy sinh viên năm nhất có ĐTB nhận thức đạo đức AI cao nhất (ĐTB = 4.38), tiếp đến là sinh viên năm tư và năm ba, trong khi sinh viên năm hai có điểm trung bình thấp nhất (ĐTB = 3.95). Kiểm định ANOVA và phân tích post hoc cho thấy chỉ có điểm trung bình của sinh viên năm nhất cao hơn có ý nghĩa thống kê so với sinh viên năm hai ở mức ý nghĩa 5%. Điều này chỉ ra rằng sinh viên năm nhất, là nhóm mới tiếp cận AI trong bối cảnh học tập đại học, có xu hướng thể hiện mức độ đồng thuận cao hơn với các phát biểu về đạo đức AI so với nhóm sinh viên năm hai.

Xét theo tần suất sử dụng AI, nhóm sinh viên sử dụng AI ở mức Thường xuyên có điểm trung bình nhận thức đạo đức cao nhất (ĐTB = 4.25), tiếp đến là nhóm Rất thường xuyên (ĐTB = 4.08) và nhóm Thỉnh thoảng (ĐTB = 4.05). Mặc dù có sự chênh lệch nhẹ về điểm trung bình giữa các nhóm nhưng kết quả kiểm định cho thấy sự khác biệt không đạt mức ý nghĩa thống kê. Điều này cho thấy việc tần suất sử dụng AI chưa đủ để tạo ra sự khác biệt rõ rệt về nhận thức đạo đức. Nó có thể còn phụ thuộc vào chất lượng trải nghiệm và bối cảnh giáo dục mà sinh viên tiếp cận các nội dung liên quan đến đạo đức AI.

Bảng 5. Mức độ nhận thức đạo đức về AI theo năm học và tần suất sử dụng AI

Biến	Nhóm	ĐTB	ĐLC
Năm học	Năm nhất	4.38 ^a	.567
	Năm hai	3.95	1.003
	Năm ba	4.12	.734
	Năm tư	4.17	.940
Tần suất sử dụng AI	Rất thường xuyên	4.08	1.066
	Thường xuyên	4.25	.655
	Thỉnh thoảng	4.05	.865

Ghi chú: a là có sự khác biệt có ý nghĩa thống kê với nhóm sinh viên năm 2 ở mức ý nghĩa 5%.

2.4. Thảo luận

Kết quả nghiên cứu cho thấy sinh viên đại học tại Việt Nam có mức độ nhận thức tương đối cao về đạo đức AI theo các nguyên tắc của UNESCO (2021). Phát hiện này cho thấy sinh viên đã hình thành một nhận thức nền tảng mang tính chuẩn mực về việc sử dụng AI trong học tập và nghiên cứu, đặc biệt ở các khía cạnh liên quan trực tiếp đến trải nghiệm cá nhân như độ tin cậy kỹ thuật, quyền riêng tư và minh bạch. Kết quả này tương đồng với nghiên cứu của Valerio (2024) [13] khi ghi nhận sinh viên tại Philippines có mức độ nhận thức cao về quyền riêng tư và các hệ quả xã hội của AI. Tuy nhiên, khác với sinh viên Nhật Bản trong nghiên cứu của Ghotbi & Ho (2021) [12], sinh viên Việt Nam thể hiện mức độ quan tâm thấp hơn đối với các rủi ro xã hội vĩ mô như thất nghiệp hay kiểm soát xã hội, mà tập trung nhiều hơn vào các nguyên tắc mang tính kỹ thuật và sử dụng trực tiếp. Sự khác biệt này có thể là do AI tại Việt Nam hiện nay chủ yếu được ứng dụng trong hoạt động học tập, thay vì gắn chặt với thị trường lao động hay quản trị xã hội.

Một kết quả đáng chú ý của nghiên cứu là không ghi nhận sự khác biệt có ý nghĩa thống kê về nhận thức đạo đức AI theo giới tính và khối ngành đào tạo. Điều này cho thấy nhận thức đạo đức AI của sinh viên mang tính phổ quát hơn là chuyên biệt theo lĩnh vực, phản ánh vai trò ngày càng rõ nét của AI như một công cụ học tập chung trong giáo dục đại học. Phát hiện này khác với một số nghiên cứu trước đây ghi nhận sự khác biệt theo giới tính trong nhận thức hoặc thái độ đối với AI (Saatci, 2025; Wang et al., 2025) [15,16]. Đồng thời, nó cũng chỉ ra rằng các chương trình giáo dục đạo đức AI tại Việt Nam có thể được thiết kế theo hướng liên ngành, thay vì phân tách cứng theo khối ngành.

Sự khác biệt theo năm học cho thấy sinh viên năm nhất có mức nhận thức đạo đức AI cao hơn sinh viên năm hai. Điều này cho thấy nhận thức đạo đức AI của sinh viên có thể hình thành ngay từ giai đoạn đầu tiếp cận công nghệ, nhất là với sinh viên năm nhất – những người bắt đầu đại học trong bối cảnh AI tạo sinh phát triển nhanh và các vấn đề đạo đức được nhấn mạnh trong xã hội. Ở giai đoạn này, sinh viên cũng có thể thể hiện xu hướng lý tưởng hóa các chuẩn mực đạo đức, dẫn đến mức độ đồng thuận cao hơn với các phát biểu mang tính nguyên tắc. Ngược lại, sinh viên năm hai – nhóm đã sử dụng AI thường xuyên hơn trong học tập – có thể xuất hiện hiện tượng “bình thường hóa” các rủi ro đạo đức, như Mittelstadt (2019) [9] từng cảnh báo. Điều này nhấn mạnh việc giáo dục đạo đức AI ở giai đoạn đầu đại học nhằm hình thành thói quen sử dụng AI có ý thức và bền vững có vai trò quan trọng. Tuy nhiên, cần lưu ý rằng sự khác biệt này cũng có thể chịu ảnh hưởng của các yếu tố gây nhiễu như sự phân bố sinh viên theo cơ sở đào tạo hoặc ngành học cụ thể trong mẫu nghiên cứu, do đó kết quả cần được diễn giải một cách thận trọng.

Cuối cùng, việc không ghi nhận sự khác biệt đáng kể theo tần suất sử dụng AI cho thấy việc sử dụng AI nhiều không đồng nghĩa với nhận thức đạo đức cao hơn. Nhận thức đạo đức AI dường như phụ thuộc nhiều hơn vào chất lượng định hướng giáo dục và các

can thiệp có chủ đích, hơn là vào mức độ tiếp xúc công nghệ thuần túy. Phát hiện này cũng củng cố lập luận của Usher & Barak (2024) [17], theo đó các mô-đun giáo dục đạo đức AI và hoạt động phân tích tình huống đóng vai trò then chốt trong việc nâng cao năng lực nhận diện và xử lý các vấn đề đạo đức của sinh viên.

3. KẾT LUẬN VÀ KHUYẾN NGHỊ

3.1. Kết luận

Nghiên cứu đã đánh giá mức độ nhận thức đạo đức về trí tuệ nhân tạo của sinh viên đại học tại Việt Nam. Kết quả khảo sát 307 sinh viên cho thấy nhận thức đạo đức AI của người học đạt mức tương đối cao trên cả bảy nguyên tắc, đặc biệt ở các khía cạnh Độ tin cậy kỹ thuật, Quyền riêng tư và Minh bạch. Điều này cho thấy sự chủ động của sinh viên trong việc nhận diện các vấn đề đạo đức liên quan đến việc sử dụng AI trong học tập và đời sống số. Phân tích theo các nhóm nhân khẩu học chỉ ra rằng giới tính và khối ngành không tạo ra sự khác biệt có ý nghĩa thống kê trong nhận thức đạo đức AI. Tuy nhiên, sự khác biệt theo năm học cho thấy mức độ nhận thức đạo đức AI có thể gắn với bối cảnh tiếp cận công nghệ và trải nghiệm học tập của từng nhóm. Bên cạnh đó, tần suất sử dụng AI không phải là yếu tố quyết định nhận thức đạo đức. Điều này chỉ ra rằng chất lượng giáo dục và sự dẫn dắt trong quá trình sinh viên sử dụng AI đóng vai trò quan trọng trong việc nhận thức đạo đức AI.

3.2. Khuyến nghị

Dựa trên kết quả của nghiên cứu, nhóm nghiên cứu đề xuất một số khuyến nghị sau: (1) Các đơn vị đào tạo cần xây dựng hoặc bổ sung học phần, chuyên đề về đạo đức AI trong các chương trình đào tạo, đặc biệt ở giai đoạn năm nhất để hình thành nhận thức nền tảng sớm cho sinh viên; cung cấp hướng dẫn và ban hành chính sách về sử dụng AI trong học tập và nghiên cứu và sáng tạo học thuật; (2) Giảng viên cần hướng dẫn sinh viên đặt câu hỏi về độ tin cậy, tính minh bạch và nguồn dữ liệu của các công cụ AI nhằm giảm nguy cơ phụ thuộc công nghệ và gia tăng khả năng đánh giá thông tin; (3) Sinh viên cần chủ động cập nhật kiến thức liên quan đến quyền riêng tư, minh bạch dữ liệu và nhận diện thiên lệch thuật toán khi sử dụng AI và sử dụng AI một cách có trách nhiệm; (4) Các nhà xây dựng chính sách cần xây dựng và ban hành khung hướng dẫn mang tính thống nhất, hỗ trợ các trường triển khai đồng bộ hoạt động giáo dục đạo đức AI.

3.3. Hạn chế nghiên cứu và gợi ý hướng nghiên cứu tiếp theo

Bên cạnh những kết quả đã đạt được, nghiên cứu vẫn còn một số hạn chế. Mẫu khảo sát được chọn theo phương pháp thuận tiện dẫn tới việc chưa phản ánh đầy đủ sự đa dạng trong toàn bộ hệ thống giáo dục đại học Việt Nam. Nghiên cứu mới dừng ở nhận thức AI mà chưa xem xét hành vi đạo đức thực tế của sinh viên khi sử dụng AI. Ngoài ra, mặc dù phân tích nhân tố khám phá (EFA) đã được thực hiện trên toàn bộ thang đo 21 biến quan sát và cho thấy cấu trúc 7 khía cạnh phù hợp với khung lý thuyết UNESCO (2021), nghiên

cứu chưa tiến hành phân tích nhân tố khẳng định (CFA) để kiểm định mức độ phù hợp của mô hình đo lường. Do đó, các nghiên cứu tiếp theo cần mở rộng quy mô mẫu và thực hiện CFA/SEM, đồng thời kết hợp phương pháp định tính để làm rõ hơn mối quan hệ giữa nhận thức – thái độ – hành vi đạo đức AI của sinh viên..

TÀI LIỆU THAM KHẢO

1. Huang, C., Zhang, Z., Mao, B., & Yao, X. (2022), *An overview of artificial intelligence ethics*, IEEE Transactions on Artificial Intelligence, 4(4), pp.799-819.
2. Singh, H. P., & Aziz, A. A. (2025), *Impact of intelligent learning assistants on creativity of university students: a self-determination theory perspective*, Future Business Journal, 11(1), p.122.
3. Chanpetch, N., & Songserm, U. (2023), *Artificial intelligence (AI): Educational applications*, Journal of Teacher Professional Development, 6(1), pp.1-13.
4. Zulfa, N. A., Na'imah, N. I., Hafni, N. D., Hibana, H., Sa'adah, D. A., Sholihah, M., & Sari, N. (2025), *Student Creativity Innovation in Completing Artificial Intelligence (AI)-Based Lecture Assignments*, AI Hikmah Indonesian Journal of Early Childhood Islamic Education, 9(1), pp.134-143.
5. Siau, K., & Wang, W. (2020), *Artificial intelligence (AI) ethics: ethics of AI and ethical AI*, Journal of Database Management (JDM), 31(2), pp.74-87.
6. Kazim, E., & Koshiyama, A. S. (2021), *A high-level overview of AI ethics*, Patterns, 2(9).
7. Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., ... & Rashidi, H. H. (2025), *Ethical and bias considerations in artificial intelligence/machine learning*, Modern Pathology, 38(3), p.100686.
8. Weidener, L., & Fischer, M. (2024), *Role of ethics in developing AI-based applications in medicine: insights from expert interviews and discussion of implications*, Jmir ai, 3(1), e51204.
9. Mittelstadt, B. (2019), *Principles alone cannot guarantee ethical AI*, Nature machine intelligence, 1(11), pp.501-507.
10. Leta, F. M., & Vancea, D. P. C. (2023), *Ethics in education: Exploring the ethical implications of artificial intelligence implementation*, Ovidius university annals, economic sciences series, 23(1), pp.413-421.
11. UNESCO. (2021), *Recommendation on the ethics of artificial intelligence*, United Nations Educational, Scientific and Cultural Organization.
12. Ghotbi, N., & Ho, M. T. (2021), *Moral awareness of college students regarding artificial intelligence*, Asian Bioethics Review, 13(4), pp.421-433.

13. Valerio, A. (2024), *Anticipating the impact of artificial intelligence in higher education: Student awareness and ethical concerns in Zamboanga City, Philippines*, Cognizance Journal of Multidisciplinary Studies, 4(6), pp.10-47760.
14. Ferhataj, A., Memaj, F., Sahatcija, R., Ora, A., & Koka, E. (2025), *Ethical concerns in AI development: analyzing students' perspectives on robotics and society*, Journal of Information, Communication and Ethics in Society, 23(2), pp.165-187.
15. Saatci, E. Y. (2025), *AI and Ethics: Scale Development for Measuring Ethical Perceptions of Artificial Intelligence Across Sectors and Countries*, International Journal of Economic Behavior and Organization, 13(1), pp.35-50.
16. Wang, Z., Chai, C. S., Li, J., & Lee, V. W. Y. (2025), *Assessment of AI ethical reflection: the development and validation of the AI ethical reflection scale (AIERS) for university students*, International Journal of Educational Technology in Higher Education, 22(1), p.19.
17. Usher, M., & Barak, M. (2024), *Unpacking the role of AI ethics online education for science and engineering students*, International Journal of STEM Education, 11(1), p.35.

MEASURING UNIVERSITY STUDENTS' AI ETHICS AWARENESS BASED ON UNESCO'S SEVEN PRINCIPLES (2021): CURRENT STATUS AND IMPLICATIONS

*Bui Thanh Thuy, Nguyen Thu Ha,
Pham Thuy Ha, Hoang Thuy An, Nguyen Thi Hai*

Abstract: *This study applies UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021) to develop and employ a principle-based measurement scale for assessing the level of AI ethics awareness among Vietnamese university students. A survey was conducted with 307 students from higher education institutions. The results indicate that students' AI ethics awareness is relatively high, particularly in the dimensions of Technical Robustness and Safety, Privacy, and Transparency. Comparative analyses show that gender, academic discipline, and frequency of AI use do not produce statistically significant differences in AI ethics awareness. By contrast, differences by year of study are observed, with first-year students demonstrating higher levels of awareness than second-year students. Based on these findings, the study recommends strengthening AI ethics education within university curricula, especially at the early stages of study, in order to foster responsible AI use habits among students.*

Keywords: *Artificial intelligence ethics, awareness, students, higher education.*

*(Ngày Tòa soạn nhận được bài: 28-11-2025; ngày phản biện đánh giá: 30-12-2025;
ngày chấp nhận đăng: 22-01-2026)*