THE RELATIONSHIP BETWEEN ETHICAL AWARENESS-ETHICAL RISK OF AI USE AND STUDENTS' INTENTION TO USE AI IN LEARNING: A QUANTITATIVE STUDY OF VIETNAMESE UNIVERSITY STUDENTS

Nguyen Xuan An*[1], Hoang Vu Linh Chi[2], Ngo Thanh Thuy[3]

Corresponding author
Email: nxan@daihochoabinh edu.vn

[1] Hoa Binh University
No. 08 Bui Xuan Phai street, Tu Liem ward,
Hanoi, Vietnam
[2] Email: linhchivh@vnu.edu.vn
VNU University of Economics and Business
144 Xuan Thuy street, Cau Giay ward,
Hanoi, Vietnam
[3] Email: thuyngothanh@vnies.edu.vn
The Vietnam National Institute of Educational Sciences
101 Tran Hung Dao street, Cua Nam ward,
Hanoi, Vietnam

Received: 01/4/2025
Revised: 06/5/2026
Accepted: 10/6/2026
Published: 15/7/2026

Abstract: This study investigates the relationships among ethical awareness of artificial intelligence (AI), perceived ethical risks (PER) of AI use, and behavioral intention to use AI in learning among Vietnamese university students, a context characterized by rapid digital transformation and emerging AI governance. Data was collected from 310 university students in Vietnam through an online questionnaire using convenience sampling. The findings show that students demonstrated relatively high levels of ethical awareness and behavioral intention, while their perceived ethical risks were moderate. Regression analysis produced a statistically significant model explaining 28.7% of the variance in students' behavioral intention to use AI. Ethical awareness was a strong positive predictor of behavioral intention ($\beta = 0.563$, $p < 0.001$), whereas perceived ethical risk did not have a statistically significant effect. Contrary to common assumptions, higher ethical awareness does not hinder but rather enhances students' intention to use AI in learning.

Keywords: *AI ethical awareness, AI perceived ethical risks, AI behavioral intention, Vietnamese students.*

MỐI QUAN HỆ GIỮA NHẬN THỨC ĐẠO ĐỨC VÀ RỦI RO ĐẠO ĐỨC TRONG VIỆC SỬ DỤNG AI VÀ Ý ĐỊNH SỬ DỤNG AI TRONG HỌC TẬP CỦA SINH VIÊN VIỆT NAM: MỘT NGHIÊN CỨU ĐỊNH LƯỢNG

Nguyễn Xuân An*[1], Hoàng Vũ Linh Chi[2], Ngô Thanh Thủy[3]

* Tác giả liên hệ
Email: nxan@daihochoabinh.edu.vn

[1] Trường Đại học Hòa Bình
Số 08 Bùi Xuân Phái, phường Từ Liêm,
Hà Nội, Việt Nam
[2] Email: linhchivh@vnu.edu.vn
Trường Đại học Kinh tế - Đại học Quốc gia Hà Nội
144 Xuân Thủy, phường Cầu Giấy,
Hà Nội, Việt Nam
[3] Email: thuyngothanh@vnies.edu.vn
Viện Khoa học Giáo dục Việt Nam
101 Trần Hưng Đạo, Phường Cửa Nam,
Hà Nội, Việt Nam

Nhận bài: 01/4/2026
Chỉnh sửa xong: 06/5/2026
Chấp nhận đăng: 10/6/2026
Xuất bản: 15/7/2026

Tóm tắt: Nghiên cứu này nhằm mục đích làm rõ mối quan hệ giữa nhận thức đạo đức về AI, nhận thức về rủi ro đạo đức của việc sử dụng AI và ý định sử dụng AI trong học tập của sinh viên Việt Nam, một bối cảnh đặc trưng bởi tốc độ chuyển đổi số cao và các quy định về AI đang trong giai đoạn hình thành. Nghiên cứu đã thu thập được 310 phản hồi hợp lệ từ sinh viên đang học tại các trường đại học tại Việt Nam bằng bảng hỏi trực tuyến với phương pháp chọn mẫu thuận tiện. Kết quả cho thấy, sinh viên có mức độ nhận thức đạo đức và ý định sử dụng AI tương đối cao, trong khi nhận thức về rủi ro đạo đức ở mức vừa phải. Phân tích hồi quy cho thấy, mô hình giải thích được 28,7% phương sai của ý định sử dụng AI. Nhận thức đạo đức được xác định là một yếu tố dự báo quan trọng, có tác động tích cực và thuận chiều đến ý định sử dụng AI ($\beta = 0,563$, $p < 0,001$). Ngược lại, nhận thức về rủi ro đạo đức không có ảnh hưởng đáng kể về mặt thống kê đến ý định sử dụng ($p = 0,308$). Nghiên cứu cho thấy, trái với giả định phổ biến, nhận thức đạo đức cao không cản trở mà thúc đẩy ý định sử dụng AI của sinh viên.

Từ khóa: *Nhận thức đạo đức về AI, nhận thức về rủi ro đạo đức của AI, ý định sử dụng AI, sinh viên Việt Nam.*

1. Đặt vấn đề

Thế kỉ XXI đang chứng kiến một cuộc Cách mạng công nghệ mang tính bước ngoặt với sự trỗi dậy của Trí tuệ nhân tạo (AI), đặc biệt là AI tạo sinh (Generative AI). Các mô hình ngôn ngữ lớn như GPT-4 của OpenAI, Gemini của Google, hay Llama của Meta đã vượt qua giới hạn của các công cụ tìm kiếm truyền thống, mang lại khả năng tương tác, sáng tạo và giải quyết vấn đề một cách linh hoạt chưa từng có. Sự thâm nhập nhanh chóng của AI tạo sinh vào mọi lĩnh vực của đời sống xã hội đã và đang định hình lại cách con người làm việc, giao tiếp và học tập. Trong số các lĩnh vực chịu ảnh hưởng sâu sắc nhất, giáo dục đại học nổi lên như một không gian vừa đón nhận những cơ hội to lớn vừa đối mặt với những thách thức phức tạp.

Đối với sinh viên, AI tạo sinh đã trở thành một công cụ hỗ trợ học tập đầy tiềm năng. Nó có thể đóng vai trò như một gia sư cá nhân hóa, một người bạn đồng hành trong quá trình động não (Brainstorming), một trợ lí hiệu quả trong việc tóm tắt các tài liệu học thuật phức tạp, gỡ lỗi mã lập trình, hay thậm chí là dịch thuật và cải thiện kĩ năng viết lách (Zhai, 2022). Tuy nhiên, song hành với những lợi ích không thể phủ nhận là một loạt các vấn đề đạo đức và rủi ro tiềm ẩn. Việc lạm dụng AI có thể dẫn đến sự xói mòn các kĩ năng tư duy phản biện và sáng tạo, vốn là nền tảng của giáo dục đại học (Lim & cộng sự, 2023). Vấn đề liêm chính học thuật (Academic integrity) trở nên cấp bách hơn bao giờ hết, với những lo ngại về hành vi đạo văn, gian lận trong thi cử và việc nộp các sản phẩm do AI thực hiện hoàn toàn mà không có sự đóng góp trí tuệ của người học (Cotton & cộng sự, 2024). Thêm vào đó, các rủi ro liên quan đến quyền riêng tư dữ liệu, sự thiên vị của thuật toán (algorithmic bias) có thể vô tình củng cố hoặc tạo ra những định kiến xã hội và sự thiếu minh bạch trong cách các mô hình AI hoạt động cũng là những mối bận tâm lớn (Zhuo & cộng sự, 2023). Sự phổ biến nhanh chóng này cũng đồng nghĩa với việc các trường đại học và bản thân người học phải đối mặt trực diện với các thách thức đạo đức nêu trên, trong khi hành lang pháp lí và các quy định học thuật cụ thể về việc sử dụng AI vẫn còn đang trong giai đoạn hình thành (Rasul & cộng sự, 2023).

Trong bối cảnh AI tạo sinh, hai cấu trúc nhận thức đặc biệt quan trọng nổi lên là nhận thức đạo đức (Ethical Awareness - EA) và nhận thức về rủi ro đạo đức (Perceived Ethical Risk - PER). Nhận thức đạo đức đề cập đến sự hiểu biết của một cá nhân về các nguyên tắc, chuẩn mực và các vấn đề đạo đức liên quan đến việc phát triển và sử dụng AI như sự công bằng, minh bạch và trách nhiệm giải trình. Ngược lại, nhận thức về rủi ro đạo đức tập trung vào những đánh giá của cá nhân về khả năng xảy ra các hậu quả tiêu cực về mặt đạo đức khi sử dụng công nghệ (Zhu và cộng sự, 2025).

Nhận thức đạo đức về AI của sinh viên là sự công nhận và thấu hiểu của một cá nhân về các vấn đề và nguyên tắc đạo đức tồn tại trong quá trình phát triển và sử dụng các sản phẩm AI (Zhu và cộng sự, 2025). Theo mô hình ra quyết định đạo đức bốn thành phần của Rest (1986), nhận thức đạo đức tương ứng với giai đoạn đầu tiên, đó là "Nhận diện đạo đức" (Moral recognition), khi người ra quyết định phải nhận ra một vấn đề đạo đức. Trong bối cảnh AI, yếu tố này không chỉ dừng lại ở việc biết đến sự tồn tại của công nghệ, mà còn bao hàm sự hiểu biết sâu sắc về tầm quan trọng của các nguyên tắc đạo đức cốt lõi như sự công bằng, minh bạch, trách nhiệm giải trình, không gây hại và quyền riêng tư (Franzke, 2022). Trong bối cảnh phát triển nhanh chóng của AI tạo sinh đã mang đến nhiều lo ngại về đạo đức như vấn đề bản quyền, sự thiên vị trong thuật toán và trách nhiệm khi có sự cố xảy ra. Khi sinh viên có nhận thức đạo đức cao, họ có xu hướng hành xử phù hợp hơn với các nguyên tắc đạo đức khi đối mặt với các tình huống liên quan đến việc sử dụng AI. Do đó, việc trang bị nhận thức đạo đức về AI ngày càng trở nên cấp thiết, đặc biệt là ở bậc giáo dục đại học, nơi đào tạo ra lực lượng lao động chủ chốt của tương lai.

Nhận thức về rủi ro đạo đức của việc sử dụng AI là sự đánh giá và phán đoán của người dùng về những ảnh hưởng tiêu cực không chắc chắn về mặt đạo đức mà việc sử dụng các sản phẩm AI tạo sinh có thể mang lại (Craft, 2013). Khái niệm này liên quan chặt chẽ đến giai đoạn thứ hai trong mô hình ra quyết định đạo đức, đó là "Phán xét đạo đức" (Moral judgement), nơi cá nhân đánh giá xem một hành động có phù hợp với các tiêu chuẩn đạo đức hay không. Các rủi ro này có thể bao gồm nhiều khía cạnh như rủi ro về hiệu suất, xã hội, tài chính và tâm lí (Featherman & Pavlou, 2003). Trong môi trường học thuật, các rủi ro đạo đức nổi bật bao gồm nguy cơ xâm phạm liêm chính học thuật, đạo văn và sự không công bằng trong đánh giá giáo dục. Việc sử dụng AI một cách thiếu cân nhắc có thể làm gia tăng sự phụ thuộc, ảnh hưởng tiêu cực đến kĩ năng tư duy phản biện và giải quyết vấn đề của sinh viên. Nhận thức về những rủi ro này đóng vai trò quan trọng vì nó có thể cản trở việc chấp nhận và sử dụng công nghệ của người dùng.

 https://doi.org/10.15625/2615-8957/12620104

Ý định sử dụng (Behavioral Intention) được coi là yếu tố dự báo trực tiếp và quan trọng nhất đối với hành vi sử dụng thực tế. Trong bối cảnh giáo dục đại học, ý định sử dụng AI thể hiện mức độ sẵn lòng của sinh viên trong việc tích hợp các sản phẩm AI vào hoạt động học tập và công việc của họ (Zhu và cộng sự, 2025).

Về mối quan hệ giữa các yếu tố đạo đức và ý định sử dụng AI của sinh viên, theo lí thuyết ra quyết định đạo đức, một người trước tiên cần nhận thức được vấn đề (nhận thức đạo đức), sau đó mới đánh giá và phán xét nó (nhận thức rủi ro) (Valentine & Hollingworth, 2012). Nghiên cứu của Zhu và cộng sự, (2025) đã chỉ ra rằng, khi nhận thức đạo đức của SV càng cao, họ càng nhận thức rõ hơn về các rủi ro đạo đức tiềm ẩn khi sử dụng AI. Điều này là hợp lí, bởi khi một người được thông tin nhiều hơn về một vấn đề, họ có xu hướng quan tâm nhiều hơn đến các rủi ro liên quan. Nghiên cứu này lại cho thấy, nhận thức về rủi ro đạo đức không có tác động tiêu cực trực tiếp đến ý định sử dụng AI của sinh viên, trong khi là nhận thức đạo đức lại có ảnh hưởng tích cực đến ý định sử dụng AI.

Do vậy, việc tìm hiểu nhận thức liên quan đến vấn đề đạo đức khi sử dụng AI của sinh viên là điều hết sức cần thiết trong bối cảnh công cụ này đang được sử dụng ngày càng phổ biến với những câu hỏi và sự quan ngại về sự lạm dụng AI. Nghiên cứu này được thực hiện với mục tiêu tổng quát là làm sáng tỏ mối quan hệ giữa các yếu tố nhận thức đạo đức, nhận thức rủi ro đạo đức và ý định sử dụng AI trong học tập của sinh viên Việt Nam.

2. Phương pháp nghiên cứu

2.1. Bộ công cụ

Nghiên cứu này sử dụng một bộ công cụ bao gồm 03 thang đo với tổng 13 biến quan sát để đo lường các yếu tố: 1) Nhận thức đạo đức về AI bao gồm 06 biến quan sát; 2) Nhận thức về rủi ro đạo đức của AI bao gồm 03 biến quan sát; 3) Ý định sử dụng AI bao gồm 04 biến quan sát. Các thang đo này được đáp ứng trong các nghiên cứu trước đó của Zhu và cộng sự (2025) và Nguyen và cộng sự (2021) (cụ thể tại Bảng 1 dưới đây).

Một bảng hỏi bao gồm 03 phần đã được thiết kế để thu thập dữ liệu với đối tượng khảo sát là SV các trường đại học tại Việt Nam với cấu trúc như sau: 1) Phần thứ nhất là phần đầu tiên của bảng hỏi với nội dung giới thiệu về nghiên cứu và nhóm nghiên cứu, cùng với đó là các tuyên bố về sự đồng thuận tham gia khảo sát của người tham gia khảo sát, sự bảo mật của khảo sát; 2) Phần thứ hai được thiết kế để thu thập các thông tin sau của sinh viên tham gia khảo sát, đó là: Giới tính, Dân tộc, Năm học, Chuyên ngành đang theo học; 3) Phần thứ ba là phần chính của bảng hỏi, đó là nội dung khảo sát bao gồm 03 câu hỏi tương ứng với 03 yếu tố được nghiên cứu. Thang đo Likert 5 mức độ được áp dụng để đo lường cho từng biến quan sát, trong đó: 1-Hoàn toàn không đồng ý; 2-Không đồng ý phần lớn; 3-Bình thường; 4-Đồng ý phần lớn; 5-Hoàn toàn đồng ý.

2.2. Quy trình thu thập dữ liệu và mẫu nghiên cứu

Nghiên cứu áp dụng phương pháp chọn mẫu thuận tiện để tuyển chọn sinh viên tham gia khảo sát, phù hợp với yêu cầu triển khai linh hoạt trong điều kiện nguồn lực hạn chế (Cohen và cộng sự, 2012). Dữ liệu được thu thập trực tuyến qua Google Forms; toàn bộ câu hỏi được đặt ở chế độ bắt buộc nhằm hạn chế thiếu hụt thông tin. Thời gian thu thập từ ngày 21 tháng 05 năm 2025 đến ngày 05 tháng 06 năm 2025, ghi nhận 385 phản hồi ban đầu. Sau khi sàng lọc và làm sạch theo các khuyến nghị về kiểm soát chất lượng dữ liệu (Hair Jr và cộng sự, 2014), bộ dữ liệu cuối cùng gồm 310 quan sát hợp lệ phục vụ phân tích với sự tham gia của các sinh viên của Trường Đại học Hòa Bình, Trường Đại học Khoa học xã hội và Nhân văn - Đại học Quốc gia Hà Nội,

Bảng 1: *Thống kê số lượng biến quan sát và các yếu tố của nghiên cứu*

STT	Mã	Yếu tố	Số lượng biến quan sát	Nguồn
1	EA	Nhận thức đạo đức về AI (Ethical Awareness about AI)	06	Zhu và cộng sự, 2025
2	PER	Nhận thức về rủi ro đạo đức của AI (Perceived Ethical Risks about AI)	03	Zhu và cộng sự, 2025
3	BI	Ý định sử dụng AI (Behavior Intention)	04	Nguyen và cộng sự, 2021

(Nguồn: Nhóm nghiên cứu tổng hợp)

Trường Đại học Kinh tế - Đại học Quốc gia Hà Nội, Trường Đại học Bách Khoa. Đặc điểm mẫu được tổng hợp ở Bảng 2.

2.3. Phương pháp phân tích

Nghiên cứu sử dụng hai phương pháp, đó là: 1) Phương pháp thống kê mô tả: giá trị trung bình của các biến tổng hợp được tính bằng trung bình cộng của các biến quan sát; 2) Phương pháp hồi quy tuyến tính bội với phương thức Enter để tìm hiểu mối quan hệ giữa nhận thức đạo đức và rủi ro đạo đức của việc sử dụng AI đối với ý định sử dụng AI trong học tập của sinh viên Việt Nam.

3. Kết quả nghiên cứu

3.1. Kiểm định độ tin cậy của các thang đo

Nghiên cứu này sử dụng hệ số Cronbach's Alpha để kiểm định độ tin cậy của các thang đo trong bộ công cụ nghiên cứu. Kết quả phân tích các độ tin cậy của các thang đo được thể hiện ở Bảng 3.

Bảng 2: *Đặc điểm của mẫu nghiên cứu*

Đặc điểm	Số lượng (người)	Tỉ lệ (%)
Giới tính	310	100.0
Nam	106	34.2
Nữ	204	65.8
Dân tộc	310	100.0
Kinh	288	92.9
Khác	22	7.1
Năm học	310	100.0
Năm thứ nhất	32	10.3
Năm thứ hai	222	71.6
Năm thứ ba	42	13.5
Năm thứ tư	5	1.6
Năm thứ năm	9	2.9
Chuyên ngành đang theo học	310	100.0
Kinh tế	243	78.4
Kĩ thuật	18	5.8
Xã hội và nhân văn	49	15.8

(Nguồn: Nhóm nghiên cứu tổng hợp từ dữ liệu khảo sát)

Bảng 3: *Kết quả phân tích độ tin cậy của các thang đo dựa trên hệ số Cronbach's Alpha*

Mã biến quan sát	Nội dung	Giá trị Cronbach Alpha	Cronbach's Alpha Based on Standardized Items
EA	Nhận thức đạo đức về AI	0.828	0.832
PER	Nhận thức về rủi ro đạo đức của AI	0.825	0.824
BI	Ý định sử dụng AI	0.833	0.832

(Nguồn: Phân tích từ dữ liệu khảo sát)

Kết quả cho thấy, các thang đo các yếu tố có giá trị Cronbach's Alpha như sau: Nhận thức đạo đức về AI (EA) đạt α = 0,828, Nhận thức về rủi ro đạo đức của việc sử dụng AI đạt α = 0,825; và Ý định sử dụng AI đạt α = 0,833. Tất cả đều vượt ngưỡng chấp nhận 0,60 và nằm trong mức "tốt" (≥ 0,80), khẳng định tính nhất quán nội bộ giữa các biến quan sát trong từng cấu trúc (Nunnally, 1994). Cùng với đó, kết quả phân tích các tương quan biến - tổng của cả ba thang đo đều lớn hơn 0,30 và hệ số Cronbach's Alpha nếu loại biến đều nhỏ hơn Cronbach Alpha của thang đo.

3.2. Nhận thức đạo đức và rủi ro đạo đức, ý định về sử dụng AI của sinh viên Việt Nam

a. Nhận thức đạo đức về AI

Về mức độ nhận thức đạo đức về AI của sinh viên Việt Nam trong nghiên cứu này, Bảng 4 dưới đây cho biết mức nhận thức đạo đức về AI trong học tập của sinh viên Việt Nam ở cả cấp độ yếu tố (EA) và cấp độ từng biến quan sát. Nhìn chung, điểm trung bình của thang EA đạt 3,91 (SD=0,666), cho thấy xu hướng đồng thuận tương đối cao của người học đối với các khía cạnh đạo đức khi phát triển và sử dụng AI trong bối cảnh giáo dục. Độ lệch chuẩn ở mức trung bình thấp gợi ý sự hội tụ ý kiến khá tốt trong mẫu khảo sát.

Xét tại cấp độ biến quan sát, các phát biểu này hướng đến những chuẩn mực cốt lõi của đạo đức AI đạt mức đồng thuận cao và ổn định. Cụ thể, EA06 (quan tâm đến đạo đức vừa mang lại lợi ích, vừa giúp tránh rủi ro) có trung bình 4,08 (SD=0,874), EA01 (bảo đảm quyền riêng tư và an toàn thông tin) đạt 4,07 (SD=0,911) và EA05 (ưu tiên yếu tố đạo đức thay

 https://doi.org/10.15625/2615-8957/12620104

Bảng 4: *Thực trạng về nhận thức đạo đức về AI của sinh viên Việt Nam trong học tập*

Mã biến quan sát	Nội dung	Giá trị trung bình	Độ lệch chuẩn (SD)
EA	Nhận thức đạo đức về AI	3.91	0.666
EA01	Tôi nhận thấy việc phát triển và sử dụng các sản phẩm AI tạo sinh cần đảm bảo quyền riêng tư và an toàn thông tin.	4.07	0.911
EA02	Tôi nhận thấy AI tạo sinh mạnh cũng có thể dẫn đến sự phân biệt đối xử và những điều không công bằng.	3.39	0.985
EA03	Tôi cho rằng, cần thiết phải minh bạch các nguyên tắc vận hành bên trong của các sản phẩm AI tạo sinh để công chúng được biết.	3.87	0.889
EA04	Tôi hiểu rằng tất cả các bên liên quan – bao gồm doanh nghiệp, nhà phát triển và người dùng – đều phải chịu trách nhiệm nếu có sự cố xảy ra khi sử dụng sản phẩm AI tạo sinh.	3.97	0.877
EA05	Tôi nhận ra rằng, các nhà phát triển AI tạo sinh cần đặt yếu tố đạo đức lên hàng đầu, thay vì chỉ chạy theo lợi nhuận kinh tế hay tiến bộ công nghệ.	4.06	0.908
EA06	Tôi nhận thấy rằng, việc quan tâm đến đạo đức trong AI không chỉ mang lại nhiều lợi ích mà còn giúp tránh được các rủi ro tiềm ẩn khi sử dụng các sản phẩm AI tạo sinh.	4.08	0.874

(Nguồn: Phân tích từ dữ liệu khảo sát)

vì chỉ lợi nhuận/tiến bộ kĩ thuật) đạt 4,06 (SD=0,908). Các giá trị này cho thấy nhận thức đạo đức của SV không chỉ dừng ở "nguyên tắc" trừu tượng mà đã gắn với các chuẩn mực vận hành cụ thể (bảo mật, an toàn, ưu tiên đạo đức) và với cách nhìn cho rằng, công cụ đạo đức như đòn bẩy tăng lợi ích và giảm thiểu rủi ro. Biến quan sát EA04 (trách nhiệm đa bên liên quan) và EA03 (minh bạch nguyên tắc vận hành) đạt lần lượt 3,97 (SD=0,877) và 3,87 (SD=0,889). Dù thấp hơn nhóm mục nêu trên, mức trung bình này vẫn tiệm cận ngưỡng "cao", phản ánh sự đồng thuận đáng kể về yêu cầu minh bạch thuật toán và cơ chế phân bổ trách nhiệm giữa doanh nghiệp, nhà phát triển và người dùng trong hệ sinh thái AI. Mức độ phân tán nhỏ đến vừa (SD khoảng 0,88) gợi ý sự nhất quán tương đối trong quan niệm của sinh viên về hai trụ cột quản trị đạo đức: Minh bạch và trách nhiệm giải trình. Đáng chú ý, EA02 (nhận diện nguy cơ phân biệt đối xử và bất công từ AI tạo sinh mạnh) có trung bình 3,39 (SD=0,985), thấp nhất trong các mục.

Như vậy, các phân tích trên tựu trung cho thấy rằng nhận thức đạo đức của sinh viên Việt Nam có sự cân bằng nhưng tập trung chủ yếu vào việc họ đánh giá cao các nguyên tắc nền tảng và yêu cầu quản trị (bảo mật, ưu tiên đạo đức, minh bạch, trách nhiệm), song vẫn còn "khoảng trống" trong việc nhận diện rủi ro thiên lệch và bất công. Đây là các khía cạnh vốn đòi hỏi kiến thức về dữ liệu, mô hình và ngữ cảnh xã hội.

b. Nhận thức về rủi ro đạo đức của việc sử dụng AI

Kết quả phân tích về nhận thức của sinh viên Việt Nam về rủi ro đạo đức của việc sử dụng AI được trình bày tại Bảng 5 sau đây. Về tổng thể, PER đạt trung bình 3,68 (SD = 0,839), thể hiện mức độ thận trọng vừa đến khá của sinh viên trước các hệ quả đạo đức khi ứng dụng AI. Độ lệch chuẩn ở mức trung bình cho thấy ý kiến trong mẫu có tính phân tán nhất định. Điều này cho thấy rằng, trải nghiệm và mức độ hiểu biết về rủi ro AI giữa các nhóm sinh viên chưa thật sự đồng đều.

Xem xét ở cấp độ biến quan sát, kết quả phân tích chỉ ra rằng, PER01 (có thể gây tổn hại khi vượt ngoài tầm kiểm soát) đạt trung bình 3,88 (SD = 0,944), là biến quan sát có điểm trung bình cao nhất. Kết quả này cho thấy sinh viên dễ dàng hình dung các rủi ro cụ thể, gần gũi về an toàn/cá nhân khi AI vận hành ngoài kiểm soát, nhất là trong bối cảnh học tập vốn gắn với dữ liệu cá nhân và các sản phẩm học thuật. Mức đồng thuận khá cao ở PER01 hàm ý

Bảng 5: *Thực trạng về nhận thức về rủi ro đạo đức của việc sử dụng AI trong học tập của sinh viên Việt Nam*

Mã biến quan sát	Nội dung	Giá trị trung bình	Độ lệch chuẩn (SD)
PER	Nhận thức về rủi ro đạo đức AI của việc sử dụng AI	3.68	0.839
PER01	Các sản phẩm AI tạo sinh có thể gây ra tổn hại và thiệt hại cho tôi và người khác nếu chúng vượt ra ngoài tầm kiểm soát của tôi.	3.88	0.944
PER02	Việc sử dụng các sản phẩm AI tạo sinh có thể dễ dàng vi phạm các chuẩn mực đạo đức và hành vi, từ đó gây ra tổn thất hoặc nguy cơ cho tôi và người khác.	3.61	1.011
PER03	Sử dụng các sản phẩm AI tạo sinh khiến tôi phải đối mặt với những rủi ro đạo đức tiềm ẩn mà tôi không thể lường trước hoặc kiểm soát được.	3.54	0.970

(Nguồn: Phân tích từ dữ liệu khảo sát)

rằng, "kiểm soát việc sử dụng" và "năng lực giám sát công cụ" là mối bận tâm trực tiếp của sinh viên khi tích hợp AI vào quá trình học tập. Biến quan sát PER02 (dễ vi phạm chuẩn mực đạo đức/hành vi) có trung bình 3,61 nhưng ghi nhận SD = 1,011, giá trị lớn nhất trong thang đo này. Sự phân tán này gợi ý dị biệt nhận thức giữa các nhóm về nguy cơ vi phạm chuẩn mực (Ví dụ: đạo văn, xuyên tạc nguồn, thiên lệch trong gợi ý). Biến quan sát có giá trị trung bình thấp nhất là PER03 (rủi ro tiềm ẩn khó lường/khó kiểm soát) (đạt 3,54 với SD = 0,970). Giá trị này phản ánh việc nhận diện rủi ro trừu tượng (hệ quả gián tiếp, dài hạn, hoặc mang tính hệ thống như thiên lệch thuật toán, xói mòn năng lực tư duy độc lập), ở mức độ thấp hơn so với các rủi ro hữu hình. Nhìn chung, sinh viên nhận thức rõ rủi ro vận hành và chuẩn mực gần gũi nhưng còn hạn chế ở rủi ro hệ thống/khó lường.

c) Ý định sử dụng AI

Về ý định sử dụng AI của sinh viên trong quá trình học tập, kết quả tại Bảng 6 đã thể hiện điều này. Kết quả phân tích cho thấy ý định sử dụng AI trong học tập (BI) của sinh viên Việt Nam ở mức khá tích cực với điểm trung bình 3,76 (SD = 0,697). Mức trung bình này cao hơn ngưỡng trung vị (3,00), hàm ý phần đông người học đã hình thành khuynh hướng sẵn sàng tích hợp AI vào hoạt động học tập. Độ lệch chuẩn tương đối thấp phản ánh sự đồng thuận vừa phải trong mẫu, không xuất hiện phân tán thái quá về xu hướng hành vi dự định.

Ở cấp độ biến quan sát, kết quả phân tích đã chỉ ra một xu hướng xem AI như một công cụ để có được những lợi ích học thuật cụ thể trong quá trình học tập. Biến quan sát BI02 (dự định sử dụng các chức năng của AI để hỗ trợ hoạt động học tập) đạt trung bình 3,95 (SD = 0,817), cao nhất thang đo. Hai biến quan sát về tần suất sử dụng, BI01 (sẽ sử dụng thường xuyên trong thời gian tới, 3,74; SD = 0,884) và

Bảng 6: *Thực trạng về ý định sử dụng AI trong học tập của sinh viên Việt Nam*

Mã biến quan sát	Nội dung	Giá trị trung bình	Độ lệch chuẩn (SD)
BI	Ý định sử dụng AI.	3.76	0.697
BI01	Tôi sẽ sử dụng AI thường xuyên trong thời gian tới.	3.74	0.884
BI02	Tôi dự định sử dụng các chức năng của AI để hỗ trợ hoạt động học tập của tôi.	3.95	0.817
BI03	Tôi sẽ đưa ra khuyến nghị của mình cho người khác sử dụng AI.	3.65	0.841
BI04	Tôi sẽ sử dụng AI một cách thường xuyên trong tương lai.	3.72	0.871

(Nguồn: Phân tích từ dữ liệu khảo sát)

 https://doi.org/10.15625/2615-8957/12620104

BI04 (sẽ sử dụng thường xuyên trong tương lai, 3,72; SD = 0,871) có giá trị tiệm cận nhau, phản ánh tính nhất quán theo thời gian của ý định: Không chỉ là sự hứng thú ngắn hạn mà có khả năng duy trì trong thời gian dài hơn. Ngược lại, biến quan sát BI03 (khuyến nghị người khác sử dụng AI) có trung bình 3,65 (SD = 0,841), thấp nhất trong nhóm. Khoảng cách này cho thấy khi chuyển từ hành vi dự định cá nhân sang hành vi mang tính xã hội/lan tỏa, mức sẵn sàng có phần giảm nhẹ. Như vậy, ý định sử dụng AI của sinh viên vào việc học tập đã hình thành rõ và thiên về mục tiêu học thuật cụ thể.

3.3. Phân tích hồi quy về mối quan hệ giữa nhận thức đạo đức về AI và rủi ro đạo đức của việc sử dụng AI đối với ý định sử dụng AI trong học tập của sinh viên Việt Nam

Kết quả phân tích hồi quy để tìm hiểu mối quan hệ giữa nhận thức đạo đức về AI và rủi ro đạo đức của việc sử dụng AI đối với ý định sử dụng AI trong học tập của sinh viên Việt Nam được trình bày tại Bảng 7.

Kết quả hồi quy cho thấy, mô hình có độ phù hợp trung bình–khá với R2=0,287 và R2 hiệu chỉnh = 0,282. Điều này tiết lộ rằng hai biến độc lập của mô hình hồi quy (EA và PER) cùng dự báo khoảng 28,7% phương sai của ý định sử dụng AI của sinh viên Việt Nam. Kiểm định F cho toàn mô hình đạt ý nghĩa cao (F=61,807; Sig.=0,000), khẳng định sự phù hợp của mô hình hồi quy này. Chẩn đoán phần dư Durbin–Watson có giá trị là 1,985 gợi ý rằng, không có hiện tượng tự tương quan nghiêm trọng và đa cộng tuyến ở mức thấp (VIF=1,371 < 3,0) (Hair Jr và cộng sự, 2014), cho phép diễn giải tin cậy các hệ số riêng phần. Phân phối phần dư xấp xỉ chuẩn do có dạng chuông, đối xứng quanh 0 (mean = 7.36E-16), không đỉnh phụ/lệch đáng kể; cùng với đó phần dư chuẩn hóa phân bổ tập trung xung quanh đường tung độ 0, do vậy giả định quan hệ tuyến tính không bị vi phạm và đồng nhất phương sai được đáp ứng. Từ đó, phương trình hồi quy chưa chuẩn hóa và chuẩn hóa được đưa ra như sau:

$$BI = 1{,}637 + 0{,}589 \times EA + \varepsilon$$

$$BI = 0{,}563 \times EA + \varepsilon$$

Theo phương trình hồi quy chưa chuẩn hóa, ý định sử dụng AI của sinh viên trong học tập chỉ chịu sự tác động của nhận thức đạo đức về AI với hệ số hồi quy chưa chuẩn hóa là 0,589, trong khi đó yếu tố nhận thức về rủi ro đạo đức của việc sử dụng AI của sinh viên lại không có ảnh hưởng gì đến ý định của sinh viên sử dụng AI cho việc học tập. Kết quả này có thể diễn giải rằng, khi EA tăng một điểm thì sẽ dự báo BI tăng thêm khoảng 0,589 điểm. Cùng với đó, phương trình hồi quy chuẩn hóa cho thấy EA có tác động thuận chiều đối với BI với hệ số hồi quy chuẩn hóa là 0,563. Với phương trình hồi quy chuẩn hóa chỉ còn một biến độc lập, EA có ảnh hưởng riêng phần mạnh và mô hình giải thích 28.7% phương sai.

4. Thảo luận

Nghiên cứu này khám phá mối quan hệ giữa nhận thức đạo đức, nhận thức về rủi ro đạo đức và ý định sử dụng AI trong học tập của sinh viên Việt Nam. Các kết quả phân tích định lượng không chỉ cung cấp một bức tranh toàn diện về thực trạng nhận thức và ý định của người học mà còn mang lại những phát hiện quan trọng, góp phần làm sâu sắc thêm hiểu biết học thuật và đưa ra các hàm ý thực tiễn giá trị.

Bảng 7: *Kết quả phân tích hồi quy*

Mô hình		Hệ số hồi quy chưa chuẩn hóa		Hệ số hồi quy chuẩn hóa	t	Sig.	Thống kê đa cộng tuyến	
		B	Sai số	β			Tolerance	VIF
1	(Sự liên tục)	1.637	0.206	-	7.926	0.000		
	EA	0.589	0.059	0.563	9.987	0.000	0.730	1.371
	PER	-0.048	0.047	-0.058	-1.021	0.308	0.730	1.371

Biến phụ thuộc: Ý định sử dụng AI (BI)
Mô hình 1: R2 = 0,287; R2 hiệu chỉnh = 0,282; Durbin-Watson = 1,985; F = 61,807 (p < 0.001) ** Có ý nghĩa ở mức 1%.
Sig: mức ý nghĩa; VIF: hệ số phóng đại phương sai; B: hệ số B; β: hệ số Beta; t: kiểm định t

(Nguồn: Phân tích từ dữ liệu khảo sát)

Kết quả nghiên cứu cho thấy, sinh viên Việt Nam thể hiện mức độ nhận thức đạo đức về AI ở mức khá cao (3,91) và có ý định sử dụng AI tích cực (3,76). Tuy nhiên, nhận thức về rủi ro đạo đức chỉ ở mức vừa phải (3,68). Đáng chú ý, trong khi các khía cạnh đạo đức nền tảng như quyền riêng tư, an toàn được nhận thức rõ, thì nguy cơ về sự phân biệt đối xử và bất công do AI gây ra lại là điểm yếu nhất trong nhận thức của sinh viên (3,39).

Phân tích hồi quy tuyến tính bội cho thấy, mô hình có độ phù hợp ở mức khá và yếu tố EA có tác động tích cực đến yếu tố BI ($\beta = 0,563$; $p < 0,001$). Kết quả này khác với tư duy thông thường khi mà người ta có thể cho rằng nhận thức cao hơn về các vấn đề đạo đức sẽ dẫn đến sự thận trọng hoặc thậm chí là do dự khi sử dụng công nghệ. Phát hiện này tương đồng với một số nghiên cứu gần đây cho rằng việc hiểu biết về các khía cạnh đạo đức giúp người dùng xây dựng lòng tin và cảm giác an toàn, từ đó thúc đẩy hành vi sử dụng công nghệ của sinh viên (Vrontis & cộng sự, 2023; Zhu & cộng sự, 2025). Mặt khác, PER không có ảnh hưởng đến BI của sinh viên ($p = 0,308$). Khám phá về mối quan hệ giữa PER và BI của nghiên cứu này lại đồng nhất với phát hiện trong nghiên cứu của Zhu và cộng sự (2025). Điều này chưa thể rút ra kết luận rằng, sinh viên Việt Nam không quan tâm đến các rủi ro về đạo đức khi sử dụng AI cho mục đích học tập và nghiên cứu mà cần có những nghiên cứu sâu hơn để có những khẳng định dựa trên bằng chứng đầy đủ hơn.

Về đóng góp ở khía cạnh lí thuyết, các kết quả nghiên cứu này đã đóng góp thêm cho vấn đề đạo đức trong việc sử dụng AI trong học tập của sinh viên, cũng như đã đề xuất yếu tố EA là một trong những yếu tố tiền đề quan trọng tác động tích cực đến BI của sinh viên. Kết quả của nghiên cứu này cũng chỉ ra rằng, không phải lúc nào "nhận thức rủi ro" luôn là một biến cản trở mang tính kháng chống đối với ý định sử dụng AI của sinh viên bằng cách chỉ ra sự thiếu vắng tác động của PER trong bối cảnh này. Với phát hiện này, nghiên cứu gợi ý rằng vai trò của rủi ro cần được xem xét một cách linh hoạt, phụ thuộc vào bản chất của công nghệ, bối cảnh sử dụng và sự cân nhắc giữa lợi ích và nguy cơ của người dùng.

Về đóng góp thực tiễn, các kết quả nghiên cứu này cũng cung cấp những căn cứ khoa học, định hướng thiết thực cho các bên liên quan trong việc thực hiện cũng như quản lí sử dụng AI trong học tập của sinh viên. Đối với các nhà hoạch định chính sách và quản lí đại học, các trường đại học nên ưu tiên các chương trình giáo dục nhằm nâng cao nhận thức đạo đức về AI hơn là cấm đoán hoặc cảnh báo về rủi ro của AI. Đối với giảng viên và các nhà phát triển chương trình, đội ngũ này cần lấp đầy "khoảng trống nhận thức" về sự thiên vị và bất công của AI để phát triển "Năng lực nhận diện thiên vị" (Bias literacy) cho sinh viên, giúp họ không chỉ là người dùng mà còn là những nhà phê bình sáng suốt đối với công nghệ này. Đối với các nhà phát triển công cụ AI trong giáo dục, họ cần thiết kế hệ thống theo định hướng đạo đức bằng việc minh bạch hóa thuật toán, giảm thiểu thiên vị.

5. Kết luận

Nghiên cứu này được thực hiện nhằm mục đích khám phá mối quan hệ giữa nhận thức đạo đức về AI, nhận thức về rủi ro đạo đức của việc sử dụng AI và ý định sử dụng AI trong học tập của sinh viên Việt Nam. Các phát hiện chính đã chỉ ra một cách thuyết phục rằng, nhận thức đạo đức là một yếu tố dự báo quan trọng, có tác động tích cực và thuận chiều đến ý định sử dụng AI của sinh viên. Đáng chú ý, nghiên cứu cũng phát hiện ra rằng, nhận thức về rủi ro đạo đức không có ảnh hưởng mang ý nghĩa thống kê đến ý định hành vi của họ. Những kết quả này không chỉ cung cấp bằng chứng thực nghiệm giá trị từ bối cảnh một quốc gia đang phát triển mà còn mang lại những hàm ý sâu sắc cả về mặt lí thuyết và thực tiễn.

Mặc dù nghiên cứu đã có những đóng góp về cả lí thuyết lẫn thực tiễn, tuy nhiên vẫn còn một số những hạn chế nhất định. *Thứ nhất*, việc áp dụng phương pháp chọn mẫu thuận tiện và đặc điểm mẫu chưa cân bằng có thể ảnh hưởng đến khả năng khái quát hóa kết quả cho toàn bộ sinh viên Việt Nam. *Thứ hai*, dữ liệu thu thập có bản chất cắt ngang, chỉ ghi nhận nhận thức và ý định tại một thời điểm duy nhất, do đó chưa thể khẳng định mối quan hệ nhân quả theo thời gian. Đây cũng là gợi ý cho hướng nghiên cứu trong tương lai đối với chủ đề này. Các nghiên cứu theo chiều dọc (Longitudinal studies) sẽ rất hữu ích để theo dõi sự thay đổi trong nhận thức và ý định của sinh viên qua các năm học. Bên cạnh đó, việc sử dụng các phương pháp nghiên cứu định tính (như phỏng vấn sâu, thảo luận nhóm) có thể giúp làm rõ hơn tại sao nhận thức rủi ro lại không ảnh hưởng đến ý định sử dụng. Cuối cùng, các nghiên cứu so sánh giữa các khối ngành khác nhau sẽ cung cấp một cái nhìn đa chiều và toàn diện hơn.

 https://doi.org/10.15625/2615-8957/12620104

Tài liệu tham khảo

Cohen, L., Manion, L. & Morrison, K. (2012). Research methods in education. In *Professional Development in Education* (6th ed., Vol. 38, Issue 3). Routledge.

Cotton, D. R. E., Cotton, P. A. & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), pp.228–239. https://doi.org/10.1080/14703297.2023.2190148

Craft, J. L. (2013). A review of the empirical ethical decision-making literature: 2004-2011. *Journal of Business Ethics*, *117*(2), pp.221–259. https://doi.org/10.1007/s10551-012-1518-9

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A. & others. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Featherman, M. S. & Pavlou, P. A. (2003). Predicting e-services adoption: a perceived risk facets perspective. *International Journal of Human-Computer Studies*, *59*(4), pp.451–474. https://doi.org/10.1016/S1071-5819(03)00111-3

Franzke, A. S. (2022). An exploratory qualitative analysis of AI ethics guidelines. *Journal of Information, Communication and Ethics in Society*, *20*(4), pp.401–423. https://doi.org/10.1108/JICES-12-2020-0125

Gujarati, D. N. & Porter, D. C. (2009). *Basic econometrics*. McGraw-hill. Hair Jr, J. F., William, Babin, B. J. & Anderson, R. E. (2014). Multivariate data analysis (MVDA). *Pharmaceutical Quality by Design: A Practical Approach*. https://doi.org/10.1002/9781118895238. ch8.

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I. & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, *21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Nguyen, X., Pho, D., Luong, D. & Cao, X. (2021). Vietnamese students'acceptance of using video conferencing tools in distance learning in COVID-19 pandemic. *Turkish Online Journal of Distance Education*, *22*(3), pp.139–162.

Nunnally, J. C. (1994). *Psychometric theory 3E*. Tata McGraw-hill education.

O'Connor, S. & ChatGPT. (2023). Open Artificial Intelligence Platforms in Nursing Education: Tools for Academic Progress or Abuse? *Nurse Education in Practice*, *66*, 103537. https://doi.org/10.1016/j.nepr.2022.103537

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A. & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, *6*(1), pp.41– 56. https://doi.org/10.37074/jalt.2023.6.1.29

Rest, J. R. (1986). *Moral development: Advances in research and theory*.

Praeger. Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), pp.1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Valentine, S. & Hollingworth, D. (2012). Moral intensity, issue importance, and ethical reasoning in operations situations. *Journal of Business Ethics*, *108*(4), pp.509–523. https://doi.org/10.1007/s10551-011-1107-3

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A. & Trichina, E. (2023). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. *Artificial Intelligence and International HRM*, pp.172–201.

Zhai, X. (2022). ChatGPT user experience: Implications for education. *Available at SSRN 4312418*.

Zhu, W., Huang, L., Zhou, X., Li, X., Shi, G., Ying, J. & Wang, C. (2025). Could AI ethical anxiety, perceived ethical risks and ethical awareness about AI influence university students' use of generative AI products? An ethical perspective. *International Journal of Human-Computer Interaction*, *41*(1), pp.742–764. https://doi.org/10.1080/10447318.2024.2323277