ẢNH HƯỞNG CỦA TÍNH MINH BẠCH VÀ CÔNG BẰNG ĐẾN Ý ĐỊNH SỬ DỤNG AI HỖ TRỢ HỌC TẬP CỦA SINH VIÊN TẠI THÀNH PHỐ HỒ CHÍ MINH: TRƯỜNG HỢP LÒNG TIN ĐÓNG VAI TRÒ TRUNG GIAN

● **VŨ THỊ PHƯƠNG ÁNH[1] - BÙI NGUYỄN MINH KHÁNH[1] - NGUYỄN TRÍ THÔNG[1]***
[1]Trường Đại học Công Thương Thành phố Hồ Chí Minh
*Email: thongnt@huit.edu.vn

TÓM TẮT:

Nghiên cứu tác động của tính minh bạch và tính công bằng đến ý định sử dụng AI trong học tập của sinh viên tại Thành phố Hồ Chí Minh, với lòng tin đóng vai trò trung gian. Dữ liệu từ 218 sinh viên được phân tích bằng phương pháp định lượng. Kết quả cho thấy, 2 yếu tố này góp phần hình thành niềm tin, qua đó thúc đẩy ý định sử dụng AI. Nghiên cứu đồng thời tích hợp khung FATE vào mô hình chấp nhận công nghệ, khẳng định vai trò trung gian của lòng tin và cung cấp cơ sở cho phát triển AI giáo dục minh bạch, có trách nhiệm.

Từ khóa: tính minh bạch, tính công bằng, niềm tin, ý định sử dụng, AI.

1. Đặt vấn đề

Sự phát triển nhanh chóng của trí tuệ nhân tạo (AI) đang làm thay đổi đáng kể giáo dục đại học thông qua các công cụ hỗ trợ học tập cá nhân hóa. Theo khảo sát của HEPI và Kortext (2025), có đến 92% sinh viên đại học đã sử dụng AI trong học tập. Tuy nhiên, việc sử dụng này vẫn mang tính tự phát khi chỉ khoảng 36% sinh viên nhận được hướng dẫn chính thức từ nhà trường (Mazaheriyan và Nourbakhsh, 2025).

Trong bối cảnh đó, việc chấp nhận AI không chỉ phụ thuộc vào tính hữu ích mà còn chịu ảnh hưởng mạnh mẽ từ các yếu tố đạo đức như tính minh bạch và tính công bằng (Karran và cộng sự, 2024). Đồng thời, niềm tin vào AI được hình thành từ sự minh bạch thuật toán và đạo đức dữ liệu được xem là yếu tố then chốt quyết định ý định sử dụng của người học (Springer Nature, 2025).

Tuy nhiên, các nghiên cứu trước đây chủ yếu tập trung vào yếu tố công nghệ mà chưa làm rõ vai trò trung gian của lòng tin trong mối quan hệ giữa các yếu tố đạo đức và ý định sử dụng AI, đặc biệt trong bối cảnh sinh viên Việt Nam (Nature Scientific Reports, 2025). Do đó, nghiên cứu này nhằm phân tích tác động của tính minh bạch và tính công bằng đến ý định sử dụng AI trong học tập của sinh viên tại

Thành phố Hồ Chí Minh, với lòng tin đóng vai trò trung gian. Kết quả nghiên cứu cho thấy, các yếu tố đạo đức này có ảnh hưởng tích cực đến niềm tin, qua đó thúc đẩy ý định sử dụng AI của sinh viên.

2. Cơ sở lý thuyết

2.1. Tổng quan các nghiên cứu về chấp nhận AI và vai trò của các yếu tố đạo đức

Nhiều nghiên cứu đã chỉ ra, việc ứng dụng AI trong giáo dục đang gia tăng nhanh chóng và đóng vai trò quan trọng trong việc nâng cao hiệu quả học tập của sinh viên. Tuy nhiên, việc sử dụng này vẫn tiềm ẩn nhiều rủi ro về đạo đức và mang tính tự phát (Tạ Tường Vi, 2025).

Về các khía cạnh đạo đức, Shin (2020, 2021) khẳng định Tính minh bạch, Tính công bằng và Trách nhiệm giải trình có ảnh hưởng trực tiếp đến việc hình thành niềm tin của người dùng đối với AI. Trái lại, Huynh và Aichner (2025) lại cho thấy, tác động của các yếu tố này không phải lúc nào cũng nhất quán, đặc biệt là sự mờ nhạt của Tính minh bạch trong bối cảnh AI tạo sinh.

Mặc dù vậy, đa số các nghiên cứu đều thống nhất Niềm tin đóng vai trò nền tảng trong việc thúc đẩy ý định sử dụng AI, đồng thời là "mắt xích" trung gian kết nối các yếu tố đầu vào với hành vi sử dụng thực tế (Zhan và cộng sự, 2024; Emon và cộng sự, 2023).

Từ các công trình trên, có thể thấy các yếu tố đạo đức và niềm tin là chìa khóa để người dùng chấp nhận AI. Tuy nhiên, cơ chế tác động giữa các yếu tố này vẫn chưa được làm rõ một cách đầy đủ, đặc biệt là đối với sinh viên Gen Z tại Việt Nam. Để giải mã cơ chế này, các lý thuyết nền tảng như Mô hình chấp nhận công nghệ (TAM), Thuyết hành vi có kế hoạch (TPB), Thuyết chấp nhận và sử dụng công nghệ (UTAUT) và Thuyết Niềm tin (Trust Theory) sẽ cung cấp cơ sở lý luận quan trọng về mối quan hệ giữa nhận thức, niềm tin và hành vi của người dùng. Trên cơ sở kế thừa các lý thuyết và nghiên cứu này, tác giả tiến hành xác định các khái niệm cũng như biến nghiên cứu phù hợp với bối cảnh đề tài, được trình bày cụ thể trong phần tiếp theo.

2.2. Các khái niệm và biến nghiên cứu

Dựa trên các lý thuyết nền tảng và mô hình nghiên cứu đề xuất, tác giả xác định 7 biến nghiên cứu chính, bao gồm: (1) Tính minh bạch; (2) Tính công bằng; (3) Nhận thức về tính dễ sử dụng; (4) Nhận thức về trách nhiệm giải trình; (5) Ảnh hưởng xã hội; (6) Niềm tin vào AI; (7) Ý định sử dụng AI.

➢ *Tính minh bạch (MB)*

Tính minh bạch là mức độ hệ thống cung cấp thông tin đầy đủ, chính xác và cho phép người dùng hiểu được quá trình ra quyết định (Rawlins, 2008). Trong bối cảnh AI, khái niệm này còn bao gồm khả năng giải thích kết quả đầu ra một cách dễ hiểu (Floridi và cộng sự, 2018).

➢ *Tính công bằng (CB)*

Tính công bằng đề cập đến việc giảm thiểu sự thiên vị trong dữ liệu và thuật toán, đảm bảo tính khách quan và không phân biệt đối xử (Binns, 2018). Khái niệm này dựa trên Thuyết Công bằng, nhấn mạnh sự so sánh giữa đầu vào và kết quả nhận được (Adams, 1965).

➢ *Nhận thức về tính dễ sử dụng (DSD)*

DSD là mức độ mà cá nhân thấy việc sử dụng hệ thống sẽ dễ học, dễ thao tác và không tốn nhiều nỗ lực (Davis, 1989). Yếu tố này ảnh hưởng trực tiếp đến thái độ và ý định sử dụng công nghệ (Venkatesh và cộng sự, 2003).

➢ *Nhận thức về trách nhiệm giải trình (GT)*

Nhận thức về trách nhiệm giải trình là mức độ cá nhân thấy hệ thống hoặc tổ chức phát triển có trách nhiệm giải trình khi xảy ra sai sót (Tetlock, 1985; Hall và cộng sự, 2017). Điều này giúp giảm cảm nhận rủi ro khi sử dụng công nghệ.

➢ *Ảnh hưởng xã hội (XH)*

Ảnh hưởng xã hội là mức độ cá nhân chịu tác động từ những người quan trọng xung quanh trong việc sử dụng hệ thống (Venkatesh và cộng sự, 2003). Ảnh hưởng này hình thành thông qua sự tuân thủ, sự đồng nhất và sự nội tâm hóa (Kelman, 1958).

➢ *Niềm tin vào AI (NT)*

Niềm tin vào AI là sự sẵn sàng chấp nhận rủi ro dựa trên kỳ vọng tích cực vào hệ thống AI sẽ hoạt động chính xác và đáng tin cậy (Mayer và cộng sự, 1995). Niềm tin phụ thuộc vào năng lực, tính chính trực và thiện chí của hệ thống (McKnight và cộng sự, 2011).

➢ *Ý định sử dụng AI (YĐ)*

Ý định sử dụng là mức độ sẵn sàng thực hiện hành vi sử dụng AI trong tương lai (Ajzen, 1991). Đây là yếu tố dự báo trực tiếp nhất cho hành vi sử dụng thực tế (Davis, 1989).

3. Phương pháp nghiên cứu

Nghiên cứu này sử dụng phương pháp định lượng nhằm kiểm định các yếu tố ảnh hưởng đến ý định sử dụng AI của sinh viên. Mô hình nghiên cứu được xây dựng dựa trên các lý thuyết TAM, UTAUT, kết hợp với thuyết niềm tin và khung đạo đức FATE. Cấu trúc mô hình bao gồm 7 biến nghiên cứu, trong đó có 5 biến độc lập là tính minh bạch (MB), tính công bằng (CB), nhận thức về tính dễ sử dụng (DSD), nhận thức về trách nhiệm giải trình (GT) và ảnh hưởng xã hội (XH); 1 biến trung gian là niềm tin vào AI (NT); và 1 biến phụ thuộc là ý định sử dụng AI (YĐ).

Dữ liệu được thu thập tại Thành phố Hồ Chí Minh thông qua Google Form từ ngày 24/02/2026 đến ngày 06/03/2026. Sau khi làm sạch, có 218 mẫu hợp lệ được sử dụng cho phân tích. Các biến được đo lường bằng thang Likert 5 mức.

Dữ liệu được xử lý bằng SPSS 27.0 với các bước gồm kiểm định độ tin cậy Cronbach's Alpha, phân tích nhân tố khám phá và phân tích tương quan Pearson. Để kiểm định giả thuyết, nghiên cứu sử dụng hồi quy tuyến tính 2 giai đoạn thay cho SEM/CFA. Cụ thể, hồi quy bội được sử dụng để đánh giá tác động của các biến độc lập lên niềm tin vào AI, sau đó hồi quy đơn được thực hiện để kiểm định ảnh hưởng của niềm tin đến ý định sử dụng AI.

4. Kết quả nghiên cứu

4.1. Thống kê mô tả mẫu khảo sát

Trong 218 quan sát thu thập tại Thành phố Hồ Chí Minh, 60,1% là nữ giới và 39,9% là nam giới, với 100% đáp viên đều nằm trong độ tuổi từ 18 - 22 tuổi. Phân theo năm học, có 40,4% là sinh viên năm 2, 22,5% là sinh viên năm 3, trong khi sinh viên năm 1 và năm 4 chiếm tỷ lệ lần lượt là 18,8% và 18,3%. Phần lớn sinh viên thường xuyên sử dụng AI với tỷ lệ lên đến 80,7%, chỉ khoảng 17% ít sử dụng và 2,3% hiếm khi sử dụng. (Bảng 1)

Bảng 1. Thống kê mô tả mẫu nghiên cứu

Thông tin mẫu		Tần số	Tỷ lệ (%)
Giới tính	Nam	87	39,9 %
	Nữ	131	60,1%
Độ tuổi	18 - 22 tuổi	218	100%
	22 tuổi trở lên	0	0%
Tần suất sử dụng	Ít sử dụng	37	17%
	Thường xuyên sử dụng	176	80,7%
	Hiếm khi sử dụng	5	2,3%
Năm học	Năm 1	41	18,8%
	Năm 2	88	40,4%
	Năm 3	49	22,5%
	Năm 4	40	18,3%

Nguồn: Kết quả xử lý từ khảo sát của nhóm nghiên cứu

4.2. Kiểm định Cronbach' Alpha đối với các thang đo lý thuyết *(Bảng 2)*

4.3. Phân tích nhân tố khám phá (EFA)

Nhóm tác giả tiến hành phân tích nhân tố khám phá (EFA) bằng phương pháp trích Principal Components kết hợp với phép xoay Varimax. Kết quả cho thấy, toàn bộ 27 biến quan sát của 5 biến

Bảng 2. Kiểm định các thang đo lý thuyết bằng Cronbach's Alpha

STT	Thang đo	Số biến quan sát	Cronbach's Alpha	Hệ số tương quan biến - tổng thấp nhất
1	Tính minh bạch (MB)	5	0,913	0,756
2	Tính công bằng (CB)	6	0,924	0,712
3	Nhận thức dễ sử dụng (DSD)	4	0,901	0,747
4	Nhận thức về trách nhiệm giải trình (GT)	5	0,908	0,722
5	Ảnh hưởng xã hội (XH)	7	0,915	0,648
6	Niềm tin vào AI (NT)	4	0,887	0,699
7	Ý định sử dụng (YĐ)	5	0,908	0,745

Nguồn: Kết quả xử lí từ khảo sát của nhóm nghiên cứu

độc lập đều đạt yêu cầu và được giữ lại, với hệ số tải nhân tố (Factor loading) > 0,5 (dao động từ 0,640 đến 0,853) và không xuất hiện hiện tượng tải chéo. (Bảng 3)

Kết quả EFA trích được 5 nhân tố với Eigenvalue > 1 (nhỏ nhất = 1,294), tổng phương sai trích đạt 72,889%, cho thấy các nhân tố giải thích tốt sự biến thiên của dữ liệu. Hệ số KMO = 0,942 (>0,5) và kiểm

Bảng 3. Kết quả phân tích EFA cho các biến độc lập

Biến quan sát	Nhân tố					
	1	2	3	4	5	
XH5	.820					Ảnh hưởng xã hội (XH)
XH2	.802					
XH3	.781					
XH1	.780					
XH6	.700					
XH4	.696					
XH7	.640					
CB4		.853				Tính công bằng (CB)
CB1		.827				
CB5		.805				
CB2		.739				
CB3		.733				
CB6		.711				
GT2			.781			Nhận thức trách nhiệm giải trình (GT)
GT4			.776			
GT3			.736			
GT1			.733			
GT5			.717			
MB2				.738		Tính minh bạch (MB)
MB4				.735		
MB1				.712		
MB3				.703		
MB5				.683		
DSD4					.804	Nhận thức về tính dễ sử dụng (DSD)
DSD2					.787	
DSD3					.780	
DSD1					.751	
Eigenvalue	12,551	2,586	1,720	1,530	1,294	
Phương sai trích	18,008	34,863	48,530	61,217	72,889	

Nguồn: Kết quả xử lý từ khảo sát của nhóm nghiên cứu

định Bartlett's có ý nghĩa thống kê (Chi-square = 4490,075; Sig. = 0,000), chứng tỏ dữ liệu phù hợp để phân tích nhân tố và các biến tương quan với nhau trong tổng thể. (Bảng 4)

5. Kết quả nghiên cứu

5.1. Các kết quả nghiên cứu chính

Nghiên cứu này phát triển và kiểm định các khái niệm gồm: Tính minh bạch, Tính công bằng, Nhận thức về tính dễ sử dụng, Nhận thức về trách nhiệm giải trình, Ảnh hưởng xã hội, Niềm tin vào AI và Ý định sử dụng AI, trên cơ sở kế thừa và điều chỉnh từ các nghiên cứu trước (Ví dụ: Shin, 2020, 2021; Huynh và Aichner, 2025). Kết quả phân tích cho thấy, cả 5 yếu tố đầu vào đều có tác động tích cực đến Niềm tin vào AI, trong đó Tính công bằng và Tính minh bạch có mức độ ảnh hưởng mạnh nhất. Đồng thời, Niềm tin vào AI đóng vai trò trung gian quan

Bảng 4. Kết quả phân tích EFA cho biến Trung gian và biến Phụ thuộc

Biến quan sát	Nhân số	Eigenvalue	Phương sai trích	
NT2	.917	2,995	74,882	Niềm tin vào AI (NT) *(Biến trung gian)*
NT4	.862			
NT1	.853			
NT3	.827			
YĐ2	.885	3,660	73,198	Ý định sử dụng AI (YĐ) *(Biến phụ thuộc)*
YĐ3	.867			
YĐ1	.845			
YĐ5	.841			
YĐ4	.839			

Nguồn: Kết quả xử lý từ khảo sát của nhóm nghiên cứu

Đối với biến trung gian (Niềm tin vào AI), kết quả EFA cho thấy, 4 biến quan sát hội tụ thành 1 nhân tố duy nhất, với hệ số tải nhân tố dao động từ 0,827 đến 0,917. Chỉ số KMO = 0,825, Bartlett's test có ý nghĩa (Sig. = 0,000), phương sai trích đạt 74,882% và Eigenvalue = 2,995, cho thấy thang đo đạt giá trị hội tụ và tính đơn hướng.

Tương tự, biến phụ thuộc (Ý định sử dụng AI) gồm 5 biến quan sát được trích thành 1 nhân tố duy nhất với hệ số tải nhân tố từ 0,839 đến 0,885. Hệ số KMO = 0,879, kiểm định Bartlett's có ý nghĩa (Sig. = 0,000), phương sai trích đạt 73,198% và Eigenvalue = 3,660, cho thấy thang đo đảm bảo độ tin cậy và giá trị.

4.4. Đánh giá độ tin cậy (Cronbach's Alpha)

Kết quả phân tích độ tin cậy cho thấy, các thang đo khái niệm đều có hệ số Cronbach's Alpha dao động từ 0,887 đến 0,924, cho thấy độ tin cậy cao (Hair et al 2010),

Bên cạnh đó, Hair et al (2010) cũng cho thấy, phương sai trích cần vượt ngưỡng 50% để đảm bảo khả năng giải thích của các nhân tố. Kết quả EFA cho thấy, phương sai trích của nhóm biến độc lập đạt 72,889%, biến trung gian (NT) đạt 74,882% và biến phụ thuộc (YĐ) đạt 73,198%, đều lớn hơn 50%, chứng tỏ các thang đo đạt giá trị hội tụ. (Bảng 5)

Bảng 5. Độ tin cậy và phương sai trích của các khái niệm

	Độ tin cậy (Cronbach's Alpha)	Phương sai trích (%)
MB	0,913.	72,889
CB	0,924	72,889
DSD	0,901	72,889
GT	0,908	72,889
XH	0,915	72,889
NT	0,887	74,882
YĐ	0,908	73,198

Nguồn: Kết quả xử lý từ khảo sát của nhóm nghiên cứu

trọng, thúc đẩy ý định sử dụng AI trong học tập. Mặc dù các kết quả này cung cấp bằng chứng thực nghiệm có giá trị, nghiên cứu vẫn còn hạn chế về phạm vi mẫu tại Thành phố Hồ Chí Minh, do đó cần có các nghiên cứu tiếp theo với phạm vi rộng hơn để kiểm chứng và khái quát hóa kết quả.

5.2. Các gợi ý nhằm nâng cao sự tin tưởng và ý định sử dụng AI trong học tập

Từ kết quả nghiên cứu, để nâng cao niềm tin và ý định sử dụng AI trong học tập, các nhà phát triển cần chú trọng thiết kế hệ thống theo hướng minh bạch, cung cấp giải thích rõ ràng về cách thức tạo ra kết quả, hiển thị nguồn thông tin và cho phép người dùng phản hồi. Đồng thời, cần đảm bảo tính công bằng thông qua việc kiểm soát và giảm thiểu thiên lệch trong dữ liệu và thuật toán. Bên cạnh đó, việc thiết kế giao diện thân thiện, dễ sử dụng và tích hợp yếu tố cộng đồng sẽ góp phần tận dụng ảnh hưởng xã hội tích cực. Về phía các cơ sở giáo dục và cơ quan quản lý, cần xây dựng các tiêu chuẩn đánh giá, yêu cầu minh bạch trong sử dụng AI và ban hành các hướng dẫn đạo đức theo khung FATE nhằm định hướng việc sử dụng AI một cách an toàn, có trách nhiệm và hiệu quả ■

Lời cảm ơn:

Nghiên cứu này do Trường Đại học Công Thương Thành phố Hồ Chí Minh bảo trợ và cấp kinh phí theo Hợp đồng số 456/HĐ-DCT ngày 15 tháng 12 năm 2025

TÀI LIỆU THAM KHẢO:

Hoàng Trọng & Chu Nguyễn Mộng Ngọc (2008). Phân tích dữ liệu nghiên cứu với SPSS. NXB Hồng Đức.

Tạ Tường Vi (2025). Sinh viên Việt Nam với việc sử dụng Trí tuệ nhân tạo: Thực trạng, nhận thức và định hướng. Truy cập tại https://www.quanlynhanuoc.vn/2025/05/27/sinh-vien-viet-nam-voi-viec-su-dung-tri-tue-nhan-tao-thuc-trang-nhan-thuc-va-dinh-huong/.

Adams J. S. (1965). Inequity in social exchange. Advances in experimental social psychology, 267-299. https://doi.org/10.1016/s0065-2601(08)60108-2.

Ajzen I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-t .

Alshammari S. H., Alrashidi M. E., Alshammari M. H., Alshammari A. E. A., & Alkhwaldi A. F. (2025). Determinants of student adoption of artificial intelligence applications in higher education. Scientific Reports, 15(1), 35921. https://doi.org/10.1038/s41598-025-19851-5

Binns R. (2017). Fairness in Machine Learning: Lessons from Political Philosophy. arXiv (Cornell University), 81. https://doi.org/10.48550/arxiv.1712.03586

Brandhofer G., & Tengler K. (2025). Acceptance of AI applications among teachers and student teachers. Discover Education, 4(1). https://doi.org/10.1007/s44217-025-00637-w

Davis F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340. https://doi.org/10.2307/249008

Emon M. M. H., Hassan F., Nahid M. H. O., & Rattanawiboonsom V. (2023). Predicting Adoption Intention of Artificial Intelligence: A Study on ChatGPT. International Journal of Technology, 14(7), 1546-1556. https://doi.org/10.14716/ijtech.v14i7.6744.

Floridi L., Cowls J., Beltrametti M., Chatila R., Chazerand P., Dignum V., Luetge C., Madelin R., Pagallo U., Rossi F., Schafer B., Valcke P., & Vayena E. (2018b). AI4 People An Ethical Framework for a Good AI Society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5.

Fraenkel J. R., & Wallen N. E. (2006). How to design and evaluate research in education (6th ed.). McGraw-Hill.

Gerbing D. W., & Anderson J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. Journal of Marketing Research, 25(2), 186. https://doi.org/10.2307/3172650,

Holmes W., Bialik M., & Fadel C. (2019). Artificial intelligence in Education: Promises and implications for teaching and learning. In Open Research Online (The Open University).

Huynh M. T., & Aichner T. (2025). In generative artificial intelligence we trust: unpacking determinants and outcomes for cognitive trust. Ai & Society, 40(8), 5849-5869. https://doi.org/10.1007/s00146-025-02378-8.

Hair J.F., Black W.C., Babin B.J. and Anderson R.E. (2010) Multivariate Data Analysis. 7th Edition, Pearson, New York.

Illia A., Ara A., Zainol Z., Duraisamy B., et al. (2025). Determinants of trust in AI-generated content. Issues in Information Systems. https://doi.org/10.48009/4_iis_2025_106.

Khoso A. K., Honggang W., & Darazi M. A. (2025). Trust and attitude towards AI as pathways to creativity: a TAM Model study of EFL students' digital literacy and AI acceptance. Humanities and Social Sciences Communications, 13(1). https://doi.org/10.1057/s41599-025-06362-x.

Kaiser H. F. (1974). An index of factorial simplicity. Psychometrika, 39(1), 31-36. https://doi.org/10.1007/bf02291575.

Kuzminov Y. I., Kruchinskaia E. V., Gruzdev I. A., & Naumov A. A. (2025). Falling behind and getting ahead: Student use of Generative AI in education. Vysshee Obrazovanie V Rossii = Higher Education in Russia, 34(6), 9-35. https://doi.org/10.31992/0869-3617-2025-34-6-9-35.

Karran A. J., Charland P., Martineau J., De Guinea Lopez D. a. a. O., Lesage A., Senecal S., & Leger P. (2024). Multi-stakeholder perspective on responsible artificial intelligence and acceptability in education. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2402.15027.

Luckin R. (2018). Machine Learning and Human Intelligence: The Future of Education for the 21st century. In CERN Document Server (European Organization for Nuclear Research). http://cds.cern.ch/record/2698126.

Mayer R. C., Davis J. H., & Schoorman F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709-734. https://doi.org/10.5465/amr.1995.9508080335

Mcknight D. H., Carter M., Thatcher J. B., & Clay P. F. (2011). Trust in a specific technology. ACM Transactions on Management Information Systems, 2(2), 1-25. https://doi.org/10.1145/1985347.1985353.

Masrek M. N., Baharuddin M. F., & Syam A. M. (2025). Determinants of behavioral intention to use generative AI: the role of trust, personal innovativeness, and UTAUT II factors. International Journal of Basic and Applied Sciences, 14(4), 378-390. https://doi.org/10.14419/44tk8615

Mazaheriyan A., & Nourbakhsh E. (2025). Beyond the hype: Critical analysis of student motivations and ethical boundaries in educational AI use in higher education. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2511.11369.

O'brien R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality & Quantity, 41(5), 673-690. https://doi.org/10.1007/s11135-006-9018-6

Rawlins B. (2008). Measuring the relationship between organizational transparency and employee trust. ScholarsArchive (Brigham Young University), 2(2). https://scholarsarchive.byu.edu/facpub/885.

Shin D. (2020). User Perceptions of Algorithmic Decisions in the Personalized AI System:Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability. Journal of Broadcasting & Electronic Media, 64(4), 541-565. https://doi.org/10.1080/08838151.2020.1843357.

Shin D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies, 146. https://doi.org/10.1016/j.ijhcs.2020.102551.

Tetlock, P. E. (1985). Accountability: a social check on the fundamental attribution error. Social Psychology Quarterly, 48(3), 227. https://doi.org/10.2307/3033683.

Venkatesh V., Morris M. G., Davis G. B., & Davis F. D. (2003). User acceptance of Information Technology: toward a unified view1. MIS Quarterly, 27(3), 425-478. https://doi.org/10.2307/3003654038.

Zhan X., Abdi N., Seymour W., & Such J. (2024). Healthcare Voice AI Assistants: Factors Influencing Trust and Intention to Use. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), 1-37. https://doi.org/10.1145/3637339

Zhang S., Meng Z., Chen B., Yang X., & Zhao X. (2021). Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants-Trust-Based Mediating Effects. Front Psychol, 12, 728495. https://doi.org/10.3389/fpsyg.2021.728495.

Zarifis A., Kawalek P., & Azadegan A. (2020). Evaluating if trust and personal information privacy concerns are barriers to using health insurance that explicitly utilizes AI. Journal of Internet Commerce, 20(1), 66-83. https://doi.org/10.1080/15332861.2020.1832817.

Ngày nhận bài: 5/03/2026
Ngày phản biện đánh giá và sửa chữa: 20/03/2026
Ngày chấp nhận đăng bài: 5/04/2026

IMPACT OF TRANSPARENCY AND FAIRNESS ON STUDENTS' INTENTION TO USE ARTIFICIAL INTELLIGENCE FOR LEARNING IN HO CHI MINH CITY: THE MEDIATING ROLE OF TRUST

● **VU THI PHUONG ANH[1]**
● **BUI NGUYEN MINH KHANH[1]**
● **NGUYEN TRI THONG[1]***
[1]Ho Chi Minh City University of Industry and Trade
*Email: thongnt@huit.edu.vn

ABSTRACT:

This study investigates the effects of transparency and fairness on students' intention to use artificial intelligence (AI) in learning in Ho Chi Minh City, with trust serving as a mediating variable. Data collected from 218 students were analyzed using quantitative methods to examine the relationships among the proposed constructs. The findings reveal that both transparency and fairness significantly enhance students' trust in AI systems, which in turn exerts a positive influence on their intention to adopt AI for educational purposes. By integrating the FATE (Fairness, Accountability, Transparency, and Ethics) framework into the Technology Acceptance Model, the study confirms the pivotal mediating role of trust and provides robust empirical evidence supporting the design and implementation of transparent, equitable, and responsible AI systems in education.

Keywords: transparency, fairness, trust, usage intention, artificial intelligence.